

## Review

## The Genetics of Multiple Sclerosis: From 0 to 200 in 50 Years

Sergio E. Baranzini<sup>1,2,3,\*</sup> and Jorge R. Oksenberg<sup>1,2</sup>

**Multiple sclerosis (MS) is a common autoimmune disease that targets myelin in the central nervous system (CNS). Multiple genome-wide association studies (GWAS) over the past 10 years have uncovered more than 200 loci that independently contribute to disease pathogenesis. As with many other complex diseases, risk of developing MS is driven by multiple common variants whose biological effects are not immediately clear. Here, we present a historical perspective on the progress made in MS genetics and discuss current work geared towards creating a more complete model that accurately represents the genetic landscape of MS susceptibility. Such a model necessarily includes a better understanding of the individual contributions of each common variant to the cellular phenotypes, and interactions with other genes and with the environment. Future genetic studies in MS will likely focus on the role of rare variants and endophenotypes.**

**MS Is a Genetic Disease**

MS is a chronic disease of the CNS that affects more than two million individuals worldwide [1,2]. It is typically diagnosed between the third and fourth decade of life, occurs more frequently in women than men, and is considered the most common cause of nontraumatic neurological disability in young adults. Although the trigger(s) of MS remains unknown, its pathogenesis is best explained by a multifactorial model that incorporates interactions between genetic, epigenetic, and infectious, nutritional, climatic, or other environmental influences, including Epstein–Barr virus (EBV) infection, sunlight exposure, and smoking [3–7]. This array of factors results in the loss of immune homeostasis and **self-tolerance** (see [Glossary](#)) manifested in brain and spinal cord infiltration by activated peripheral mononuclear cells, and the development of unregulated pathologic inflammatory responses against structural components of the CNS. Myelin loss, **gliosis**, and the resulting axonal pathology culminate in progressive, often severe, neurological dysfunction. Emerging data show that the neuroinflammation and neurodegeneration that occur in MS are overlapping and have a complex dependence [8–11] but the autoimmune model of pathogenesis has set early the tone for immunotherapy as the primary clinical management strategy, first by global immune suppression using aggressive anti-inflammatory drugs, and, more recently, by selectively targeting specific elements of the immune response [12–15]. The recognition that axonal damage unresponsive to immunotherapy is the main driver of disability in MS highlights the need to emphasize the genetics of gray and white matter cell death and repair as pressing research frontiers in this disease.

Heritable contributions to MS risk are undisputable. Pivotal epidemiological data consistent with a prominent role of genetic factors in MS include the high disease prevalence in distinct ancestral groups (particularly those of northern European origin) compared with others

## Trends

More than 200 loci have been associated with MS susceptibility to date (half of them in the past 4 years alone).

There is extensive sharing of genetic risk variants between MS and other autoimmune diseases. This suggests a model in which a general risk for autoimmunity is inherited. Additional genetic (and epigenetic) determinants and environmental exposures are compounded to ultimately define the target organ of the autodestructive process.

Efforts to characterize the biological consequences of reported associations are ongoing.

Cell-specific pathways are being developed to understand disease heterogeneity and individualized risk assessment.

<sup>1</sup>Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, San Francisco, CA, USA

<sup>2</sup>Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA

<sup>3</sup>Graduate Program in Bioinformatics, University of California San Francisco, San Francisco, CA, USA

\*Correspondence: [Sergio.baranzini@ucsf.edu](mailto:Sergio.baranzini@ucsf.edu) (S.E. Baranzini).

(e.g., Africans and Asians), in some cases irrespective of geographic location. Remarkably, the prevalence of MS appears to have steadily increased over the past century, primarily in women [16–18] and in regions previously considered low incidence [19,20]. The higher prevalence rates (140–250 per 100 000) are still found in northern Europeans and in whites living in the northern USA and Canada [2,21]. By contrast, low prevalence rates are found in Asian countries (e.g., six per 100 000 in Japan) and native populations across the Americas and Oceania [22]. Notwithstanding difficulties in surveillance, MS is almost nonexistent in black Africans and early estimates in the USA suggested that the disease is significantly less prevalent in admixed African Americans than in European Americans (relative risk of 0.64 [23]). However, more recent studies are challenging this long-held belief, suggesting that MS incidence in African Americans is equal to, or potentially even higher than in European Americans [24,25]. Furthermore, compared with Europeans, African Americans also carry a greater risk of ambulatory disability that may be, at least in part, genetically determined [26,27].

Analogous to other autoimmune diseases, individuals with MS tend to cluster in families. Multiple studies have shown, for example, that monozygotic twins have a higher concordance rate (20–30%) compared with dizygotic twins (2–5%), providing strong support for a significant but complex genetic etiology in MS [28–30]. Equally important, siblings of an affected individual are at least seven times more likely to develop MS than the general population [31,32]. Second- and third-degree relatives, but not spouses, carry a modest increased risk [33,34], also in agreement with genetic factors causing the familial clustering. Altogether, the population and family data lead to a broad acceptance that the MS-prone genotype results from multiple independent DNA variants that are relatively frequent in the population [35]. The recognition during the early 1970s of the influence of specific **human leukocyte antigen** (HLA) variants within the **major histocompatibility complex** (MHC) gene complex (chromosome 6p21) in MS susceptibility represented the first empirical demonstration linking disease risk with common genetic variation, but it was the polygenic model of MS **heritability** that provided the theoretical justification for assembling multicenter large DNA data sets to pursue **GWAS** (Box 1).

### Ten Years of GWAS in MS

The first concerted efforts to identify susceptibility genome-wide were done through linkage analysis in multicaser families. These studies failed to identify any consistent statistically significant signal [36–38]. Yet, a suggestive **logarithm of the odds ratio** (OR) (LOD) score found in the *HLA* region was later confirmed in larger studies, providing a statistical road map to take full advantage of the emerging microarray technology [39,40]. The first GWAS in MS was completed by the International Multiple Sclerosis Genetics Consortium (IMSGC) on subjects from the USA and UK using an unorthodox family-based (patients with MS and both parents) study design for discovery coupled with case-control replication [41]. Several GWAS and large-scale targeted studies followed in the ensuing 10 years [42–48], culminating in the collation of a genetic data set representing 47 351 MS subjects and 68 284 controls that showed unequivocal statistical evidence for the association of 200 autosomal susceptibility variants outside the MHC, one chromosome X variant, and 32 independent associations within the extended MHC [49]. Altogether, different measures of the heritability explained by all associations are still in the 20–30% range, of which a substantial proportion can be assigned to the MHC region. While the explanation for this missing heritability may be rooted in gene×gene and gene×environment (G×E) interactions and epigenetic factors, it is also conceivable that the early measures of heritability were overestimated.

