Title
Chapter 9 Multiple sclerosis genetics

Author
Cree, Bruce AC

Publication Date
2014

DOI
10.1016/b978-0-444-52001-2.00009-1

Peer reviewed

Handbook of Clinical Neurology, Vol. 122 (3rd series)
Multiple Sclerosis and Related Disorders
D.S. Goodin, Editor

Chapter 9

Multiple sclerosis genetics

BRUCE A.C. CREE*
Department of Neurology, University of California, San Francisco, USA

INTRODUCTION

The pathogenesis of multiple sclerosis (MS) is complex and involves both an individual's genetic makeup as well as environmental exposures. Evidence for the importance of genetics comes from epidemiologic studies of race, geography, familial aggregation, and more recently, from genome-wide association studies (GWAS). For example, it is notable that the lifetime risk of disease in individuals who have an affected family member increases roughly in proportion to the amount of genetic information shared between the affected relative and the individual (Ebers et al., 1995, 2004; Robertson et al., 1996; Sadovnick et al., 1996; Compston and Coles, 2002; Nielsen et al., 2005). Thus, in the northern hemisphere, persons in the general population have a risk of approximately 0.1%. Third-degree relatives, such as first cousins (12.5% genetic similarity), have a risk less than 1%; second-degree relatives, such as aunts and uncles (25% genetic similarity), have a risk of approximately 1–2%; first-degree relatives, such as siblings, parents, and children of an MS proband (50% genetic similarity), have a risk of approximately 2–5%; and monozygotic twins (100% genetic similarity) have a risk of about 25–30%. Because the concordance between identical twins is 25–30%, factors other than genetic similarity, such as environmental exposures or postgenomic modifications to immune function, must also contribute to MS susceptibility to explain the relatively low penetrance of the genetic risk. Despite the probably important role of environmental events in MS pathogenesis, genetic susceptibility is critical. It is the purpose of this chapter to review this evidence and to explore the genetic basis of this complex disease.

GEOGRAPHY AND ETHNICITY

The fact that both ethnicity and geography influence the prevalence of MS suggests that heritable factors contribute to MS pathogenesis (Davenport, 1922). Compared to other ethnic groups residing at the same latitudes, people of northern European ancestry are at higher risk for MS (Dean et al., 1976; Pugliatti et al., 2002; Alter et al., 2006; Smestad et al., 2008). This increased susceptibility might be due to genetic differences between ethnic groups. For example, some studies have shown that MS is approximately 50% less common in African Americans compared to whites (Kurtzke et al., 1977; Wallin et al., 2004). In addition, MS is still less common in both native Japanese and Japanese Americans (5 per 100 000) compared to northern European populations (100–150 per 100 000) (Detels et al., 1977). Similarly, in both the United States and Canada, MS has been reported to be relatively less common among Native Americans (Oger et al., 1975; Kurtzke et al., 1979; Hader, 1982; Svenson et al., 1994). These observed racial patterns of variation in MS prevalence lead to the hypothesis that certain genetic traits may be enriched in populations that are at higher risk for MS and that the same traits may be underrepresented in populations at lower risk for MS. Nevertheless, there are other factors (both ethnic and environmental) that differ between the races (e.g., lifestyle or diet) and could potentially confound the genetic interpretation of these racial variations.

FAMILIAL AGGREGATION

Although MS was initially thought to be a sporadic disease, the familial occurrence of MS was recognized by the late 19th century (Gowers, 1893; Eichorst, 1896). Systematic studies of familial aggregation in MS support a genetic contribution to the disease (Curtius, 1938; McAlpine, 1946; Pratt et al., 1951; Millar and Allison, 1954; Sadovnick et al., 1988; Robertson et al., 1996;

*Correspondence to: Bruce A.C. Cree, M.D., Ph.D., M.C.R., Associate Professor of Clinical Neurology, Clinical Research Director, UCSF Multiple Sclerosis Center, 675 Nelson Rising Lane, Suite 221, San Francisco, CA 94158, USA. Tel: +1-415-514-2466, Fax: +1-415-514-2470, E-mail: bruce.cree@ucsf.edu

Table 9.1

Familial risks for + multiple sclerosis (MS)

Relationship to patient	Recurrence risk (%)	Risk relative to population	Proportion of genetic sharing (%)
Adopted first-degree relative	0.2	Identity	0
Sibling with MS	3.0–5.0	15–25-fold increase	50
Dizygotic twin	3.0–5.0	15–25-fold increase	50
Monozygotic twin	34.0	170-fold increase/	100
One parent with MS	3.0–5.0	15-25-fold increase	50
Two parents with MS	6.0–10.0	30-50-fold increase	50 with each parent

Assumes lifetime population prevalence of 0.2% (Ebers et al., 2000; Dyment et al., 2004a).

Carton et al., 1997). These studies found that approximately 15–20% of MS patients reported a family history of MS, a proportion that is significantly higher than that which would be expected based on the low population prevalence of MS.

Siblings, as well as other first-degree relatives of MS patients (probands), are at increased risk for MS. This can be demonstrated by calculating the ratio of the relative risk of disease in relatives of affected individuals compared to the relative risk of disease in the overall population. This ratio is referred to as λ_s (Risch, 1990). If there were no increased risk for relatives of MS patients then the λ_s ratio would be 1. If relatives of patients are at higher risk of disease then the ratio increases. For example, in cystic fibrosis (a highly penetrant autosomal-recessive trait) the value of λ_s ratio is 500. For MS, the λ_s ratio is approximately 15–40, indicating a moderately strong familial influence on MS risk (Compston, 1997). To place this value in the context of other heritable complex diseases, the λ_s ratio for MS is higher than that for schizophrenia, about the same as that for type 1 diabetes and less than that for autism (Merikangas and Risch, 2003). Nevertheless, it is important to recognize that a λ_s ratio greater than 1 does not necessarily indicate a genetic cause for the trait in question. Thus, similar environmental factors are also shared among family members that might explain such familial aggregation (Guo, 2002).

TWIN STUDIES

Some of the most compelling evidence indicating that MS susceptibility has a genetic component comes from twin studies. The percentage of concordance for monozygotic twins is approximately 25–30% whereas the concordance percentage for dizygotic twins is 3–5% (Ebers et al., 1986, 1995, 2004; Mumford et al., 1994; Willer et al., 2003; Hansen et al., 2005; Islam et al., 2006; Ristori et al., 2006). This concordance rate for fraternal twins is greater than that of other first-degree relatives of MS patients but is still far less than that of identical twins (Willer et al., 2003). Because twins share the same intrauterine environment and similar postnatal environments, the large difference in concordance between monozygotic and dyzgotic twins must, at least in part, be due to genetic factors.

Moreover, conjugal pair studies also show that the risk of MS increases substantially if both parents have MS, again implying a strong heritable component to MS susceptibility (Table 9.1; Robertson et al., 1997; Ebers et al., 2000; Dyment et al., 2004a). Taken together, these familial and population-based studies indicate that some component of MS risk is heritable. However, the fact that the majority of MS patients have no family history suggests that either environmental factors outweigh genetic risks, or that genetic risk is due to the influence of multiple genetic traits that, by themselves, result in a low disease penetrance. Indeed, the concept that MS risk is inherited as a complex (multigenic) trait instead of following simple single-gene Mendelian rules (either as a recessive trait such as cystic fibrosis or dominant trait such as Huntington disease) is central to our understanding of the genetic contributions to MS risk (Fig. 9.1).

HUMAN LEUKOCYTE ANTIGENS

The first studies to establish a link between MS heredity and specific genetic variations compared human leukocyte antigen (HLA) protein polymorphisms between MS cases and healthy controls. These early studies found that certain cell surface proteins (antigens), that are present on the membranes of peripheral blood mononuclear cells, were overrepresented in MS patients compared to unaffected controls. The first such antigens to be reported were HLA-A3 (Bertrams and Kuwert, 1972; Bertrams et al., 1972; Naito et al., 1972), followed by HLA-B7 and then HLA-DRw2 (Jersild et al., 1972, 1973; Winchester et al., 1975; Compston et al., 1976; Terasaki et al., 1976). As it turned out, these HLA associations were not independent of each other but rather reflected a common shared haplotype due to the fact that the chromosomal region coding for these proteins was in

