

# A review of genome-wide association studies for multiple sclerosis: classical and hypothesis-driven approaches

V. V. Bashinskaya<sup>1,2</sup> · O. G. Kulakova<sup>1,2</sup> · A. N. Boyko<sup>1</sup> · A. V. Favorov<sup>3,4</sup> ·  
O. O. Favorova<sup>1,2</sup>

Received: 13 July 2015 / Accepted: 10 September 2015 / Published online: 25 September 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Multiple sclerosis (MS) is a common complex neurodegenerative disease of the central nervous system. It develops with autoimmune inflammation and demyelination. Genome-wide association studies (GWASs) serve as a powerful tool for investigating the genetic architecture of MS and are generally used to identify the genetic factors of disease susceptibility, clinical phenotypes, and treatment response. This review considers the main achievements and challenges of using GWAS to identify the genes involved in MS. It also describes hypothesis-driven studies with extensive genome coverage of the selected regions, complementary to GWASs. To date, over 100 MS risk loci have been identified by the combination of both approaches; 40 of them were found in at least two GWASs and meet genome-wide significance threshold ( $p \leq 5 \times 10^{-8}$ ) in at least one GWAS, whereas the threshold for the rest of GWASs was set in our review at  $p < 1 \times 10^{-5}$ . Yet, MS risk loci identified to date explain only a part of the total heritability, and the reasons of “missing heritability” are discussed. The functions of MS-associated genes are described briefly;

the majority of them encode immune-response proteins involved in the main stages of MS pathogenesis.

## Introduction

Multiple sclerosis (MS) is a chronic demyelinating disease of the central nervous system (CNS); it develops with progressive neurodegeneration mainly due to autoimmune inflammation (Nylander and Hafler 2012). MS is diagnosed in about 2.3 million people worldwide (Browne et al. 2014), affects mostly young people leading to early disability, and reduces the lifespan by approximately 7 years as compared with the general population (Marrie et al. 2015; Kaufman et al. 2014; Leray et al. 2015). MS is a complex disease: several causes underlie its development and are still incompletely understood (Marian 2012; Kilpinen and Barrett 2013).

MS is characterized by familial aggregation; namely, the risk to develop MS is higher in patient’s relatives than in the total population and is in negative correlation with genetic distance to the proband (Oksenberg 2013). Its transmission in families, however, disagrees with Mendelian inheritance. The concordance rates are about 24–30 % in monozygotic twins and 3–5 % in dizygotic twins (Lin et al. 2012). This type of inheritance actually is typical for polygenic diseases and results from integral action of many independent or interacting polymorphic genes, each making only a minor contribution to the disease (Bomprezzi et al. 2003; IMSSC et al. 2010). Other risk factors act on the background of genetic susceptibility and are currently thought to include the epigenetic regulation and environmental factors (Oksenberg 2013; Gourraud et al. 2012).

MS is clinically a heterogenic disease with several courses, different manifestations and progression rates. The

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-015-1601-2) contains supplementary material, which is available to authorized users.

---

✉ V. V. Bashinskaya  
vitalina.bashinskaya@gmail.com

- <sup>1</sup> Pirogov Russian National Research Medical University, Moscow 117997, Russia
- <sup>2</sup> Russian Cardiology Research and Production Complex, Moscow 121552, Russia
- <sup>3</sup> Vavilov Institute of General Genetics, Moscow 119333, Russia
- <sup>4</sup> Oncology Biostatistics and Bioinformatics, John Hopkins School of Medicine, Baltimore, MD 21205, USA

most frequent is bout-onset MS, that is, relapsing–remitting MS (RRMS) (85 % of cases) and secondary progressive MS (SPMS), which develops in the majority of RRMS patients following a more or less long period of the disease. Primary progressive MS (PPMS), characterized by progressive disability augment, is less frequent (10–15 % of cases) (Nylander and Hafler 2012). There is growing evidence that different pathogenic mechanisms prevail in RRMS and PPMS (Bradl and Lassmann 2009; Iwanowski and Losy 2015). The genetic basis of MS clinical heterogeneity is also complex (Baranzini et al. 2009) and remains the important unresolved problem.

Studies of the molecular basis underlying the pathogenesis of MS are an important part of research in the field, helping to design new strategies for its prevention and treatment by identifying the genetic, epigenetic, and environmental risk factors of MS that might provide a prognostic tool to estimate the individual predisposition to MS. In this review, only studies of genetic architecture of MS are discussed.

