Title
The genetics of multiple sclerosis: an up-to-date review

Journal
Immunological Reviews, 248(1)

ISSN
0105-2896

Authors
Gourraud, Pierre-Antoine
Harbo, Hanne F
Hauser, Stephen L
et al.

Publication Date
2012-07-01

DOI
10.1111/j.1600-065x.2012.01134.x

Peer reviewed

HHS Public Access

Author manuscript

Immunol Rev. Author manuscript; available in PMC 2018 May 24.

Published in final edited form as:

Immunol Rev. 2012 July ; 248(1): 87–103. doi:10.1111/j.1600-065X.2012.01134.x.

The genetics of multiple sclerosis: An up-to-date review

Pierre-Antoine Gourraud[1], **Hanne F. Harbo**[1,2], **Stephen L. Hauser**[1], and **Sergio E. Baranzini**[1]

[1]Department of Neurology, University of California San Francisco. 513 Parnassus Ave. Room S-256. San Francisco, CA. 94143-0435'

[2]Department of Neurology, Oslo University Hospital and University of Oslo, Oslo, Norway

Abstract

Multiple sclerosis (MS) is a prevalent inflammatory disease of the central nervous system that often leads to disability in young adults. Treatment options are limited and often only partly effective. The disease is likely caused by a complex interaction between multiple genes and environmental factors, leading to inflammatory-mediated central nervous system deterioration. A series of genomic studies have confirmed a central role for the immune system in the development of MS, including genetic association studies which have now dramatically expanded the roster of MS susceptibility genes beyond the longstanding HLA association in MS first identified nearly 40 years ago. Advances in technology together with novel models for collaborative across research groups have enabled the discovery of more than 50 non-HLA genetic risk factors associated with MS. However, with a large proportion of the disease heritability still unaccounted for, current studies are now geared towards identification of causal alleles, associated pathways, epigenetic mechanisms, and gene-environment interactions. This article will review recent efforts in addressing the genetics of MS and the challenges posed by an ever increasing amount of analyzable data which is spearheading development of novel statistical methods necessary to cope with such complexity.

A brief history of MS genetics

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS) that leads to demyelination and axonal degeneration and accompanying neurological impairment and disability. MS is a prevalent neurological disease in the northern hemisphere, where prevalence rates of 0.5–1.5 per 1,000 inhabitants are reported (1). The genetic contribution to the susceptibility of developing multiple sclerosis (MS) is undeniable (2–6). While the largest effect is located in the HLA class II region (discussed in more detail below), additional independent risk loci within the HLA class I region have been also recently confirmed (7, 8). The mechanisms for how these HLA-alleles affect MS susceptibility is, however, not clear. Furthermore, the implicated HLA-associated alleles are neither necessary nor sufficient to cause or predict the development of MS, thus other factors must also contribute to the disease.

Corresponding author: Stephen L. Hauser, hausers@neurology.ucsf.edu.

The search for non-HLA genetic factors in MS started in the early 1970s by several groups who performed small sized candidate gene studies, typically analyzing a few hundred cases and controls. While several positive and negative associations were reported, no associations were convincingly replicated for many years. In the early 1980s, the success of linkage studies using microsatellite markers in the identification of genes for monogenetic disorders in the early 1980s (9), fueled its application to complex disorders like MS (10–13). Unfortunately, those studies, employing multiple-affected member (multiplex) MS families, only confirmed linkage to the HLA region at chromosome 6p.21, and it soon became clear that linkage analysis had insufficient statistical power to identify risk loci in complex genetic disorders, where several genes with modest effects may contribute to the disease. This fact was emphasized in 2005 when a most ambitious genome-wide microsatellite screen, employing the largest multinational collection of MS families collected until then (n=730), also failed to identify linkage outside of the HLA region (14).

Subsequent advances in technology, including the development of novel chemistries, miniaturization, and automation made possible the effective screening of thousands of single nucleotide polymorphisms (SNPs) in thousands of samples at an affordable price. These technologies ushered in a new era of more fruitful genome wide association studies (GWAS), as these were theoretically capable of detecting susceptibility genes of modest effects in multifactorial, complex genetic disorders such as MS. In contrast to previous candidate gene studies, the GWAS strategy is hypothesis-free, and aims to screen the whole genome by tagging linkage disequilibrium (LD) blocks. In contrast to linkage studies, GWAS utilizes readily available case-control datasets rather than multiplex family-based sets, permitting collection of much larger datasets. Compared with linkage, GWAS has less power but better resolution; the former can be overcome by testing datasets of sufficient size and controlling for any population stratification (differences in genetic structure between disease and control groups). With all of these approaches, there is an additional challenge that any associated marker may not itself be the causal variant but is in linkage disequilibrium (a non-random statistical association of the variants due to physical linkage on the chromosomes) with the causal variant. Therefore, subsequent fine-mapping studies of the associated genetic region and studies of the functional relevance of the genetic variant are typically needed to validate initial findings of association.

The GWAS strategy relies on the common disease-common variant hypothesis (CD-CV), by searching for statistical differences between cases and controls for relatively “common” variants, i.e. alleles that are relatively frequent in the population (> 5–10% frequency in control chromosomes) (15, 16). Under this scenario, because many statistical tests are performed in parallel, the actual level of significance needs to be corrected, and typically only p-values beyond $10^{-7} – 10^{-8}$ are considered statistically significant. In this context, only when using a large enough sample collection (i.e. thousands of cases and controls), is the power to detect significant allele frequency differences between cases and controls adequate (17). Indeed, in 2007 the first MS GWAS identified the first non-HLA regions with genome-wide significance ($p< 10^{-8}$) using a two-step approach (18). After screening of 931 family trios samples using 334,923 SNPs and replication in 609 family trios, 2322 case subjects and 789 control subjects, the genes encoding the IL7Rα and IL2Rα were found to be significantly associated with relatively modest odds ratios (OR) (< 1.35). Simultaneously,

the IL7Rα gene was confirmed to be associated in other MS cohorts (19, 20). In the years that followed, a new series of GWAS and meta-analyses were performed in different MS cohorts, steadily adding more regions to the list of confirmed MS associated loci, adding up to 26 by the summer of 2011 (21–29). However, it soon became clear that most of the early MS GWAS studies, as well as series of GWAS studies in other multifactorial diseases, were hampered by inadequate samples sizes, in the same way than most early candidate gene and linkage studies were similarly underpowered. Some power estimates showed that to reach the genome wide significance level for risk loci with OR as modest as 1.2 or less (as expected from earlier studies) the sample size should ideally include approximately 10,000 cases and controls (17). Those estimates may also suggest that GWAS are also intrinsically biased to reveal common variants rather than rarer variants that could have larger effect sizes.

The largest GWAS in MS required International collaboration

To achieve this number of MS samples as well as the funding for such an experiment, the International MS Genetic Consortium (IMSGC) set out to perform a GWAS study in close to 10,000 MS cases and 20,000 controls in collaboration with the Welcome Trust Case Control Consortium 2 (WTCCC2) (6). From an analytical standpoint, the computational resources and expertise needed to analyze a dataset of this size (more than 20 billion data points) are far from trivial. In addition to sample quality control, the analytical pipeline needed to consider potential sources of bias related to differential genotyping success rates, Hardy-Weinberg equilibrium, and population stratification before the final analysis was conducted. Twenty-two centers spanning Europe, North America and Oceania participated in this study thus creating the challenge to effectively control for genetic heterogeneity in the populations under study. Although all individuals were of European ancestry, the inclusion of Scandinavian, British, German, French, Spanish, and Italian samples among others resulted in a non-negligible degree of genetic heterogeneity that needed to be accounted for. The main parameter that reflects underlying population structure (in addition to the polygenic architecture of MS) is the genomic inflation factor (λ), a statistical index reflecting the inflation of the test statistic attributable to population stratification (λ is estimated using markers that are not linked with the trait in question, and can be used to correct for the population stratification effect). While little evidence for inflation was observed when analysis was restricted to the UK samples alone (as expected), substantial inflation was apparent when the rest of the data were included in the analysis, thus prompting additional efforts towards minimizing λ. This was achieved by computing a linear mixed model in which the covariance of the pair-wise correlation matrix over all individuals and genotypes was included as a term in the equation. Computing pair-wise correlations of hundreds of thousands of SNPs in tens of thousands of individuals is challenging, even for modern and powerful computer clusters. To cope with this hurdle, the entire dataset was split in two and merged at the end. This approach successfully reduced the inflation factor to close to 1, thus virtually eliminating any spurious effects due to population structure.

In all, this latest screen identified 52 MS risk loci with genome wide significance (Table 1), of which 29 were novel (6). All loci except HLA showed modest OR in the range 1.1–1.3, as could be expected from previous studies (21–29). Approximately, half of the associated

variants are common in Europeans (Figure 1). Remarkably, the vast majority of these confirmed MS associated loci are located close or inside genes encoding immune related molecules, strongly supporting the hypothesis that MS is primarily an immune mediated disease (Figure 2). These recent studies also underscored that more than one SNP may be needed to account for the risk of the same region. As an example 2 SNPs (rs3118470 rs7090512) are located 9 Kb from one another in the IL2RA gene and independently contribute to MS risk. Furthermore, allelic heterogeneity has been described at this locus, in which the same allele that predisposes to MS protects to type 1 diabetes (30). Of interest, one third of the identified loci were reportedly associated with at least one other autoimmune disease, strengthening the notion that common disease mechanism(s) may underlie most, if not all autoimmune diseases. (31, 32). Available data suggests that a core of genes is shared among multiple autoimmune diseases, and some are disease specific. Of note, these studies take into account the possibility of gene sharing by chance, and present convincing statistical evidence to support the fact that the number of shared susceptibility genes is significantly higher than what it would be expected by chance.

Of note, CYP27B1, a gene associated with vitamin D metabolism was among the new associations identified by this paper. Previous work had suggested that vitamin D regulates >80% of MS-associated genes (33) -including the HLA-DRB1*15:01 allele (34). A recent article described the whole exome sequencing (i.e. sequencing of all protein-coding regions of the genome) on 43 probands from MS families with multiple affected individuals, and reported that a non-synonymous variant in CYP27B1 segregating in one family in a dominant fashion with incomplete penetrance (35). The variant causes an arginine to histidine change at position 389 of the protein (R389H) and leads to complete loss of enzyme activity of CYP27B1, which converts 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D, the biologically active form of vitamin D (calcitriol).