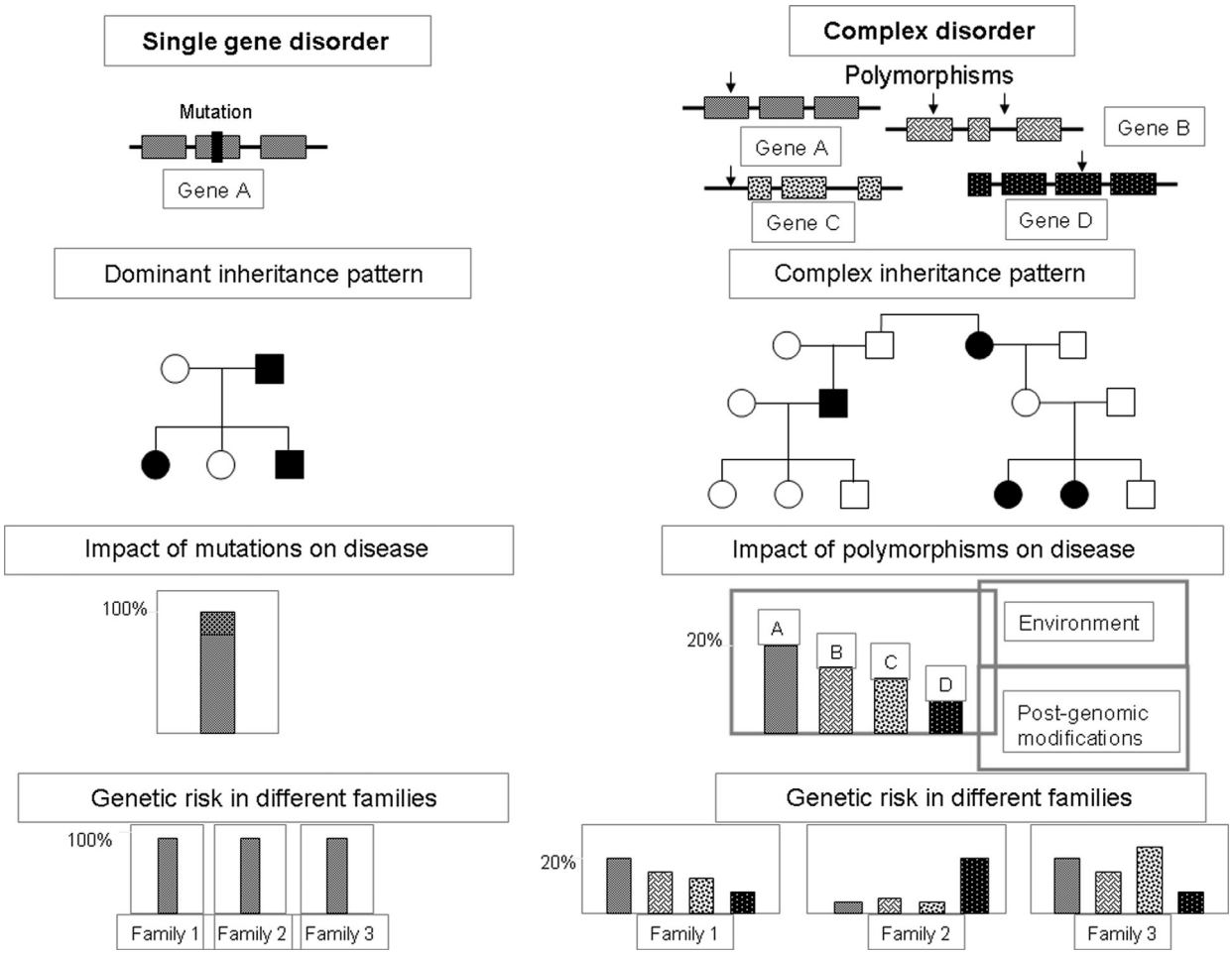

Fig. 9.1. Inheritance of single gene and polygenic (complex) disorders. (Modified from Peltonen and McKusick, 2001.)

strong linkage disequilibrium. Linkage disequilibrium refers to the observation that alleles of certain neighboring genes tend to be inherited together as a set of traits. This may be the consequence of natural selection, although the basis for such selection (if it occurs) is unknown. It could also be due to the small number of individuals who migrated out of Africa (with limited genetic repertoires) and an insufficient number of generations to disassociate genetic loci by crossover events. For whatever reason, the antigens HLA-A3, HLA-B7, and HLA-DRw2 are closely coupled in populations of European descent. The elucidation as to which of these linked genes in the major histocompatibility complex (MHC) was responsible for MS susceptibility was initially not possible and required the development of modern molecular techniques.

Linkage analysis

In the 1980s, new DNA-based technology was developed for studying Mendelian patterns of inheritance, first with the use of restriction fragment length polymorphisms (RFLPs) and then followed by the analysis of microsatellite repeats (Botstein et al., 1980; Weber and May, 1989). RFLPs and microsatellite repeats are variations in DNA sequence that can be identified by molecular techniques. These DNA variations can be used to determine whether the DNA locus that harbors the variant has more than a chance association with an inherited trait of interest. These new molecular techniques allowed for dissection of linked genes at the HLA locus as well as screening the entire genome. By studying the linkage between an inherited trait and DNA markers in families with some members affected by a heritable disease it was possible to identify the chromosomal location of disease-causing genes (Fig. 9.2). Markers that are physically near the disease-causing gene are likely to be inherited together with the disease trait because recombination between genetic loci occurs less frequently between neighboring genes compared to genes at greater distances. Thus, the phenomenon of linkage disequilibrium that confounded efforts to discriminate between alleles

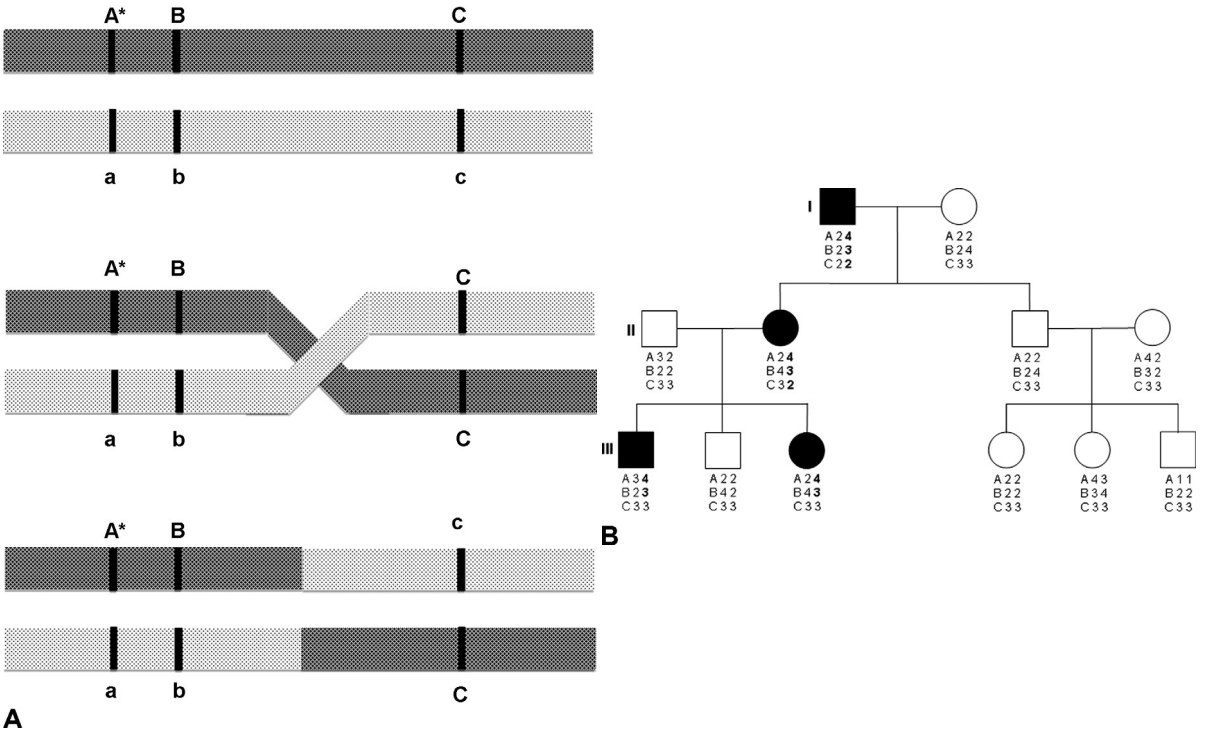

Fig. 9.2. (**A**) Meiotic recombination is the underlying genetic principle of linkage analysis. Paternal (dark gray) and maternal (light gray) chromosomes are aligned in a germ cell (cell that gives rise to sperm or ova). Sequence A* is a disease-causing allele, whereas a is the normal allele. Alleles for nearby DNA sequences on the same chromosome are depicted as B and C. Paternal alleles are represented in capital letters and maternal alleles are represented in lower case, and the lower-case letters represent the maternal alleles. During meiotic recombination paired chromosomal DNA strands cross over. The crossover event results in a break in the paternal DNA strand that is recombined with the maternal DNA strand, resulting in recombined chromosomes. The mixed chromosomes are passed to the sperm or ova. If the disease gene is A*, then recombination is more likely to occur between the disease gene and alleles of C than alleles of B. By following the segregation of the disease gene in families along with the segregation of genetic markers, the disease-causing gene A* can be mapped relative to the markers B and C.

(**B**) Pedigree analysis showing segregation of markers with a dominantly inherited trait. In the second generation the marker combination A4 B3 C2 is inherited by the affected daughter. Due to recombination between markers B and C, in the third generation affected individuals carry the marker combination A4 B3 C3, showing that the trait is linked to the A4 B3 haplotype. Although these markers are linked to the trait, they are can also be found in the general population. Linkage analysis relies on segregation of markers that are linked to a trait taking into account family structure. (Copyright Bruce Cree.)

of genes at HLA could be exploited to identify previously unknown disease-associated genes. These new techniques were first applied to Mendelian inherited diseases and culminated in the identification of many single-gene mutations, such as those for cystic fibrosis and Huntington's disease. In MS, RFLPs were used to dissect the molecular contributions of the HLA locus to MS susceptibility. Thus, it became clear for the first time that alleles of *HLA-DR2* were the major contributors to MS risk (Cohen et al., 1984).

MS as a complex trait

This linkage-based approach had the potential for application not only to single-gene disorders but also to complex multigenic traits. However, the hurdles for identifying the genetic basis for such traits is considerably higher because the increase in penetrance of disease resulting from each gene is expected to be far less than that for single disease-causing mutations. Penetrance refers to the likelihood that a particular genotype will manifest as a phenotype. For some Mendelian traits (e.g., the dominant trait of Huntington's disease) the penetrance is very high, meaning that nearly all individuals who carry the disease-causing genotype will ultimately develop the disease. However, for complex traits, because the penetrance is low and because the disease-associated polymorphisms may be common in the population, the penetrance of each gene may be very low. This is clearly the case for the HLA locus: none of the HLA alleles associated with MS are, by themselves, disease-causing mutations. All these alleles are commonly found

in healthy controls, despite being overrepresented among MS patients. Initially, the importance of this observation was not fully appreciated by investigators. Heritable MS risk was present in some families who lacked the MS-associated HLA alleles. This suggested that other loci elsewhere in the genome could be present that accounted for MS inheritance. It was hypothesized that these non-MHC loci might even contribute to MS risk even more than the HLA locus. If so, then a systematic study focused on the genomes of families affected by MS was anticipated to readily identify these other loci.