Genes involved in MS have long been sought, and several approaches to this problem have been applied more or less successfully. A candidate gene approach was employed for several decades in numerous studies, wherein genes potentially associated with MS were chosen on the basis of the presumable MS pathogenesis. Several loci involved in MS were identified, including certain HLA class I (Naito et al. 1972) and class II (Compston et al. 1976) genes, *IL7RA* (Lundmark et al. 2007), *CIITA* (Rasmussen et al. 2001), and *SOCS1* (Zuvich et al. 2011; Vanderbroeck et al. 2012). Data on the association of common allele groups of the HLA class II *DRB1* gene with MS in various populations of the world and the non-HLA genes that showed a significant association with MS in more than two independent studies were summarized in (Ramagopalan and Ebers 2009; Bahreini et al. 2010). At the same time, negative results were obtained in many association studies; this is not surprising because improper genes can easily be chosen as candidates given that the MS pathogenesis is complex and still incompletely understood. Several studies were focused on possible interactions of HLA-*DRB1* alleles and non-HLA genes or between non-HLA genes. For example, associations of *CCR5*  $\Delta 32$  (Favorova et al. 2002) and of VNTR polymorphism in *MBP* (Guerini et al. 2003) were found only in DR-stratified individuals. This strategy of searching for combinations resulted in creation of the APSampler algorithm (Favorov et al. 2005) that was successfully applied for MS cohorts and allowed finding the two three-allelic combinations associated with MS, both including HLA-*DRB1* alleles (Favorova et al. 2006).

Genome-wide linkage analysis was used in the 1990s–2000th to investigate the MS inheritance in families and to identify the genome regions that deviate from

independent segregation and cosegregate with the disease. Linkage analyses with several hundreds of highly polymorphic microsatellite repeats showing a relatively uniform distribution through the genome were carried out in more than 30 studies involving families with several MS patients (Ebers et al. 1996; Sawcer et al. 1996; Ban et al. 2002). Genome-wide linkage analyses performed in different ethnic groups, as well as meta-analyses of the pooled data (Hermanowski et al. 2007; GAMES and Transatlantic Multiple Sclerosis Genetics Cooperative 2003), yielded inconsistent results in the majority of cases, pointing to potential genetic heterogeneity of MS in different populations. The HLA class II locus in the major histocompatibility complex (MHC) region on chromosome 6p21 was the only exception, being already identified as a MS risk factor in earlier association studies. Later, linkage to MS was tested for single nucleotide polymorphisms (SNPs) (Sawcer et al. 2005), which usually consist of two alleles and are consequently less informative than microsatellites, but are more abundant in the genome. In spite of the higher genome coverage density, the study did not identify any region with statistically significant linkage to MS outside the HLA locus, pointing to a low sensitivity of the method (Cree 2014).

In the last decade, genome-wide association study (GWAS) has become the most common method to search for new genes involved in predisposition to MS. GWAS provides a powerful tool for investigating the genetic architecture of human polygenic diseases and is based on comparing the allele frequencies of SNPs distributed throughout the genome between samples of unrelated patients and control individuals. The analysis is performed using microarrays or more advanced techniques that allow a simultaneous genotyping at several tens of thousands to several millions of SNPs per genome (Cree 2014). Such experiments are comparatively inexpensive, allowing hundreds or thousands of patients to be genotyped in one study. This review considers the main achievements and challenges of using GWAS and complementary hypothesis-driven studies to identify the genes involved in MS. We overview the biological functions of the loci replicated between GWASs and discuss some aspects of “missing heritability.”

### Gwas data for multiple sclerosis

The GWAS data for MS, as well as for other complex diseases, can be found in the regularly updated NHGRI-EBI GWAS catalog (Hindorf et al. 2009). Established in 2008 on the website of the National Human Genome Research Institute (NHGRI, United States), it migrated in March 2015 to the European Bioinformatics Institute (EMBL-EBI) and is now available at (<http://www.ebi.ac.uk/gwas>; Burdett et al. 2015). This catalog includes data of all published

GWASs assaying at least 100,000 SNPs and all SNP-trait associations with  $p$  values  $<1 \times 10^{-5}$  (Hindorf et al. 2009). This catalog includes not only studies of genetic susceptibility to MS but also GWASs of important MS clinical features: oligoclonal bands' status, IgG levels, response to interferon beta therapy, MS severity, brain lesion load, age of onset, brain glutamate levels, normalized brain volume.

In Table 1, we summarized the data on all independent GWASs that investigated MS susceptibility in various (mostly Caucasian) populations by comparing the allele frequencies between MS patients and control subjects. Detailed information, including a description of the individual MS-associated loci at level of significance less than  $1 \times 10^{-5}$ , is presented in Table S1.

The leading role in these studies belongs to International Consortia, which possess individual DNA samples from various clinics of the world; the main leaders are the International Multiple Sclerosis Genetic Consortium (IMSGC), Welcome Trust Case Control Consortium 2 (WTCCC2), and Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). IMSGC was formed in 2003 with the aim of adopting a collaborative approach to the challenge of identifying all the inherited risk factors for MS. It has combined resources of 25 research groups from 14 countries and possesses DNA collected from more than 20,000 subjects. WTCCC2 performs GWASs of 13 common diseases including MS. Together with the IMSGC and the Genetic Analysis of Multiple sclerosis in EuropeanS (GAMES) collaborative group, WTCCC2 was able to recruit about 10,000 independent MS cases. The ANZgene Consortium, a collaborative project formed in 2007, comprises eight centers possessing over 3500 DNA samples for use in MS research.