Although next generation sequencing has been applied before to the study of MS genomes (36), the work described above is the first report that identifies mutations likely involved in MS pathogenesis. The analytical pipeline required to analyze these datasets is not a minor aspect. On average, each exome is composed of 3.7×10^9 nucleotides of sequence, thus covering the coding genome with about 70-fold redundancy. Once fluorescence intensities are obtained for each short read (approx. 52.8 million), they are translated into sequence (i.e. base-calling) by the manufacturer's software. The next step was to align each of the 53 million reads to the human reference genome and determine the differences. On average, 58,400 SNPs and 5,060 insertion/deletion polymorphisms (indels) were called per individual, of which 58% and 14% were already annotated in public databases, respectively.

Subsequent genotyping of this variant in more than 3000 parent-affected child trios, resulted in its identification in 19 parents, all of whom transmitted the variant to their affected offspring. These findings strongly suggest that that lower levels of calcitriol (as a result of CYP27B1 mutations) lead to a disruption to critical gene–environment interactions important for the developing immune or nervous system, predisposing to MS.

HLA associations in MS

The discovery of HLA association in MS

The first reports of general associations between genetic markers and autoimmune diseases were published in the 1960s, and identified the human leucocyte antigens (HLA) locus as a critical region. Susceptibility to multiple diseases like diabetes type I, rheumatoid arthritis and ankylosing spondylitis were mapped to this region (37–39). The HLA genes are located in the 3.5 Mb major histocompatibility complex (MHC) region on the short arm of chromosome 6 (6p21), consisting of HLA class I, II and III genes. HLA genes encode highly polymorphic cell surface glycoproteins, several of which play a crucial role in the self - and non-self antigen recognition by the immune system. While different autoimmune diseases show association to a diverse array of HLA loci and alleles (39), some degree of overlap was detected even in these early studies, and later confirmed by larger and more accurate mapping.

The earliest associations between MS and HLA were described for the class I alleles A3 (3) and B7 (2) in the 1970s using serological based measurements. The association with HLA-A3 was reported to be secondary to HLA-B7, which in turn was shown to be secondary to a suggested primary association with class II antigens HLA-DR2 and -DQw6 (40–42). It was originally suggested that this indirect association of HLA class I genes was due to linkage disequilibrium (LD) within the HLA region, a puzzle that would take several more decades to clarify. With the development of genomic genotyping, better resolution and a detailed, accurate nomenclature of the alleles in the very gene-rich and extraordinarily polymorphic MCH region was finally established (http://hla.alleles.org/). Thus, the association of the HLA class II haplotype DRB1*15:01, DQA1*01:02, DQB1*06:02 in MS (genotypes corresponding to the serologic alleles HLA-DR2, DQ6) was then firmly established [24]. The HLA DRB1*15:01 haplotype is carried by 28–33% of Northern Caucasian MS patients compared to 9–15% of healthy controls, corresponding to an average OR of 3.08 (6), making this the single strongest susceptibility locus in MS genomewide (Table 2).

HLA class II allele associations and HLA-DR heterogeneity

Due to the strong LD between alleles on the HLA DRB1*15:01-haplotype it has been difficult to identify which of the class II alleles cause the primarily association in MS. Some studies have supported a primary association to DQB1*06:02 (43–46), but it wasn’t until studies were performed in African Americans populations that strong evidence supporting the selective association with HLA-DRB1*15 (encompassing HLA-DRB1*15:01 and HLA-DRB1*15:03 in this population) was presented, indicating a primary role for the DRB1 locus in MS independent of DQB1*06:02 (47, 48). The major risk MS loci in African Americans was found to be HLA-DRB1*15:03, highly related to the *15:01 allele of whites. Interestingly, the HLA-DRB5 gene, located at the telomeric boundary of the HLA class II region, and only expressed by HLA-DRB1*15 haplotypes is reported to attenuate MS severity (49) (Table 2). Interestingly, HLA-DRB1*15:01 has a low frequency even among healthy individuals in the indigenous Sami (natives of the northern part of the Scandinavian countries) which coincides with a low MS prevalence, suggesting a general reduced genetic risk for MS in this population (50, 51).

A high degree of allelic heterogeneity of the HLA-DRB1 locus in Caucasians has more recently been described, showing that variation in this region influences MS susceptibility in a complex manner (5) (Table 2) and highlighting the need for in-depth follow-up studies that help refine these categories. A dominant HLA-DRB1*15 dose effect has been identified as well as a modest recessive dose effect for the HLA-DRB1*03 allele (i.e. two copies of either of these susceptibility haplotypes further increases disease risk) [34]. Also a high risk genotype composed of HLA-DRB1*15/HLA-DRB1*08 alleles has been identified (OR 7.7), thus suggesting a trans-HLA-DRB1 allelic interaction in MS (5). In the large combined multinational cohort of the recent MS IMSGC GWAS study, the HLA-DRB1*13:03 allele was also identified as being associated with MS (OR 2.43). Furthermore, HLA-DRB1*01:08 (OR =1.18) and HLA-DRB1*03:01 (which is strongly linked to HLA-DQB1*02:01; OR = 1.26) showed significant associations. Finally, evidence of an additive effect for each additional allele was recently described (6). In the Sardinia region of Italy, where MS prevalence is high, HLA-DRB1*04, *03:01 and HLA-DRB1*13:01 (in addition to HLA-DRB1*15:01) have shown positive associations to MS (52–57).

Extended MHC region associations in MS and epistasis

In recent years there has been a growing body of evidence suggesting that other HLA loci outside the HLA class II region can modulate MS susceptibility (7, 8, 56, 58–61). It is now well established that HLA- A*02 has an independent protective effect (OR 0.73) [25], and that both HLA-Cw*5 (60) and HLA-B*44 (62) as well as their combination are protective (63). Whether these effects are independent remain unknown due to the high frequency of the HLA-A*02:01 alleles and that of the common HLA-A*02:01~ HLA-B*44:02 ~ HLA-C*05:01 ~ HLA-DRB1*04:01 haplotype. Interestingly, the corresponding HLA-Bw*04 is a ligand for the inhibitory KIR3DL1 receptor, found to protect against MS in an HLA-DRB1-“independent” manner (64).

Recently, MHC-wide single-nucleotide polymorphism (SNP) studies have made it possible to further explore associations in the extended MCH region. In a study of white Americans In which 958 SNPs spanning the MHC region were typed, a polymorphism located in the non-classical HLA-G gene was shown to be independently associated with MS (65). A Japanese population genotyped for 3534 SNPs in the MHC region showed independent associations to both an HLA class III marker in the NOTCH4 gene and to an HLA class II SNP. Interestingly, both HLA class II and an HLA class I SNPs in the NRM gene were associated with MS susceptibility in aquaporin-4 negative (AQP4-) subjects. Haplotype analysis in AQP4- MS further revealed a large association which included the HLA-DRB1*04 haplotype, but excluded HLA-DRB1*15:01[47]. More recently, SNP typing of the MHC region has also been used for imputation of alleles as defined by classical genomic HLA typing. While useful when detailed typing is not available, HLA imputation is an indirect method, which cannot replace classical HLA typing of HLA alleles and haplotypes to completely assess their functional relevance in MS.

Haplotype analyses indicate that there might be epistatic effects between associated co-dominantly expressed HLA alleles in MS (66). This is also supported by animal studies. By using a functional assay in humanized mice that characterized two MS-associated HLA-DR

alleles at separate loci, LD between the two alleles could be explained by a functional epistatic interaction (67). One allele modified the T-cell response activated by the second allele through activation-induced cell death (AICD), an indication that epistatic interaction might be a mechanism for modifying immune responses. Interplay between both predisposing and protective alleles and their encoded molecules will have impact on the immune response and thereby on the phenotypic expression in the MS patients.

HLA association and MS subtypes

A series of studies have explored phenotype-genotype correlation for associated HLA alleles in MS with somewhat conflicting results, probably due to a modest statistical power in many of the reported studies. The best-powered study is the multinational MS GWAS, which used linear mixed models with HLA-tagging SNPs (see supplement to (6)) in up to 8,715 MS patients. A significant (indirect) correlation was found between age at onset of MS and the HLA-DRB1*15:01 allele (through a SNP in high LD with this allele). Each copy of the minor allele decreased the age at onset by 10.6 months. No other locus showed genome wide significance with age at onset. No robust evidence was found for genetic association with gender, clinical course (including MS subtype), disease severity or month of birth neither for HLA- nor non-HLA markers.

The association of HLA-DRB1*1501 with age at onset has also previously been reported in other, smaller studies, some of which also indicated that this effect was most pronounced in females (68–71). Some studies reported increased disability progression in HLA-DRB1*1501-positive MS subjects [54], but others pointed to HLA-DRB1*01 and HLA-DRB1*04 alleles (72). Both carriage of HLA-DRB1*15 and the presence of oligoclonal bands (OCB) in the cerebrospinal fluid (CSF) have been reported to hasten disease progression (73). In African American MS patents, subjects with HLA-DRB1*15 alleles were twice as likely to develop typical MS (as opposed to optico-spinal MS), and HLA-DRB1*15 alleles were associated with a 2.1-year earlier age at onset (74). Several studies have also examined whether specific HLA alleles are correlated with outcome as measured by brain MRI examination. HLA-DRB1*1501 has been shown to increase disease severity in MS by facilitating the development of more T2-foci, leading to a decline in brain volume and an impact on cognitive performance (75). HLA B*44 may moderate disease course, preserving brain volume and reducing the burden of T2 hyperintense lesions in subjects with MS (62).

In conclusion, available evidence suggests that that clinical phenotype in MS correlates only modestly with the strongest genetic risk factor, i.e. HLA alleles, illustrating the challenge of translating genetics of susceptibility into clinically meaningful associations. This is in contrast to some other autoimmune diseases, for instance myasthenia gravis, in which the HLA-associated loci and alleles clearly differ in patients with early vs. late disease onset (76). However, a well-designed genetic study of fully characterized MS sub-phenotypes in MS has not to date been undertaken. Identification of genetic effects on disease expression, including response to therapy, will require the standardized, prospective collection of clinical, imaging, and immunologic data from adequately sized cohorts.

Assessing the genetic burden of MS

By the end of 2011, the number of genomic regions associated with MS susceptibility was more than 50. With additional GWAS currently being performed, and meta-analysis compiling an ever growing number of studies, the count of MS associated variants will most likely surpass 100 in 2013. Furthermore, work on the well-established MHC genomic region suggest that assessment of additional associations will require more elaborate models.

The discovery of additional susceptibility genes highlights the need for summary metrics of the disease-specific genetic assessment of patients. While previous genetic discoveries were necessarily derived from population based studies, it is now possible to focus on single patients by computing summaries of their individual genetic risk factors.

