COMPLEX APPROACHES TO STUDY COMPLEX TRAIT GENETICS IN MULTIPLE SCLEROSIS

Bernadette KÁLMÁN
Center for Research and Education, Markusovszky University Hospital, Szombathely

Multiple sclerosis (MS) is a complex trait disorder defined by several genes and their interactions with environmental factors. A comprehensive exploration of the susceptibility variants had not been feasible until recently when new developments in biotechnology and bioinformatics made possible sequencing of the whole human genome, cataloguing of nucleotide variants and alignments of these variants in haplotypes.
Earlier observations from epidemiological, candidate gene and linkage studies provided ample evidence to support a complex genetic determination of MS. New biotechnology and bioinformatics resources have been recently applied to further successful explorations of the disease. These efforts were paralleled by more careful and reliable ascertainments of disease phenotypes, collaborations among specialized centers to generate sufficient sample size and involvement of clinician-scientists capable of working both on the clinical and scientific study sides. Data obtained from the whole genome association studies (GWAS) elevated our understanding of MS genetics to a new level by identifying an extensive list of genetic determinants. Pathway analyses of MS-associated variants provided evidence to support the immune etiology of the disease. Future research will likely explore how environmental factors interact with the genome, and contribute to the abnormal immune activation and inflammation.
This review summarizes the outcomes of MS genetic explorations including those of recent GWAS, and highlights practical consequences of genetic and genomic studies by pointing out as to how the derived data facilitate further elucidation of MS pathogenesis. A better understanding of disease processes is necessary for future advancements in therapeutics and the development of disease prevention strategies.

Keywords: *complex trait, single nucleotide polymorphism, haplotype, genome wide association study, pathway analyses*

KOMPLEX MEGKÖZELÍTÉSEK A SCLEROSIS MULTIPLEX ÖSSZETETT GENETIKAI JELLEGÉNEK TANULMÁNYOZÁSÁHOZ

Kálmán B, MD, PhD, DSc

Ideggyogy Sz 2014;67(9–10):309–321.

Kapcsolódó cikk online

A sclerosis multiplex (SM) komplex genetikai betegség, melyet több gén határoz meg környezeti faktorokkal történő kölcsönhatásban. A hajlamosító variánsok átfogó feltárására csak nemrégen kerülhetett sor, amikor az új biotechnológiai és bioinformatikai fejlődés eredményeként lehetővé vált a teljes humán genom szekvenciájának meghatározása, a polimorf variánsok katalogizálása és a variánsok haplotípusokba való sorolása.
Korábbi epidemiológiai, jelölt gén és kapcsoltsági tanulmányok alátámasztották az SM komplex genetikai meghatározottságát. Az új biotechnológiai és bioinformatikai eszközök kiaknázása az SM további sikeres feltárásához vezetett. Ezeket az erőfeszítéseket kiegészítette a betegség fenotípusának megbízható meghatározása, a specializált SM-központok közötti kollaborációk kialakítása a megfelelő mintaszám elérése érdekében és olyan alapkutató-klinikusok bevonása a tanulmányokba, akik képesek klinikai és alaptudományi feladatokat is ellátni. A teljes genom asszociációs tanulmányokból nyert adatok az SM megértését új szintre emelték a genetikai determinánsok átfogó azonosításával. A betegséggel asszociálódó variációk biológiai összefüggéseinek elemzéséből meggyőző bizonyítékot nyertünk az SM immuneredetét illetően. A közeljövőben valószínűleg olyan megközelítések vezethetnek az SM további megértéséhez, melyek feltárják, hogy a környezeti faktorok milyen módon állnak kölcsönhatásban a genommal és eredményeznek abnormális immunaktivációt, valamint gyulladást.
Ez az áttekintés összegzi az SM genetikai tanulmányok, köztük a teljes genom asszociációs tanulmányok eredményeit, és kiemeli a genetikai/genomikai tanulmányok gyakorlati jelentőségét abban, milyen módon segíthetik ezen adatok az SM kialakulásának jobb megértését. A betegség kórfolyamatának jobb megértése a terápiás megközelítések jövőbeli fejlődésének és új megelőző stratégiák kialakításának kulcsa lehet.

Kulcsszavak: *komplex trait, nukleotida-pontpolimorfizmus, haplotípus, teljes genom asszociációs tanulmány, biológiai kórfolyamat elemzése*

Correspondence (levelezési cím): Prof. Dr. med. Bernadette KÁLMÁN, PhD, DSc, PTK, ETK, Markusovszky University Hospital; H-9700 Szombathely, Markusovszky Str. 5. E-mail: Kalman.bernadett@markusovszky.hu

Érkezett: 2013. június 11. Elfogadva: 2013. augusztus 22.

www.elitmed.hu

Multiple sclerosis (MS) is characterized by inflammation, demyelination and neurodegeneration of the central nervous system (CNS). The onset of the disease is typically in young adulthood, but may occur at any age from small childhood to late adulthood[1]. Presenting symptoms of motor, visual, sensory, coordination, autonomous and cognitive impairments may occur alone or in combination. The diagnosis is established today based on the McDonald's criteria and its updated versions[2, 3], all with the core requirement to demonstrate dissemination of pathology in time and space. About 85% of MS patients have relapsing-remitting course during the early phase the disease, while the remaining 15% of patients have progressive course with or without superimposed relapses from disease onset. These two latter forms are referred to as progressive-relapsing and primary progressive MS, respectively. Most patients with relapsing-remitting disease onset will later develop secondary progressive course. The older the age of relapsing-remitting disease onset, the higher the probability and the shorter the latency of conversion to secondary progressive disease[1, 4].

Magnetic resonance (MR) imaging not only made feasible to establish the diagnosis earlier and with higher degrees of certainty, but also facilitated gaining much insight into the disease pathogenesis. MS is considered today a disease of the whole CNS, with extensive microscopic pathology present not only within but also outside of plaques, in the so called "normal appearing" white and gray matter[5]. This microscopic pathology includes small foci of inflammation, demyelination, astrogliosis, microglial activation and neuroaxonal degeneration along with complex molecular alterations, and gives signal alterations when using specific MR sequences[6]. These microscopic lesions are the precursors of future macroscopic lesions. Plaques, that had been considered the pathological hallmarks of MS, represent the most severely affected regions of inflammatory demyelination, neurodegeneration and reactive gliosis. Most plaques are visible by the naked eye in brain sections and visualized by conventional MR sequences in clinical routine. As a plaque evolves, its appearances change from inflammatory demyelination and tissue loss to a severe non-inflammatory but rather degenerative and astrogliotic (sclerotic) lesion. Because of the progressive loss of white and gray matter regions secondary to inflammation and demyelination in the CNS, tissue atrophy accumulates in the brain and spinal cord from the earliest stages of MS[5, 6]. Brain atrophy and lesion load strongly predict long term clinical disability[7].

There has been a debate as to whether there is an inter-individual heterogeneity (but intra-individual homogeneity) of pathology or the heterogeneity evolves over time during progressive lesion development[8, 9]. Investigators representing the former view suggest that acute demyelinating lesions of MS may be grouped into four subclasses each consistent in individual patients. These four subgroups include lesions with 1) mononuclear inflammation, 2) mononuclear inflammation and immunoglobulin plus complement deposition, 3) dying back oligodendrocytopathy and apoptosis, and 4) other forms of oligodendrocytopathy and necrosis[8]. Investigators representing the second view argue that the earliest histological lesion involves oligodendrocyte apoptosis and microglial activation in still myelinated tissue with sparse mononuclear infiltration[9]. Based on this observation, demyelination is preceded by oligodendrocyte injury of unknown etiology, and the infiltration by blood-borne mononuclear cells may be secondary to the evolving CNS pathology.