Affymetrix or Illumina genome-wide platforms were mostly employed in GWASs, and target SNPs more or less regularly distributed throughout the genome increased in number with progress in technology. The number of SNPs covered by arrays widely varies in different studies (with range from 262 K to approximately 600 K) (Table 1). Importantly, only SNPs whose minor alleles are common (minor allele frequency, or MAF, more than 1 %) were included in the GWAS panels; in some studies, only SNPs with MAF even more than 5 % were studied.

Apart from increasing the average density of genome coverage, the genotype imputation method, i.e., the statistical inference of unobserved genotypes, was used in studies (Aulchenko et al. 2008; ANZgene et al. 2009; Nischwitz et al. 2010; Sanna et al. 2010; IMSGC et al. 2011, 2013; Patsopoulos et al. 2011; Matesanz et al. 2012; Isobe et al. 2015) to improve the efficiency of analyses. Genotype imputation allows evidence to be accurately evaluated for association of genetic markers that are not directly genotyped. The method is based on estimating the linkage

disequilibrium level for polymorphic sites of a particular haploblock; as a result, a genotype at a specific SNP is predicted with a high probability (Li et al. 2009). Imputation makes it possible, first, to achieve consistency in the polymorphic site sets assessed in different samples of patients and controls (often genotyped using different platforms); second, to improve the statistical power of the study, thus detecting more significant associations than studies with a particular platform; third, to increase coverage at the putatively associated loci, allowing fine mapping of such regions. Genotype imputation was applied in the reviewed GWASs for all these purposes. In particular, due to this technique, datasets genotyped by different platforms were analyzed jointly in (Matesanz et al. 2012); the number of tested SNPs increased from 555,335 to 6,607,266 in (Sanna et al. 2010); and several previously identified loci were replicated by (Nischwitz et al. 2010), if not available in the used microarray.

The associations revealed in genome-wide studies need validation. For this purpose, GWAS is nearly always performed in two phases, or stages. The first, or discovery phase, aims at detecting associations, while the second, or replication phase, verifies the associations on new independent samples. As a rule, independent genotyping methods, such as real-time PCR or mass spectrometry, are used in replication phase. As seen from Tables 1 and S1, all studies except (Baranzini et al. 2009; Nischwitz et al. 2010; Patsopoulos et al. 2011) included these two phases.

Early MS GWASs included fewer than 1000 patients at discovery phase (Table 1), whereas up to 10,000 subjects were enrolled in the GWAS conducted in 2011 (IMSGC et al. 2011). On the whole, the number of associations increased with increasing sample size; this observation is in line with analogous data for other complex traits (Visscher et al. 2012). According to simulation studies, GWASs with fewer than about 2000 cases and 2000 controls at discovery phase will have low power to detect associations typical for common variants (Spencer et al. 2009). Ideally, to maximize the number of genuine associations, such studies should exceed the effective population size of ~10,000 individuals per group (Pèer et al. 2008; Sawcer et al. 2014). Of 13 GWASs performed for MS, only four (De Jager et al. 2009a, b, c; ANZgene et al. 2009; Patsopoulos et al. 2011; IMSGC et al. 2011) comply with the first requirement and only one (IMSGC et al. 2011) corresponds to the second. This is not surprising taking into account the median global MS prevalence of 33 per 100,000 (Browne et al. 2014).

In fact, the increase in the sample size in the majority of GWASs is achieved by including in the same study patients with different MS forms or even with the clinically isolated syndrome, a distinct etiologic form that may, but does not always, develop in clinically diagnosed MS (Kuhle et al. 2015). Patients with a certain MS form were examined only