In 2010, our group published a study that combined genotype information of the 16 non-HLA replicated MS susceptibility variants known at the time into one genetic score. The score (the Multiple Sclerosis Genetic Burden, MSGB) summarized how a large number of genetic markers confers susceptibility to MS patients and their families (77). The MSGB is based on an algorithm that incorporates each risk variant for a given individual and weighs each SNP according to its reported effect size. Studying over 1,200 families, we could confirm the hypothesis that MS patients from multi-case families carry a higher MS genetic burden than patients from single-case families or controls. Likewise, parents of probands from multi-case families exhibited a higher MSGB than those from sporadic families.

Interestingly, we were not able to detect any departure from the assumption of independent contribution of gender, HLA and non-MHC SNPs to the MSGB. While most of similar genetic risk score approaches seek to predict development of a complex disease in a population, the MSGB is not predictive of disease status, even within families whose members are likely exposed to similar environmental influences, thus underscoring the high variability of genetic contribution and indicating that our current knowledge of the interplay of genetic factors in MS is still incomplete. Although finding that a low number of markers cannot deliver an accurate prediction of MS course even within families of MS patients was not surprising, an updated MSGB computation that accounts for all newly identified genetic associations with MS and rare genetic variants, and may include high-risk environmental exposures (vitamin D, smoking, viral exposure) that will ultimately facilitate translational application.

In addition, the score can be used as a summary of an individual’s genetic load (Figure 3). For example, this score was used to provide evidence that the relapsing from onset and primary progressive forms of MS share the same genetic architecture (77). We anticipate that the MSGB scoring method will also facilitate the integration of genetics with other metrics acquired in MS clinics and imaging centers. For example, patients with severe disease were found to have a slightly higher genetic load; highly loaded patients have an earlier age of onset (in line with the previously mentioned HLA association); highly loaded patients have more oligoclonal bands; and in patients who experience an initial attack of demyelination (a.k.a. a “clinically isolated syndrome (CIS)”, often a precursor to MS), those

with a higher genetic load tend to convert to chronic MS more rapidly than do those with lower scores (unpublished observations).

Functional studies and the importance of the environment

Association and linkage studies are useful strategies for the identification of genetic risk loci, but they need to be interpreted appropriately. For example and as noted above, neither the variant that gives the most significant association signal nor the one with largest effect size in a GWAS, might be the one that actually causes the phenotype. Similarly, several untyped and potentially unknown variants in close physical proximity could represent independent signals, but could also represent the same information by virtue of being in LD. Follow-up studies are typically necessary to provide a more detailed picture of the make-up of genetic risk loci identified in GWAS.

A detailed study of the CIITA-CLEC16A-SOCS1 region was recently carried out to explore the nature of loci with multiple association signals. Three genes in this region (DEXI, CLEC16A, and SOCS1) showed correlated expression in three lymphoblastoid cell lines, a finding supported by chromatin immunoprecipitation (ChIP) experiments which demonstrated an active chromatin state spanning those genes. Several SNPs in this region could be applied as quantitative trait loci (eQTLs) associated with the expression of nearby genes (78).

Another study interrogated the IL2RA locus, within which SNP rs2104286 has been associated with MS (79). Interestingly, this SNP was not the variant that most accurately predicted the risk of MS. Indeed, by targeted genotyping of 161 SNPs in over 500 families and almost 250 sib pairs the strongest association of the phenotype was with a different pair of SNPs (rs2256774 and rs3118470), both of which correlated with IL2RA expression (79). This study highlights the need for fine mapping of GWAS signals.

A recent study investigated the effect of the established MS risk allele CD6 ($rs17824933^{G}$) on immune function (80). Using T cells from healthy controls, individuals carrying the risk allele expressed less CD6 in its long form and higher levels of a spliced version of the molecule, lacking exon 5. Functional assays in $CD4^{+}$ T cells showed that the alternatively spliced CD6 is deficient in transmitting proliferative signals upon stimulation with CD6. Given that activated T lymphocytes that infiltrate the CNS play a major role in MS pathology, it seems counterintuitive that carriers of the CD6 risk allele would show reduced activation of $CD4^{+}$ T cells. However, the authors suggest that the phenotype might result from the reduced activation of regulatory $CD4^{+}$ T cells, which suppress excessive T lymphocyte activation (80).

The function of another established MS risk variant, TYK2 ($rs34536443^{G}$), was recently examined (81). This variant causes an amino acid substitution in the protein that, while not affecting gene expression levels in T cells, it results in enhanced activation of the kinase function upon stimulation with IFN-beta, IL-6 and IL-10. The lower kinase activity in carriers of the alternative, protective C allele results in increased release of cytokines involved in the polarization of T lymphocytes into Th2 cells. Hence, protection against MS

might be conferred by disfavoring differentiation of activated T cells into Th1 cells, critical drivers of the autoimmune process in MS (81).

Another environmental factor (in addition to Vitamin D reported above) shown to increase the risk for MS is smoking (82–84). Recently, Hedström et al. assessed interactions between smoking and the genetic risk factors HLA-DRB1*15:01 and HLA-A in a Swedish case-control study (85). Interestingly, a significant interaction of two loci, the presence of the HLA-DRB1*15:01 risk allele and absence of the protective HLA-A*02 allele, increased the risk of developing MS only in smokers by a ratio of 13.5. Non-smokers with the same genotype showed an odds ratio of 4.9, but no interaction was detected between the two genetic loci. This finding let the authors hypothesize that smoking-induced irritations or infections in the lung might boost autoreactive immune responses, especially in genetically predisposed individuals (85). However, a number of statistical shortcomings were later identified that casted doubts about the validity of that study (86). For example, this study obtained smoking data in 70% of cases but in only 40% of controls, an issue that may represent an important bias. Furthermore, questions were raised about the validity of using a linear model to test whether HLA-DRB1*15, HLA-A*X (i.e. not A*02) and smoking are independent causes that determine MS susceptibility. Thus, even though many studies have now shown that smoking affects MS susceptibility (87), there are still many practical and statistical challenges for analyses of how genes and smoking or other environmental factors potentially interact to confer MS risk.

In contrast to proteins and nucleic acids, production of complex carbohydrates is not template driven. Rather, *N*-glycan biosynthesis depends on the nutrient environment of the cell, metabolic supply of substrates, and enzymatic activities of the ER/Golgi enzymes. Cell surface *N*-glycans serve as ligands for a number of carbohydrate-binding protein families, including galectins, siglecs, and selectins, all of which play important roles in immunity. A recent study explored the role of N-glycosylation and its dependence on genetic and environmental factors in MS (88). Golgi N-glycosylation may play a role in MS since mice lacking the Golgi glycosylating proteins Mgat1 and/or Mgat5 develop spontaneous inflammatory demyelination (89). MGAT1 and MGAT5 regulate the extent of N-glycan branching on glycosylated proteins; in activated T cells, highly glycosylated CTLA-4 on the cell surface inhibits cell proliferation. This study was the first to establish a connection between two known MS risk variants (IL2RA and IL7RA), and the extent of N-glycan branching in vitro by showing that in resting T cells, IL-2 and IL-7 reduce branching, while the opposite was observed in T cell blasts upon activation of TCR signaling. Authors of this study next assessed whether further genetic variants in other members of the N-glycosylation pathway (MGAT1 and CTLA-4) confer additional susceptibility to MS in a targeted association study. They found that one particular MGAT1 haplotype (rs7726005, rs2070924, rs2070925) was associated with MS in two Caucasian MS case-control cohorts. Furthermore, they also reported that cells from individuals with this haplotype showed a 20% reduction in N-glycan branching. Under specific conditions, an interaction of the MGAT1 haplotype with the IL2RA and IL7RA risk alleles as well as with CTLA-4 was also observed. The latter (but not the former) interaction was replicated in an African-American cohort. Since only a selection of variants was tested in these association studies, more significant p-values were obtained than would be in a GWAS, typically subjected to strict

correction for multiple testing, underscoring the power of the presented biology-driven approach. In addition, an interaction of the IL2RA and IL7RA risk variants and vitamin D was observed, thus reinforcing the association between this sunlight-activated vitamin and development of MS. VitD3 induced MAGT1 expression in T cell blasts and enhanced branching in cells with two or more copies of the risk alleles (but not in cells with the protective alleles).

Biomarkers in MS

Stimulated by the widespread availability of high-throughput technologies, advances in statistical methods of analysis, and a growing consensus that individual genetic variation likely influences therapeutic responses, the search for biomarkers in MS (as well as in other diseases) is a very active field of research. Indeed, only 122 articles included the term “biomarker” in the title in 2001, while more than 1200 fulfilled that criterion in 2011. This, in part, highlights the enormous interest that biomarker research has stimulated, particularly in light of the new paradigm of personalized medicine (90, 91).

Interest in biomarker development is focused on two main areas: disease progression and therapeutic response to disease-modifying therapies (DMTs). Identification of a predictive biomarker of disease progression (particularly when measured early in the disease process) would have an enormous impact on the decision-making of physicians treating patients with MS. For example, DMTs may not be a priority for a patient with clinically isolated syndrome (CIS) who is predicted to have a very slow progression over the next few years. On the other hand, if a biomarker predicts with high accuracy that a patient is very likely to convert to clinically definite MS within months, an early start of a first-line DMT regimen might be the appropriate course of action. This hypothetical scenario is supported by recent work in which CD4+ T cells from patients with CIS shortly after diagnosis were subjected to transcriptional profiling and statistical analysis (92). The expression signature from one quarter of the patients was clearly distinct from the rest, and this was highly predictive of their future disease progression, with 92% of these subjects converting to clinically definite MS within 9 months, while only 20% of the subjects with a different signature converted during the same period of time. Genes in that signature were enriched in key functions such as cell cycle and T-cell activation. Interestingly, gadolinium enhancing lesions determined by magnetic resonance imaging were present in only 58% of the patients predicted to be at high risk by the expression signature, indicating – potentially - a possible higher sensitivity of transcriptional profiling compared to a single MRI scan to predict disease activity.

In another study with similar design, pre-existing RNA signatures in CD8+ T cells were used to predict risk of progression in antineutrophil cytoplasmic antibody–associated vasculitis and in systemic lupus erythematosus (93). Investigators in this study showed that the subset of genes defining the poor prognosis group was enriched for genes involved in the interleukin-7 receptor (IL-7R) pathway and T-cell receptor (TCR) signaling and those expressed by memory T cells. Furthermore, the poor prognosis group tended to show an expanded CD8+ T-cell memory population.