GENOMIC LINKAGE SCREENS

The first series of genome-wide screens using several hundred microsatellite DNA markers in approximately 100 affected sib pairs (pairs of non-twin siblings in which one was affected by MS and the other was not) was undertaken in the 1990s (Ebers et al., 1996; Haines et al., 1996; Sawcer et al., 1996). Assuming that other loci in the genome would have had similar effects on MS risk as that of, or even greater than, the MHC locus, these studies were expected to identify novel loci. However, despite this expectation, no statistically significant additional loci were found. Furthermore, one of these studies was unable to detect a signal even from the MHC (Ebers et al., 1996). Follow-up studies using multiply affected families also failed to detect any convincing new MS susceptibility loci (Kuokkanen et al., 1997; Coraddu et al., 2001; Akesson et al., 2002; Ban et al., 2002; Eraksoy et al., 2003). Adding more microsatellite markers to the initial genome screens also failed to produce new MS-associated genes (Hensiek et al., 2003; Dyment et al., 2004b; Kenealy et al., 2004). Pooling data for meta-analysis was similarly of no help in identifying loci other than the MHC (Dyment et al., 2001; GAMES, Transatlantic Multiple Sclerosis Genetics Cooperative, 2003). Consequently, it became clear that the identification of the other effects of genetic variations impacting MS susceptibility would require not only better markers but also a substantially increased number of families to reach necessary statistical power. The way forward required access to data from a larger number of affected families than was possible for any single group. Thus, the International Multiple Sclerosis Genetics Consortium (IMSGC) was founded in 2003 to overcome these barriers. This consortium brought together many, previously competing, investigators in a collaborative effort to decode MS heritability (Sawcer et al., 2004; http://www.neurodiscovery.harvard.edu/research/imsgc.html).

The first large-scale linkage study with sufficient statistical power to detect loci that had similar effects to that of the MHC across the genome came from populations in Australia, Scandinavia, the United Kingdom and the United States. Disappointingly, however, this study identified only the well-known association between the MHC and MS susceptibility (Fig. 9.3) (Sawcer et al., 2005). Moreover, other loci whose associations with MS had been suggested from smaller studies were not replicated. Nevertheless, this study was an important milestone for the field of MS genetics because, for the first time, a large number of markers and a substantial number of MS-affected families were brought together through an international collaborative effort. Furthermore, the markers used were sufficiently numerous and evenly spaced across the genome that there was confidence that the majority of the genome was adequately represented for linkage analysis. Perhaps most importantly, 730 families were studied, thus providing adequate power to detect genetic effects that increased the odds of MS risk by more than twofold. Only MHC was found to increase risk of MS by this degree, which indicated that the other loci, which influence MS risk, must have an effect lower than this. Thus study gave the first quantitative estimate for the possible impact of these other genes. Identification of such loci was effectively not possible using linkage analysis unless tens of thousands of families were analyzed (Risch and Merikangas, 1996). This inherent limitation of linkage methodology potentially could be overcome by a different genetic analysis: the GWAS.

GENOME-WIDE ASSOCIATION SCREEN

Further technological innovation led to identification of single nucleotide polymorphisms (SNPs) throughout the genome. SNPs are genetic variants that occur at a single basepair position within the genome (Fig. 9.4). The remarkable achievement of sequencing of the human genome in conjunction with mapping hundreds of thousands of these SNP variants led to the realization that 99.9% of the human genome is invariant (International HapMap Consortium, 2003; International Human Genome Sequencing Consortium, 2004). Nevertheless, there are still many millions of genetic variations, some of which are quite common. By mapping traits that are linked to common SNP variants (i.e., SNPs that are present in at least 5% of the general population) and by coupling informative SNP variants with microchip-based miniaturization, it became possible to interrogate hundreds of thousands of SNP variants from thousands of individuals (Fig. 9.4). If heritable traits such as MS susceptibility are linked to common SNP variants, then genotyping these variants in both case and control populations would lead to the identification of those loci that were associated with the heritable trait. This hypothesis is referred to as the common disease–common variant hypothesis.

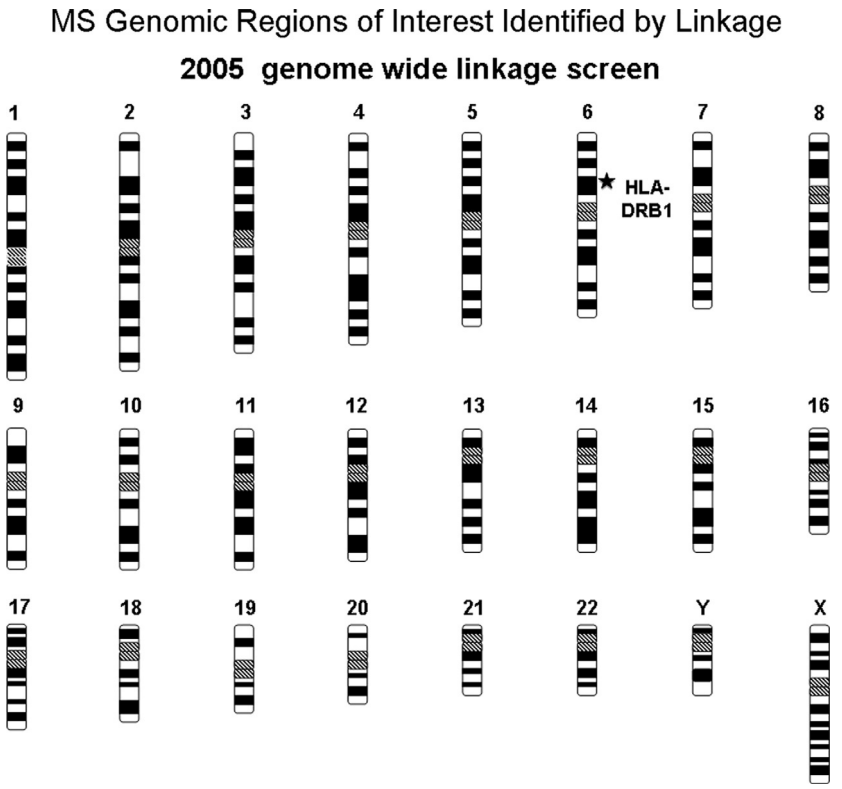

Fig. 9.3. Mutiple sclerosis genomic regions of interest identified by linkage. (Reproduced from Sawcer et al., 2005.)

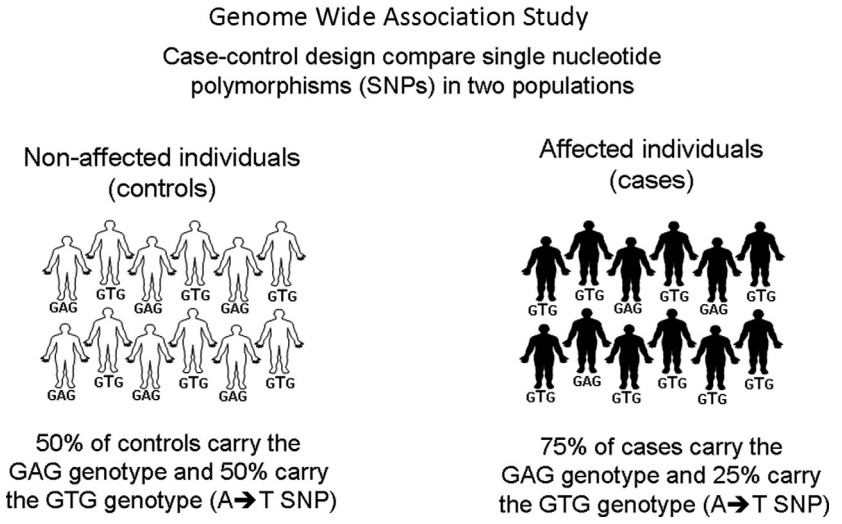

Fig. 9.4. Genome-wide association study. Association analysis compares the prevalence of markers in two populations. In this example the marker of interest, a single nucleotide polymorphism at position 2, is present in 75% of cases and 25% of controls. The odds ratio for the association of this marker with the disease state is therefore 3.0. (Copyright Bruce Cree.)

Unlike linkage analysis that required relatively large effect sizes for tracking heritable traits in families, the newer SNP-based technology was capable of detecting smaller individual genetic effects by increasing the numbers of affected individuals and unaffected controls. Association testing, which compares the prevalence of any given SNP marker between two populations, has the capability of detecting very small genetic effects provided that the numbers of affected and unaffected individuals are sufficiently large. Moreover, the samples for

association screens do not require DNA from family members. Although family structure can still be taken into account in GWAS statistics, doing so is not a prerequisite. This approach simply compares the prevalence of any given SNP marker in cases and controls, which is equivalent to a chi-square statistic. As long as the controls are from the same genetic background as the cases, then statistically significant differences in the prevalence of a particular SNP allele would presumably be due to a disease-related trait.

GWAS identifies the first genes outside the MHC

MS was one of the first diseases to be studied using this new GWAS technique. The IMSGC conducted the first GWAS in 2007 using 334 923 SNPs in 930 MS trio families (a trio family is a MS patient and both parents) with replication datasets consisting of another 609 family trios and an additional 2322 case subjects and 789 unrelated controls (Hafler et al., 2007). It was hoped that this massive, and costly, effort would finally determine the genetic architecture of MS, especially with regard to the non-MHC contributions. As anticipated, the MHC was definitively associated with MS susceptibility; however, beyond the MHC only two other loci were identified with a statistically significant level of confidence. These loci (the first non-MHC loci that were definitely associated with MS risk) encoded genes involved in immune regulation: the interleukin-2 receptor (IL2Rα) and the interleukin-7 receptor (IL7Rα). Associations with MS susceptibility for both loci were subsequently validated in other populations (Matesanz et al., 2001; Gregory et al., 2007; Lundmark et al., 2007a, b; Rubio et al., 2008; Weber et al., 2008).