Oligodendrocyte and neuronal apoptosis as well as the neurodegenerative changes, however, have been interpreted in most studies as lesions developing secondary to inflammation, and may involve oxidative damage to mitochondrial macromolecules leading to mitochondrial mechanisms of tissue loss[10, 11].

While recent development in *in vivo* and *in vitro* imaging technologies facilitated to gain much insight into the temporal evolution and molecular characteristics of MS lesions, the causes of immune activation and inflammation are not well understood. Based on the abnormalities observed in the peripheral immune system, MS has been considered an autoimmune disease. The main immune effectors of lesion development appear to be CD4+ T helper (TH) subpopulations, specifically the TH-17 and TH-1 cells with distinct pro-inflammatory cytokine profiles (interleukine[IL]-17and IL2, interferon-gamma, respectively)[12]. In contrast, TH-2 cells with IL-4 and IL-10 cytokine profiles may exert protective regulatory effects. Monocytes, macrophages, dendritic cells and B lymphocytes present various antigens and contribute to the differentiation and activation of TH-17/TH-1 cells[13]. Immunoglobulins specific for myelin, axonal and neuronal molecular components also may contribute to the development of subtypes of lesions[8, 14]. While the disease-relevant antigenic determinants remain to be determined, a deficiency in the regulatory $CD4^+$, $CD25^+$, $Foxp3^+$ Treg cells in the peripheral circulation as well as in the brain seems to play important roles in the abnormal immune

response[15]. Besides abnormalities in the adaptive immune system (e.g. antigen specific CD4+ T cells), functional deficiencies in the innate immune system (e.g. natural killer cells) have also been noted[16]. Evidence suggests the involvements of several environmental factors (early microbial exposures, vitamin D, UV, smoking, etc…) in the pathogenesis of MS, which likely act through interactions with the immune system and contribute to the abnormal activation of T cells[17]. Nevertheless, the exact role and temporal sequence of host-environment interactions in MS have not been sufficiently clarified to unequivocally explain the etiology of the disease.

With the advancement in molecular technologies and the successful exploration of the human genome the hope arose that gaining deeper insights into the characteristics of the genome, transcriptome and epigenome of MS patients will clarify ambiguities concerning the etiology of the disease and facilitate our understanding of its pathogenesis. This survey of genetic and genomic literature, therefore, aims to summarize as to what we have learnt from studying the genome and its products to better explain the pathogenesis of MS and to identify new treatment targets.

Genetic and genomic studies in multiple sclerosis

EPIDEMIOLOGY

Epidemiological studies investigate occurrence (prevalence and incidence), geographic and ethnic distribution, familial recurrence and environmental determinants of a disease, and explore interactions among host and environmental factors. Extensive epidemiological studies have been conducted for over a century in MS, and provided ample evidence to support its genetic etiology, but with significant environmental involvement. The genetic determination of the disease is supported by several observations. First, there are ethnic differences in the occurrence of the disease even among populations living next to each other in the same environment[18–21]. Second, the risk for monozygotic twin-pairs is 300-times, and for first degree relatives of MS patients is 20-40-times higher than for someone in the general population[22–24]. With the decrease of relatedness, the risk for MS is also decreasing. Third, half-sib studies, conjugal pairs and adoption studies confirm that this increased risk exists only for biological relatives[25–27]. Fourth, the inheritance of MS is not compatible with mendelian (autosomal dominant, recessive or X-linked) or mitochondrial transmission patterns. MS is a complex trait defined by interactions among several genes and environmental factors[11, 28].

CASE-CONTROL CANDIDATE GENE ASSOCIATION STUDIES

Since the early 1970-ies, case-control association studies have been conducted in increasing numbers, as improving methods for defining polymorphic alleles at the protein, RNA and DNA levels revealed more and more variants in candidate genes of inflammation and demyelination[11]. These hypothesis-driven association studies selected candidate genes based on the autoimmune hypothesis of inflammatory demyelination in MS, and statistically tested the allele frequency differences within these candidate genes in populations of sporadic cases (patients) and controls, matched for age, gender and ethnicity. Similar to most autoimmune diseases, the involvement of the Human Leukocyte Antigen (HLA) locus in MS is now unambiguously established[11, 29]. Due to a strong linkage disequilibrium in this region, however, it has been difficult to sort out which allele of the HLA-DRB5*0101-HLA-DRB1*1501-HLA-DQA1*0102-HLA-DQB1*0602 MS-associated extended haplotype has the primary role in the disease[30, 31]. This extended haplotype confers a relative risk of =3 in individuals of Northern-European descent (meaning that carriers of the haplotype have a 3X risk for MS compared to non-carriers), but other DRB1 alleles also play important roles in Northern-Europeans and in other ethnic groups as well[31]. At the individual level, epistatic interactions occur among DRB1 susceptibility and resistance allele combinations and contribute to a fine tuning of the overall risk for the disease[31]. Alleles of non-DRB1 HLA genes (e.g. HLA A3 predisposing and HLA A2 protective alleles) also may modify susceptibility to the disease[32–35]. Thus, both HLA allelic (e.g. alleles of the DRB1 gene) and locus (e.g. HLA A locus or HLA DR locus) heterogeneity seem to play roles in defining MS susceptibility. Hypothesis driven, case-control studies also tested allele frequency differences in non-HLA candidate genes involved in immune response and demyelination/remyelination, but revealed less unambiguous associations due to the small effects of these genes, heterogeneity of the disease and often insufficient sample sizes[11].

LINKAGE ANALYSES

The method of linkage became very successful in identifying disease loci with subsequent cloning of

disease-relevant genes in mendelian disorders (caused by highly pathogenic mutation[s] in single genes) from the early 1990-ies. Because of the strong effect of such a mutation, analyzing one or a few large families with multiple affected members usually was sufficient in mendelian disorders. In contrast, the polymorphic genetic variants (not pathogenic mutations) that define complex trait disorders exert small effects, and therefore, several hundreds of families are needed for a linkage study to have statistical outcome. Generally in linkage studies, polymorphic microsatellite markers are genotyped in multiple affected and unaffected individuals in families with more than one patient, and the segregation of microsatellite alleles with the disease trait is determined. Microsatellites are short, 2-6 base pairs sequence repeats. The numbers of these repeats result in length variations and define the microsatellite alleles, which are inherited. The first genome-wide linkage analyses using microsatellite markers in MS were published in 1996[36–38]. These studies identified several loci, each with moderate lod scores, suggesting potential relevance to the disease, and those which were identified in more than one study, particularly attracted attention for follow up. Since then, numerous linkage analyses, with varying degrees of coverage and in families of varying ethnicity, have been carried out[29]. However, this approach revealed that linkage is not the ideal tool to study complex disorders (as opposed to mendelian disorders), due to the weak signals of genes and the genetic heterogeneity among families in complex traits. The numbers of available families were also relatively low suggesting that the weak results were in part related to the lack of power. In addition, the identified susceptibility loci were very large encompassing millions of base pairs often coding for hundreds of genes in the identified chromosomal regions, and linkage as a method was unsuitable for further narrowing down these large susceptibility loci. Nevertheless, linkage data have confirmed the involvement of the HLA Class II region on chromosome 6p21.3 in MS, even though the lod score has not reached the level of formal statistical significance of 3 that is conventionally considered the lowest threshold suggesting linkage (at least in mendelian disorders). All other "loci of interest" had lod scores between 1 and 2 (or below). Linkage as a hypothesis free approach confirmed that multiple loci play a role, MS is a polygenic disease, and the HLA Class II region may have the strongest influence on disease susceptibility. Based on the analyses of HLA allele sharing by descent in sib pairs, the HLA region was estimated to define 20-60% of genetic susceptibility in MS[29, 34, 39].