The second avenue by which biomarker research is likely to impact MS care is in the prediction of therapeutic response. This is an area of intense research and represents a shift in the previously widespread notion (particularly in the pharmaceutical industry) that all patients are equally likely to respond to a given drug. Recent data on the architecture of the human genome have unequivocally shown that there is abundant genetic variation between individuals that could account for differential ADME profiles, drug metabolism, transport, safety, and ultimately, response to pharmacological agents (94, 95).

Several groups have attempted to identify biomarkers of therapeutic response using transcriptional profiling. In 2005 our group reported that certain combinations of blood transcripts were highly predictive of the response to interferon (IFN)-beta treatment (96). Subsequently, Comabella and colleagues reported that perturbations of the type I IFN signaling pathway in monocytes are related to a lack of response to IFN-beta, and type I IFN-regulated genes may be used as response markers in IFN-beta treatment. Another study from the same group identified a subpopulation of natural killer cells (CD56bright) that was predictive of an IFN-beta response (97). This is of interest in relation to an earlier study in which expansion of CD56bright cells were related to the positive response exerted by daclizumab in MS patients (98). More recently, the involvement of the TLR4 and type I IFN signaling pathways in the response to IFN-beta was investigated with intriguing results (99). Here, peripheral blood mononuclear cells from non-responders were characterized by increased baseline expression levels of endogenous IFN-beta and elevated IFN receptor 1 (IFNAR1) expression in monocytes. Furthermore, the capacity of IFN-beta to induce its own expression was deficient in cells from non-responders. Finally, baseline expression of a negative regulator of TLR4 signaling (IRAK3) was decreased in IFN-beta responders.

Other studies have attempted to identify biomarkers of IFN-beta response (100–111). While these approaches are highly promising, a number of challenges must still be overcome for this field to fully develop its potential. The first challenge relates to sample sizes in reported studies. In most cases, studies are significantly underpowered and lack replication, thus any reported results remain preliminary. Part of this is due to only modest sample availability from a single (or a few) research center(s). In this regard, samples collected during industry-sponsored phase III clinical trials would be ideally suited for these studies. Still, regulatory, ethical, and operational constraints that currently restrict availability of these samples for pharmacogenomics research will need to be addressed (112). Another challenge is related to the biological source for testing. An ideal biomarker would be tested from a minimally invasive procedure (e.g., blood draw) and would require little processing so as to allow quick and universal determinations. However, a trade-off exists between the informativeness of unprocessed (e.g., whole blood) vs. processed (e.g., cell type separation) samples. While whole blood can be tested faster and more cheaply, the signal to noise ratio of any measurement is typically compromised by the complexity of this fluid, which contains thousands of proteins, sugars, and lipids in largely variable concentrations spanning several orders of magnitude. On the other hand, time and resources invested in sample manipulation to obtain pure cell subpopulations can result in more robust signatures that potentially offset small, heterogeneous sample sizes (113).

The heterogeneity and biological complexity of diseases like MS may make the identification of single predictive biomarkers an unattainable goal. A more likely scenario will be the utilization of advanced computational methods such as supervised classification (e.g., neural networks, Bayesian classifiers) that combine different types of information (clinical, imaging, genotype, molecular, etc.) into a single model that maximizes predictive ability and minimizes misclassifications (114). The application of such tools in the integrative context of systems biology approaches will become a critical step in the development of useful biomarkers for MS.

The role of data integration and systems biology in biomarker development

Unfortunately, most biomarker studies have not been yet validated and therefore, they are not clinically useful at present. The lack of validation is a common problem with biomarkers of complex diseases. This may reflect a bias in statistical analysis or a lack of available data, but it may also indicate difficulties in performing clinical validation studies (115, 116). Moreover, in addition to the need for validation of such candidates, integration of the information provided by each biomarker is also needed for a comprehensive representation of the disease process. Biomarker discovery would be greatly enhanced by applying systems biology principles such as the multi-scale integration of information and the analysis of dynamic patterns with the help of computational tools.

Systems biology aims to increase understanding of biological and medical problems from a holistic perspective by considering all processes involved and their dynamics (117, 118). The application of systems biology to biomarker discovery implies the integration of the molecular data generated by -omics studies within models of disease pathogenesis, signaling pathways, and biological networks, ideally arriving at a physiological explanation of the findings and identifying how they are related across the different levels of biological complexity (genes, molecules, cells, tissues, and the organisms) and to the clinical phenotype (119). For this reason, the biomarker discovery process (searching for molecules associated with a given disease phenotype) is only the first step. It will ideally be followed by validation of a set of biological pathways related to disease pathogenesis and associated with the disease course at the clinical level.

Determining whether a given biomarker is causal or reactive to a specific disease is likely to provide information on the underlying pathogenic process and ultimately impact its general applicability. In recent years, the availability of high-throughput datasets from a variety of biological sources has prompted the creation of a multitude of databases that significantly facilitate biomedical research. However, these datasets are frequently analyzed only in the context of similar data types (e.g., gene expression datasets are usually analyzed and compared to each other but rarely to protein abundance studies), and they remain in their own silos. Integration of these vast but heterogeneous sources of information will allow the systematic and comprehensive analysis of molecular and clinical datasets, spanning hundreds of dimensions and thousands of individuals. This integration is essential to capitalize on the value of current and future molecular- and cellular-level data on humans to gain novel insights about health and disease.

In recent years, network biology has emerged as a powerful paradigm to visualize and analyze large data ensembles in novel ways with unparalleled flexibility (120). More recent applications of this approach have enabled a detailed look at the genetic landscape of complex human phenotypes (121–124). Overall, the availability of large-scale datasets has prompted efforts to integrate data, with the ultimate goal of providing systematic insights into complex traits. For example, identification of genetic similarities among complex diseases, particularly autoimmune diseases, is a topic of intense research (32, 125, 126). With the goal of bridging this gap in analytical tools that allow integrating data from disparate sources, a novel approach was recently developed (127). Through this tool (called iCTNet), multiple relationships between diseases, genes, proteins, organs and therapeutic drugs that were joined into a common dedicated database can be visualized together in a network environment (using Cytoscape, a popular software for network visualization). With iCTNet, users can download genetic associations for more than 200 complex traits and display shared risk factors for any given set of traits. This exercise allows identification of hidden patterns, such as the extent of the genetic similarity between MS and other autoimmune diseases, and the number of MS-associated genes expressed in the brain. Furthermore, it allows the contextualization of therapeutic drugs and their molecular targets together with the diseases they are indicated for, thus enabling researchers to identify potential off-label use of drugs in a rational and controlled fashion.

Integrating genomic research data into a single analytical framework will unify medical practice with cutting-edge genetic research, and pave the way to inform and personalize therapeutic decision-making in MS. By profiling patients using the wide array of data typically collected for research project, embracing the complexity of MS will make the exponentially growing wealth of information in the field more accessible to scientists, physicians and ultimately patients.

Acknowledgments

S.E.B. is a Harry Weaver Neuroscience Scholar of the National Multiple Sclerosis Society. Authors will like to acknowledge Alannah Tomich for assisting with preparation of Tables for this manuscript.

References

1. Hauser, SL., Goodin, DS. Multiple Sclerosis and other demyelinating diseases. In: Longo, DI., editor. Harrison’s principles of internal medicine. 18. New York: McGraw-Hill; 2012. p. 3395-409.
2. Jersild C, Svejgaard A, Fog T. HL-A antigens and multiple sclerosis. Lancet. 1972; 1:1240–1.
3. Naito S, Namerow N, Mickey MR, Terasaki PI. Multiple sclerosis: association with HL-A3. Tissue Antigens. 1972; 2:1–4. [PubMed: 5077731]
4. Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens. 1991; 38:1–15. [PubMed: 1926129]
5. Barcellos LF, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet. 2006; 15:2813–24. [PubMed: 16905561]
6. Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–9. [PubMed: 21833088]
7. Brynedal B, et al. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PLoS ONE. 2007; 2:e664. [PubMed: 17653284]
8. Harbo HF, et al. Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis. Tissue Antigens. 2004; 63:237–47. [PubMed: 14989713]

9. Gusella JF, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983; 306:234–8. [PubMed: 6316146]
10. Ebers GC, et al. A full genome search in multiple sclerosis. Nat Genet. 1996; 13:472–6. [PubMed: 8696345]
11. Haines JL, et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. Nat Genet. 1996; 13:469–71. [PubMed: 8696344]
12. Kuokkanen S, et al. A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus Eae2. Nat Genet. 1996; 13:477–80. [PubMed: 8696346]
13. Sawcer S, et al. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nat Genet. 1996; 13:464–8. [PubMed: 8696343]
14. Sawcer S, et al. A high-density screen for linkage in multiple sclerosis. Am J Hum Genet. 2005; 77:454–67. [PubMed: 16080120]
15. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature genetics. 2003; 33:177–82. [PubMed: 12524541]
16. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends in genetics : TIG. 2001; 17:502–10. [PubMed: 11525833]
17. Sawcer S. Bayes factors in complex genetics. European journal of human genetics : EJHG. 2010; 18:746–50. [PubMed: 20179745]
18. Hafler DA, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007; 357:851–62. [PubMed: 17660530]
19. Gregory SG, et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet. 2007; 39:1083–91. [PubMed: 17660817]
20. Lundmark F, et al. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet. 2007; 39:1108–13. [PubMed: 17660816]
21. ANZgene. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009; 41:824–8. [PubMed: 19525955]
22. 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. 2010; 86:285–91. [PubMed: 20159113]
23. Sanna S, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nat Genet. 2010; 42:495–7. [PubMed: 20453840]
24. Nischwitz S, et al. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. J Neuroimmunol. 2010; 227:162–6. [PubMed: 20598377]
25. Briggs FB, Shao X, Goldstein BA, Oksenberg JR, Barcellos LF, De Jager PL. Genome-wide association study of severity in multiple sclerosis. Genes and immunity. 2011; 12:615–25. [PubMed: 21654844]
26. De Jager PL, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009; 41:776–82. [PubMed: 19525953]
27. Baranzini SE, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet. 2009; 18:767–78. [PubMed: 19010793]
28. 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. 2008; 3:e3490. [PubMed: 18941528]
29. Burton PR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nature genetics. 2007; 39:1329–37. [PubMed: 17952073]
30. Maier LM, et al. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet. 2009; 5:e1000322. [PubMed: 19119414]
31. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Curr Opin Immunol. 2009; 21:596–605. [PubMed: 19896815]