**Table 1** Summary of genome-wide association studies of susceptibility to multiple sclerosis published up to June 2015

Reference	Number of tested SNPs	Initial sample size, MS patients <sup>a</sup> /controls	Replication sample size, MS patients <sup>a</sup> /controls	Number of reported loci <sup>b</sup> at $p < 1 \times 10^{-5}$	$p$ values range	$p$ value for HLA locus
IMSGC et al. (2007)	334,923	931 trios/2431 controls	609 trios, 2322/2987	6	$10^{-5}$ – $8.9 \times 10^{-81}$	$9 \times 10^{-81}$ (HLA-DRA)
Comabella et al. (2008)	428,867	242 (RRMS + SPMS)/242	375 (RRMS + SPMS)/375	1	$9 \times 10^{-11}$	$9 \times 10^{-11}$ (HLA-DRB1)
Aulchenko et al. (2008)	~262,000	45/195	2634/2930	1	$2.5 \times 10^{-10}$	NA
Baranzini et al. (2009)	551,642	978/883	–	11	$9 \times 10^{-6}$ – $10^{-32}$	$<1 \times 10^{-32}$ (HLA-DRB1)
De Jager et al. (2009a, b, c)	~2.56 mln	2624/7220	2215/2116	16	$10^{-5}$ – $4 \times 10^{-225}$	$4 \times 10^{-225}$ (HLA-DRB1), $1 \times 10^{-17}$ (HLA-B)
ANZgene et al. (2009)	302,098	1618/3413	2256/2310	9	$2.5 \times 10^{-6}$ – $1 \times 10^{-184}$	$10^{-184}$ (HLA-DRB1)
Jakkula et al. (2010)	297,343	68/136	4570/10143	2	$3 \times 10^{-10}$ – $2 \times 10^{-25}$	$2 \times 10^{-25}$ (HLA)
Sanna et al. (2010)	6,607,266	882 (RRMS)/872	1775/2005	2	$2 \times 10^{-10}$ – $1 \times 10^{-20}$	$1 \times 10^{-20}$ (HLA-DRB1, HLA-DQB1)
Nischwitz et al. (2010)	~300,000	590/825	–	4	$1 \times 10^{-6}$ – $4 \times 10^{-17}$	$4 \times 10^{-17}$ (HLA-DQA1)
Patsopoulos et al. (2011)	2,529,394	5545/12153	–	19	$4.9 \times 10^{-6}$ – $1 \times 10^{-206}$	$1 \times 10^{-206}$ (HLA-DRB1), $1 \times 10^{-11}$ (HLA-B)
IMSGC et al. (2011)	~660,000 SNPs; 60,000 CNVs	9772/17376	4218/7296	60	$3.7 \times 10^{-6}$ – $1.0 \times 10^{-132}$	$1.0 \times 10^{-132}$ (HLA-DRB1), $9 \times 10^{-23}$ (HLA-A)
Matesanz et al. (2012)	262,264	296/801	2785/2940	3	$2.1 \times 10^{-7}$ – $8.2 \times 10^{-34}$	$8.2 \times 10^{-34}$ (HLA)
Martinielli-Boneschi et al. (2012)	277,866	197 (PPMS)/234	379 (PPMS)/398	1	$6.7 \times 10^{-16}$	$6.7 \times 10^{-16}$ (HLA-DRB1, HLA-DQB)

Trios are families consisting of an affected individual and his/her parents. Those studies, in which genotype imputation was used, are given in italics

NA data not available

<sup>a</sup> If not indicated, MS patients with different MS courses are included

<sup>b</sup> Loci are understood as the position on a chromosome of a gene or other chromosome marker; also, the DNA at that position (the glossary in Human Genome Project information archive, available at [http://web.ornl.gov/sci/techresources/Human\\_Genome/glossary.shtml](http://web.ornl.gov/sci/techresources/Human_Genome/glossary.shtml))

in two GWASs, one involving patients with bout-onset MS (RRMS or SPMS) (Comabella et al. 2008), and the other with PPMS (Martinelli-Boneschi et al. 2012). This allowed searching for SNPs specifically involved in the given MS form development, but at the same time restricted potential sample size, thereby decreasing the power to identify associations. As a result, unfortunately, only *HLA* loci were identified in these studies.

Another way to increase the sample size that was applied in the majority of GWASs (see Table S1) is to enroll individuals from different ethnic groups. In such cases, the problem of population stratification arises, that is, the risk of false-positive/-negative results due to allele frequency variations among different ethnic groups. Indeed, the enrichment of specific disease-susceptibility alleles in more genetically homogeneous populations was noted in several studies of complex diseases (Lu et al. 2014). It has been observed that associations of some SNPs are more consistent among populations than associations of other SNPs. This may depend on SNPs allelic frequencies, localization in genes and chromosomes, different LD patterns in populations, different evolutionary history of genes affecting complex disease, etc. (Marigorta et al. 2011). A possible way to partially obviate the problem is to include subjects of closely related ethnic groups in the study (Baranzini et al. 2009). Statistical adjustment, wherein data are analyzed separately for each ethnic group, is used to avoid the problem, but additional parameters are thereby introduced in the analysis, and the significance level grows lower (Bush and Moore 2012). More intricate statistical methods in data analysis are also used (IMSGC et al. 2011) to cope with population stratification, relying on comparison of allelic frequencies with HapMap data and using modified data on ethnicity as covariates (Bush and Moore 2012).

The ethno-specific nature of some associations may be a reason for lack of reproducibility of GWASs. So *KIF1B* found in Danish isolate (Aulchenko et al. 2008) was not replicated in datasets from other GWASs and, moreover, was shown by equivalence analysis not to be associated in IMSGC samples (Gourraud and IMSGC 2011). In cases where the results are replicated in several ethnoses (e.g., *IL2RA*, *IL7RA*, *CLEC16A*), associations are likely to be universal.

Gender-specific effects were also analyzed in two GWASs (Baranzini et al. 2009; IMSGC et al. 2011). MS associations identified in these studies differ between men and women, though the majority of gender-specific associations are not significant at the genome-wide level. Three SNPs in the most well-powered GWAS to date (IMSGC et al. 2011) were associated with MS at the genome-wide significance either in women (rs1800693 in *TNFRSF1A*) or in men (rs2293370 in *TIMMDC1* and rs13333054 downstream *IRF8*). Taking into account the existing GWAS data

and differences in prevalence and clinical phenotype of MS in men and women, it seems reasonable to pay more attention to gender-specific associations to shed light on potentially divergent gender mechanisms of MS pathogenesis.