The power of genetic association studies mostly depends on three variables: effect size, sample size, and allele frequency. Initial power calculations for MS estimated that, to detect an association with **OR** >1.2 investigating common variants (those with minor allele frequency

### Glossary

**Expression quantitative trait loci (eQTL):** regions of the genome containing DNA sequence variants that influence the expression level of one or more genes.

**Genome-wide association study (GWAS):** examination of a genome-wide set of genetic variants in different individuals to see whether any variant is associated with a trait. GWASs typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits such as major human diseases, but can equally be applied to any other organism.

**Gliosis:** a process leading to scars in the CNS that involves the production of a dense fibrous network of neuroglia (supporting cells) in areas of damage. In MS, gliosis ensues after axonal loss.

**Heritability:** a statistical term representing the proportion of phenotypic variance attributable to genetic variance.

**Human leukocyte antigen (HLA):** a genomic locus encoding the MHC proteins in humans.

**Interactome (protein):** the entire set of molecular interactions of a cell or target protein.

**Linkage disequilibrium (LD):** in population genetics, LD is the nonrandom association of alleles at different loci in a given population.

**Logarithm of the odds ratio (OR) (LOD):** in genetics, the OR is one way to quantify how strongly the presence or absence of a given phenotype is associated with the frequency of a particular allele in a specific population.

**Major histocompatibility complex (MHC):** a set of cell surface proteins essential for the acquired immune system to recognize foreign molecules in vertebrates, which in turn determines histocompatibility. Failure to match HLA types between donor and recipient in an organ or tissue transplant can result in rejection.

**Metagenomics:** the study of genetic material recovered directly from environmental samples. Here, it refers to the genomes of bacteria and other microorganisms around or inhabiting the human body.

**Roadmap Epigenomics Project (REP):** an NIH-sponsored initiative to produce a public resource of human epigenomic data to catalyze basic

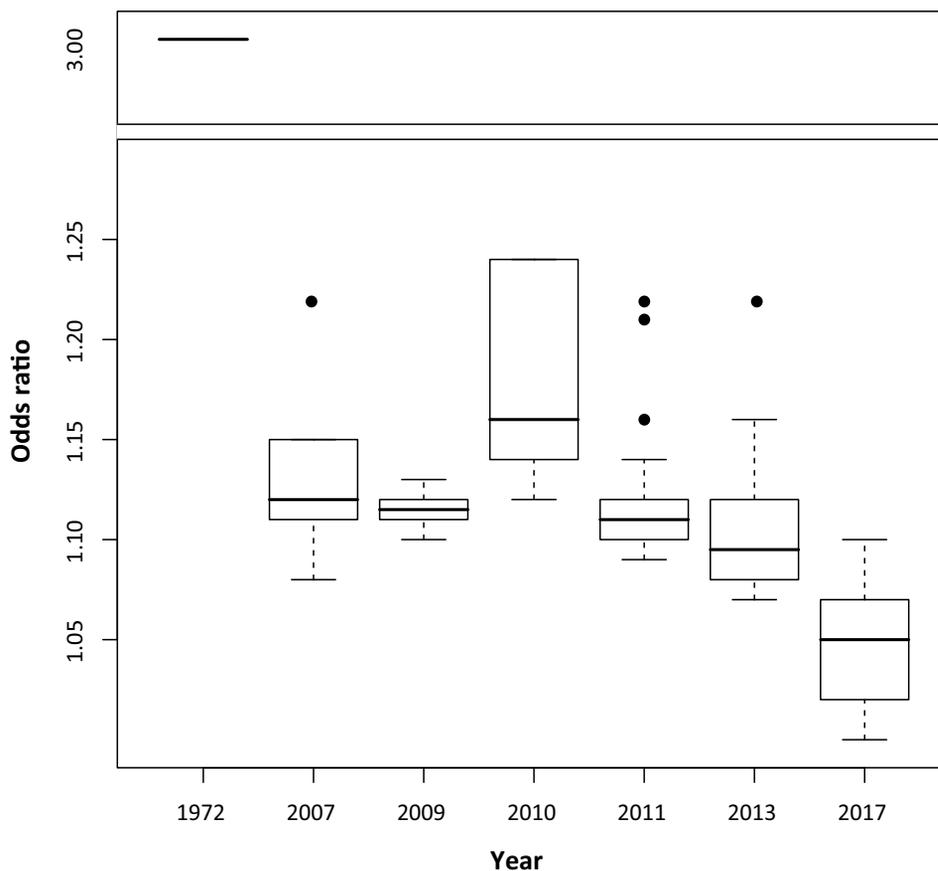
>5%), 10 000 samples would be required to ensure that a  $P$  value of  $10^{-7}$  is a hundred times more likely to be true than false (by contrast, the posterior odds of the same  $P$  value with only 2000 patients would be in favor of a false discovery) [35]. While in later years, a  $P$  value threshold of  $5 \times 10^{-8}$  for GWAS emerged as a consensus [50], a systematic analysis of published studies suggested a further relaxation of this  $P$  value would result in the discovery of additional genuine associations [51].

When all MS associations are analyzed by year of discovery (a proxy for study size), they show an ever-decreasing median odds ratio (OR; a measure of effect size) (Figure 1). For example, the association with HLA-DRB1\*15:01 has an OR  $\sim 3.5$  and was first reported in 1972 with just a few hundred cases, while the first discovered non-HLA loci (in 2007) have OR  $\sim 1.15$  and nearly 1000 cases were needed. The most recent studies (2011, 2013, and 2017) analyzed tens of thousands of cases (10 000, 15 000, and 40 000, respectively), which enabled the reporting of statistically replicated signals down to OR = 1.05. This asymptotic trend toward OR  $\sim 1$  also means that the cost:benefit incentive for performing additional GWAS searching for additional common susceptibility signals in Europeans may no longer be justified or realistic. Associations in that low effect size-range mean that those variants are almost as frequent in cases as they are in controls, thus raising the argument of whether they are biologically relevant. Quantitative risk scores have been developed to assess the significance of the 'genetic load' of individual

biology and disease-oriented research.

**Self-tolerance:** immunological tolerance is the failure to mount an immune response to an antigen. Self-tolerance refers to the (normal) failure to respond to one own's antigens, thus avoiding avoid autoimmunity.