This landmark achievement established, unquestionably, that genes outside the MHC contributed to MS susceptibility. However, variations at these two alleles, along with those of the MHC, could not account for all of MS heritability. Moreover, the alleles identified were, by definition, present in at least 5% of the population. However, the frequency of MS susceptibility-associated SNPs in both MS cases and controls was surprising. For example, the *IL2R*α variant was present in 88% of patients and 85% of controls. Similarly, the MS-associated *IL7R*α variant was present in 78% of patients and 75% of controls. The finding of these variants in the large majority of controls had two very important implications. First, these variants are the most common polymorphisms of each receptor. Thus, unlike either recessive or dominant mutations, the causal variant does not produce either a loss of function or a gain of an abnormal function. Rather, the protein associated with the polymorphism functions normally. Second, because there was only a very slight overrepresentation of these polymorphisms in MS cases, the effect that this polymorphism has on MS risk is miniscule. Indeed, the odds ratios for these alleles were <1.5. If the other non-HLA MS risk alleles were similarly linked to common SNP variants, then the sample size calculations indicated that variants associated with a 1.1-fold or higher odds of MS risk would require a minimum of 10 000 MS cases and a similar number of controls to be studied (Sawcer, 2008, 2010). Thus, despite the tremendous collaborative effort needed to conduct this study, it was significantly underpowered to identify the genetic variants that contribute to MS.

Although this GWAS identified only two non-HLA loci with a genome-wide level of statistical significance, there were other loci associated with MS susceptibility that just missed the statistical cutoff for definite association. GWAS performed by other groups, as well as meta-analyses that combined GWAS data from different studies, identified multiple other MS susceptibility loci (Burton et al., 2007; Comabella and Martin, 2007; Australia and New Zealand Multiple Sclerosis Genetics Consortium, 2009; Baranzini et al., 2009; De Jager et al., 2009; Jakkula et al., 2010; Nischwitz et al., 2010; Sanna et al., 2010; IMSGC, 2011).

In order to expand the statistical power needed for the next round of GWAS studies, the IMSGC expanded its membership, ultimately including 23 research groups from 15 countries. The IMSGC also partnered with the Wellcome Trust Case Control Consortium 2 (WTCCC2) to make use of the most up-to-date GWAS technology (Sawcer et al., 2011). In the end, 9772 MS cases and 17 376 control DNA samples passed stringent quality control assessments and 441 547 autosomal SNPs were genotyped. During analysis it became clear that the problem of population stratification might bias the analysis. Because this GWAS did not use a family-based approach, the comparison of cases to controls was predicated on the assumption that cases and controls shared a common genomic structure except at the MS susceptibility loci. However, if cases and controls were somewhat different from each other in their genomic structure, then the differences found between cases and controls could be due to either disease-causing loci or to spurious differences in genomic structure. When cases and controls from a single country, such as the United Kingdom, were compared there was no evidence of population stratification. However, because cases and controls were not perfectly matched by country of origin, the entire dataset showed evidence of genomic inflation. Thus, there was a systematic difference for genomic markers between these two groups that would bias the GWAS results. Therefore, several methods to control for genomic inflation were employed but ultimately a novel approach (variance component method) was able to adjust for this genomic inflation bias effectively.

The IMSGC and WTCCC2's MS GWAS identified 52 loci that were definitively associated with MS susceptibility (Table 9.2, Fig. 9.5). This study not only replicated the known MHC, *IL2Rα*, and *IL7Rα* associations but also found 20 loci that had been implicated in MS risk through other GWAS studies as well as meta-analyses. Furthermore, 29 novel loci were identified. All non-MHC loci had only a minor influence on MS susceptibility, with odds ratios ranging from 1.07 to 1.21. Perhaps the most important observation from this study was that the majority of SNPs identified were located near genes encoding immune functions. This observation strongly supported the hypothesis that MS is an autoimmune disease. Furthermore, many of the implicated genes share common pathways involved in immune regulation, providing important clues as to how normal immune function might become dysregulated in MS. Moreover, 23 of the identified loci are known to be involved in other autoimmune diseases, suggesting that common mechanisms are likely to underlie autoimmune diseases in general. However, the identification of these common loci did not lead immediately to an understanding as to why the central nervous system (CNS) is the primary target of autoimmune injury in MS. Nevertheless, several identified genes are expressed in the CNS, and some, such as *GALC*, were previously implicated in MS (Menge et al., 2005).

It is important to understand that, for the majority of the loci, multiple neighboring genes are linked to the MS-associated SNPs. Therefore, with the current level of resolution of this GWAS the exact genetic variant involved in MS susceptibility cannot be determined. Although it is possible that the MS-associated variants are the SNPs identified by the GWAS, it is also possible that the identified SNPs are in linkage disequilibrium with the true MS-associated allele. Additional SNPs or resequencing of these regions of interest will be necessary to refine the map of the causal variants. Perhaps even more puzzling was the finding that two SNP-identified loci do not have any neighboring genes. While it is possible that these SNPs are false-positive results, it is perhaps more likely that these regions contain transcriptional regulatory elements such as promoters or enhancers for distant genes or even are transcribed regulatory RNAs that are not translated into proteins. The recent ENCODE project's remarkable discovery that 80% of the human

Table 9.2

Multiple sclerosis risk-associated non-major histocompatibility complex common genetic variants

Chromosome	SNP	Gene of interest	Immune disease	Known immune function	Neighboring genes	Odds ratio*	Population frequency of risk allele (%)
1	rs4648356	MMEL1	RA, CeD		7	1.16	66.8
1	rs11810217	EVI5			15	1.15	25.7
1	rs11581062	VCAM1		Yes	5	1.07	29.2
1	rs1335532	CD58		Yes	2	1.18	86.3
1	rs1323292	RGS1	CeD	Yes	1	1.12	80.1
1	rs7522462	KIF21B	UC, CeD, CrD		4	1.11	67.3
2	rs12466022	No gene			0	1.16	74.8
2	rs7595037	PLEK	CeD	Yes	4	1.15	54.9
2	rs17174870	MERTK		Yes	7	1.15	73.5
2	rs10201872	SP140	CLL		3	1.15	19.6
2	rs6718520†	THADA			5	1.17	48.0
3	rs11129295	EOMES		Yes	1	1.09	36.3
3	rs669607	No gene			0	1.15	48.7
3	rs2028597	CBLB		Yes	1	1.13	90.7
3	rs2293370	TMEM39A			7	1.16	85.0
3	rs9282641	CD86		Yes	5	1.20	90.2
3	rs2243123	IL12A	CeD	Yes	3	1.09	29.2
5	rs6897932	IL7RA	T1D	Yes	7	1.11	75.7
5	rs4613763	PTGER4	CrD	Yes	1	1.21	16.8
5	rs2546890	IL12B	PS, CrD	Yes	4	1.15	56.2
6	rs12212193	BACH2	CeD, T1D		1	1.08	47.8

Table 9.2

Continued

Chromosome	SNP	Gene of interest	Immune disease	Known immune function	Neighboring genes	Odds ratio*	Population frequency of risk allele (%)
6	rs802734	THEMIS	CeD		5	1.13	70.8
6	rs13192841	OLIG3				1.10	23.5
6	rs11154801	MYB			3	1.09	39.7
6	rs17066096	IL22RA2			3	1.14	18.1
6	rs1738074	TAGAP	CeD		2	1.14	53.5
7	rs354031	ZNF767			4	1.14	23.5
8	rs1520333	IL7		Yes	3	1.11	24.1
8	rs4410871	MYC			2	1.09	71.2
8	rs2019960	PVT1			1	1.16	24.3
9	rs2150702†	MLANA			10	1.16	49.0
10	rs3118470	IL2RA	RA	Yes	4	1.12	31.0
10	rs1250542	ZMIZ1	CeD, IBD		3	1.10	37.0
10	rs7923837	HHEX	T2D	Yes	3	1.09	63.3
11	rs650258	CD6			4	1.12	63.8
12	rs1800693	TNFRSF1A			4	1.12	48.2
12	rs10466829	CLECL1	T1D		9	1.12	46.9
12	rs12368653	CYP27B1	RA	Yes	33	1.11	44.7
12	rs949143	MPHOSPH9			13	1.08	33.2
14	rs4902647	ZFP36L1	CeD, T1D		3	1.13	56.2
14	rs2300603	BATF			3	1.08	70.4
14	rs2119704	GALC			3	1.12	93.2
16	rs7200786	CLEC16A	T1D		8	1.15	54.0
16	rs13333054	IRF8		Yes	1	1.12	20.8
17	rs9891119	STAT3	CrD	Yes	25	1.10	38.9
18	rs7238078	MALT1		Yes	2	1.14	79.6
19	rs1077667	TNFSF14		Yes	3	1.14	78.6
19	rs8112449	TYK2	T1D		12	1.10	69.5
19	rs874628	MPV17L2			11	1.07	71.7
19	rs2303759	DKKL1			9	1.11	29.6
20	rs2425752	CD40	RA	Yes	13	1.10	27.0
20	rs2248359	CYP24A1			2	1.11	58.8
22	rs2283792	MAPK1		Yes	9	1.12	52.7
22	rs140522	SCO2			15	1.12	34.5

*All p-values associated with each odds ratio are $<1 \times 10^{-8}$, the genomic level of significance, i.e., Bonferroni correction for one million possible variants across the genome, a current estimate for all current genomic variants.

†These loci identified by recent meta-analysis (Patsopoulos et al., 2011).

CeD, celiac disease; CLL, chronic lymphocytic leukemia; CrD, Crohn's disease; IBD, inflammatory bowel disease; PS, psoriasis; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism; T1D, type 1 diabetes; T2D, type 2 diabetes; UC, ulcerative colitis.

genome contains elements linked to biologic processes underscores that DNA regions without open reading frames can be biologically important and in fact do not contain what previously had been disregarded as "junk" DNA (Djebali et al., 2012; Bernstein et al., 2012).