Results of a meta-analysis that N.A. Patsopoulos et al. (2011) reported in 2011 are also included in Table 1 along with GWAS data obtained on relatively or totally independent samples. In this meta-analysis, data from seven earlier genome-wide studies (the sample was substantially extended in one of them), which were carried out using different platforms and included approximately 750,000 SNPs, were integrated via genotype imputation to produce one panel of 2.5 million SNPs. Considering that various methods of genotype imputation are widely used in genome-wide studies, including both their first and second phases (Matesanz et al. 2012), the study (Patsopoulos et al. 2011) not only statistically analyzed the total wealth of GWAS data available for MS at that time, but also demonstrated a strategy for further development of such studies.

As is seen in Table 1, virtually all GWASs showed a highly significant association of MS with the MHC region, whose most important role is the genetic control of the immune response. The most significant signals mapped to *HLA-DRB1* class II gene (IMSGC et al. 2011; Patsopoulos et al. 2011; De Jager et al. 2009a, b, c). The significance level of the associations reaches  $p < 4 \times 10^{-225}$  for rs3135388 (De Jager et al. 2009a, b, c) and  $p < 1 \times 10^{-320}$  for non-attributed SNP tagging the same *DRB1* gene (IMSGC et al. 2011). As a rule, risk alleles at these SNPs tag for *HLA-DRB1\*15:01*. In several studies *HLA* class II is the only MS-associated locus (Comabella et al. 2008; Martinelli-Boneschi et al. 2012); in others, the number of significantly associated loci varies from 2 (Jakkula et al. 2010; Sanna et al. 2010) to 60 (IMSGC et al. 2011).

Eighty-six loci associated with MS with the significance level less than  $1 \times 10^{-5}$  were identified in GWASs conducted from 2007 to 2014 (Table S1). Inconsistency for a majority of identified loci in different GWASs is seen, possibly reflecting clinical and ethnic heterogeneity of samples as well as the presence of false-positive and/or false-negative results.

It is clear that with increasing genome coverage, a growing number of hypotheses are tested simultaneously, thus making false-positive associations more likely. Currently, the stringent requirements are put on the significance level, which is set at  $p < 5 \times 10^{-8}$  to take into account the Bonferroni correction for 1,000,000 comparisons in GWASs (Pèer et al. 2008; Stranger et al. 2011). Figure 1 (modified from <http://www.ebi.ac.uk/fgpt/gwas/>) shows the chromosomal positions for all the loci associated with MS at the given significance level (black circles) on the background of the loci that had been associated with any CD at the same significance level by the time of diagram construction.



**Fig. 1** Chromosome distribution of the loci that were associated with MS susceptibility in any GWAS included in the NHGRI-EBI GWAS Catalog as of April 17, 2015 at  $p < 5 \times 10^{-8}$  (black circles). Background circles, loci that were associated with other CDs at the same significance level by the time of figure preparation [modified from

the EBI website (<http://www.ebi.ac.uk/fgpt/gwas/>)]. One locus on the first chromosome is absent from this figure comparing with the original GWAS diagram because it is irrelevant for our study (rs533259, associated with MS lesion distribution from the study of Gourraud et al. 2013)

It is seen that MS risk loci were found on many, but not all, chromosomes, and that they are irregularly distributed through the chromosomes. MS-associated SNPs were the most numerous on chromosomes 1 and 6 (seven and six loci, respectively) and absent from autosomes 4, 13, 15, 21 and sex chromosomes. In total, 47 loci were associated with MS at  $p < 5 \times 10^{-8}$ .

### Hypothesis-driven genome-wide association studies for multiple sclerosis

In the hypothesis-driven studies described below, which are complementary to GWASs, the genome coverage density is increased for the regions most likely involved in the disease.

For the first time, this approach was applied for MS, along with ankylosing spondylitis, autoimmune thyroid disease (Graves' disease) and breast cancer (1500 shared controls and 1000 patients for each disease) in the study of Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium, et al. (WTCCC et al. 2007). Genotyping was performed with a custom-made Infinium array (Illumina) and involved 14,436 non-synonymous SNPs (nsSNPs) in protein-coding genome regions. In addition, because three of the diseases were of autoimmune etiology, a dense set of 897 SNPs throughout the MHC region was genotyped. The strongest association observed in the study was between SNPs in the MHC region and the three autoimmune diseases studied, with  $p$  values of  $< 1 \times 10^{-20}$  for each disease. In case of MS, the maximum signal is

centered near *HLA-DRB1*, a well-known MS-associated gene.

To exploit the hypothesis-driven GWAS approach, Illumina has used the available data on genetic susceptibility to 11 autoimmune and inflammatory disorders to design a new microarray (ImmunoChip) (Parkes et al. 2013). The ImmunoChip includes more than 195,000 SNPs, with 2000 SNPs known to display the most significant associations per autoimmune disorder and a high coverage density for 186 loci. In fact, this array allows deep replication of formerly identified associations with MS and other immune-mediated diseases. An analysis with the ImmunoChip is not only more efficient, but also less expensive.