**Splicing acceptor site:** in the splicing of RNA, the site at the 3' end of an intron.



Trends in Genetics

Figure 1. Decreasing Odds Ratio (OR) as Study  $n$  Increases. Box plots represent the median and quartiles of the OR of studies performed at the indicated years. With exception of 2010, each new study revealed associations with smaller OR.

patients and whether all associations count equally towards risk [52,53]. However, at least in these early versions, these algorithms are not able to accurately predict risk of individual subjects and, thus, are not yet an option to be used in the clinic for precision medicine purposes.

However, this does not mean the GWAS era is over in MS. Due to inherently different **linkage disequilibrium** (LD) structures, studies of other populations (African-American/European, Hispanics, or Asians) is a meritorious proposal for fine-mapping and identification of population-specific effects, notwithstanding the difficulties in assembling the required large DNA sample sets [54,55]. Likewise, the study of disease heterogeneity (clinical subtypes or endo-phenotypes among cases) by GWAS is a valid approach with high potential for discoveries of disease modifiers, particularly with the large cohorts that have been already assembled. Here, accurate and quantitative description of the phenotypes, beyond whether a subject has the disease or not, is crucial to success. Finally, the search for rare(r) variants by exome sequencing or similar approaches is an active field of investigation.

Nearly all the identified associations map to noncoding regions of the genome, either in intragenic or intergenic regions. Naturally, those mapping to intragenic regions were explored first (IL7RA, TNFSF10, etc.) as their relationship to function is thought to relate to splicing, mRNA stability, or promoter activity of the gene in question. However, most associations lie in genomic regions distant from any known gene. In those cases, the most likely explanation is that they modify regulatory sites and indirectly influence gene activity. Examples of regulatory regions include open chromatin, histone modifications, enhancers, and repressors. Efforts to integrate statistical evidence of association with their potentially regulatory effect are underway. In the most recent MS GWAS, an analysis of open chromatin regions was performed using DNase I hypersensitivity data collated from 56 tissues from the NIH **Roadmap Epigenomics Project** (REPI) [56]. For each tissue, the proportion of SNPs in the list that were located on DHS regions to the proportion of all SNPs located on DHS regions active in that tissue were compared. Notably, statistically significant enrichment was observed only in nine types of cell from the immune system, including CD3, CD4, CD8 (T cells), CD14 (monocytes), and CD19 (B cells) [49]. This finding highlights the immune nature of genetic MS susceptibility.

### Sharing with Other Autoimmune Diseases

One salient characteristic of the MS susceptibility map is the large proportion of the associations (either the exact same variant, or within the same gene or locus) that are shared with other autoimmune diseases, such as type 1 diabetes mellitus (T1DM), rheumatoid arthritis (RA),

#### Box 1. The MHC Is the Most Robust MS Genetic Susceptibility Locus

GWAS have confirmed that the main MS susceptibility association signal genome-wide maps to the *HLA-DRB1* locus in the MHC. This region is located on the short arm of chromosome 6 at p21.3, spans almost 4000 kb of DNA, and contains approximately 165 closely linked genes, some 50% of which have pivotal roles in the immune system [82]. These include the *HLA* genes, which have been extensively studied as the principal determinants of allogeneic transplantation outcomes [83]. The polymorphism of many of the *HLA* genes is extraordinary, with more than 14 000 alleles identified to date. More than 100 infectious, autoimmune, and inflammatory disease phenotypes, including MS, as well as drug reactions and cancers, are associated with *HLA* gene variations [84,85]. The association of the *HLA* locus with MS risk was first described several decades ago [86] and has since been observed across all populations studied [87]. *HLA-DRB1* allelic heterogeneity *vis-à-vis* MS risk and copy number effects have also been described [88]. Individuals who are *HLA-DRB1\*15:01* homozygotes carry a high-risk genotype with ORs exceeding 7.0, compared with a range between 3.5 and 5.0 for heterozygotes *HLA-DRB1\*15:01/X*. In addition to risk, *HLA-DRB1\*15:01* has been associated with phenotypic markers of disease severity [89]. A decade-long effort with genomic screens provided statistical evidence for over 30 independent allelic and genetic associations within the extended MHC region. Hierarchical allelic lineages and epistatic effects affecting risk are described but poorly understood [90,91]. It is sobering to recognize that, despite the fact that examination of *HLA* variation at this level has been performed in MS cohorts comprising many thousands of individuals, a complete understanding of the role of the MHC in disease pathogenesis remains elusive.

systemic lupus erythematosus (SLE), and Crohn disease (CD), among others. Interestingly, some of these shared associations confer risk for one disease but protection for another. For example, allelic heterogeneity was reported in the *IL2RA* locus, where rs11594656-A was shown to be associated with susceptibility to one disease (MS) and protection against another (T1DM). Furthermore, rs41295061-A only conferred susceptibility to T1DM but not MS. Finally, rs2104286-G was only associated with T1DM [57]. While somewhat surprising at first, genetic and allelic heterogeneity across autoimmune diseases is now recognized to be a distinctive feature of this group of disorders.