Conclusions of the microsatellite-based linkage analyses were further confirmed in a single nucleotide polymorphisms (SNP) – based linkage study that used 4500 SNP markers in 750 multiplex MS families and showed a peak lod score of 11.7 in the HLA region (SNPs are sequence variations at certain positions as opposed to the length polymorphism of microsatellite repeats; see more specific definition of SNPs below)[40]. However, even in this highly powered SNP-based linkage study, no other locus showed lod score with genome-wide significance. Linkage as a method had clearly reached its limits in MS, and it was necessary to turn to new methods based on information from the Human Genome Project.

SINGLE NUCLEOTIDE POLYMORPHISMS GENOTYPING AND LINKAGE DISEQUILIBRIUM ANALYSES

Recent explorations of the human genome included not only full sequence determinations, but also systematically collected information on genetic variations by the SNP Consortium Allele Frequency Project and the HapMap Project[41, 42]. The SNP and Hapmap databases are publically available (http://www.ncbi.nlm.nih.gov/snp/ and http://hapmap.ncbi.nlm.nih.gov/). We have learnt from data of the Human Genome Project (http://www.ncbi.nlm.nih.gov/genome) and the SNP and Hapmap studies that the human genome is composed of 3.2 billion base pairs, and comparing the genomic sequences of any two today living individuals, 99.9% similarity may be observed. The 0.1% differences come from genetic variations, the most frequent of them being the single nucleotide polymorphisms (SNPs). As the name suggests, alleles of SNPs are defined by single nucleotide variations at a given position (at such position, pyrimidine-pyrimidine i.e. cytosine/thymine, or purine-purine i.e. adenine/guanine alternate alleles may be detected; or pyrimidine/purine alleles i.e. cytosine/adenine, cytosine/guanine, thymine/adenine, thymine/guanine alternates may also be present). These genetic variations determine phenotypic differences and susceptibility to common diseases in the population. The distribution of SNPs is not even in the genome. Nevertheless, roughly one SNP may be expected for every 1000 nucleotides[43–45].

SNP alleles align in haplotypes that are chromosomal blocks within which recombinations do not occur during meiosis. An offspring carries one copy of each chromosome from his/her mother and father. During meiosis, multiple recombinations or crossing over exchanges occur between the homologous maternal and paternal chromosomes. These

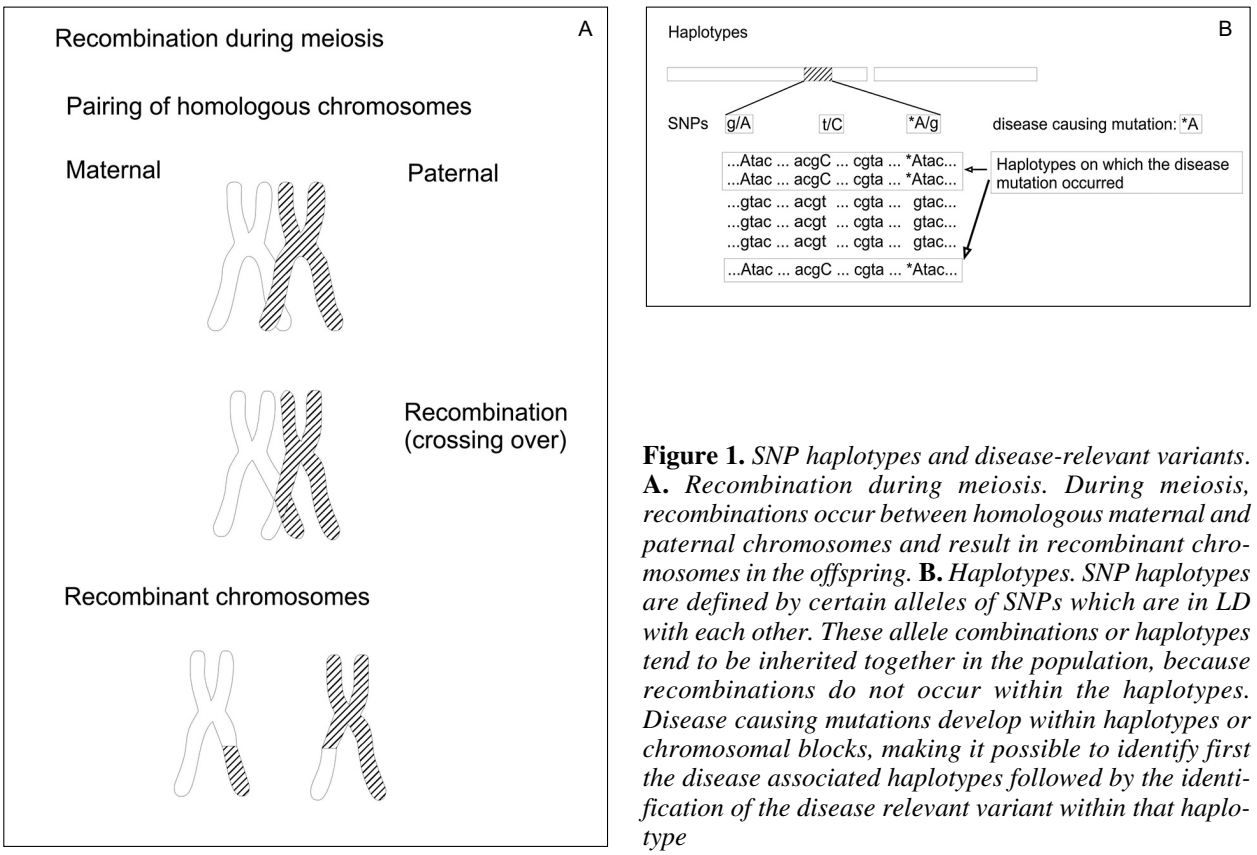

Figure 1. *SNP haplotypes and disease-relevant variants.* **A.** *Recombination during meiosis. During meiosis, recombinations occur between homologous maternal and paternal chromosomes and result in recombinant chromosomes in the offspring.* **B.** *Haplotypes. SNP haplotypes are defined by certain alleles of SNPs which are in LD with each other. These allele combinations or haplotypes tend to be inherited together in the population, because recombinations do not occur within the haplotypes. Disease causing mutations develop within haplotypes or chromosomal blocks, making it possible to identify first the disease associated haplotypes followed by the identification of the disease relevant variant within that haplotype*

exchanges result in mosaics of maternally and paternally derived chromosomal blocks in the offspring's chromosomes **(Figure 1A)**. SNP alleles are in linkage disequilibrium (LD) with each other within particular chromosomal blocks, and these chromosomal blocks, also called haplotypes, tend to be inherited together from parents to offspring through generations in the population. The size distribution of haplotype blocks is related to the population/meiotic history and thus, characteristic of each ethnic group. Older populations (e.g. African Yorubans) are characterized by shorter haplotype blocks than younger populations (e.g. Asians and Europeans). The longer the population history, the more meiosis and more recombinations occurred through subsequent generations, which resulted in shorter haplotype blocks in the chromosomes of the today representatives of that population[41].