The ImmunoChip was used in MS as a typical autoimmune inflammatory disorder (IMSGC et al. 2013). More than 14,000 MS patients and more than 24,000 healthy individuals were included in the study at the first stage, and independent samples from earlier genome-wide studies were additionally used at the second stage, the total sample size thus exceeding 80,000 (all of the subjects were Caucasians). A large panel of highly significant associations was found in this global study. Particularly, association with MS at  $p < 5 \times 10^{-8}$  was demonstrated for 49 polymorphic loci that had been found earlier (IMSGC et al. 2011), and 48 new loci were found (Table S2) that had not been detected in any earlier GWAS and still need replication.

Recently another ImmunoChip MS study was published (Isobe et al. 2015). All subjects were African Americans. The first stage of the study was performed on 803 MS patients and 1516 controls. 21 non-MHC MS risk loci out of 110 established earlier in Europeans (Oksenberg 2013) were replicated. The second phase was conducted on 620 MS cases and 1565 controls and was aimed toward replication of new associations. None of the SNPs achieved the level of significance less than  $1 \times 10^{-5}$ . The power of the study is relatively low (MS is rare in persons of African ancestry); another problem is that it is not quite clear to what extent the ImmunoChip is applicable to investigating MS susceptibility in African Americans, keeping in mind that haploblocks are shorter in more ancient ancestries (Lambert and Tishkoff 2009).

### Ms-associated loci replicated in independent genome-wide studies

Table 2 summarizes results of GWASs and hypothesis-driven genome-wide studies. Taking into account inconsistency of different genome-wide results, we present here only those MS-associated loci that were identified in at least two independent studies. Depending on the used microarrays, which include different SNPs in panels, a number of linked SNPs have been identified as conferring MS risk in the

same loci. The following thresholds were set: the genome-wide significance level ( $p$  value  $< 5 \times 10^{-8}$ ) for the combined discovery and replication datasets should be reached at least in one study, while in the other(s)  $p$  value(s) should be less than  $1 \times 10^{-5}$ . Different thresholds are used because of different numbers of simultaneously tested hypotheses in an initial and replication GWAS studies.

The loci corresponding to these criteria are summarized in Table 2. In total, up to May 11, 2015, 40 loci and 52 genes mapped to them are characterized, and the association of their alleles with MS can be accounted for as validated at least in Caucasians. They are located in all autosomes except 4, 7, 9, 13, 15, 18, 21.

As seen from Table 2, most associated SNPs are located in introns, intergenic regions; associations found in exons or UTRs are considerably less frequent. Keeping in mind that GWAS allows to identify MS-associated genomic regions, each of the MS-associated SNPs can be either functional (causative) SNP or tag SNP for the causative polymorphism, which was not tested directly. Causative polymorphism may affect the functional activity, level, timing or location of the gene's product. So far, functional significance was demonstrated for several polymorphisms. For example, rs6897932, rs2104286, and rs1800693 of the cytokine receptor genes *IL7RA*, *IL2RA*, and *TNFRSF1A* influence proportions of the membrane-bound and soluble respective receptor forms (Gregory et al. 2007, 2012; Maier et al. 2009).

As for tag SNPs, association significance is in straight correlation with LD between causative SNP and tag SNP included in the panel. To identify the causative SNPs, several other SNPs not tested in the original GWAS are additionally analyzed in the associated loci; otherwise, the regions of interest are resequenced. This approach is known as fine-mapping study. The thorough analysis often involves logistic regression analysis to search for the most associated SNP as well as to detect independent associations within one locus. As an example, rs2300747 was identified as a primary association in *CD58* in fine-mapping study (De Jager et al. 2009b) and therefore tested in the replication phase and in the meta-analysis of the simultaneous GWAS (De Jager et al. 2009a, b, c). Independent effects were assumed for two SNPs of *TNFRSF1A* (De Jager et al. 2009a, b, c).

Substantial differences in  $p$  values for a given locus (Table 2) can be seen. This can be due to the statistical power of the study (the lowest  $p$  values were obtained for the most well-powered studies), ethnic specificity of associations with MS, used SNP panels, etc. It is important to note that among loci withstanding the suggested criteria of replication, the majority could fulfill even more stringent requirements, having been identified in at least two studies at the genome-wide significance level. This particularly

**Table 2** Multiple sclerosis-associated alleles of polymorphic loci, which have been identified in at least two independent studies out of conducted genome-wide studies and complementary to GWAS hypothesis-driven studies (WTCCC et al. 2007; IMMSGC et al. 2013), and showed  $p < 5 \times 10^{-8}$  in at least one of them