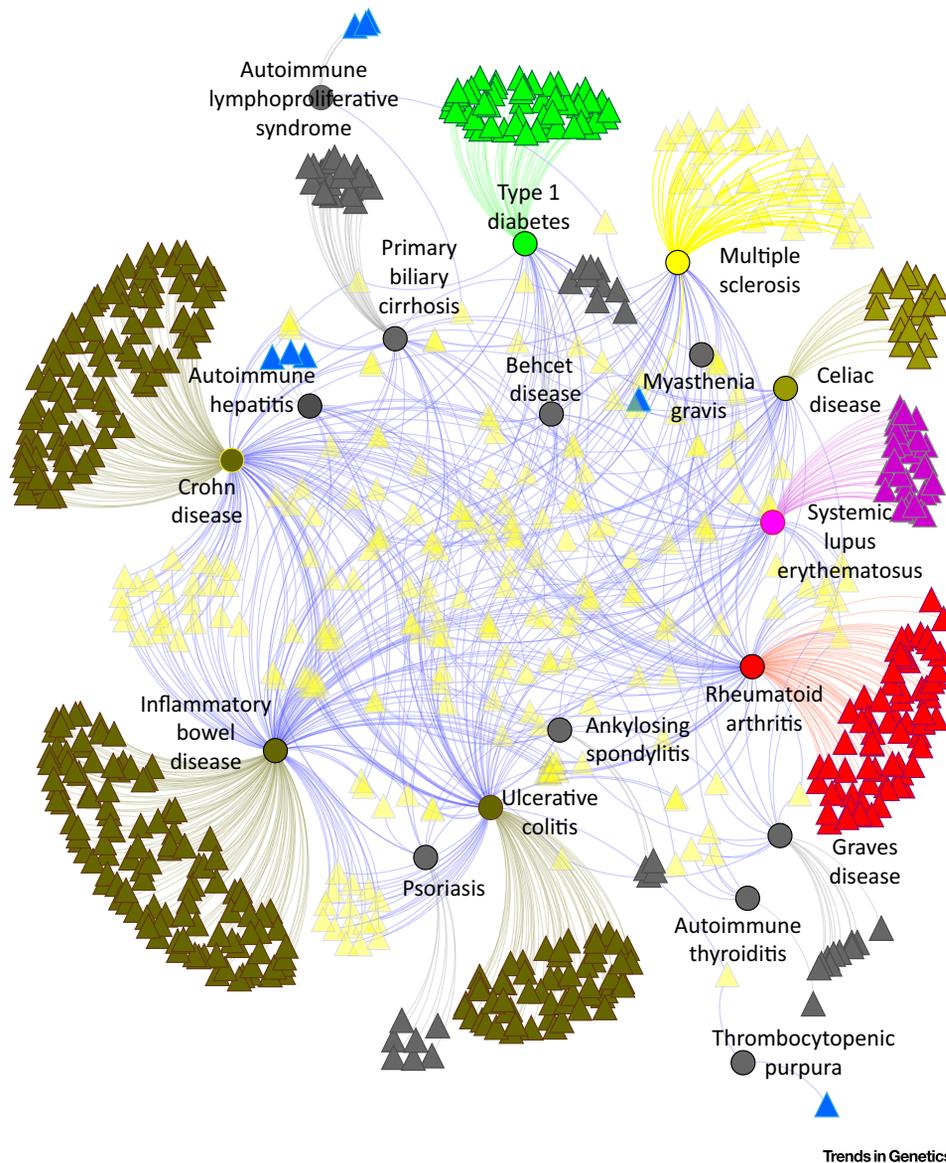
Figure 2 shows associated genes for most autoimmune diseases with GWAS data (obtained from the GWAS Catalog). It clearly shows a core of shared loci (blue triangles) and several disease-specific ones represented by different colors (MS in yellow). This remarkable sharing is the highest among the complex diseases, including cancers, neurological, and metabolic (data not shown). This particular architecture confirms that MS is at its core an autoimmune disease and suggests that susceptibility to autoimmunity is an inherited risk that later in life may be compounded and shaped by additional genetic (and epigenetic) determinants and environmental exposures to ultimately define the predominant effector mechanism (cellular versus humoral) and target organ of the autodestructive process.

### From Mapping to Function

One of the most difficult steps in any GWAS is to infer the biologically relevant consequences of each statistical association. Given that most associations map to noncoding regions, in the absence of obvious functional signals, earlier studies simply reported as candidate causative the closest gene(s) to the most significant independent effect. However, with continuous improvement in our understanding of the molecular machinery of gene regulation, it has now become clear that this crude approach needs further refinement. One way to confirm the functional effect of genetic associations is to conduct *in vitro* or *in vivo* experiments in which the presence of the variant is studied in the context of expression of the putative regulated gene target(s). The first such demonstration in MS was the discovery that rs6897932, a polymorphism located in chromosome 5p13.2 and previously associated with disease susceptibility (OR = 1.18) [41], influenced the function of the gene encoding interleukin 7 receptor  $\alpha$  (*IL7RA*). Specifically, inheritance of the risk allele (C) disrupts a **splicing acceptor site** and results in transcriptional skipping of exon 6 of the gene, thus altering the relative amounts of soluble and membrane-bound isoforms of the gene [58]. More recently, several single nucleotide polymorphisms (SNPs) within the gene encoding RNA helicase DEAD box polypeptide 39B (*DDX39B*) were shown to be associated with MS risk in a meta-analysis [59]. In particular, rs2523506-A, located within the 5'-untranslated region (UTR) of this gene, was shown to reduce mRNA translation of *DDX39B*, which is in turn a potent activator of *IL7R* exon 6 transcription. This example provides functional evidence that two associated loci can work together to confer susceptibility to MS. Given the large number of independent associations, and the networked molecular pathways that underlie cellular behavior, this example may only be the tip of the iceberg in unraveling the complex genetic architecture of MS.

Another well-characterized example is the intronic SNP rs1800693 in the gene encoding tumor necrosis factor receptor super family 1A (*TNFRSF1A*). In this case, the risk allele also results in skipping of exon 6 and the production of a novel soluble form of the TNF receptor. This soluble protein is able to inhibit TNF signaling inside the cells, mirroring somehow the exacerbating effects of TNF-blocking drugs on MS course [60].

One of the most complex regions in the MS susceptibility map is 1p22.1, which has been the subject of several reports with seemingly contradictory results; that is, different genes in the locus were linked to the association signal [61–64]. In this locus, GWAS meta-analysis identified



**Figure 2. An Autoimmunity Gene Network.** All known associations with common autoimmune diseases were obtained from the Genome-Wide Association Study (GWAS) Catalog [92] and mapped to the closest or most likely affected gene, as reported by the authors of each publication. Then, data were displayed graphically as a network using the software Cytoscape [93]. Nodes represent either diseases (circles) or genes (triangles), while edges indicate genetic associations. Genes associated with more than one disease are shown in yellow and lie at the center of the network. Genes associated with only one disease are displayed with the same color as their corresponding disease and organized in an outward fashion.

multiple statistically independent genome-wide effects, three of which were found under the same association peak. In 2011, the disease-associated SNP rs11804321, which is located in the last intron of the gene encoding ecotropic viral integration site 5 (*EVI5*), was shown to overlap with an insulator element that modulates the expression of the neighboring *GFI1* via the transcriptional repressor CCCTC-binding factor (CTCF) [64]. More recently, it was reported that the nonsynonymous, strongly associated exonic SNP rs11808092 in the *EVI5* gene itself induces changes in superficial hydrophobicity patterns of the coiled-coil domain of *EVI5* protein, which, in turn, affects the *EVI5* **interactome**. In particular, this work showed that

the EVI5 protein bearing the risk allele selectively interacts with sphingosine 1-phosphate lyase (SGPL1), an enzyme important for the creation of the S1P gradient, which is relevant to the adaptive immune response and the therapeutic management of MS [62]. Thus, it is conceivable that the association peak in this region truly represents multiple independent functional effects.