Haplotypes, defined by specific allelic combinations in a population may facilitate performing research studies. When a disease-causing mutation arises, it does so within a haplotype **(Figure 1B)**. Defining first the haplotype that is associated with a disease may subsequently help to narrow down and identify the specific disease-causing mutation or SNP variant within that haplotype. Since the marker allele is in LD with the disease allele, this method is often referred to as LD mapping. LD mapping has been extensively used for exploring genetic susceptibility in complex trait disorders.

Our three-stage study focused on the 17q11 chromosomal region in MS exemplifies how this strategy works. We selected the 17q11 region because 1) A meta-analysis of three genome scans revealed the highest nonparametric linkage score (NPL=2.58) at 17q11[46] and two case-control studies suggested involvement of this region in MS[47, 48]; 2) This chromosomal region encodes two clusters of β-chemokine ligands (or CC chemokine ligands - CCLs) within a 1.85 MB chromosomal segment of 17q11; CCLs have been implicated in the pathogenesis of MS as well as in experimental autoimmune encephalomyelitis (EAE, an induced inflammatory demyelinating model of MS)[49]; and 3) EAE studies showed that mouse and rat chromosomal regions synthenic to the human 17q11 confer susceptibility to EAE (just like 17q11, the synthenic mouse and rat chromosomal regions also encode CCL gene clusters)[50, 51].

In the first phase of this three-parts study, we genotyped 32 SNPs within 17q11, in 1085 individuals of 257 MS families[52]. The results suggested

SNP alleles and haplotypes within the CCL2, CCL11-CCL8-CCL13, CCL3 and CCL15 genes likely relevant to MS. In the second phase, we genotyped 232 SNP markers in the same region in a new set of 1369 individuals in 361 MS families. This second study confirmed and refined the previously suggested associations[53]. In the third phase, we selected from the MS-associated haplotypes one within the CCL3 gene, and sequenced this haplotype along with its flanking regions to reveal which polymorphisms (SNP allele) have the strongest association with MS and likely be the disease-relevant variant within that haplotype. Sequencing revealed several new SNPs within this CCL3 haplotype, which we genotyped in an independent set of MS cases and controls, and thus defined which variant within this haplotype showed the strongest association with the disease[54]. This three-stage study based on LD analyses first identified MS-associated SNP variants and haplotypes within specific genes at 17q11, which were then confirmed and refined in the second study. We postulated that these marker variants were in LD with specific disease-relevant variants. Therefore, in the third study we defined the SNPs that showed the strongest association within one of the MS-associated haplotypes (in the CCL3 gene), which likely have direct relevance to the disease. This study shows how we may get sequentially closer from a large (1.8 millions of base-pairs) chromosomal susceptibility region (17q.11) to a few hundred-base-pair size haplotype and then to single disease-relevant genetic variants by relying on LD.

GENOME-WIDE ASSOCIATION STUDIES

Similar to the above outlined 17q11 analyses, genome-wide association studies (GWAS) studies also rely on LD between a study-marker and a disease-relevant variant to define susceptibility in a complex trait, but instead of characterizing a selected region, the entire genome is scanned to reveal the genetic architecture of a disease. Typically from 3-6X10^5 to a few millions of markers are tested in large patient (case) and control populations to identify allele frequency differences suggesting associations with the disease. This association-based approach is far more powerful to reveal disease variants with small effects than the method of linkage. However, some key requirements need to be satisfied in order to yield a successful outcome in genome-wide association analyses: 1) Genome-wide significance cannot be reached if the sample size is inadequate and the signal of the disease gene is small. To minimize the risk of false negative observations (Type I error), inclusion of large numbers (typically several thousands) of study subjects may be necessary which may only be available through collaborations within consortia; 2) Because of the ethnic-specific nature of SNP frequency distributions, matching ethnically the control and the patient populations is a key requirement to avoid population stratification and false positive observations (Type II error); 3) Due to the large numbers of markers, corrections for multiple testing are necessary. However, using overly stringent corrections, a genome-wide significance may get lost (Type I error). Therefore, proper selection of methods to correct for multiple testing and taking LD between markers into consideration (acknowledging that associations with the markers within a haplotype are not independent) are necessary[55, 56].

The first highly powered MS GWAS study with dense coverage of the entire genome was published in 2007[57]. This study used DNA microarray technology to test almost 335,000 SNP markers in over 12,000 total subjects in MS trio families and case-control cohorts in the initial and replication study parts. The results showed that two SNPs within the interleukin-2 receptor alpha gene (IL2RA) ($P=2.96\times10^{-8}$), a nonsynonymous SNP in the interleukin-7 receptor alpha gene (IL7RA) ($P=2.94\times10^{-7}$) and multiple SNPs in the HLA-DRA locus ($P=8.94\times10^{-81}$) were associated with MS. Numerous studies accompanied or followed this GWAS to confirm and refine the involvement of the IL2RA and the IL7RA variants in MS susceptibility, and to sort out their biological significance[58–63]. It was shown that the MS-associated SNP variant in the IL7RA gene regulates alternative splicing by augmenting the exonic silencer of exon 6. Skipping of (splicing out) exon 6 increases the soluble and decreases the membrane-bound isoform of the receptor, which impacts on T and B cell development and differentiation[59]. A similar influence of the MS-associated IL2RA variant on the soluble vs. membrane-bound forms was also suggested[56]. A SNP variant in the CD58 gene, encoding the LFA-3 costimulatory molecule, was negatively associated with MS (Odd ratio or OR=0.82), suggesting that this allele exerts protective effect[64]. The protective allele was shown to increase CD58 expression, which upon engagement with its CD2 receptor, upregulates the expression of transcription factor FoxpP3. Increased FoxP3 is associated with an enhanced function of $CD4^+CD25^{high}$ Treg cells that, as indicated above, are defective in MS patients[64].

In addition to these well characterized MS-associated non-HLA genes, further non-HLA MS-susceptibility variants were revealed by analyses of

data from focused or comprehensive SNP genotyping[60, 64–69].

As of today, 7 independent GWAS studies have been reported in MS, but numerous studies with narrower scopes were also published[57, 63, 68, 70–75]. The HLA and non-HLA genetic variants associated with MS in the 7 GWAS are listed in **Table 1.**[29]. The International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium 2 [WTCCC-2][76] reported a GWAS on 9772 cases of European descent collected by 23 groups in 15 countries. This study replicated 23 of 26 previously known MS-associated variants and identified 29 novel susceptibility variants. Within the HLA DR1 gene, this group refined the risk alleles and showed that certain HLA A alleles may provide protective effect in MS.

Meta-analyses of GWAS data can be controversial, because it mixes results from various ethnic groups, but since both cases and controls are groupwise included, it can also be extremely powerful. Two meta-analyses of MS GWAS studies refined previous discoveries and added new susceptibility markers to the previous lists[68, 77]. They also revealed that the investigated variants altered gene expression in peripheral blood mononuclear cells and were involved in inflammatory processes suggesting their functional significance in MS[68, 77].

A proportion of susceptibility markers identified in MS overlaps with those found in other complex trait disorders such as rheumatoid arthritis, Crohn's disease, Type I diabetes, celiac disease, SLE and psoriasis, suggesting that these variants define an overarching autoimmune trait and are not specific for MS (reviewed in 29). Some of these autoimmune variants are located within genes of PTP22, IL23R, NRXN1, KIAA1109, EPHA7, TRIM27, TNFAIP3, TNKS, and C20orf42[29].