32. Cotsapas C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS genetics. 2011; 7:e1002254. [PubMed: 21852963]
33. Ramagopalan SV, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. Genome research. 2010; 20:1352–60. [PubMed: 20736230]
34. Ramagopalan SV, et al. Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1*1501 is regulated by vitamin D. PLoS genetics. 2009; 5:e1000369. [PubMed: 19197344]
35. Ramagopalan SV, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Annals of neurology. 2011; 70:881–6. [PubMed: 22190362]
36. Baranzini SE, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature. 2010; 464:1351–6. [PubMed: 20428171]
37. Amiel, JC. Study of the leukocyte phenotypes in Hodkin's disease. In: Curtoni, ES.Mattiuz, PL., Tosi, RM., editors. Histocompatibility testing 1967 Report of a conference and workshop; Torino and Saint-Vincent, Italy. 14–24 June 1967; Munksgaard: København; 1967. p. 458
38. McDevitt HO, Chinitz A. Genetic control of the antibody response: relationship between immune response and histocompatibility (H-2) type. Science. 1969; 163:1207–8. [PubMed: 5765335]
39. Thorsby E. Invited anniversary review: HLA associated diseases. Human immunology. 1997; 53:1–11. [PubMed: 9127141]
40. Jersild C, Fog T, Hansen GS, Thomsen M, Svejgaard A, Dupont B. Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. Lancet. 1973; 2:1221–5. [PubMed: 4128558]
41. Bertrams HJ, Kuwert EK. Association of histocompatibility haplotype HLA-A3-B7 with multiple sclerosis. Journal of immunology. 1976; 117:1906–12.
42. Compston DA, Batchelor JR, McDonald WI. B-lymphocyte alloantigens associated with multiple sclerosis. Lancet. 1976; 2:1261–5. [PubMed: 63743]
43. Vartdal F, Sollid LM, Vandvik B, Markussen G, Thorsby E. Patients with multiple sclerosis carry DQB1 genes which encode shared polymorphic amino acid sequences. Human immunology. 1989; 25:103–10. [PubMed: 2737928]
44. Spurkland A, Celius EG, Knutsen I, Beiske A, Thorsby E, Vartdal F. The HLA-DQ(alpha 1*0102, beta 1*0602) heterodimer may confer susceptibility to multiple sclerosis in the absence of the HLA-DR(alpha 1*01, beta 1*1501) heterodimer. Tissue Antigens. 1997; 50:15–22. [PubMed: 9243750]
45. Caballero A, Alves-Leon S, Papais-Alvarenga R, Fernandez O, Navarro G, Alonso A. DQB1*0602 confers genetic susceptibility to multiple sclerosis in Afro- Brazilians. Tissue Antigens. 1999; 54:524–6. [PubMed: 10599893]
46. Kaushansky N, Altmann DM, David CS, Lassmann H, Ben-Nun A. DQB1*0602 rather than DRB1*1501 confers susceptibility to multiple sclerosis-like disease induced by proteolipid protein (PLP). Journal of neuroinflammation. 2012; 9:29. [PubMed: 22316121]
47. McElroy JP, et al. Refining the association of MHC with multiple sclerosis in African Americans. Human molecular genetics. 2010; 19:3080–8. [PubMed: 20466734]
48. Oksenberg JR, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet. 2004; 74:160–7. [PubMed: 14669136]
49. Caillier SJ, et al. Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. J Immunol. 2008; 181:5473–80. [PubMed: 18832704]
50. Harbo HF, et al. Norwegian Sami differs significantly from other Norwegians according to their HLA profile. Tissue Antigens. 2010; 75:207–17. [PubMed: 20047643]
51. Harbo HF, et al. Low frequency of the disease-associated DRB1*15-DQB1*06 haplotype may contribute to the low prevalence of multiple sclerosis in Sami. Tissue Antigens. 2007; 69:299–304. [PubMed: 17389012]
52. Brassat D, et al. The HLA locus and multiple sclerosis in Sicily. Neurology. 2005; 64:361–3. [PubMed: 15668443]
53. Coraddu F, et al. A genome screen for multiple sclerosis in Sardinian multiplex families. European journal of human genetics : EJHG. 2001; 9:621–6. [PubMed: 11528508]

54. Marrosu MG, et al. Sardinian multiple sclerosis is associated with HLA-DR4: a serologic and molecular analysis. Neurology. 1988; 38:1749–53. [PubMed: 2903464]
55. Marrosu MG, Murru MR, Costa G, Murru R, Muntoni F, Cucca F. DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population. Human molecular genetics. 1998; 7:1235–7. [PubMed: 9668164]
56. Marrosu MG, et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Hum Mol Genet. 2001; 10:2907–16. [PubMed: 11741834]
57. Muntoni F, et al. Different HLA DR2-DQw1 haplotypes in Sardinian and northern Italian populations: implications for multiple sclerosis susceptibility. Tissue Antigens. 1991; 38:34–6. [PubMed: 1926132]
58. Fogdell-Hahn A, Ligers A, Gronning M, Hillert J, Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. Tissue Antigens. 2000; 55:140–8. [PubMed: 10746785]
59. Rubio JP, et al. Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. Am J Hum Genet. 2002; 70:1125–37. [PubMed: 11923913]
60. Yeo TW, et al. A second major histocompatibility complex susceptibility locus for multiple sclerosis. Ann Neurol. 2007; 61:228–36. [PubMed: 17252545]
61. Link J, et al. Two HLA class I genes independently associated with multiple sclerosis. Journal of neuroimmunology. 2010; 226:172–6. [PubMed: 20678810]
62. Healy BC, et al. HLA B*44: protective effects in MS susceptibility and MRI outcome measures. Neurology. 2010; 75:634–40. [PubMed: 20713950]
63. Bergamaschi L, et al. Association of HLA class I markers with multiple sclerosis in the Italian and UK population: evidence of two independent protective effects. J Med Genet. 2011; 48:485–92. [PubMed: 21441263]
64. Lorentzen AR, et al. Killer immunoglobulin-like receptor ligand HLA-Bw4 protects against multiple sclerosis. Annals of neurology. 2009; 65:658–66. [PubMed: 19630074]
65. Cree BA, et al. A major histocompatibility Class I locus contributes to multiple sclerosis susceptibility independently from HLA-DRB1*15:01. PLoS ONE. 2010; 5:e11296. [PubMed: 20593013]
66. Lincoln MR, et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:7542–7. [PubMed: 19380721]
67. Gregersen JW, et al. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. Nature. 2006; 443:574–7. [PubMed: 17006452]
68. Hensiek AE, et al. HLA-DR 15 is associated with female sex and younger age at diagnosis in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2002; 72:184–7. [PubMed: 11796767]
69. Masterman T, Ligers A, Olsson T, Andersson M, Olerup O, Hillert J. HLA-DR15 is associated with lower age at onset in multiple sclerosis. Annals of neurology. 2000; 48:211–9. [PubMed: 10939572]
70. Celius EG, Harbo HF, Egeland T, Vartdal F, Vandvik B, Spurkiand A. Sex and age at diagnosis are correlated with the HLA-DR2, DQ6 haplotype in multiple sclerosis. Journal of the neurological sciences. 2000; 178:132–5. [PubMed: 11018705]
71. Smestad C, et al. The impact of HLA-A and -DRB1 on age at onset, disease course and severity in Scandinavian multiple sclerosis patients. Eur J Neurol. 2007; 14:835–40. [PubMed: 17662002]
72. Romero-Pinel L, et al. HLA-DRB1: genetic susceptibility and disability progression in a Spanish multiple sclerosis population. Eur J Neurol. 2011; 18:337–42. [PubMed: 20629714]
73. Imrell K, Greiner E, Hillert J, Masterman T. HLA-DRB115 and cerebrospinal-fluid-specific oligoclonal immunoglobulin G bands lower age at attainment of important disease milestones in multiple sclerosis. Journal of neuroimmunology. 2009; 210:128–30. [PubMed: 19327846]
74. Cree BA, et al. Modification of Multiple Sclerosis Phenotypes by African Ancestry at HLA. Arch Neurol. 2009; 66:226–33. [PubMed: 19204159]
75. Okuda DT, et al. Genotype-Phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. Brain. 2009; 132:250–9. [PubMed: 19022862]

76. Vandiedonck C, Giraud M, Garchon HJ. Genetics of autoimmune myasthenia gravis: the multifaceted contribution of the HLA complex. Journal of autoimmunity. 2005; 25(Suppl):6–11. [PubMed: 16260117]
77. Gourraud PA. When is the absence of evidence, evidence of absence? Use of equivalence-based analyses in genetic epidemiology and a conclusion for the KIF1B rs10492972*C allelic association in multiple sclerosis. Genetic epidemiology. 2011
78. Zuvich RL, et al. Interrogating the complex role of chromosome 16p13.13 in multiple sclerosis susceptibility: independent genetic signals in the CIITA-CLEC16A-SOCS1 gene complex. Human molecular genetics. 2011; 20:3517–24. [PubMed: 21653641]
79. Babron MC, et al. Determination of the real effect of genes identified in GWAS: the example of IL2RA in multiple sclerosis. European journal of human genetics : EJHG. 2011
80. Kofler DM, Severson CA, Mousissian N, De Jager PL, Hafler DA. The CD6 multiple sclerosis susceptibility allele is associated with alterations in CD4+ T cell proliferation. Journal of immunology. 2011; 187:3286–91.
81. Couturier N, et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. Brain : a journal of neurology. 2011; 134:693–703. [PubMed: 21354972]
82. Hernan MA, Olek MJ, Ascherio A. Cigarette smoking and incidence of multiple sclerosis. Am J Epidemiol. 2001; 154:69–74. [PubMed: 11427406]
83. Biran I, Steiner I. Smoking is a risk factor for multiple sclerosis. Neurology. 2004; 63:763. author reply.
84. Riise T, Nortvedt MW, Ascherio A. Smoking is a risk factor for multiple sclerosis. Neurology. 2003; 61:1122–4. [PubMed: 14581676]
85. Hedstrom AK, et al. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. Brain : a journal of neurology. 2011; 134:653–64. [PubMed: 21303861]
86. Sawcer S, Hellenthal G. The major histocompatibility complex and multiple sclerosis: a smoking gun? Brain : a journal of neurology. 2011; 134:638–40. [PubMed: 21310724]
87. Hawkes CH. Smoking is a risk factor for multiple sclerosis: a metanalysis. Multiple Sclerosis. 2007; 13:610–5. [PubMed: 17548439]
88. Mkhikian H, et al. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. Nat Commun. 2011; 2:334. [PubMed: 21629267]
89. Lee SU, et al. N-glycan processing deficiency promotes spontaneous inflammatory demyelination and neurodegeneration. The Journal of biological chemistry. 2007; 282:33725–34. [PubMed: 17855338]
90. Auffray C, Charron D, Hood L. Predictive, preventive, personalized and participatory medicine: back to the future. Genome medicine. 2010; 2:57. [PubMed: 20804580]
91. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol. 2011; 8:184–7. [PubMed: 21364692]
92. Corvol JC, et al. Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. Proc Natl Acad Sci U S A. 2008; 105:11839–44. [PubMed: 18689680]
93. McKinney EF, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. Nature medicine. 2010; 16:586–91. 1p–91.
94. Ma Q, Lu AY. Pharmacogenetics, pharmacogenomics, and individualized medicine. Pharmacol Rev. 2011; 63:437–59. [PubMed: 21436344]
95. Meyer MR, Maurer HH. Absorption, distribution, metabolism and excretion pharmacogenomics of drugs of abuse. Pharmacogenomics. 2011; 12:215–33. [PubMed: 21332315]
96. Baranzini SE, et al. Transcription-based prediction of response to IFNbeta using supervised computational methods. PLoS Biol. 2005; 3:e2. [PubMed: 15630474]
97. Martinez-Rodriguez JE, et al. Natural killer cell phenotype and clinical response to interferon-beta therapy in multiple sclerosis. Clinical immunology. 2011; 141:348–56. [PubMed: 21992960]
98. Bielekova B, et al. Regulatory CD56(bright) natural killer cells mediate immunomodulatory effects of IL-2Ralpha-targeted therapy (daclizumab) in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:5941–6. [PubMed: 16585503]