Chromosome	Locus	Gene	rs ID	SNP localization	Risk allele	$p$ value for combined datasets	OR	References
1	1p36.32	<i>MME1L1</i>	rs4648356	Intergenic	C	$3.1 \times 10^{-14}$	1.16	IMMSGC et al. (2011)
			rs3748817	Intronic	A	$1.3 \times 10^{-26}$	1.14	IMMSGC et al. (2013)
	1p22.3	<i>DDAHI</i>	rs233100	Intergenic	G	$1.0 \times 10^{-6}$	1.08	IMMSGC et al. (2011)
			rs11587876	Intronic	A	$4.4 \times 10^{-9}$	1.09	IMMSGC et al. (2013)
	1p22.1	<i>EVIS, RPL5</i>	rs6604026	Intronic	G	$8.0 \times 10^{-6}$	1.15	IMMSGC et al. (2007)
						$2.5 \times 10^{-6}$	1.17	ANZgene et al. (2009)
	1p21.2	<i>VCAMI</i>	rs11810217	Intronic	A	$6.5 \times 10^{-12}$	1.15	IMMSGC et al. (2011)
			rs41286801	3' UTR	A	$1.4 \times 10^{-26}$	1.19	IMMSGC et al. (2013)
			rs11581062	Intronic	G	$2.5 \times 10^{-10}$	1.13	IMMSGC et al. (2011)
			rs7552544	Intergenic	A	$1.9 \times 10^{-16}$	1.10	IMMSGC et al. (2013)
rs2300747			Intronic	A	$3.0 \times 10^{-10}$	1.30 <sup>a</sup>	De Jager et al. (2009a, b, c)	
1p13.1	<i>CD58</i>	rs1335532	Intronic	A	$9.6 \times 10^{-8}$	1.28	ANZgene et al. (2009)	
							IMMSGC et al. (2011)	
1q31.2	<i>RGS1</i>	rs6677309	Intronic	A	$2.0 \times 10^{-9}$	1.18	IMMSGC et al. (2011)	
		rs2760524	Intergenic	G	$5.4 \times 10^{-42}$	1.29	IMMSGC et al. (2013)	
					$9.8 \times 10^{-6}$	1.15 <sup>a</sup>	De Jager et al. (2009a, b, c)	
							IMMSGC et al. (2011)	
1q32.1	<i>C1orf106</i>	rs1323292	Intergenic	A	$2.3 \times 10^{-8}$	1.12	IMMSGC et al. (2011)	
		rs1359062	Intergenic	C	$4.8 \times 10^{-20}$	1.15	IMMSGC et al. (2013)	
		rs7522462	Intronic	G	$1.9 \times 10^{-9}$	1.11	IMMSGC et al. (2011)	
		rs55838263	Intronic	A	$4.0 \times 10^{-19}$	1.13	IMMSGC et al. (2013)	
		rs7595037	Intergenic	A	$5.1 \times 10^{-11}$	1.10	IMMSGC et al. (2011)	
2p13.3	<i>PLEK</i>	rs7595717	Intergenic	A	$1.2 \times 10^{-13}$	1.10	IMMSGC et al. (2013)	
		rs10201872	Intronic	A	$1.8 \times 10^{-10}$	1.13	IMMSGC et al. (2011)	
2q37.1	<i>SP140</i>		rs9989735	Intronic	C	$4.2 \times 10^{-23}$	1.16	IMMSGC et al. (2013)



Table 2 continued

Chromosome	Locus	Gene	rs ID	SNP localization	Risk allele	<i>p</i> value for combined datasets	OR	References		
3	3p24.1	<i>EOMES</i>	rs170934	Intergenic	T	$1.6 \times 10^{-8}$	1.17	Patsopoulos et al. (2011)		
			rs11129295	Intergenic	A	$1.2 \times 10^{-9}$	1.11	IMSGC et al. (2011)		
	3q13.33	<i>CD86</i> , <i>TIMMDC1</i> , <i>IQCB1</i>	rs2371108	Intergenic	A	$1.5 \times 10^{-15}$	1.10	IMSGC et al. (2013)		
			rs9282641	Exonic	G	$1.0 \times 10^{-11}$	1.21	IMSGC et al. (2011)		
			rs2252114	Intergenic	C	$1.2 \times 10^{-24}$	1.13	IMSGC et al. (2013)		
			rs2293370	Intronic	G	$3.0 \times 10^{-9}$	1.13	IMSGC et al. (2011)		
			rs1131265	Exonic	C	$1.4 \times 10^{-23}$	1.17	IMSGC et al. (2013)		
			rs1920296	Intronic	C	$6.5 \times 10^{-22}$	1.12	IMSGC (2013); IMSGC et al. (2013)		
			3q25.33	<i>IL12A</i>	rs4680534	Intronic	C	$6.0 \times 10^{-6}$	1.12	De Jager et al. (2009a, b, c)
					rs2243123	Intronic	G	$7.0 \times 10^{-6}$	1.09	IMSGC et al. (2011)
5p13.2	<i>IL7RA</i>	rs1014486	Intergenic	G	$1.1 \times 10^{-18}$	1.11	IMSGC et al. (2013)			
		rs6897932	Exonic	C	$3.0 \times 10^{-7}$	1.18	IMSGC et al. (2007)			
					$2.0 \times 10^{-6}$	1.12	De Jager et al. (2009a, b, c)			
		rs6881706	Intergenic	C	$1.7 \times 10^{-8}$	1.11	IMSGC et al. (2011)			
5p13.1	<i>PTGER4</i>	rs6896969	Intergenic	C	$4.3 \times 10^{-17}$	1.12	IMSGC et al. (2013)			
		rs4613763	Intergenic	G	$6.9 \times 10^{-16}$	1.21	IMSGC et al. (2011)			
5q31.1	<i>TCF7</i>	rs9292777	Intergenic	T	$9.8 \times 10^{-9}$	1.16 <sup>a</sup>	Matesanz et al. (2012)			
		rs756699	Intergenic	A	$6.0 \times 10^{-7}$	1.12	IMSGC et al. (2011)			
5q31.3	<i>NDFIP1</i>	rs1062158	Intronic	A	$8.8 \times 10^{-11}$	1.12	IMSGC et al. (2013)			
		NA (proxy rs1036207)	Intronic	C	$2.0 \times 10^{-6}$	1.08	IMSGC et al. (2011)			
5q35.3	<i>RGS14</i>	rs4075958	Intergenic	A	$5.0 \times 10^{-7}$	1.09	IMSGC et al. (2011)			
		rs4976646	Intronic	G	$4.4 \times 10^{-18}$	1.12	IMSGC et al. (2013)			