There are a few technical issues to take into consideration concerning GWAS studies. It is important to keep in mind that not all the MS-associated SNP alleles have direct relevance to the disease. As outlined above, some of these alleles may only be in LD with a disease variant. In addition, most of the GWAS studies used microarray technologies that preferentially test common SNP variants (population frequency >5%)[29]. Therefore, evolutionarily newer, rare variants have not been systematically interrogated in MS by GWAS, and therefore, non-array-based, focused association studies, even though smaller than GWAS (such as our 17q11 scan), may reveal information missed by GWAS. The challenge is to identify methods suitable for testing not only common but also rare variants with genome-wide coverage for the refinement of current knowledge concerning the genomic landscape in MS and other complex diseases.

PATHWAY ANALYSES

The above discussed hypothesis-free GWAS **(Table 1.)** reveal that the identified variants are mainly within genes of immune response and thereby support without preconceived assumption that immune-mediated pathways play key roles in the development of MS. This finding has major significance, since the views concerning a primary immune vs. CNS etiology of MS have long been debated[78], and objective observations to unequivocally support either position had been missing.

Formal pathway analyses of the MS-associated genetic variants identified by various GWAS studies have been conducted with similar outcomes. In the study by the WTCCC-2[76], statistical analyses were performed to identify genes with similar function. First, the authors identified the nearest gene to the SNP of interest in each region of association and used the so called Gene Ontology (GO) database to test whether or not the groups of nearest genes were enriched for functional relatedness. This analysis showed significant enrichment for lymphocyte function, particularly for T cell proliferation and activation, T helper cell differentiation, and intercellular immune communication through cell surface receptors and cytokines, and signal transduction. This study, however, did not find evidence for genetic associations with clinical course, disease severity, month of birth or gender.

More recently, a protein-interaction-network-based pathway analysis (PINBPA) of data from two large MS studies (WTCCC2 and meta2.5,)[68, 76] was reported, which involved a total of 15,000 cases and 31,000 controls[75]. PINBPA is even a more advanced and more powerful analytical approach than the previously used pathway analysis methods. This study assumed that even modestly associated genes may participate in biologically relevant pathways, and therefore considered all genes with $p<0.05$ after p-values were computed from individual SNP-wise summary-level data using a special program (VEGAS) (This low threshold of significance needs to be viewed in the context of GWAS where a genome-wide significance of $p<10^{-8}$ is typically required assuming independence among markers). After this more liberal inclusion of candidate genes, another program (the so called Manhattan visualization plot) was used to define association peaks. The distribution of nominally associated loci was largely identical in the two studies (WTCCC2 and meta2.5)[68, 76]. The association blocks matched the

Table 1. *GWAS studies and meta-analysis in MS*

Study	Year	Platform (SNP passing QC)	Initial sample size	Replication sample size	Region	Gene	SNPs	p-values	OR
Nischwitz et al.[73]	2010	Illumina [300,000]	590 cases 825 controls	NR	10q11.21	Intergenic	rs2503875	2.00E-07	1.66
					19p13.2	ZNF433	rs3745672	1.00E-06	7.39
					9q34.2	VAV2	rs3780792	1.00E-06	1.6
					6p21.3	DQA1	rs9271366	4.00E-17	2.62
Sanna et al. [74]	2010	Affymetrix [6,607,266] (imputed)	882 cases 872 controls	1,775 cases 2,005 controls	6p21.32	HLA-DRB, HLA-DQB1	rs2040406	1.00E-20	2.05
					3q13.11	CBLB	rs9657904	2.00E-10	1.4
Jakkula et al.[63]	2010	Illumina [297,343]	68 cases 136 controls	4,570 cases 10,143 controls	17q21.2	STAT3	rs744166	3.00E-10	1.15
					6p21.32	HLA	rs3135338	2.00E-25	3.43
Bahlo et al.(AZN) [71]	2009	Illumina [302,098]	1,618 cases 3,413 controls	2,256 cases 2,310 controls	10p15.1	IL2RA	rs2104286	7.00E-06	1.16
					20q13.12	CD40	rs6074022	1.00E-07	1.2
					16p12.1	NR	rs8049603	1.00E-06	1.19
					1p22.1	EVI5, RPL5	rs6604026	3.00E-06	1.17
					1p13.1	CD58	rs1335532	1.00E-07	1.28
					6p21.32	HLA-DRB1	rs9271366	7.00E-184	2.78
					8q24.21	ASAP1, DDEF1	rs6984045	2.00E-06	1.59
					12q14.1	METTL1, CYP27B1	rs703842	5.00E-11	1.23
De Jager et al. * (meta2.5)[68]	2009	Affymetrix & Illumina [~2.56 million] (imputed)	2,624 cases 7,220 controls	2,215 cases 2,116 control	6p21.32	HLA-DRB1	rs3135388	4.00E-225	2.75
					1p13.1	CD58	rs2300747	3.00E-10	1.3
					6p22.1	HLA-B	rs2523393	1.00E-17	1.28
					3q25.33	IL12A	rs4680534	6.00E-06	1.12
					10p15.1	IL2RA	rs2104286	9.00E-08	1.15
					12p13.31	TNFRSF1A	rs4149584	5.00E-06	1.58
					2q22.1	CXCR4	rs882300	1.00E-07	1.19
					12p13.31	TNFRSF1A	rs1800693	2.00E-11	1.2
					16p13.13	CLEC16A	rs11865121	2.00E-07	1.15
					5p13.2	IL7R	rs6897932	2.00E-06	1.12
					16q24.1	IRF8	rs17445836	4.00E-09	1.25
					11q12.2	CD6	rs17824933	4.00E-09	1.18
					5p13.1	PTGER4	rs6896969	2.00E-07	1.1
					12q24.31	MPHOSPH9	rs1790100	7.00E-07	1.11
					10q22.3	ZMIZ1	rs1250540	2.00E-06	1.12

Table 1. *Continued*

Study	Year	Platform (SNP passing QC)	Initial sample size	Replication sample size	Region	Gene	SNPs	p-values	OR
Baranzini et al.[72]	2009	Illumina [551,642]	978 cases 883 controls	NR	13q31.3	GPC5	rs9523762	1.00E-06	1.36
					9p22.2	SH3GL2	rs1755289	3.00E-06	1.35
					8p23.2	CSMD1	rs1529316	2.00E-06	1.36
					12q12	PDZRN4	rs1458175	2.00E-06	1.34
					3q23	SLC25A36	rs908821	3.00E-06	1.37
					4q35.1	MGC45800	rs7672826	8.00E-06	1.37
					20p13	C20orf46	rs397020	8.00E-07	1.41
					2p25.1	DDEF2	rs1109670	9.00E-06	1.38
					3q24	ZIC1	rs1841770	8.00E-06	1.34
					2q14.2	EN1	rs651477	7.00E-06	1.38
Comabella et al.[70]	2008	Affymetrix [428,867] (pooled)	242 cases 242 controls	375 cases 375 controls	6p21.32	HLA-DRB1	rs3129934	9.00E-11	3.3
Hafler et al.[57]	2007	Affymetrix [334,923]	931 trios 2,431 controls	609 trios 2,322 cases 2,987 controls	9q33	DBC1	rs10984447	8.00E-06	1.17
					6p21.32	HLA-DRA	rs3135388	9.00E-81	1.99
					16p13.13	KIAA0350	rs6498169	4.00E-06	1.14
					10p15.1	IL2RA	rs12722489	3.00E-08	1.25
					5p13.2	IL7RA	rs6897932	3.00E-07	1.18
					1p22.1	RPL5	rs6604026	8.00E-06	1.15