99. Bustamante MF, et al. Implication of the toll-like receptor 4 pathway in the response to interferon-beta in multiple sclerosis. Annals of neurology. 2011; 70:634–45. [PubMed: 22028223]
100. Koike F, et al. Microarray analysis identifies interferon beta-regulated genes in multiple sclerosis. Journal of neuroimmunology. 2003; 139:109–18. [PubMed: 12799028]
101. Wandinger KP, et al. Complex immunomodulatory effects of interferon-beta in multiple sclerosis include the upregulation of T helper 1-associated marker genes. Annals of neurology. 2001; 50:349–57. [PubMed: 11558791]
102. Wandinger KP, et al. TNF-related apoptosis inducing ligand (TRAIL) as a potential response marker for interferon-beta treatment in multiple sclerosis. Lancet. 2003; 361:2036–43. [PubMed: 12814715]
103. van Baarsen LG, et al. Pharmacogenomics of interferon-beta therapy in multiple sclerosis: baseline IFN signature determines pharmacological differences between patients. PLoS ONE. 2008; 3:e1927. [PubMed: 18382694]
104. Sturzebecher S, et al. Expression profiling identifies responder and non-responder phenotypes to interferon-beta in multiple sclerosis. Brain : a journal of neurology. 2003; 126:1419–29. [PubMed: 12764062]
105. Sellebjerg F, et al. Identification of new sensitive biomarkers for the in vivo response to interferon-beta treatment in multiple sclerosis using DNA-array evaluation. Eur J Neurol. 2009; 16:1291–8. [PubMed: 19558503]
106. Sellebjerg F, et al. Gene expression analysis of interferon-beta treatment in multiple sclerosis. Multiple Sclerosis. 2008; 14:615–21. [PubMed: 18408020]
107. Muraro PA, et al. Decreased integrin gene expression in patients with MS responding to interferon-beta treatment. Journal of neuroimmunology. 2004; 150:123–31. [PubMed: 15081256]
108. Killestein J, Polman CH. Determinants of interferon beta efficacy in patients with multiple sclerosis. Nature reviews Neurology. 2011; 7:221–8. [PubMed: 21364522]
109. Gilli F. Role of differential expression of interferon receptor isoforms on the response of multiple sclerosis patients to therapy with interferon beta. J Interferon Cytokine Res. 2010; 30:733–41. [PubMed: 20874250]
110. Bertolotto A, et al. Biological activity of interferon betas in patients with multiple sclerosis is affected by treatment regimen and neutralising antibodies. Journal of neurology, neurosurgery, and psychiatry. 2004; 75:1294–9.
111. Bertolotto A, Gilli F. Interferon-beta responders and non-responders. A biological approach. Neurol Sci. 2008; 29(Suppl 2):S216–7. [PubMed: 18690496]
112. Warner AW, et al. Challenges in obtaining adequate genetic sample sets in clinical trials: the perspective of the industry pharmacogenomics working group. Clinical pharmacology and therapeutics. 2011; 89:529–36. [PubMed: 21346757]
113. Lyons PA, et al. Microarray analysis of human leucocyte subsets: the advantages of positive selection and rapid purification. BMC Genomics. 2007; 8:64. [PubMed: 17338817]
114. Bejarano B, et al. Computational classifiers for predicting the short-term course of Multiple sclerosis. BMC Neurol. 2011; 11:67. [PubMed: 21649880]
115. Ioannidis JP. Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. Circulation Cardiovascular genetics. 2009; 2:7–15. [PubMed: 20031562]
116. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2009; 181:488–93.
117. Foster KR. The sociobiology of molecular systems. Nature reviews Genetics. 2011; 12:193–203.
118. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nature reviews Genetics. 2011; 12:56–68.
119. Baranzini SE. Systems-based medicine approaches to understand and treat complex diseases. The example of multiple sclerosis. Autoimmunity. 2006; 39:651–62. [PubMed: 17178562]
120. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101–13. [PubMed: 14735121]

121. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008; 18:644–52. [PubMed: 18381899]

122. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007; 104:8685–90. [PubMed: 17502601]

123. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. Nat Rev Drug Discov. 2009; 8:286–95. [PubMed: 19337271]

124. Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. Mamm Genome. 2007; 18:389–401. [PubMed: 17653589]

125. Gutierrez-Achury J, Coutinho de Almeida R, Wijmenga C. Shared genetics in coeliac disease and other immune-mediated diseases. Journal of internal medicine. 2011; 269:591–603. [PubMed: 21401738]

126. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics. 2009; 10:43–55.

127. Wang L, Khankhanian P, Baranzini SE, Mousavi P. iCTNet: a Cytoscape plugin to produce and analyze integrative complex traits networks. BMC Bioinformatics. 2011; 12:380. [PubMed: 21943367]

128. Patsopoulos NA, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Annals of neurology. 2011; 70:897–912. [PubMed: 22190364]

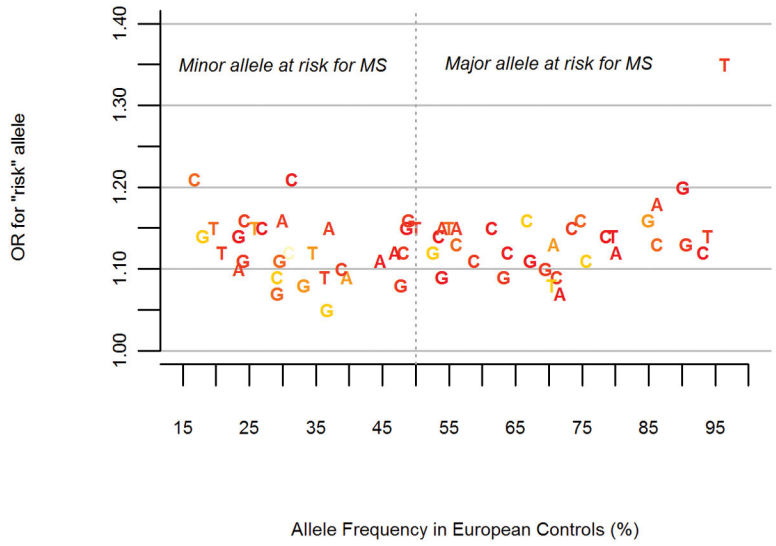
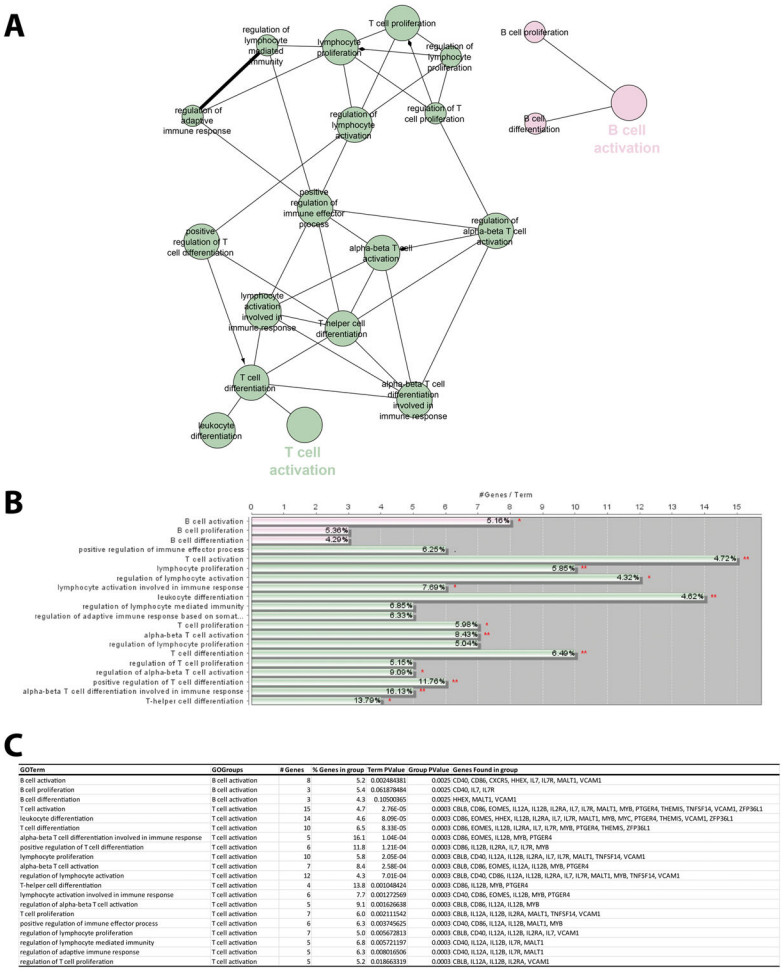


Figure 1. Non-MHC genetic variants associated MS susceptibility

This Figure depicts the allele frequencies in European controls of all MS associated variants and their respective OR for the risk allele (Table 1). It can be clearly seen that the effects of most risk alleles is modest. Almost half of the MS-associated alleles are common in Europeans (i.e. have a frequency higher than 50% in healthy individuals).