A specific DNA polymorphism in *TNFSF113B*, insertion-deletion GCTGT→A (BAFF-var), was recently reported to be associated with MS risk in Sardinians [65]. Whereas the frequency in Sardinians is approaching 30%, as a result of environmental pressures, BAFF-var frequency in other ancestral groups, including Europeans, ranges from <1% to 6%. *TNFSF113B* encodes the cytokine BAFF, which belongs to the TNF ligand superfamily. BAFF is expressed in B lineage cells and acts as a potent B cell differentiator and activator. Noteworthy, BAFF-var results in a shorter transcript that escapes miRNA inhibition, resulting in increased production of soluble BAFF. In addition to disease risk, BAFF-var was shown to be significantly associated with several endophenotypic traits, such as B cell and monocyte counts and total IgG, IgA, and IGM levels. The authors hypothesized that BAFF-var carriers would have a weaker response to B cell-depleting therapies due to an accelerated resurgence of memory B cells.

Additional studies with individual risk variants have been performed but current understanding of the function of each bona fide susceptibility variant in the molecular pathogenesis of MS is superficial. Moreover, since DNA variants are likely to act in concert to confer risk, studies on their interaction and on how biological pathways are affected by them are needed.

### Gene×Environment Interactions in MS

A reasonable interpretation of the incidence increase and distinctive geoprevalence distribution of MS implicates precipitating environmental or lifestyle triggers and/or the disappearance of protective environmental factors. The biological interactions between these environmental factors and genetic variance could also explain the inherent complexity of MS as a genetic trait. In recent years, several nongenetic risk factors for MS have been identified. These include vitamin D deficiency, exposure to EBV after early childhood and manifestations of infectious mononucleosis, and cigarette smoking [66,67]. Although the causality of these associations remains controversial, there is justifiable interest in formally assessing the additive interactions between exposure and the emerging MS-susceptibility DNA map [68]. These studies have been limited to HLA genes so far, with positive results reported for active and passive smoking, EBV infection and infectious mononucleosis, and childhood and/or adolescent obesity (reviewed in [5]). For example, in a pooled analysis of six data sets, smokers carrying HLA-DRB1\*15 and lacking HLA-A\*02 had a 13-fold increased disease risk compared with never-smokers without these genetic risk factors (OR 12.7, 95% CI 10.8–14.9) [69]. Novel modeling approaches will afford the analysis of gene×environment interactions genome-wide [70], potentially addressing an important knowledge gap in our understanding of MS heritability.

### From Individual Loci to Pathways and Beyond

The multiple efforts in mapping MS susceptibility loci have succeeded in identifying more than 200 independent associations. Without an elaborated analysis, the overall flavor of the associations can be assumed to be heavily immunological. Furthermore, several associated SNPs have been shown to be **expression quantitative trait loci** (eQTL) in different cells of the immune system. For example, associated variants that are eQTLs for genes such as *IFITM3*, *CD37*, and *CD6* in CD4<sup>+</sup> T cells strongly suggest that they have an effect in modulating adaptive immune responses. Furthermore, eQTL variants in or near *CLECL1*, *RGS1*, and *MERTK* in CD14<sup>+</sup> monocytes suggest that the innate immune system is also involved. Notably, variants in genes expressed in the central nervous system (CNS) are less common in these studies, thus

supporting a model in which genetic susceptibility to MS is mediated primarily by sustained dysregulation of immune responses over several decades before clinical symptoms appear.

However, as described above, the functional consequences have been described for only a handful of loci. Another way to contribute to the interpretability of genetic associations is to conduct pathway analysis. In its simplest form, pathway analysis is the search for enrichment of associated signals at or near genes with similar or complementary functions. Analogous to how large gene sets from transcriptomics data are analyzed, associated polymorphisms can be mined for their effects on genes from common predefined processes, such as gene ontologies or biological pathways. The main challenge with GWAS signals is that the affected gene(s) is not always obvious. While some analyses focus on the closest gene to the risk-associated SNP, it has now become clear that conserved motifs from promoters, enhancers, and regulatory elements can be far away from the target gene(s). Modern approaches will likely incorporate this type of information from projects such as REP [56], Encyclopedia of DNA elements (ENCODE) [71,72], and the International Human Epigenome Consortium (IHEC) [73]. Once the SNP-to-gene mappings are performed in a biologically meaningful way, enrichment of the putative genes in specific pathways can be evaluated [74,75].

Instead of performing enrichment analysis with canonical pathways (an approach that could lead to biased results, based on our current knowledge) it was proposed to merge the statistical evidence of association with the biological evidence of interaction at the protein level by virtue of a protein interaction network (PIN) [76–79]. The main question posed by this approach is whether genes known to be regulated by risk SNPs are more likely to interact among them than what would be expected by chance. A significantly higher number of interactions would indicate that those genes are likely participating in a common biological process (independently of whether that pathway was described in a database). This approach was termed ‘PIN-based pathway analysis’ (PINBPA) and a user-friendly analysis tool as been created (as an App for the software Cytoscapeii), which currently has more than 4000 users worldwide [80,81].

To provide a more accurate representation of the overall risk posed by MS susceptibility loci, we propose an approach that integrates PINBPA with cell-specific regulatory signals (Figure 3, Key Figure). Furthermore, this approach can also be used to compute personalized risk profiles, an area of much interest within the precision medicine community.

### Concluding Remarks

It is hard to argue against the contribution of common genetic variation to the susceptibility of MS. While discovery of non HLA-related loci lagged for some time, MS is one of the great successes of the GWAS era in human genetics. The impressive pace of discovery over the past 10 years highlights the need for collaboration, because the number of cases necessary to achieve statistical power is only possible in the context of international consortia.

While additional susceptibility GWAS focusing on common variants will likely be more sporadic in the future, there is still much to be done (see Outstanding Questions). As efforts on precision medicine evolve, we anticipate that the study of non-European populations will take a more prominent role. Also, studies of rare variants will likely be undertaken, either by whole-genome or exome sequencing, or targeted resequencing efforts. Similarly the contribution of genetic variation to endophenotypes has not yet been fully explored and, with the advent of wearables and other digital technology that can precisely sense and record internal variables (gait, vision, MRI, micro-movement detection, etc.), we anticipate that these studies will see much development.