*Meta-analysis

This table summarizes the results of 7 GWAS studies and a meta-analysis (meta2.5) in MS, and is reproduced with modifications from Publication: Curr Opin Genet Dev. 2011; 21 (3):317-324. Baranzini SE. Revealing the genetic basis of multiple sclerosis: Are we there yet?[29] - with permission from Elsevier.

boundaries of the association regions of the 57 MS susceptibility loci[76]. Then, the authors combined statistical evidence of gene association and physical evidence of protein-protein interaction. The PINBPA analyses revealed that protein products of genes that showed genome-wide significance were more likely to physically interact and to belong to the same or similar pathways. The sub-networks or modules of genes (and their protein products) enriched with nominally associated loci were also investigated within each of the two studies (WTCCC2 and meta2.5)[68, 76]. These analyses demonstrated that the modules were more likely to include genes with "bona fide" susceptibility variants and identified several new, biologically relevant candidates[75]. Finally, to define biological significance of associated and candidate genes, the group conducted Gene Ontology analyses, which revealed three main categories of biological processes: leukocyte activation, apoptosis and positive regulation of macromolecular metabolic processes. Another, the so called KEGG pathway analysis highlighted the JAK-STAT signaling pathway, acute myeloid leukemia, and T cell signaling pathways. (KEGG [Kyoto Encyclopedia of Genes and Genomes] is an assembly of online databases of genomes, enzymatic pathways and biomaterials. The PATHWAY database of KEGG includes networks of intracellular molecular interactions in various organisms). Tissue specificity analyses suggested that 2/3 of the genes were highly expressed in immune cells, and half of the genes were expressed in the CNS. However, further analyses excluded the involvement of a neural pathway in MS. These complex, multi-step analyses provided strong evidence to further support the involvement of immune pathways in MS[75].

MITOCHONDRIAL GENOME

The mitochondrial (mt)DNA is a small, extranuclear part of the genome, and often omitted from reviews of MS genetics/genomics. As the author of this paper carried out the first comprehensive mtDNA polymorphism and full sequence analyses in MS cases and controls, and the investigations of nuclear and mtDNA-defined variants of Complex I in MS families, the results are briefly discussed here[11, 79–81]. The involvement of mtDNA in MS susceptibility was suggested by the observations that MS occurred more frequently than expected by chance in Leber's Hereditary Optic Neuropathy (LHON) pedigrees, MS also co-occurred intra-individually with LHON in some patients, and subsequently, primary LHON mutations were detected in some MS patients (Reviewed in 11). Our comprehensive analyses of mtDNA excluded, however, the involvement of mtDNA mutations with pathogenic significance in MS, but revealed the associations of certain mtDNA haplotypes and variants with the disease[79, 80]. These findings were later reproduced by several groups (reviewed in 11). We also found that variants within nuclear and mtDNA-encoded genes of Complex I are associated with MS and likely interact with each other at the protein level[81]. Complex I is the first enzyme complex involved in mitochondrial oxidative phosphorylation and encoded by 38 nuclear and 7 mitochondrial genes. We found that the activity of Complex I is negatively affected by oxidative damage developing secondary to inflammation in MS lesions[10]. However, we were unable to detect mitochondrial genetic markers of phenotypic subgroups of MS. Our comprehensive studies identified mtDNA variants and haplotypes that confer modest risks to the disease, and revealed that Complex I, a mitochondrial enzyme with MS-associated genetic variants, also has functional involvement in lesion development. These studies laid the ground works for establishing that a mitochondrial mechanism leads to neurodegeneration developing secondary to inflammation in MS.

Conclusion and future directions

Data from genetic and genomic studies have consistently identified the HLA locus as a major susceptibility locus, and established that HLA defines more than 50% of genetic susceptibility in MS. GWAS studies revealed thus far 57 non-HLA susceptibility variants with moderate and small effects. These genetic variants are within genes that define pathways of immune response, intercellular communication, macromolecular metabolic processes, apoptosis and signal transduction. Additional genes (and their protein products) may be identified by approaches that reveal candidates with smaller effects and by newer genome-wide genotyping strategies that may capture very rare variants in the population. Gene ontology and pathway analyses have greatly facilitated to view genomic results in functional context. Protein-protein interaction analyses replicated and confirmed most of the previous conclusions concerning pathways and networks defined by GWAS data, and further refined our knowledge regarding MS pathogenesis. Altogether, genomic data, first in MS history, provided objective evidence to support the immune etiology of the disease. Studies with narrower focus on selected chromosomal regions may still be justified to reveal variants with small effects. Such information may be relevant to additional MS pathways, and complementary to the GWAS data, if used in the appropriate context. MtDNA analyses also complemented the large genome-wide efforts, and identified genetic variants and functional pathways involved in the down-stream process of neurodegeneration developing secondary to inflammation.

A future extension of these investigations will likely explore as to how environmental factors interact with the genetic susceptibility variants and contribute to immune activation and inflammation. Gene-environmental interactions can be probed by methods of epigenetics, which investigate modifications of histone by biochemical processes (methylation, acetylation, phosphorylation, ubiquitination and sumoylation) at certain amino acids, methylation of CpG sequences within DNA, and microRNA expression. All these epigenetic processes are involved in gene expression regulation. The epigenetic modifications of histones, DNA and microRNA expression often result from interactions with environmental factors such as microbial agents, UV irradiation, toxic molecules, and smoking, previously identified in MS epidemiologic studies. While there have been a great number of epigenetic investigations in MS[82–85], none was comprehensive enough to allow unambiguous conclusions. After the rewarding outcomes of GWAS studies, epigenetic approaches may represent the next step to gain further insights into the complex etiology of the disease. Integration of genomic, mitochondrial genetic, epigenetic and transcriptomic data may provide an additional dimension to the analyses, and elucidate not only cross-sectional but also longitudinal aspects of biological processes important in MS pathogenesis. The emphasis on specifying the pathogenesis is very important for future pharmaceutical discoveries and targeting, or for the ultimate goal of preventing the disease.