Missing heritability

Despite the remarkable achievement of the IMSGC–WTCC2 GWAS, the estimate of the total contribution to MS heritability by the polymorphisms identified in this study was only 25%. Given that the MHC itself accounts for ~20% of MS heritability, the total contribution of the other 51 genetic loci to MS risk is only 5%. This suggests that 75% of MS genetic risk will be accounted for by variants that cannot be identified using SNP chips that are designed to test the common allele–common variant hypothesis. The identification of rare disease-causative alleles, that have weak or modest effects, poses an additional challenge for MS genetic analysis. First, the

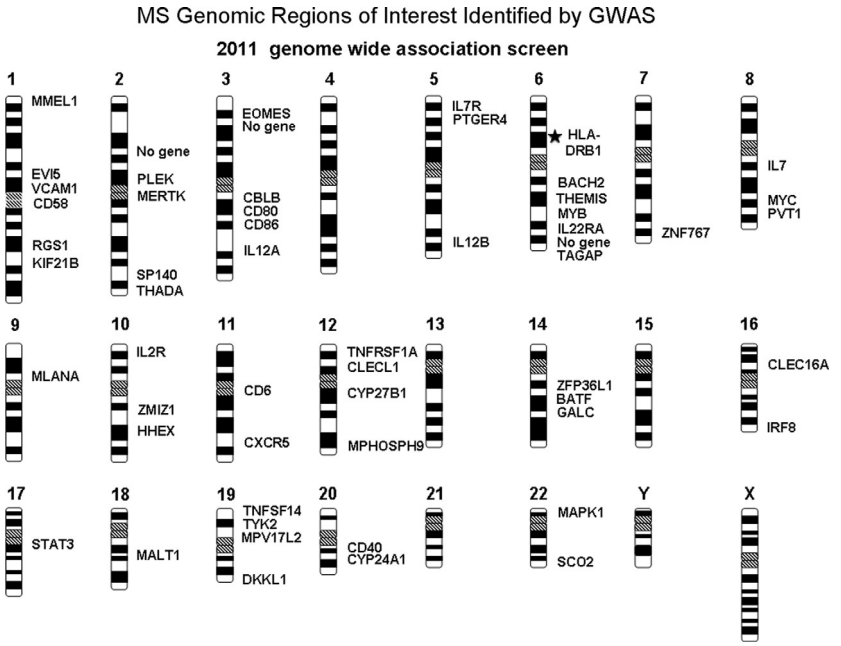

Fig. 9.5. Multiple sclerosis (MS) genomic regions of interest identified by genome-wide association study (GWAS). (Reproduced from IMSGC, 2011.)

number of rare variants is much greater than the number of common variants. Second, the majority of rare variants have not yet been described in publicly available databases. Identification and cataloguing of these rare variants will require sequencing many more genomes. Finally, optimal methods for typing an individual's DNA for rare variants are still being developed. It remains to be determined whether rare variant SNP chips or other techniques such as individual exome or whole-genome sequencing will be the most effective method for identifying rare disease-associated alleles.

Vitamin D genetics

Vitamin D deficiency is a risk factor for MS susceptibility and likely accounts for some of the differences in MS's geographic prevalence. Multiple studies showed that 25-OH vitamin D levels are lower in MS cases compared to controls in both European-descended and African American populations (Munger et al., 2006; Islam et al., 2007; Gelfand et al., 2011; Ramagopalan et al., 2011b). The IMSGC–WTCC2 GWAS identified two genes involved in vitamin D metabolism as conferring increased susceptibility to MS. *CYP27B1* encodes an enzyme that catalyzes the synthesis of 1,25-dihydroxyvitamin D (1,25-dihydroxyvitamin D is the biologically active form of the vitamin). *CYP24A1* encodes an enzyme that degrades 1,25-dihydroxyvitamin D. Given that vitamin D deficiency is a risk factor for MS and that two genes (*CYP27B1* and *CYP24A1*), that regulate vitamin D synthesis and degradation, confer susceptibility to MS, it seems likely that these genes contribute to MS risk by decreasing levels of active vitamin D. *CYP27B1* was also shown to influence MS risk in MS families with multiply affected individuals (Ramagopalan et al., 2011a). By systematically sequencing all genomic protein-encoding regions in 43 MS patients from multiplex families, a non-synonymous variant of *CYP27B1* (R389H) was identified segregating in one family with an incompletely penetrant dominant inheritance pattern. This variant leads to complete loss of *CYP27B1* activity and therefore causes low levels of 1,25-dihydroxyvitamin D. The R389H *CYP27B1* variant was genotyped in 3000 parent-affected trios and was transmitted from parent to affected offspring in 19 trios. Other studies have not yet found an MS association for this rare allele. Nonetheless, taken together, these results underscore the important role of vitamin D in MS and show that not only environmental factors but also genetic factors influence vitamin D levels.

Vitamin D itself has important regulatory roles in gene expression. RNA expression level of the major MS susceptibility gene *HLA-DRB1*15:01* is regulated by vitamin D, albeit somewhat paradoxically (i.e., expression of the allele is upregulated by vitamin D) (Ramagopalan et al., 2009). Vitamin D receptor-binding elements have been identified in the majority of MS-associated genes, implying that expression of many of these genes could also be controlled by vitamin D (Ramagopalan et al., 2010). Although the details of

the network of interactions between genes that regulate vitamin D synthesis and MS susceptibility genes, whose expression is in turn regulated by vitamin D, have yet to be established, these studies illustrate the importance of vitamin D in MS pathogenesis. Low levels of vitamin D, either because of environmental factors such as decreased sunlight exposure or low dietary intake of vitamin D, or because of genetic traits that reduce levels of 1,25-dihydroxyvitamin D, clearly contribute to MS susceptibility and may also contribute to disease activity (Smolders et al., 2008; Soilu-Hanninen et al., 2008; Mowry et al., 2010; Simpson et al., 2010).

EXOME AND GENOME SEQUENCING

Another example of exome sequencing's power to detect rare alleles of MS susceptibility genes is the discovery of a missense mutation in the *TYK2* gene (Dyment et al., 2012). An allele of the *TYK2* gene was previously found to be protective in GWAS (Australia and New Zealand Multiple Sclerosis Genetics Consortium, 2009; Ban et al., 2009; Mero et al., 2010). In contrast, the missense allele identified by exome sequencing modestly increases the risk of MS. Similar to the study of *CYP27B1*, the rare allele of *TYK2* (rs557627444) was first identified in a multiply-affected large MS pedigree and then replicated in 2104 trios.

The studies of *CYP27B1* and *TYK2* showcase the powerful advantage of exome sequencing. Unlike the SNP chip-based technology that is restricted by the genetic variants imprinted on the chip, sequencing the exome has the potential to identify any variant present within an individual's coding DNA. Therefore exome-sequencing technology has the potential to identify rare coding variants that would not be identified using SNP chips.

Individual genome sequencing is also now possible. The cost of high-throughput sequencing has dramatically decreased since the first human genome was sequenced (~US$1 billion) and currently runs approximately $3500/genome. It is anticipated that in the next few years the price will fall further to less than $1000/genome. The advantage of genome sequencing over exome sequencing is that the entire genome is sequenced, which includes all the non-coding DNA that may contain important regulatory elements in addition to the sequences used to encode specific proteins. Not accounted for by the relatively low cost of sequencing is the added cost of data management and analysis for the additional non-coding sequences.

Although the technology for determining genetic sequences has rapidly progressed such that it will soon be commercially feasible to sequence any individual's entire genome, the analytic techniques for interpreting the massive amounts of data are still being developed. The technology for deriving the primary sequence has temporarily outpaced the technology for genomic data analysis. Both software and hardware computational technologies are being developed that will enable desktop analysis of the human genome's 3 billion basepairs.

Preliminary studies suggest that every individual's DNA contains over 50 000 SNP variants and over 5000 insertion/deletion polymorphisms (Baranzini et al., 2010). Importantly, 42% of the SNPs and 86% of the insertion/deletion polymorphisms are novel, meaning that they had not been previously recorded in publicly available databases. Given the very large numbers of rare polymorphisms contained in every individual's DNA, assigning disease-causative roles for these variants poses considerable methodologic challenges. The identification of the rare R389H *CYP27B1* allele was made possible by the additional information imparted by the multiplex family structure and was validated by large-scale trio analysis. It is likely that additional rare coding variants will be identified by this approach. That whole-genome sequencing identifies many novel rare polymorphisms raises the question as to whether some of these variants could have relatively strong disease associations within a given family. Current technology does not offer an obvious solution to proving that so-called "private" disease-causing mutations, if such variants even exist, influence MS risk. If such variants exist the strategy of proving the disease-causative association for the specific allele of interest will fail because the specific allele will be found only within a very small minority of affected individuals. It may be nearly impossible to prove such associations using standard genetic methods. Nonetheless, unique alleles might aggregate within the same genes across many affected individuals. If the intergenic aggregation of unique alleles occurs more often than chance then the within-gene clustering may suggest association of the gene with the trait of interest.

The MHC and MS susceptibility

The MHC is 3.5 million bases (Mb) of DNA located on the short arm of chromosome 6. It is the most genetically dense area of the human genome and encodes over 3000 genes. The HLA genes are grouped into three structurally related classes from the telomere to the centromere: class I, class III, and class II (Fig. 9.6). The HLA genes encode for glycoproteins that are expressed on the cell surface and play critical roles in recognition of self-antigens by the immune system. Many of these genes are highly polymorphic, adding additional complexity to this locus. Multiple autoimmune diseases have risk alleles that map to this region, thereby

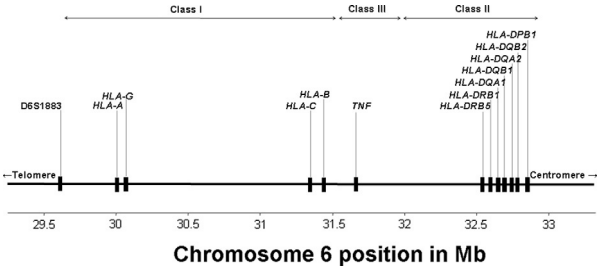

Fig. 9.6. Human leukocyte antigen (HLA) genes. (Copyright Bruce Cree.)

underscoring its importance in the regulation of immune function (Rioux et al., 2009).

As a consequence of selective pressures or other factors, the MHC (especially in northern European populations) is characterized by extensive linkage disequilibrium that can span the MHC and confounds mapping studies. Thus, alleles of class I genes can be genetically linked to distant alleles of class II genes. Many of the HLA alleles were first identified serologically and gave rise to a complex and often inconsistent nomenclature. Recent extensive efforts were made at mapping the serologic types to DNA sequences and the genetic architecture of the MHC is now much better understood (Allcock et al., 2002; Stewart et al., 2004; Miretti et al., 2005; de Bakker et al., 2006; Horton et al., 2008). As a consequence, the genetic basis of the serotypes has been defined and a consistent nomenclature has finally come into focus that will help advance further study of the MHC (http:/hla.alleles.org/).