GOTerm	GOGroups	# Genes	% Genes in group	Term PValue	Group PValue	Genes Found in group
B cell activation	B cell activation	8	5.2	0.002484381		0.0025 CD40, CD86, CXCR5, HHEX, IL7, IL7R, MALT1, VCAM1
B cell proliferation	B cell activation	3	5.4	0.061879484		0.0025 CD40, IL7, IL7R
B cell differentiation	B cell activation	3	4.3	0.10500365		0.0025 HHEX, MALT1, VCAM1
T cell activation	T cell activation	15	4.7	2.76E-05		0.0003 CIB1B, CD86, EOMES, IL12A, IL12B, IL2RA, IL7, IL7R, MALT1, MYB, PTGER4, THEMIS, TNFSF14, VCAM1, ZFP36L1
leukocyte differentiation	T cell activation	14	4.6	8.09E-05		0.0003 CD86, EOMES, HHEX, IL12B, IL2RA, IL7, IL7R, MALT1, MYB, MYC, PTGER4, THEMIS, VCAM1, ZFP36L1
T cell differentiation	T cell activation	10	6.5	8.33E-05		0.0003 CD86, EOMES, IL12B, IL2RA, IL7, IL7R, MYB, PTGER4, THEMIS, ZFP36L1
alpha-beta T cell differentiation involved in immune response	T cell activation	5	16.1	1.04E-04		0.0003 CD86, EOMES, IL12B, MYB, PTGER4
positive regulation of T cell differentiation	T cell activation	6	11.8	1.21E-04		0.0003 CD86, IL12B, IL2RA, IL7, IL7R, MYB
lymphocyte proliferation	T cell activation	10	5.8	2.05E-04		0.0003 CIB1B, CD40, IL12A, IL12B, IL2RA, IL7, IL7R, MALT1, TNFSF14, VCAM1
alpha-beta T cell activation	T cell activation	7	8.4	2.58E-04		0.0003 CIB1B, CD86, EOMES, IL12A, IL12B, MYB, PTGER4
regulation of lymphocyte activation	T cell activation	12	4.3	7.01E-04		0.0003 CIB1B, CD40, CD86, IL12A, IL12B, IL2RA, IL7, IL7R, MALT1, MYB, TNFSF14, VCAM1
T-helper cell differentiation	T cell activation	4	13.8	0.001048424		0.0003 CD86, IL12B, MYB, PTGER4
lymphocyte activation involved in immune response	T cell activation	6	7.7	0.001272569		0.0003 CD40, CD86, EOMES, IL12B, MYB, PTGER4
regulation of alpha-beta T cell activation	T cell activation	5	9.1	0.001626638		0.0003 CIB1B, CD86, IL12A, IL12B, MYB
T cell proliferation	T cell activation	7	6.0	0.002111542		0.0003 CIB1B, IL12A, IL12B, IL2RA, MALT1, TNFSF14, VCAM1
positive regulation of immune effector process	T cell activation	6	6.3	0.003745625		0.0003 CD40, CD86, IL12A, IL12B, MALT1, MYB
regulation of lymphocyte proliferation	T cell activation	7	5.0	0.005672813		0.0003 CIB1B, CD40, IL12A, IL12B, IL2RA, IL7, VCAM1
regulation of lymphocyte mediated immunity	T cell activation	5	6.8	0.005721197		0.0003 CD40, IL12A, IL12B, IL7R, MALT1
regulation of adaptive immune response	T cell activation	5	6.3	0.008036506		0.0003 CD40, IL12A, IL12B, IL7R, MALT1
regulation of T cell proliferation	T cell activation	5	5.2	0.018663319		0.0003 CIB1B, IL12A, IL12B, IL2RA, VCAM1

Figure 2. Biological pathways involved in MS susceptibility

Gene ontology (GO) analysis of genes associated with MS from the WTCCC2 study (6) and subsequent meta-analysis in 5,545 independent cases and 12,153 controls (128). **A.** Significant GO categories and their relationships. Categories (GO terms) are linked with an edge if they share at least two genes. Categories are colored by groups, defined by their overall similarity. The green group represents T cell function (most significant category is *T cell activation*), while the pink group represents B cell function (most significant category is *B cell activation*). **B.** The number of genes assigned to each term and their enrichment (red asterisks denotes significance after Bonferroni correction). **C.** Genes assigned to each term and their nominal significance for enrichment for both GO term and group.

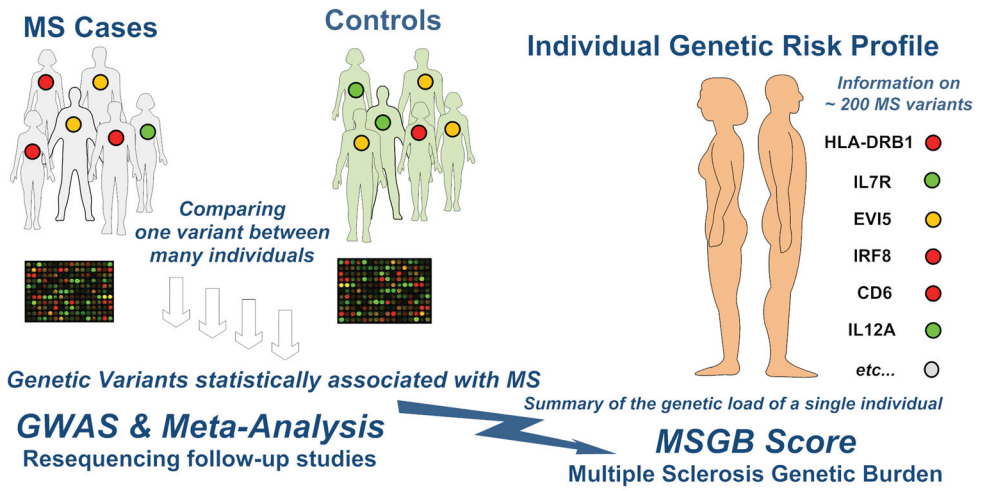


Figure 3. The Multiple Sclerosis Genetic Burden (MSGB) Score

The GWAS approach opens possibilities for individual genetic risk score computations. The MSGB is based on an algorithm that incorporates each risk variant for a given individual and weighs each SNP according to its reported effect size. The MSGB score quantitatively represents the known MS genetic risk for each individual. This provides an opportunity to analyze patients in the context of whole populations allowing personalized care.

Immunol Rev. Author manuscript; available in PMC 2018 May 24.

Table 1

The 2011 non-MHC genetic variants associated MS susceptibility

Chromosome	SNP rs#	Effect Size	P-value	Risk Allele	Risk Allele frequencies in European Population	Alleles	Gene of Interest
1	rs4648356	1.16	3.10E-14	C	66.8%	A/C	MMEL1
1	rs11810217	1.15	6.50E-12	T	25.7%	C/T	EVI5
1	rs11581062	1.07	2.50E-10	G	29.2%	A/G	VCAM1
1	rs1335532	1.18	2.00E-09	A	86.3%	A/G	CD58
1	rs1323292	1.12	8.80E-07	A	80.1%	A/G	RGS1
1	rs7522462	1.11	9.20E-07	G	67.3%	A/G	KIF21B
2	rs12466022	1.16	6.20E-10	C	74.8%	A/C	X
2	rs7595037	1.15	5.10E-11	T	54.9%	C/T	PLEK
2	rs17174870	1.15	1.30E-08	C	73.5%	C/T	MERTK
2	rs10201872	1.15	1.80E-10	T	19.6%	C/T	SP140
3	rs11129295	1.09	1.20E-09	T	36.3%	C/T	EOMES
3	rs2028597	1.13	2.10E-04	G	90.7%	A/G	CBLB
3	rs2293370	1.16	1.10E-09	G	85.0%	A/G	TMEM39A
3	rs9282641	1.2	1.00E-11	G	90.2%	A/G	CD86
3	rs2243123	1.09	3.70E-06	C	29.2%	C/T	IL12A
3	rs669607	1.15	1.90E-15	G	48.7%	T/G	intergenic
4	rs228614	1.09	1.40E-07	G	54.0%	A/G	NFKB1
5	rs6897932	1.11	2.60E-06	C	75.7%	C/T	IL7R
5	rs4613763	1.21	6.90E-14	C	16.8%	C/T	PTGER4
5	rs2546890	1.15	1.20E-11	A	56.2%	A/G	IL12B
6	rs12212193	1.08	3.80E-08	G	47.8%	A/G	BACH2
6	rs802734	1.13	5.50E-09	A	70.8%	A/G	THEMIS
6	rs11154801	1.09	1.00E-13	A	39.7%	A/C	MYB
6	rs17066096	1.14	6.00E-13	G	18.1%	A/G	IL22RA2
6	rs1738074	1.14	6.80E-15	C	53.5%	C/T	TAGAP
6	rs13192841	1.1	2.30E-06	A	23.5%	A/G	OLIG3

Chromosome	SNP rs#	Effect Size	P-value	Risk Allele	Risk Allele frequencies in European Population	Alleles	Gene of Interest
7	rs354031	1.14	4.70E-09	G	23.5%	A/G	ZNF767
8	rs1520333	1.11	6.10E-07	G	24.1%	A/G	IL7
8	rs4410871	1.09	7.70E-09	C	71.2%	C/T	MYC
8	rs2019960	1.16	5.20E-09	C	24.3%	C/T	PVT1
9	rs2150702	1.16	3.28E-08	G	49.0%	A/G	MLANA
10	rs3118470	1.12	2.00E-09	C	31.0%	C/T	IL2RA
10	rs7090512	1.21	4.60E-20	C	31.4%	C/T	IL2RA
10	rs1250542	1.15	3.97E-07	A	37.0%	G/A	ZMIZ1
10	rs7923837	1.09	4.90E-09	G	63.3%	A/G	HHEX
11	rs650258	1.12	1.70E-09	C	63.8%	C/T	CD6
11	rs630923	1.13	2.80E-07	C	86.3%	A/C	CXCR5
12	rs1800693	1.12	1.80E-10	C	48.2%	C/T	TNFRSF1A
12	rs10466829	1.12	1.40E-08	A	46.9%	A/G	CLECL1
12	rs12368653	1.11	2.00E-07	A	44.7%	A/G	CYP27B1
12	rs949143	1.08	1.50E-04	G	33.2%	A/G	MPHOSPH9
13	rs9596270	1.35	7.00E-07	T	96.4%	T/C	Intergenic
14	rs4902647	1.13	9.30E-12	C	56.2%	C/T	ZFP36L1
14	rs2300603	1.08	2.00E-08	T	70.4%	C/T	BATF
14	rs2119704	1.12	2.20E-10	C	93.2%	A/C	GALC
16	rs7200786	1.15	6.30E-14	A	54.0%	A/G	CLEC16A
16	rs7191700	1.15	6.40E-07	C	61.5%	C/T	TNP2
16	rs13333054	1.12	7.00E-08	T	20.8%	C/T	IRF8
17	rs9891119	1.1	4.60E-07	C	38.9%	A/C	STAT3
17	rs8070463	1.15	9.55E-08	T	50.0%	T/C	KPNB1/TBKBP1/TBX21
17	rs180515	1.05	8.80E-08	G	36.7%	A/G	RPS6KB1
18	rs7238078	1.14	2.50E-09	T	79.6%	G/T	MALT1