Finally, the development of affordable and massively parallel sequencing technologies has enabled the exploration of genomes other than our own, giving birth to the field of

### Outstanding Questions

**Rare variants:** the role of rare variation in MS susceptibility has not yet been systematically explored. However, exome sequencing in other common diseases (T2DM, etc.) has met with only moderate success.

**Endophenotypes:** in addition to identification of susceptibility (risk) alleles, genetic association studies can also focus on the presence of modifiers of disease severity, response to therapy, or other traits that alter disease expressivity. For these studies, availability of quantitative, reproducible measurements is critical. Such endophenotypes include imaging metrics (MRI/OCT), electrophysiological parameters, biomarkers, and high-definition, computerized gait analysis, among others.

**Gene×gene and gene×environment interactions:** more-sophisticated models of interacting molecular pathways and cellular networks and their interaction with the environment will need to be developed.

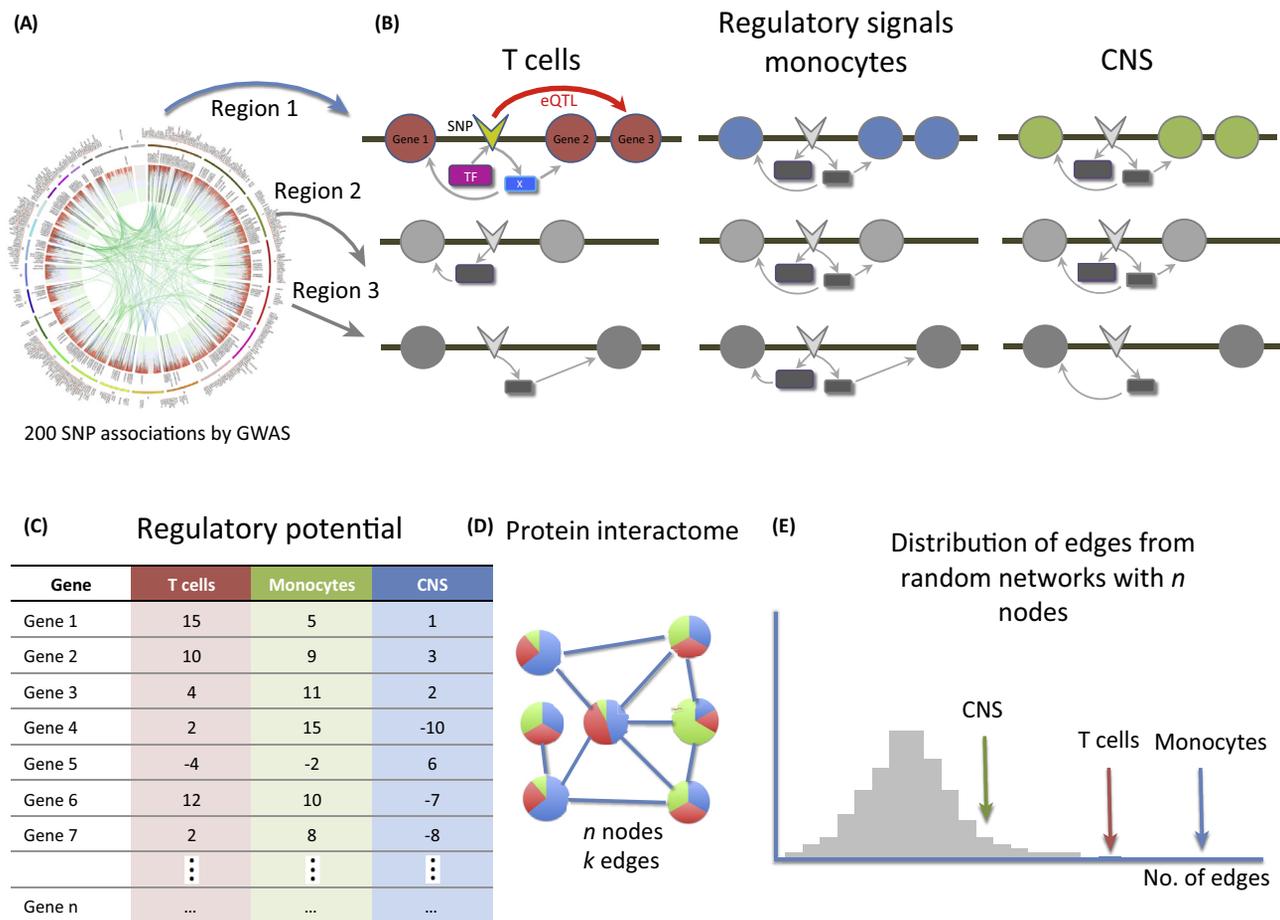
**Metagenomics:** studies of the cross-talk between the genomes of the host and those of human symbionts (bacteria, viruses, Archaea, etc.) in MS are now underway. Initial findings suggest that the presence of certain gut bacteria can influence susceptibility.

**Non-whites:** mounting on the success in Europeans, the next round of studies will likely focus on non-white populations such as African, Asian, Native Americans, Hispanics, etc. Emerging evidence suggests that, while some risk loci are conserved, ethnic-specific variants may contribute to shaping disease susceptibility in different populations [50,69–72].

**Risk prediction/precision medicine:** efforts to use risk variants either alone or in combination with other clinical or environmental variables to define an individual-specific susceptibility profile are underway.

## Key Figure

## Strategy for Cell-Specific Protein Interaction Network-Based Pathway Analysis (PINBPA)



Trends in Genetics

**Figure 3.** (A) Each of the 200 multiple sclerosis (MS)-associated single-nucleotide polymorphisms (SNPs) and those in linkage disequilibrium (LD) (regions) is queried for regulatory features, such as expression quantitative trait loci (eQTL), enhancer, DNase hypersensitivity region, histone modification, and so on, from Encyclopedia of DNA Elements (ENCODE), Roadmap Epigenomics Project (REP), and International Human Epigenome Consortium (IHEC). (B) All signals are integrated in a cell-specific manner. (C) Regulatory potentials are computed for each gene in proximity to associated SNPs for different cell types. (D) The regulatory potentials are incorporated as node attributes into a PIN and the number of edges among genes with score  $>0$  is computed. (E) The number of edges among genes for each cell type is displayed along with the distribution of edges from random networks of similar size. Abbreviations: CNS, central nervous system; GWAS, genome-wide association study.