As described above, the first genetic associations identified in MS were found for HLA class I alleles using serologic typing of HLA antigens on leukocytes (Bertrams and Kuwert, 1972; Bertrams et al., 1972; Jersild et al., 1972, 1973; Naito et al., 1972). When HLA class II alleles were also associated with MS susceptibility, it was proposed that the class I associations were accounted for by linkage disequilibrium with the class II loci (Winchester et al., 1975; Compston et al., 1976; Terasaki et al., 1976). Clarifying the associations of HLA loci with MS susceptibility was ultimately made possible by DNA-based typing of HLA polymorphisms in multiple datasets.

It is now clear that the primary MS susceptibility signal at HLA stems from the MHC class II locus. In European-descended populations the primary risk allele is *HLA-DRB1*15:01*, that is, part of a haplotype: *DRB1*15:01*, *DQA1*01:02*, *DQB1*0602*. This haplotype encodes for cell surface glycoproteins that can present antigen peptides to T cells. Together, these genes correspond to the serologic markers known as HLA-DR2, DQ6.

Fine mapping studies indicate that the most important contributors to MS susceptibility are polymorphisms in the *HLA-DRB1* gene (Oksenberg et al., 2004). Neighboring polymorphisms in the *HLA-DQB1* gene, although tightly linked to DRB1, do not contribute to MS risk, thus establishing a centromeric boundary for MS risk at *HLA-DRB1*.

Multiple polymorphisms within *HLA-DRB1* influence MS susceptibility in populations (Fig. 9.6). *HLA-DRB1*1501* contributes to MS susceptibility with a dominant, dose-dependent effect (Barcellos et al., 2003). In African-descended populations, the closely related *HLA-DRB1*15:03* allele contributes to MS risk (Oksenberg et al., 2004; Cree et al., 2009). *HLA-DRB1*03* contributes to MS risk as a recessive trait (Ramagopalan et al., 2009). *HLA-DRB1*13:03* also contributes to MS risk (Sawcer et al., 2011). In the presence of *HLA-DRB1:15*, the *HLA-DRB1*08* allele further increases MS risk (Dyment et al., 2005; Barcellos et al., 2006; Chao et al., 2007), whereas the *HLA-DRB1*14* and *HLA-DRB1*10* alleles attenuate the risk of MS transmitted by *HLA-DRB1*15* (Fig. 9.7). To add to the complexity, certain alleles seem to contribute to MS in some, but not all, populations. In Sardinia, in addition to *HLA-DRB1*15* and *HLA-DRB1*03*, *HLA-DRB1*04* alleles contribute to MS susceptibility (Marrosu et al., 1988, 1998, 2001; Brassat et al., 2005). Although the allelic interactions at this class II MHC locus are remarkably complex, similar principles will likely apply to other risk loci.

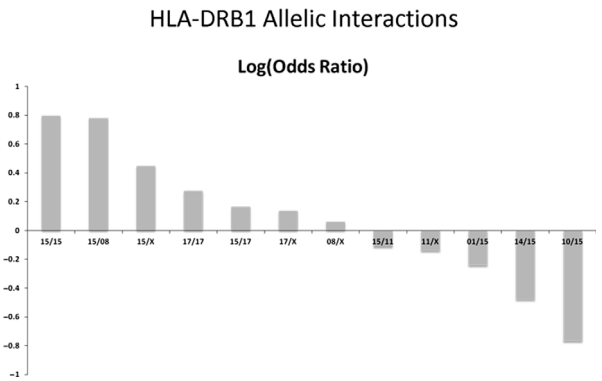

Fig. 9.7. Combinations of *HLA-DRB1* alleles and risk of multiple sclerosis (MS). Odds ratios for MS risk and various combinations of *HLA-DRB1* alleles are depicted graphically. The *y*-axis shows the log of the odds ratio, with positive values increasing the risk of MS and negative values decreasing the risk of MS. The highest odds ratios are for *HLA-DRB1*15* homozygotes and for *HLA-DRB1*15/HLA-DRB1*08* heterozygotes. In contrast, *HLA-DRB1*14* and *HLA-DRB1*10* alleles are associated with a lower risk of MS in *HLA-DRB1*15* heterozygotes. The graph depicts that the impact of the *HLA-DRB1*08* allele on increasing MS risk is proportionately as strong as that of *HLA-DRB1*10* in lowering MS risk in *HLA-DRB1*15* heterozygotes. (Adapted from data presented in Dyment et al., 2005.) (Copyright Bruce Cree.)

Fine mapping studies of the telomeric boundary of MS susceptibility found that the MHC class I locus independently contributes to MS susceptibility, although the exact risk gene, or genes, in this region has not been precisely mapped (Brynedal et al., 2007; Yeo et al., 2007; Cree et al., 2010; Sawcer et al., 2011). Alleles of the class I gene *HLA-A* are proposed to have a protective effect for MS susceptibility. However, as with the class II locus, extensive linkage disequilibrium is present in the class I region and therefore the class I signal might stem from linked alleles of neighboring genes, including *HLA-B* (Healy et al., 2010). *HLA-C* (Yeo et al., 2007) and *HLA-G* (Cree et al., 2010). Alleles of these genes form a linked haplotype that spans the MHC: *HLA-A*02:01–HLA-B*44:02–HLA-C*05:01–HLA-DRB1*04:01*. Thus far, no study to date has had a sufficiently large number of subjects to establish definitively the precise location of the class I MS susceptibility signal. MHC class I locus influences MS susceptibility, which implies a role for innate immune function in MS.

Genotype–phenotype correlations

In addition to influencing MS risk, HLA alleles may contribute to the MS phenotype. The most consistent effect of HLA on MS phenotype is for the *HLA-DRB1*15:01* allele on the age of onset. Several studies showed that this allele decreases the age of onset and does so in a dose-dependent manner (Celius et al., 2000; Masterman et al., 2000; Hensiek et al., 2002; Smestad et al., 2007; Cree et al., 2009; Sawcer et al., 2011). *HLA-DRB1*15:01* may also contribute to the radiographic burden of disease on T2-weighted brain magnetic resonance imaging (MRI) and impact cognitive performance (Okuda et al., 2009). In contrast, the *HLA-B*44:02* allele, that is proposed to be protective for MS susceptibility, may reduce the radiographic burden of disease on T2-weighted brain MRI (Healy et al., 2010). A consistent impact of HLA alleles on MS neurologic impairments, as measured by expanded disability status scale progression, has not been found (Marrosu et al., 1988; Romero-Pinel et al., 2011). Interestingly, a spontaneously occurring null allele of the *HLA-DRB5* gene located telomeric to *HLA-DRB1* might contribute to MS severity (Caillier et al., 2008). However, the IMSGC GWAS efforts have thus far found no definite associations between non-HLA loci and disease severity, age of onset, or disease course (relapsing versus primary progressive).

CURRENT DIRECTIONS AND LIMITATIONS

Several genetic loci were not identified by the recent IMSGC GWAS but only narrowly missed the cutoffs for genome-wide levels of statistical significance (*NFKB1*, *CXCR5*, *SOX8*, *RPS6KB1*, and *TNFRSF6B*). A recent meta-analysis also found several candidate loci with suggestive evidence of association (*TBX21*, *EPS15L1*, *TNP2*, and an intergenic SNP rs9596270) (Patsopoulos et al., 2011). One previously identified gene (*CXCR4*) may have been missed due to a genotyping error (i.e., this might be a false negative). Several of these genes have known functions in immune regulation (e.g., *NFKB1*, *CXCR5*, *TNFRSF6B*, *CXCR4*, and *TBX21*) and efforts are underway using meta-analyses to establish whether these loci contribute to MS risk.

Saturated SNP studies are being performed to reduce the number of possible genes at each loci identified by the tagging SNPs. The results of the first large-scale targeted saturation SNP analysis (the IMMUNOCHIP) are expected soon.

Rare variants with low mean allelic frequencies (MAFs) are being examined by GWAS and exome and genome sequencing studies in multiplex families to identify novel rare variants. The success of exome sequencing in identifying rare variants of *CYP27B1* and *TYK2* illustrates the limitations of the common variant hypothesis. These rare alleles may only be the tip of the iceberg, with many more rare MS variants yet to be found by exploiting this strategy. However, these rare variants were identified in families with multiply affected members. Multiply affected families are relatively uncommon in MS and not all such families have informative structures for identification of rare variants. Furthermore, risk alleles in multiply affected families might be expected to have stronger effect sizes than variants that contribute to sporadic MS. At this time, no solution is apparent that does not require extremely large sample sizes using the case-control approach.

In populations of non-European descent, studies have replicated some, but not all, non-MHC SNPs (Johnson et al., 2010). It is encouraging that at least some MS risk alleles identified in European-descended populations replicate in other racial groups and further substantiates that such risk alleles are genuine. Several explanations for the lack of replication in diverse populations are possible. These include the relatively smaller sample sizes resulting in lack of power and an inadequate control for population stratification, epistatic interactions, and/or gene–environment interactions. Moreover, European-descended alleles are present in the control groups for non-European populations, which indicates that genetic diversity alone does not account for the divergence in genotype to phenotype correlation.