Chromosome	SNP rs#	Effect Size	P-value	Risk Allele	Risk Allele frequencies in European Population	Alleles	Gene of Interest
19	rs1077667	1.14	9.40E-14	C	78.6%	C/T	TNFSF14
19	rs8112449	1.1	1.50E-06	G	69.5%	A/G	TYK2
19	rs10411936	1.16	2.04E-07	A	30.0%	G/A	EPS15L1
19	rs874628	1.07	1.30E-08	A	71.7%	A/G	MPV17L2
19	rs2303759	1.11	5.20E-09	G	29.6%	G/T	DKKL1
20	rs6074022	1.15	4.91E-06	C	27.0%	T/C	CD40
20	rs2248359	1.11	2.50E-11	C	58.8%	C/T	CYP24A1
20	rs6062314	1.14	1.30E-07	T	93.8%	C/T	TNFRSF6B
22	rs2283792	1.12	4.70E-09	G	52.7%	G/T	MAPK1
22	rs140522	1.12	1.70E-08	T	34.5%	C/T	SCO2

This summary table refers to published variants only (6, 128)

Table 2

Population frequencies of genetic variants associated with multiple sclerosis in the MHC region

MHC region	Locus	Allele or SNP	Odds Ratio	Reference for OR	Most frequent haplotypes *(1)*	(Rank) Haplotype frequency	
					HLA-A, -B, -C, -DRB1	European	African American
Class I Region	**HLA-A**	A*02	0.73	Sawcer et al 2011	**02**01g-4402g-0501g-0401	(1) 0.0259	(1) 0.0055
					0201g-0702g-0702-1501	(2) 0.0234	(5) 0.0022
					0201g-1501g-0304-0401	(3) 0.0124	(35) 0.0011
					0201g-4001g-0304-1302	(4) 0.0099	(16) 0.0016
					0201g-4501g-1601-1302	Not observed	(2) 0.0032
					0201g-4501g-1601-1101	Not observed	(3) 0.0024
					0201g-5301-0401g-1303	Not observed	(4) 0.0023
					0201g-0801g-0701g-0301	(5) 0.0097	(6) 0.0022
	HLA-B	B*44	0.62	Healy et al 2010	0201g-**44**02g-0501g-0401	(1) 0.0259	(3) 0.0055
					2902-**44**03-1601-0701	(2) 0.0184	(2) 0.0056
					2301g-**44**03-0401g-0701	(3) 0.0069	(5) 0.0028
					0201g-**44**02g-0501g-1301	(4) 0.0054	Not observed
					0201g-**44**02g-0501g-1501	(5) 0.0052	(6) 0.0017
					3402-**44**03-0401g-1503	(217) 0.0001	(1) 0.0056
					2301g-**44**03-0401g-1503	Not observed	(4) 0.0030
	HLA-C	Cw*5 /C*05	0.63*	Yeo et al 2007, *Healy et al 2010	0201g-4402g-**05**01g-0401	(1) 0.0258	(1) 0.0055
					0201g-4402g-**05**01g-1301	(2) 0.0054	Not observed
					0201g-4402g-**05**01g-1501	(3) 0.0052	(2) 0.0017
					0201g-4402g-**05**01g-0101	(4) 0.0046	(4) 0.0013
					3002-1801g-**05**01g-0301	(5) 0.0043	(7) 0.0009
					0201g-1801g-**05**01g-0301	(6) 0.0023	(3) 0.0014
					0201g-4402g-**05**01g-0701	(10) 0.0015	(5) 0.0011
	HLA-G	rs4959039 (2)	1.59	Cree et al, 2010	NC_000006.11:g.29957069A>G	A=0.655 G=0.345 (CEU)	A=0.920 G=0.080 (YRI)
	NRM	rs2269704 (2)		McElroy et al, 2011	Japanese population C=0.878 T=0.122 (JPT)	C= 0.996 T=0.004 (CEU)	C= 0.996 T=0.004 (YRI

MHC region	Locus	Allele or SNP	Odds Ratio	Reference for OR	Most frequent haplotypes (1)	(Rank) Haplotype frequency	
					HLA-A, -B, -C, -DRB1	European	African American
Class II Region	**HLA-DRB1**	DRB1*15:01	3.08	Sawcer et al, 2011	0301g-0702g-0702-**1501**	(1) 0.0357	(1) 0.0070
					0201g-0702g-0702-**1501**	(2) 0.0234	(3) 0.0022
					2402g-0702g-0702-**1501**	(3) 0.0083	(8) 0.0006
					0101g-0702g-0702-**1501**	(4) 0.0064	(6) 0.0011
					2501-1801g-1203-**1501**	(5) 0.0058	(2) 0.0023
					0201g-4402g-0501g-**1501**	(7) 0.0052	(4) 0.0017
					1101g-0702g-0702-**1501**	(8) 0.0033	(5) 0.0017
		DRB1*15:03	1.5	Oksenberg et al, 2004	3402-4403-0401g-**1503**	(1) 0.0001	(1) 0.0056
					3402-0702g-1801g-**1503**	(2) 0.0001	Not observed
					6801g-5702-1801g-**1503**	(3) 0.0001	Not observed
					6802-5301-0401g-**1503**	Not observed	(2) 0.0040
					3002-1402-0802-**1503**	Not observed	(3) 0.0040
					2902-4901-0701g-**1503**	Not observed	(4) 0.0035
					6802-0702g-0702-**1503**	Not observed	(5) 0.0035
		DRB1*13:03	2.43	Sawcer et al, 2011	6802-1402-0802-**1303**	(1) 0.0034	Not observed
				DRB1*13:03 in non-Ashkenazi Jews : Kwon et al 1999	6601-4102-1701g-**1303**	(2) 0.0010	(82) 0.0002
					2501-5801g-0701g-**1303**	(3) 0.0004	Not observed
					2402g-4102-1701g-**1303**	(4) 0.0003	Not observed
					0201g-4102-1701g-**1303**	(5) 0.0003	(25) 0.0003
					7401g-5703-0701g-**1303**	Not observed	(1) 0.0035
					6802-5301-0401g-**1303**	Not observed	(2) 0.0025
					0201g-5301-0401g-**1303**	Not observed	(3) 0.0023
					0201g-5101g-1601-**1303**	Not observed	(4) 0.0020
					2301g-5101g-1601-**1303**	Not observed	(5) 0.0013
		DRB1*13:01	>3	Kwon et al, 1999	0201g-1501g-0303g-**1301**	(1) 0.0062	Not observed
					0201g-4402g-0501g-**1301**	(2) 0.0054	Not observed
					0301g-0702g-0702-**1301**	(3) 0.0028	Not observed
					2601g-3801-1203-**1301**	(4) 0.0026	Not observed

MHC region	Locus	Allele or SNP	Odds Ratio	Reference for OR	Most frequent haplotypes (1)	(Rank) Haplotype frequency	
					HLA-A, -B, -C, -DRB1	European	African American
					0101g-0801g-0701g-**1301**	(5) 0.0025	Not observed
					6601-5802-0602-**1301**	Not observed	(1) 0.0029
					0201g-4501g-1601-**1301**	Not observed	(2) 0.0021
					3002-5703-1801g-**1301**	(145) 0.0001	(3) 0.0021
					0301g-5802-0602-**1301**	Not observed	(4) 0.0019
					3601-5301-0401g-**1301**	Not observed	(5) 0.0012
		DRB1*03:01	1.26	Sawcer et al 2011 DRB1*03:01, HLA-DQB1*02:01 frequent in Sardinia, Italy	0101g-0801g-0701g-**0301**	(1) 0.0745	(1) 0.0118
					0201g-0801g-0701g-**0301**	(2) 0.0097	(3) 0.0022
					3002 1801g 0501g-**0301**	(3) 0.0043	(4) 0.0014
					0301g-0801g-0701g-**0301**	(4) 0.0042	Not observed
					2402g-0801g-0701g-**0301**	(5) 0.0039	Not observed
					6802-1510-0304-**0301**	Not observed	(2) 0.0062
					3002-0801g-0701g-**0301**	Not observed	(5) 0.0014
		DRB1*04:05		Marrosu et al 1997, In Sardinia, Italy	Rare allele in Northern European and African	<0.005	
					DRB1*04:05 is common in Sardinia and East-Asian populations		
		DRB1*01:08	1.18	Sawcer et al, 2011	Rare allele	<0.001	
	HLA-DQB1	DQB1*06:02	3.08	Sawcer et al, 2011- same haplotype, carrying DRB1*15:01	0301g-0702g-0702-1501-0602	(1) 0.0356	(1) 0.0091
					0201g-0702g-0702-1501-0602	(2) 0.0235	(10) 0.003
					2402g-0702g-0702-1501-0602	(3) 0.0080	(38) 0.0013
					0101g-0702g-0702-1501-0602	(4) 0.0061	Not observed
					2501-1801g-1203-1501-0602	(5) 0.0057	(11) 0.003
					3402-4403-0401g-1503-0602	(220) 0.0001	(2) 0.0061
					6802-0702g-0702-1503-0602	Not observed	(3) 0.0056
					3601-5301-0401g-1101-0602	Not observed	(4) 0.0052
					2301g-4403-0401g-1503-0602	Not observed	(5) 0.0042
	HLA-DRB5	Presence	1.4	Caillier et al, 2008	Most commonly encountered in African Americans		

MHC region	Locus	Allele or SNP	Odds Ratio	Reference for OR	Most frequent haplotypes (1)	(Rank) Haplotype frequency	
					HLA-A, -B, -C, -DRB1	European	African American
	NOTCH4	rs422951 (2)	0.4	McElroy et al, 2011	A=0.738 G=0.262 (JPT)	A=0.580 G=0.420 (CEU)	A=0.735 G=0.265 (YRI)

(1)When available, haplotype frequencies were listed from Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol. 2007 Sep;68(9):779–88.

(2)When variant is a SNP, frequencies of the alleles are reproted from HAPMAP relevant populations, details can be obtained at http://www.ncbi.nlm.nih.gov/projects/SNP/