REFERENCES

1. *Scalfari A, Neuhaus A, Daumer M, Ebers GC, Muraro PA*. Age and disability accumulation in multiple sclerosis. Neurology. 2011 Sep 27;77(13):1246-52.
2. *McDonald WI, Compston A, Edan G, et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 2001;50:121-7.
3. *Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al*. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011 Feb;69(2):292-302. doi: 10.1002/ana.22366.
4. *Lublin FD, Reingold SC*. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology 1996;46(4):907-11.
5. *Minagar A, Barnett MH, Benedict RH, Pelletier D, Pirko I, Sahraian MA, et al.* The thalamus and multiple sclerosis: modern views on pathologic, imaging, and clinical aspects. Neurology 2013;80(2):210-9. doi: 10.1212/WNL.0b013e31827b910b. Review.
6. *Kolasinski J, Stagg CJ, Chance SA, Deluca GC, Esiri MM, Chang EH, et al*. A combined post-mortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology. Brain. 2012 Oct;135(Pt 10):2938-51
7. *Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP, et al; on behalf of the MAGNIMS Study Group*. Brain atrophy and lesion load predict long term disability in multiple sclerosis. J Neurol Neurosurg Psychiatry 2013 Mar 23. [Epub ahead of print]
8. *Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H*. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann Neurol 2000 Jun;47(6):707-17.
9. *Barnett MH, Prineas JW*. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. Ann Neurol. 2004 Apr;55(4):458-68.
10. *Lu F, Selak M, O'Connor J, Croul S, Lorenzana C, Butunoi C, Kalman B*. Oxidative damage to mitochondrial DNA and activity of mitochondrial enzymes in chronic active lesions of multiple sclerosis. J Neurol Sci 2000;177(2):95-103.
11. *Kalman B, Albert RH, Leist TP*. Genetics of multiple sclerosis: Determinants of autoimmunity and neurodegeneration. Autoimmunity 2002;35(4):225-34.
12. *Rostami A, Ciric B*. Role of Th17 cells in the pathogenesis of CNS inflammatory demyelination. J Neurol Sci 2013 Apr 8. doi: pii: S0022-510X(13)00108-1.
13. *Stoeckle C, Tolosa E.* Antigen processing and presentation in multiple sclerosis. Results Probl Cell Differ 2010;51: 149-72.
14. *Vyshkina T, Kalman B*. Autoantibodies and neurodegeneration in multiple sclerosis. Lab Invest 2008;88:796-807.
15. *Fritzsching B, Haas J, König F, Kunz P, Fritzsching E, Pöschl J, et al*. Intracerebral human regulatory T cells: analysis of CD4+ CD25+ FOXP3+ T cells in brain lesions and cerebrospinal fluid of multiple sclerosis patients. PLoS One 2011;6(3):e17988. doi: 10.1371/journal.pone.0017988.
16. *Gyódi E, Benczur M, Pálffy G, Tálas M, Petrányi G, Földes I, Hollán SR*. Association between HLA B7, DR2 and dysfunction of natural- and antibody-mediated cytotoxicity without connection with the deficient interferon production in multiple sclerosis. Hum Immunol 1982;4(3):209-17.
17. *Koch MW, Metz LM, Agrawal SM, Yong VW*. Environmental factors and their regulation of immunity in multiple sclerosis. J Neurol Sci. 2013 Jan 15;324(1-2):10-6. doi: 10.1016/j.jns.2012.10.021. Epub 2012 Nov 13. Review.
18. *Pálffy G, Gyódi E, Petrányi G*. Multiple sclerosis in 3 gipsy and 2 partly gipsy subjects (a clinical and immunogenetic study). Acta Med Hung. 1986;43(1):31-7.
19. *Takacs K, Kalman B, Gyodi E, Tauszik T, Palffy G, Kuntar L, et al.* Association between the lack of HLA-DQw6 and low incidence of multiple sclerosis in Hungarian Gypsies. Immunogenetics 1990;31:383-5.
20. *Kalman B, Takacs K, Gyodi E, Kramer J, Fust G, Tauszik T, et al*. Sclerosis multiplex in Gypsies. - Acta Neurol Scand 1991; 84:181-5.
21. *Wallin MT, Culpepper WJ, Coffman P, Pulaski S, Maloni H, Mahan CM, et al; Veterans Affairs Multiple Sclerosis Centres of Excellence Epidemiology Group*. The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service. Brain 2012;135(Pt 6):1778-85.
22. *Ebers GC, Bulman DE, Sadovnick AD, Paty DW, Warren S, Hader W, et al.* A population-based study of multiple sclerosis in twins. N Engl J Med 1986;315(26):1638-42.
23. *Sadovnick AD, Baird PA*. The familial nature of multiple sclerosis: age-corrected empiric recurrence risks for children and siblings of patients. Neurology 1988 Jun;38(6): 990-1.
24. *Sadovnick AD, Armstrong H, Rice GP, Bulman D, Hashimoto L, Paty DW, et al.* A population-based study of multiple sclerosis in twins: update. Ann Neurol 1993;33(3): 281-5.
25. *Ebers GC, Sadovnick AD, Risch NJ*. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. Nature 1995;377(6545):150-1.
26. *Ebers GC, Yee IM, Sadovnick AD, Duquette P*. Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group. Ann Neurol 2000;48(6):927-31.
27. *Ebers GC, Sadovnick AD, Dyment DA, Yee IM, Willer CJ, Risch N*. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet 2004;363(9423):1773-4.
28. *Sawcer S*. The complex genetics of multiple sclerosis: pitfalls and prospects. Brain 2008 Dec;131(Pt 12):3118-31.
29. *Baranzini SE*. Revealing the genetic basis of multiple sclerosis: Are we there yet? Curr Opin Genet Dev 2011;21(3): 317-24.
30. Fogdell A, Hillert J, Sachs C, Olerup O: The multiple sclerosis- and narcolepsy-associated HLA class II haplotype includes the DRB5*0101 allele. Tissue Antigens 1995; 46:333-6.
31. *Ramagopalan S, Ebers GC.* Multiple sclerosis: major histocompatibility complexity and antigen presentation. Genome Medicine 2009;1:105.1-105.5
32. *de Jong BA, Huizinga TW, Zanelli E, Giphart MJ, Bollen EL, Uitdehaag BM, et al*. Evidence for additional genetic risk indicators of relapse-onset MS within the HLA region. Neurology 2002;59(4):549-55.
33. *Rubio JP, Bahlo M, Butzkueven H, van Der Mei IA, Sale MM, Dickinson JL, 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(5): 1125-37. Epub 2002 Mar 29.
34. *Haines JL, Terwedow HA, Burgess K, Pericak-Vance MA, Rimmler JB, Martin ER, et al*. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group. Hum Mol Genet 1998;7(8):1229-34.
35. *McMahon RM, Friis L, Siebold C, Friese MA, Fugger L, Jones EY*. Structure of HLA-A*0301 in complex with a