As impressive as the advances in recent years have been for MS risk gene discovery, several limitations of genetic research in MS have become clear. First, it is

highly unlikely that genotyping will yield clinically useful diagnostic tools. The effect sizes for genetic effects on MS risk are far too small to find application in diagnosing MS patients. Second, as ever larger sample sizes became necessary for identifying the genetic differences between MS cases and controls, it seems likely that equally large (or even larger) sample sizes will be necessary to identify those genetic factors that influence the MS clinical phenotype (e.g., whether the disease is relapsing or progressive at onset, whether the disease is predominantly spinal or cerebral, the rate of progression). As intriguing as the preliminary studies of MS risk genes on MRI and clinical correlates may be, the studies performed to date are underpowered and reported associations are likely false-positive results (similar to the experience in early studies of MS risk). Given that multiple factors influence the MS phenotype, including the widespread use of disease-modifying therapies, future studies of genotype–phenotype correlations may face even greater challenges than studies of MS risk. Similarly, although use of genetics as a tool for individualizing treatment selection holds intrinsic appeal, proof that a genetic marker is associated with a particular outcome will require careful study of large populations, especially because the contribution of genetics to the overall treatment effect of highly potent immune therapies is likely to be small. These daunting challenges raise the question as to whether genetic study in MS has therapeutic or prognostic value. Although MS genetics is unlikely to yield concrete clinical utility in the near future, genetics remains an invaluable tool for defining the key constituents underlying the complex biology of the disease. Only through identifying the genes involved in determining MS susceptibility can we hope to understand the role of heritability in the disease pathogenesis, which in turn may point to new therapeutic opportunities.

References

Akesson E, Oturai A, Berg J et al. (2002). A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis. Genes Immun 3: 279–285.

Allcock RJ, Atrazhev AM, Beck S et al. (2002). The MHC haplotype project: a resource for HLA-linked association studies. Tissue Antigens 59: 520–521.

Alter M, Kahana E, Zilber N et al. (2006). Multiple sclerosis frequency in Israel's diverse populations. Neurology 66: 1061–1066.

Australia and New Zealand Multiple Sclerosis Genetics Consortium (2009). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet 41: 824–828.

Ban M, Stewart GJ, Bennetts BH et al. (2002). A genome screen for linkage in Australian sibling-pairs with multiple sclerosis. Genes Immun 3: 464–469.

Ban M, Goris A, Lorentzen AR et al. (2009). Replication analysis identifies TYK2 as a multiple sclerosis susceptibility factor. Eur J Hum Genet 17: 1309–1313.

Baranzini SE, Wang J, Gibson RA et al. (2009). Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet 18: 767–778.

Baranzini SE, Mudge J, van Velkinburgh JC et al. (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464: 1351–1356.

Barcellos LF, Oksenberg JR, Begovich AB et al. (2003). HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. Am J Hum Genet 72: 710–716.

Barcellos LF, Sawcer S, Ramsay PP et al. (2006). Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet 15: 2813–2824.

Bernstein BE, Birney E, Dunham I et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.

Bertrams J, Kuwert E (1972). HL-A antigen frequencies in multiple sclerosis. Significant increase of HL-A3, HL-A10 and W5, and decrease of HL-A12. Eur Neurol 7: 74–78.

Bertrams J, Kuwert E, Liedtke U (1972). HL-A antigens and multiple sclerosis. Tissue Antigens 2: 405–408.

Botstein D, White RL, Skolnick M et al. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331.

Brassat D, Salemi G, Barcellos LF et al. (2005). The HLA locus and multiple sclerosis in Sicily. Neurology 64: 361–363.

Brynedal B, Duvefelt K, Jonasdottir G et al. (2007). HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PLoS One 2: e664.

Burton PR, Clayton DG, Cardon LR et al.; Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium (TASC) (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 39: 1329–1337.

Caillier SJ, Briggs F, Cree BA et al. (2008). Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. J Immunol 181: 5473–5480.

Carton H, Vlietinck R, Debruyne J et al. (1997). Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. J Neurol Neurosurg Psychiatry 62: 329–333.

Celius EG, Harbo HF, Egeland T et al. (2000). Sex and age at diagnosis are correlated with the HLA-DR2, DQ6 haplotype in multiple sclerosis. J Neurol Sci 178: 132–135.

Chao MJ, Barnardo MC, Lui GZ et al. (2007). Transmission of class I/II multi-locus MHC haplotypes and multiple sclerosis susceptibility: accounting for linkage disequilibrium. Hum Mol Genet 16: 1951–1958.

Cohen D, Cohen O, Marcadet A et al. (1984). Class II HLA-DC beta-chain DNA restriction fragments differentiate among HLA-DR2 individuals in insulin-dependent diabetes and multiple sclerosis. Proc Natl Acad Sci U S A 81: 1774–1778.

Comabella M, Martin R (2007). Genomics in multiple sclerosis – current state and future directions. J Neuroimmunol 187: 1–8.

Compston A (1997). Genetic epidemiology of multiple sclerosis. J Neurol Neurosurg Psychiatry 62: 553–561.

Compston A, Coles A (2002). Multiple sclerosis. Lancet 359: 1221–1231.

Compston DA, Batchelor JR, McDonald WI (1976). B-lymphocyte alloantigens associated with multiple sclerosis. Lancet 308: 1261–1265.

Coraddu F, Sawcer S, D'Alfonso S et al. (2001). A genome screen for multiple sclerosis in Sardinian multiplex families. Eur J Hum Genet 9: 621–626.

Cree BA, Reich DE, Khan O et al. (2009). Modification of multiple sclerosis phenotypes by African ancestry at HLA. Arch Neurol 66: 226–233.

Cree BA, Rioux JD, McCauley JL et al. (2010). A major histocompatibility class I locus contributes to multiple sclerosis susceptibility independently from HLA-DRB1*15:01. PLoS One 5: e11296.

Curtius FSH (1938). Multiple sklerose und erbanlage. Z Gesamte Neurol Psychiatrie 160: 226–245.

Davenport CB (1922). Multiple sclerosis from the standpoint of geographic distribution and race. Arch Neurol 8: 51–58.

de Bakker PI, McVean G, Sabeti PC et al. (2006). A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 38: 1166–1172.

De Jager PL, Jia X, Wang J 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.

Dean G, McLoughlin H, Brady R et al. (1976). Multiple sclerosis among immigrants in Greater London. Br Med J 1: 861–864.

Detels R, Visscher BR, Malmgren RM et al. (1977). Evidence for lower susceptibility to multiple sclerosis in Japanese-Americans. Am J Epidemiol 105: 303–310.

Djebali S, Davis CA, Merkel A et al. (2012). Landscape of transcription in human cells. Nature 489: 101–108.

Dyment DA, Willer CJ, Scott B et al. (2001). Genetic susceptibility to MS: a second stage analysis in Canadian MS families. Neurogenetics 3: 145–151.

Dyment DA, Ebers GC, Sadovnick AD (2004a). Genetics of multiple sclerosis. Lancet Neurol 3: 104–110.

Dyment DA, Sadovnick AD, Willer CJ et al. (2004b). An extended genome scan in 442 Canadian multiple sclerosis-affected sibships: a report from the Canadian Collaborative Study Group. Hum Mol Genet 13: 1005–1015.

Dyment DA, Herrera BM, Cader MZ et al. (2005). Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. Hum Mol Genet 14: 2019–2026.

Dyment DA, Cader MZ, Chao MJ et al. (2012). Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene. Neurology 79: 406–411.

Ebers GC, Bulman DE, Sadovnick AD et al. (1986). A population-based study of multiple sclerosis in twins. N Engl J Med 315: 1638–1642.

Ebers GC, Sadovnick AD, Risch NJ; Canadian Collaborative Study Group (1995). A genetic basis for familial aggregation in multiple sclerosis. Nature 377: 150–151.

Ebers GC, Kukay K, Bulman DE et al. (1996). A full genome search in multiple sclerosis. Nat Genet 13: 472–476.

Ebers GC, Yee IM, Sadovnick AD et al. (2000). Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group. Ann Neurol 48: 927–931.

Ebers GC, Sadovnick AD, Dyment DA et al. (2004). Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet 363: 1773–1774.

Eichorst H (1896). Über infantile und heriditare multiple Sklerose. Virchow's Arch Pathol Anat Berl 146: 173–192.

Eraksoy M, Hensiek A, Kurtuncu M et al. (2003). A genome screen for linkage disequilibrium in Turkish multiple sclerosis. J Neuroimmunol 143: 129–132.

GAMES; Transatlantic Multiple Sclerosis Genetics Cooperative (2003). A meta-analysis of whole genome linkage screens in multiple sclerosis. J Neuroimmunol 143: 39–46.

Gelfand JM, Cree BA, McElroy J et al. (2011). Vitamin D in African Americans with multiple sclerosis. Neurology 76: 1824–1830.

Gowers WR (1893). A manual of diseases of the nervous system. 2nd edn. Churchill, London.

Gregory SG, Schmidt S, Seth P et al. (2007). Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet 39: 1083–1091.

Guo SW (2002). Sibling recurrence risk ratio as a measure of genetic effect: caveat emptor! Am J Hum Genet 70: 818–819.

Hader WJ (1982). Prevalence of multiple sclerosis in Saskatoon. Can Med Assoc J 127: 295–297.

Hafler DA, Compston A, Sawcer S et al.; International Multiple Sclerosis Genetics Consortium (2007). Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 357: 851–862.

Haines JL, Ter-Minassian M, Bazyk A et al. (1996). A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. Nat Genet 13: 469–471.

Hansen T, Skytthe A, Stenager E et al. (2005). Risk for multiple sclerosis in dizygotic and monozygotic twins. Mult Scler 11: 500–503.

Healy BC, Liguori M, Tran D et al. (2010). HLA B*44: protective effects in MS susceptibility and MRI outcome measures. Neurology 75: 634–640.

Hensiek AE, Sawcer SJ, Smilie B et al. (2002). HLA-DR 15 is associated with female sex and younger age at diagnosis in multiple sclerosis. J Neurol Neurosurg Psychiatry 72: 184–187.

Hensiek AE, Roxburgh R, Smilie B et al. (2003). Updated results of the United Kingdom linkage-based genome screen in multiple sclerosis. J Neuroimmunol 143: 25–30.

Horton R, Gibson R, Coggill P et al. (2008). Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 60: 1–18.

IMSGC; Briggs FB, Shao X, Goldstein BA et al. (2011). Genome-wide association study of severity in multiple sclerosis. Genes Immun 12: 615–625.