**metagenomics.** It is expected that, in the near future, large studies will be carried out to characterize the genomes of all microscopic entities that coexist with humans (e.g. bacteria, Archaea, viruses, etc.) and their potential relationship to disease triggering or perpetuation.

In summary, the genetics architecture of MS has been largely unraveled by international, collaborative efforts. The next frontier appears to be the identification of their functional consequence, the pathways affected by these variants, and finally, the integration of multi-modal data sources into a subject's medical record, a process that could bring the promise of precision medicine in MS one step closer.

### Acknowledgments

We thank the National Institute of Health's (NIH) Institute for Neurological Diseases and Stroke (NINDS) (grants R01NS088155 and R01NS26799). S.E.B. is the Heidrich Family and Friends Endowed Chair in Neurology at UCSF. J.R.O. is the G. Zimmerman Endowed Chair in Neurology at UCSF.

### Resources

<sup>i</sup>[www.roadmappigenomics.org](http://www.roadmappigenomics.org)

<sup>ii</sup>[www.cytoscape.org](http://www.cytoscape.org)

### References

- Hauser, S.L. and Goodin, D.S. (2015) Multiple sclerosis and other demyelinating diseases. In *Harrison's Principles of Internal Medicine* (19th edn) (Longo, D.J., ed.), pp. 3395–3409, McGraw-Hill
- Compston, A. and Coles, A. (2008) Multiple sclerosis. *Lancet* 372, 1502–1517
- Hedstrom, A.K. *et al.* (2016) Environmental factors and their interactions with risk genotypes in MS susceptibility. *Curr. Opin. Neurol.* 29, 293–298
- Hedstrom, A.K. *et al.* (2016) Smoking is a major preventable risk factor for multiple sclerosis. *Mult. Scler.* 22, 1021–1026
- Olsson, T. *et al.* (2017) Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat. Rev. Neurol.* 13, 25–36
- Rhead, B. *et al.* (2016) Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. *Neurol. Genet.* 2, e97
- Sundqvist, E. *et al.* (2014) Cytomegalovirus seropositivity is negatively associated with multiple sclerosis. *Mult. Scler.* 20, 165–173
- Baxi, E.G. *et al.* (2015) Transfer of myelin-reactive th17 cells impairs endogenous remyelination in the central nervous system of cuprizone-fed mice. *J. Neurosci.* 35, 8626–8639
- Dendrou, C.A. *et al.* (2015) Immunopathology of multiple sclerosis. *Nat. Rev. Immunol.* 15, 545–558
- Devier, D.J. *et al.* (2015) Increase in NF- $\kappa$ B-sensitive miRNA-146a and miRNA-155 in multiple sclerosis (MS) and pro-inflammatory neurodegeneration. *Front. Mol. Neurosci.* 8, 5
- Yshii, L. *et al.* (2015) Neurons and T cells: understanding their interaction for inflammatory neurological diseases. *Eur. J. Immunol.* 45, 2712–2720
- Kappos, L. *et al.* (2016) The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. *Neurology* 87, 978–987
- Giovannoni, G. *et al.* (2016) Alemtuzumab improves preexisting disability in active relapsing-remitting MS patients. *Neurology* 87, 1985–1992
- Hauser, S.L. *et al.* (2017) Ocrelizumab versus interferon  $\beta$ -1a in relapsing multiple sclerosis. *N. Engl. J. Med.* 376, 221–234
- Blinkenberg, M. and Soelberg Sorensen, P. (2017) Monoclonal antibodies for relapsing multiple sclerosis: a review of recently marketed and late-stage agents. *CNS Drugs* 31, 357–371
- Orton, S.M. *et al.* (2006) Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol.* 5, 932–936
- Sellner, J. *et al.* (2011) The increasing incidence and prevalence of female multiple sclerosis – a critical analysis of potential environmental factors. *Autoimmun. Rev.* 10, 495–502
- Koch-Henriksen, N. and Sorensen, P.S. (2010) The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol.* 9, 520–532
- Eskandarieh, S. *et al.* (2016) Multiple sclerosis epidemiology in East Asia, South East Asia and South Asia: a systematic review. *Neuroepidemiology* 46, 209–221
- Benito-Leon, J. (2011) Are the prevalence and incidence of multiple sclerosis changing? *Neuroepidemiology* 36, 148–149
- Dilokthornsakul, P. *et al.* (2016) Multiple sclerosis prevalence in the United States commercially insured population. *Neurology* 86, 1014–1021
- Rosati, G. (2001) The prevalence of multiple sclerosis in the world: an update. *Neurol. Sci.* 22, 117–139
- Wallin, M.T. *et al.* (2004) Multiple sclerosis in US veterans of the Vietnam era and later military service: race, sex, and geography. *Ann. Neurol.* 55, 65–71
- Wallin, M.T. *et al.* (2012) The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service. *Brain* 135, 1778–1785
- Langer-Gould, A. *et al.* (2013) Incidence of multiple sclerosis in multiple racial and ethnic groups. *Neurology* 80, 1734–1739
- Weinstock-Guttman, B. *et al.* (2003) Multiple sclerosis characteristics in African American patients in the New York State Multiple Sclerosis Consortium. *Mult. Scler.* 9, 293–298
- Cree, B.A. *et al.* (2004) Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology* 63, 2039–2045
- Willer, C.J. *et al.* (2003) Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12877–12882
- Hansen, T. *et al.* (2005) Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. *Mult. Scler.* 11, 504–510
- Kuusisto, H. *et al.* (2008) Concordance and heritability of multiple sclerosis in Finland: study on a nationwide series of twins. *Eur. J. Neurol.* 15, 1106–1110
- Robertson, N.P. *et al.* (1996) Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain* 119, 449–455
- Westerlind, H. *et al.* (2014) Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. *Brain* 137, 770–778
- Ebers, G.C. (1983) Genetic factors in multiple sclerosis. *Neurol. Clin.* 1, 645–654
- Ebers, G.C. *et al.* (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* 377, 150–151
- Sawcer, S. (2008) The complex genetics of multiple sclerosis: pitfalls and prospects. *Brain* 131, 3118–3131
- Haines, J.L. *et al.* (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. The Multiple Sclerosis Genetics Group. *Nat. Genet.* 13, 469–471
- Ebers, G.C. *et al.* (1996) A full genome search in multiple sclerosis. *Nat. Genet.* 13, 472–476
- Sawcer, S. *et al.* (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat. Genet.* 13, 464–468
- Ligers, A. *et al.* (2001) Evidence of linkage with HLA-DR in DRB1\*15-negative families with multiple sclerosis. *Am. J. Hum. Genet.* 69, 900–903
- Sawcer, S. *et al.* (2005) A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* 77, 454–467
- International Multiple Sclerosis Genetics Consortium (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* 357, 851–862
- Jakkula, E. *et al.* (2010) 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
- Comabella, M. *et al.* (2008) 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

44. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZGene) (2009) Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 41, 824–828
45. Baranzini, S.E. *et al.* (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.* 18, 767–778
46. De Jager, P.L. *et al.* (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.* 41, 776–782
47. Consortium, I.M.S.G. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214–219
48. Beecham, A.H. *et al.* (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 45, 1353–1360
49. Patsopoulos, N. *et al.* (2017) The Multiple Sclerosis Genomic Map: Role of peripheral immune cells and resident microglia in susceptibility. *bioRxiv* 2017, 143933
50. Jannot, A.S. *et al.* (2015)  $P < 5 \times 10^{-8}$  has emerged as a standard of statistical significance for genome-wide association studies. *J. Clin. Epidemiol.* 68, 460–465
51. Panagiotou, O.A. *et al.* (2012) What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int. J. Epidemiol.* 41, 273–286
52. Gourraud, P.A. *et al.* (2011) Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. *Ann. Neurol.* 69, 65–74
53. De Jager, P.L. *et al.* (2009) Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* 8, 1111–1119
54. Isobe, N. *et al.* (2015) An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain* 138, 1518–1530
55. Langefeld, C.D. *et al.* (2017) Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* 8, 16021
56. Roadmap Epigenomics Consortium (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330
57. Maier, L.M. *et al.* (2009) IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* 5, e1000322
58. Gregory, S.G. *et al.* (2007) Interleukin 7 receptor  $\alpha$  chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat. Genet.* 39, 1083–1091
59. Galarza-Munoz, G. *et al.* (2017) Human epistatic interaction controls IL7R splicing and increases multiple sclerosis risk. *Cell* 169, 72–84
60. Gregory, A.P. *et al.* (2012) TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 488, 508–511
61. Alcina, A. *et al.* (2010) Tag-SNP analysis of the GF11-EVI5-RPL5-FAM69 risk locus for multiple sclerosis. *Eur. J. Hum. Genet.* 18, 827–831
62. Didonna, A. *et al.* (2015) A non-synonymous single-nucleotide polymorphism associated with multiple sclerosis risk affects the EVI5 interactome. *Hum. Mol. Genet.* 24, 7151–7158
63. Hoppenbrouwers, I.A. *et al.* (2008) EVI5 is a risk gene for multiple sclerosis. *Genes Immun.* 9, 334–337
64. Martin, D. *et al.* (2011) Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat. Struct. Mol. Biol.* 18, 708–714
65. Steri, M. *et al.* (2017) Overexpression of the cytokine BAFF and autoimmunity risk. *N. Engl. J. Med.* 376, 1615–1626
66. Ascherio, A. and Munger, K.L. (2016) Epidemiology of multiple sclerosis: from risk factors to prevention – an update. *Semin. Neurol.* 36, 103–114
67. Amato, M.P. *et al.* (2017) Environmental modifiable risk factors for multiple sclerosis: report from the 2016ECTRIMS focused workshop. *Mult. Scler.* Published online January 6, 2017. <http://dx.doi.org/10.1177/1352458516686847>
68. Ebers, G. (2013) Interactions of environment and genes in multiple sclerosis. *J. Neurol. Sci.* 334, 161–163
69. Hedstrom, A.K. *et al.* (2017) The interaction between smoking and HLA genes in multiple sclerosis: replication and refinement. *Eur. J. Epidemiol.* Published online June 8, 2017. <http://dx.doi.org/10.1007/s10654-017-0250-2>
70. Lu, M. *et al.* (2014) Logistic principal component analysis for rare variants in gene-environment interaction analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11, 1020–1028
71. Consortium, E.P. (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* 306, 636–640
72. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
73. Stunnenberg, H.G. *et al.* (2016) The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* 167, 1145–1149
74. Horikoshi, M. *et al.* (2016) Genome-wide associations for birth weight and correlations with adult disease. *Nature* 538, 248–252
75. Lamparter, D. *et al.* (2016) Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* 12, e1004714
76. Baranzini, S.E. *et al.* (2009) Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* 18, 2078–2090
77. Baranzini, S.E. (2009) The genetics of autoimmune diseases: a networked perspective. *Curr. Opin. Immunol.* 21, 596–605
78. Gustafsson, M. *et al.* (2015) A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. *Sci. Transl. Med.* 7, 313ra178
79. Rossin, E.J. *et al.* (2011) Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 7, e1001273
80. Wang, L. *et al.* (2015) PINBPA: cytoscape app for network analysis of GWAS data. *Bioinformatics* 31, 262–264
81. Wang, L. *et al.* (2015) iPINBPA: an integrative network-based functional module discovery tool for genome-wide association studies. *Pac. Symp. Biocomput.* 2015, 255–266
82. Campbell, R.D. and Trowsdale, J. (1993) Map of the human MHC. *Immunol. Today* 14, 349–352
83. Opelz, G. and Lenhard, V. (1983) Immunological factors influencing renal graft survival. *Annu. Rev. Med.* 34, 133–144
84. Chapman, S.J. and Hill, A.V. (2012) Human genetic susceptibility to infectious disease. *Nat. Rev. Genet.* 13, 175–188
85. Tsai, S. and Santamaria, P. (2013) MHC class II polymorphisms, autoreactive T-cells, and autoimmunity. *Front. Immunol.* 4, 321
86. Bertrams, J. *et al.* (1972) HL-A antigens and multiple sclerosis. *Tissue Antigens* 2, 405–408
87. Hollenbach, J.A. and Oksenberg, J.R. (2015) The immunogenetics of multiple sclerosis: a comprehensive review. *J. Autoimmun.* 64, 13–25
88. Barcellos, L.F. *et al.* (2003) HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am. J. Hum. Genet.* 72, 710–716
89. Isobe, N. *et al.* (2016) Association of HLA genetic risk burden with disease phenotypes in multiple sclerosis. *JAMA Neurol.* 73, 795–802
90. Patsopoulos, N.A. *et al.* (2013) Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* 9, e1003926
91. Moutsianas, L. *et al.* (2015) Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* 47, 1107–1113
92. Welter, D. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42 (Database issue), D1001–D1006
93. Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504