peptide of proteolipid protein: insights into the role of HLA-A alleles in susceptibility to multiple sclerosis. Acta Crystallogr D Biol Crystallogr. 2011 May;67(Pt 5):447-54.
36. *Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, et al*. full genome search in multiple sclerosis. Nat Genet 1996;13(4):472-6.
37. *Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, 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(4):469-71.
38. *Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, et al*. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nat Genet 1996;13(4):464-8.
39. *Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F, et al*. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet 2006;15(18):2813-24. Epub 2006 Aug 11.
40. *Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A, et al; International Multiple Sclerosis Genetics Consortium.* A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 2005;77(3):454-67. Epub 2005 Jul 29.
41. *Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al.* The structure of haplotype blocks in the human genome. Science 2002;296(5576):2225-9.
42. *Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al.* Linkage disequilibrium in the human genome. Nature 2001;411(6834):199-204.
43. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J,et al; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. Erratum in: Nature 2001 Aug 2;412(6846):565. Nature 2001 Jun 7;411(6838):720. Szustakowki, J [corrected to Szustakowski, J].
44. *Istrail S, Sutton GG, Florea L, Halpern AL, Mobarry CM, Lippert R, et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. Proc Natl Acad Sci U S A 2004;101(7):1916-21. Epub 2004 Feb 9.
45. *Kruglyak L, Nickerson DA*. Variation is the spice of life. Nature Genetics 2001;27:234-6.
46. The Transatlantic Multiple Sclerosis Genetics Cooperative. A meta-analysis of genomic screens in multiple sclerosis. Mult Scler 2001;7(1):3-11.
47. *Jonasdottir A, Thorlacius T, Fossdal R, Jonasdottir A, Benediktsson K, Benedikz J, et al*. A whole genome association study in Icelandic multiple sclerosis patients with 4804 markers. J Neuroimmunol 2003;143(1-2):88-92.
48. *Goertsches R, Villoslada P, Comabella M, Montalban X, Navarro A, de la Concha EG, et al; Spanish MS Genetics Group*. A genomic screen of Spanish multiple sclerosis patients reveals multiple loci associated with the disease. J Neuroimmunol 2003;143(1-2):124-8.
49. *Banisor I, Leist TP, Kalman B*. Involvement of -chemokines in the development of inflammatory demyelination. J Neuroinflammation 2005;2:7. epub
50. *Teuscher C, Butterfield RJ, Ma RZ, Zachary JF, Doerge RW, Blankenhorn EP*. Sequence polymorphisms in the chemokines Scya1 (TCA-3), Scya2 (monocyte chemoattractant protein (MCP)-1), and Scya12 (MCP-5) are candidates for eae7, a locus controlling susceptibility to monophasic remitting/nonrelapsing experimental allergic encephalomyelitis. J Immunol 1999;163:2262-6.
51. *Jagodic M, Becanovic K, Sheng JR, Wu X, Bäckdahl L, Lorentzen JC, et al*. An advanced intercross line resolves Eae18 into two narrow quantitative trait loci syntenic to multiple sclerosis candidate loci.J Immunol 2004;173(2): 1366-73.
52. *Vyshkina T, Yao SY, Birnbaum G, Leist TP, Kalman B*. Association of haplotypes in the -chemokine locus with multiple sclerosis. Eur J Hum Genet 2005;13:240-247.
53. *Vyshkina T, Kalman B*. Haplotypes within genes of beta-chemokines are associated with multiple sclerosis: a second phase study. Human Genetics 2005;118:67-75.
54. *Vyshkina T, Kalman B*. Analyses of a MS-associated haplotype encompassing the CCL3 gene. J Neuroimmunol 2006;176:216-218.
55. *Kemppinen A, Sawcer S, Compston A*. Genome-wide association studies in multiple sclerosis: lessons and future prospects. Brief Funct Genomics 2011;10(2):61-70.
56. *Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, Severson C, et al*. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet 2009;5(1):e1000322. doi: 10.1371/journal.pgen.1000322.
57. *International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, et al*. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 2007;357(9): 851-62. Epub 2007 Jul 29.
58. *Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallström E, Khademi M, et al*. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet 2007;39(9):1108-13.
59. *Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, et al; Multiple Sclerosis Genetics Group.* Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet 2007;39(9):1083-91.
60. *Rubio JP, Stankovich J, Field J, Tubridy N, Marriott M, Chapman C, et al*. Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. Genes Immun 2008;9(7):624-30.
61. *Weber F, Fontaine B, Cournu-Rebeix I, Kroner A, Knop M, Lutz S, et al.* IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. Genes Immun 2008;9(3):259-63.
62. *Zuvich RL, McCauley JL, Oksenberg JR, Sawcer SJ, De Jager PL; International Multiple Sclerosis Genetics Consortium, Aubin C, et al*. Genetic variation in the IL7RA/IL7 pathway increases multiple sclerosis susceptibility. Hum Genet 2010;127(5):525-35.
63. *Jakkula E, Leppä V, Sulonen AM, Varilo T, Kallio S, Kemppinen A, 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(2): 285-91.
64. *De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L, et al*. The role of the CD58 locus in multiple sclerosis. Proc Natl Acad Sci U S A 2009;106 (13):5264-9. doi: 10.1073/pnas.0813310106.
65. *Mero IL, Lorentzen AR, Ban M, Smestad C, Celius EG, Aarseth JH, et al*. A rare variant of the TYK2 gene is confirmed to be associated with multiple sclerosis. Eur J Hum Genet 2010;18(4):502-4. doi: 10.1038/ejhg.2009.195.
66. *Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, Broer L, Jafari N, Hillert J, et al*. Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. Nat Genet 2008;40(12):1402-3. doi: 10.1038/ng.251.
67. *Hoppenbrouwers IA, Aulchenko YS, Ebers GC, Ramagopalan SV, Oostra BA, van Duijn CM, et al.* EVI5 is a risk gene for multiple sclerosis. Genes Immun 2008;9(4):334-7. doi: 10.1038/gene.2008.22.
68. *De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L,*

Aggarwal NT, Piccio L, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet 2009;41 (7):776-82. doi: 10.1038/ng.401. Epub 2009 Jun 14.
69. *International Multiple Sclerosis Genetics Consortium (IMSGC).* Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies KIF21B and TMEM39A as susceptibility loci. Hum Mol Genet 2010;19(5):953-62. doi: 10.1093/hmg/ddp542. Epub 2009 Dec 9.
70. *Comabella M, Craig DW, Camiña-Tato M, Morcillo C, Lopez C, Navarro A, Rio J; BiomarkerMS Study Group, Montalban X, Martin R*. 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(10):e3490. doi: 10.1371/journal.pone. 0003490.
71. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet 2009;41(7):824-8.
72. *Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al*. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet 2009;18(4):767-78.
73. *Nischwitz S, Cepok S, Kroner A, Wolf C, Knop M, Müller-Sarnowski F, et al*. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. J Neuroimmunol 2010 Oct 8;227(1-2):162-6.
74. *Sanna S, Pitzalis M, Zoledziewska M, Zara I, Sidore C, Murru R, et al.* Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nat Genet 2010;42(6):495-7.
75. *International Multiple Sclerosis Genetics Consortium*. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. The American Journal of Human Genetics 2013; http://dx.doi.org/10.1016/j.ajhg.2013.04.019.
76. *International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium 2*. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011;476:214-8.
77. *Patsopoulos NA; Bayer Pharma MS Genetics Working Group; Steering Committees of Studies Evaluating IFNbeta-1b and a CCR1-Antagonist; ANZgene Consortium; GeneMSA; International Multiple Sclerosis Genetics Consortium, et al.* Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol 2011;70(6):897-912. doi: 10.1002/ana.22609.
78. *Mastronardi FG, Moscarello MA*. Molecules affecting myelin stability: a novel hypothesis regarding the pathogenesis of multiple sclerosis. J Neurosci Res 2005;80(3):301-8. Review.
79. Kalman B, Lublin FD, Alder H. Characterization of the mitochondrial DNA in multiple sclerosis. J. Neurol. Sci. 1996; 140:75-89.
80. *Kalman B, Li S, Chatterjee D, O'Connor J, Voehl MR, Brown MD, Alder H*. Large scale screening of the mitochondrial DNA reveals no pathogenic or polymorphic mutations but a haplotype associated with multiple sclerosis in Caucasians. Acta Neurol Scand 1999; 99:16-25.
81. *Vyshkina T, Banisor I, Yao Shugart Y, Leist TP, Kalman B.* Genetic variants of Complex I in multiple sclerosis. J Neurol Sci 2005;228:55-64.
82. *Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, et al*. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 2010;464(7293):1351-6.
83. *Kumagai C, Kalman B, Middleton FA, Vyshkina T, Massa PT*. Increased promoter methylation of the immune regulatory gene SHP-1 in leukocytes of multiple sclerosis subjects. J Neuroimmunol 2012;246(1-2):51-7.
84. *Koch MW, Metz LM, Kovalchuk O*. Epigenetics and miRNAs in the diagnosis and treatment of multiple sclerosis. Trends Mol Med 2013;19(1):23-30.
85. *Reijerkerk A, Lopez-Ramirez MA, van Het Hof B, Drexhage JA, Kamphuis WW, Kooij G, et al*. MicroRNAs regulate human brain endothelial cell-barrier function in inflammation: implications for multiple sclerosis. J Neurosci 2013;33(16):6857-63.