International HapMap Consortium (2003). The International HapMap Project. Nature 426: 789–796.
International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.
Islam T, Gauderman WJ, Cozen W et al. (2006). Differential twin concordance for multiple sclerosis by latitude of birthplace. Ann Neurol 60: 56–64.
Islam T, Gauderman WJ, Cozen W et al. (2007). Childhood sun exposure influences risk of multiple sclerosis in monozygotic twins. Neurology 69: 381–388.
Jakkula E, Leppä V, Sulonen AM 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.
Jersild C, Svejgaard A, Fog T (1972). HL-A antigens and multiple sclerosis. Lancet 1: 1240–1241.
Jersild C, Fog T, Hansen GS et al. (1973). Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. Lancet 2: 1221–1225.
Johnson BA, Wang J, Taylor EM et al. (2010). Multiple sclerosis susceptibility alleles in African Americans. Genes Immun 11: 343–350.
Kenealy SJ, Babron MC, Bradford Y et al. (2004). A second-generation genomic screen for multiple sclerosis. Am J Hum Genet 75: 1070–1078.
Kuokkanen S, Gschwend M, Rioux JD et al. (1997). Genomewide scan of multiple sclerosis in Finnish multiplex families. Am J Hum Genet 61: 1379–1387.
Kurtzke JF, Beebe GW, Nagler B et al. (1977). Studies on the natural history of multiple sclerosis – 8. Early prognostic features of the later course of the illness. J Chron Dis 30: 819–830.
Kurtzke JF, Beebe GW, Norman JE Jr (1979). Epidemiology of multiple sclerosis in U.S. veterans: 1. Race, sex, and geographic distribution. Neurology 29: 1228–1235.
Lundmark F, Duvefelt K, Hillert J (2007a). Genetic association analysis of the interleukin 7 gene (IL7) in multiple sclerosis. J Neuroimmunol 192: 171–173.
Lundmark F, Duvefelt K, Iacobaeus E et al. (2007b). Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet 39: 1108–1113.
Marrosu MG, Muntoni F, Murru MR et al. (1988). Sardinian multiple sclerosis is associated with HLA-DR4: a serologic and molecular analysis. Neurology 38: 1749–1753.
Marrosu MG, Murru MR, Costa G et al. (1998). DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population. Hum Mol Genet 7: 1235–1237.
Marrosu MG, Murru R, Murru MR et al. (2001). Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Hum Mol Genet 10: 2907–2916.
Masterman T, Ligers A, Olsson T et al. (2000). HLA-DR15 is associated with lower age at onset in multiple sclerosis. Ann Neurol 48: 211–219.
Matesanz F, Fedetz M, Collado-Romero M et al. (2001). Allelic expression and interleukin-2 polymorphisms in multiple sclerosis. J Neuroimmunol 119: 101–105.
McAlpine D (1946). The problem of disseminated sclerosis. Brain 69: 233–250.
Menge T, Lalive PH, von Büdingen HC et al. (2005). Antibody responses against galactocerebroside are potential stage-specific biomarkers in multiple sclerosis. J Allergy Clin Immunol 116: 453–459.
Merikangas KR, Risch N (2003). Genomic priorities and public health. Science 302: 599–601.
Mero IL, Lorentzen AR, Ban M et al. (2010). A rare variant of the TYK2 gene is confirmed to be associated with multiple sclerosis. Eur J Hum Genet 18: 502–504.
Millar JH, Allison RS (1954). Familial incidence of disseminated sclerosis in Northern Ireland. Ulster Med J 23 (Suppl. 2): 29–92.
Miretti MM, Walsh EC, Ke X et al. (2005). A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am J Hum Genet 76: 634–646.
Mowry EM, Krupp LB, Milazzo M et al. (2010). Vitamin D status is associated with relapse rate in pediatric-onset multiple sclerosis. Ann Neurol 67: 618–624.
Mumford CJ, Wood NW, Kellar-Wood H et al. (1994). The British Isles survey of multiple sclerosis in twins. Neurology 44: 11–15.
Munger KL, Levin LI, Hollis BW et al. (2006). Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA 296: 2832–2838.
Naito S, Namerow N, Mickey MR et al. (1972). Multiple sclerosis: association with HL-A3. Tissue Antigens 2: 1–4.
Nielsen NM, Westergaard T, Rostgaard K et al. (2005). Familial risk of multiple sclerosis: a nationwide cohort study. Am J Epidemiol 162: 774–778.
Nischwitz S, Cepok S, Kroner A et al. (2010). Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. J Neuroimmunol 227: 162–166.
Oger JF, Arnason BG, Wray SH et al. (1975). A study of B and T cells in multiple sclerosis. Neurology 25: 444–447.
Oksenberg JR, Barcellos LF, Cree BA et al. (2004). Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet 74: 160–167.
Okuda DT, Srinivasan R, Oksenberg JR et al. (2009). Genotype–phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. Brain 132: 250–259.
Patsopoulos NA; Bayer Pharma MS Genetics Working Group; Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist; ANZgene Consortium; GeneMSA; International Multiple Sclerosis Genetics Consortium, Esposito F, Reischl J et al. (2011). Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol 70: 897–912.
Peltonen L, McKusick VA (2001). Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291: 1224–1229.
Pratt RT, Compston ND, McAlpine D (1951). The familial incidence of disseminated sclerosis and its significance. Brain 74: 191–232.

Pugliatti M, Sotgiu S, Rosati G (2002). The worldwide prevalence of multiple sclerosis. Clin Neurol Neurosurg 104: 182–191.

Ramagopalan SV, Maugeri NJ, Handunnetthi L et al. (2009). Expression of the multiple sclerosis-associated MHC class II allele HLA-DRB1*1501 is regulated by vitamin D. PLoS Genet 5: e1000369.

Ramagopalan SV, Heger A, Berlanga AJ et al. (2010). A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. Genome Res 20: 1352–1360.

Ramagopalan SV, Dyment DA, Cader MZ et al. (2011a). Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Ann Neurol 70: 881–886.

Ramagopalan SV, Handel AE, Giovannoni G et al. (2011b). Relationship of UV exposure to prevalence of multiple sclerosis in England. Neurology 76: 1410–1414.

Rioux JD, Goyette P, Vyse TJ et al.; International MHC and Autoimmunity Genetics Network (2009). Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proc Natl Acad Sci U S A 106: 18680–18685.

Risch N (1990). Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46: 222–228.

Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. Science 273: 1516–1517.

Ristori G, Cannoni S, Stazi MA et al. (2006). Multiple sclerosis in twins from continental Italy and Sardinia: a nationwide study. Ann Neurol 59: 27–34.

Robertson NP, Fraser M, Deans J et al. (1996). Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. Brain 119: 449–455.

Robertson NP, O'Riordan JI, Chataway J et al. (1997). Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. Lancet 349: 1587–1590.

Romero-Pinel L, Pujal JM, Martínez-Yélamos S et al. (2011). HLA-DRB1: genetic susceptibility and disability progression in a Spanish multiple sclerosis population. Eur J Neurol 18: 337–342.

Rubio JP, Stankovich J, Field J et al. (2008). Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. Genes Immun 9: 624–630.

Sadovnick AD, Baird PA, Ward RH (1988). Multiple sclerosis: updated risks for relatives. Am J Med Genet 29: 533–541.

Sadovnick AD, Ebers GC, Dyment DA et al. (1996). Evidence for genetic basis of multiple sclerosis. Lancet 347: 1728–1730.

Sanna S, Pitzalis M, Zoledziewska M et al. (2010). Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nat Genet 42: 495–497.

Sawcer S (2008). The complex genetics of multiple sclerosis: pitfalls and prospects. Brain 131: 3118–3131.

Sawcer S (2010). Bayes factors in complex genetics. Eur J Hum Genet 18: 746–750.

Sawcer S, Jones HB, Feakes R et al. (1996). A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nat Genet 13: 464–468.

Sawcer SJ, Maranian M, Singlehurst S et al. (2004). Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. Hum Mol Genet 13: 1943–1949.

Sawcer S, Ban M, Maranian M et al. (2005). A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 77: 454–467.

Sawcer S, Hellenthal G, Pirinen M et al.; International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2 (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476: 214–219.

Simpson S Jr, Taylor B, Blizzard L et al. (2010). Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. Ann Neurol 68: 193–203.

Smestad C, Brynedal B, Jonasdottir G et al. (2007). The impact of HLA-A and -DRB1 on age at onset, disease course and severity in Scandinavian multiple sclerosis patients. Eur J Neurol 14: 835–840.

Smestad C, Sandvik L, Holmoy T et al. (2008). Marked differences in prevalence of multiple sclerosis between ethnic groups in Oslo, Norway. J Neurol 255: 49–55.

Smolders J, Menheere P, Kessels A et al. (2008). Association of vitamin D metabolite levels with relapse rate and disability in multiple sclerosis. Mult Scler 14: 1220–1224.

Soilu-Hänninen M, Laaksonen M, Laitinen I et al. (2008). A longitudinal study of serum 25-hydroxyvitamin D and intact parathyroid hormone levels indicate the importance of vitamin D and calcium homeostasis regulation in multiple sclerosis. J Neurol Neurosurg Psychiatry 79: 152–157.

Stewart CA, Horton R, Allcock RJ et al. (2004). Complete MHC haplotype sequencing for common disease gene mapping. Genome Res 14: 1176–1187.

Svenson LW, Woodhead SE, Platt GH (1994). Regional variations in the prevalence rates of multiple sclerosis in the province of Alberta, Canada. Neuroepidemiology 13: 8–13.

Terasaki PI, Park MS, Opelz G et al. (1976). Multiple sclerosis and high incidence of a B lymphocyte antigen. Science 193: 1245–1247.

Wallin MT, Page WF, Kurtzke JF (2004). Multiple sclerosis in US veterans of the Vietnam era and later military service: race, sex, and geography. Ann Neurol 55: 65–71.

Weber JL, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 388–396.

Weber F, Fontaine B, Cournu-Rebeix I et al. (2008). IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. Genes Immun 9: 259–263.

Willer CJ, Dyment DA, Risch NJ et al. (2003). Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci U S A 100: 12877–12882.

Winchester R, Ebers G, Fu SM et al. (1975). B-cell alloantigen Ag 7a in multiple sclerosis. Lancet 2: 814.

Yeo TW, De Jager PL, Gregory SG et al. (2007). A second major histocompatibility complex susceptibility locus for multiple sclerosis. Ann Neurol 61: 228–236.