Table 2 continued

Chromosome	Locus	Gene	rs ID	SNP localization	Risk allele	<i>p</i> value for combined datasets	OR	References		
6	6p21.32	<i>HLA-DRB1</i>	NA	Intronic		$<1.0 \times 10^{-20}$	NA	WTCCC et al. (2007)		
			rs3129934			$9 \times 10^{-11}$	3.30	Comabella et al. (2008)		
								$<1.0 \times 10^{-32}$	NA	Baranzini et al. (2009)
								$6.7 \times 10^{-16}$	2.34	Martinelli-Bonneschi et al. (2012)
								$9 \times 10^{-81}$	1.99	IMSGC et al. (2007)
								$4.0 \times 10^{-225}$	2.75	De Jager et al. (2009a, b, c)
								$7.0 \times 10^{-184}$	2.78	ANZgene et al. (2009)
								$4.0 \times 10^{-17}$	2.62	Nischwitz et al. (2010)
								$1.0 \times 10^{-20}$	2.05	Sanna et al. (2010)
								$1.0 \times 10^{-206}$	2.97	Patsopoulos et al. (2011)
								$<1.0 \times 10^{-320}$	3.10	IMSGC et al. (2011)
			8	6q15	<i>BACH2</i>	rs12212193	Intronic	G	$3.8 \times 10^{-8}$	1.09
rs72928038	Intronic	A				$1.5 \times 10^{-15}$	1.14	IMSGC et al. (2013)		
rs9321619	Intergenic	G				$1.7 \times 10^{-5}$	1.12	De Jager et al. (2009a, b, c)		
rs13192841	Intergenic	A				$1.0 \times 10^{-8}$	1.10	IMSGC et al. (2011)		
rs17066096	Intergenic	G				$6.0 \times 10^{-13}$	1.14	IMSGC et al. (2011)		
rs11154801	Intronic	A				$1.6 \times 10^{-23}$	1.13	IMSGC et al. (2013)		
rs67297943	Intergenic	A				$1.0 \times 10^{-13}$	1.12	IMSGC et al. (2011)		
rs1738074	5' UTR	G				$1.8 \times 10^{-20}$	1.12	IMSGC et al. (2013)		
rs212405	Intergenic	T				$5.5 \times 10^{-13}$	1.11	IMSGC et al. (2013)		
rs4410871	Intergenic	G				$5.3 \times 10^{-11}$	1.13	IMSGC et al. (2011)		
rs759648	Intergenic	C				$8.0 \times 10^{-21}$	1.12	IMSGC et al. (2013)		
rs2019960	Intergenic	G				$7.7 \times 10^{-9}$	1.11	IMSGC et al. (2011)		
rs759648	Intergenic	C	$4.3 \times 10^{-16}$	1.12	IMSGC et al. (2013)					
rs2019960	Intergenic	G	$5.2 \times 10^{-9}$	1.12	IMSGC et al. (2011)					
rs759648	Intergenic	C	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)					
8	6q25.3	<i>TAGAP</i>	rs67297943	Intergenic	A	$1.0 \times 10^{-8}$	1.10	IMSGC et al. (2011)		
			rs1738074	5' UTR	G	$6.0 \times 10^{-13}$	1.14	IMSGC et al. (2011)		
			rs212405	Intergenic	T	$1.6 \times 10^{-23}$	1.13	IMSGC et al. (2013)		
			rs4410871	Intergenic	G	$1.0 \times 10^{-13}$	1.12	IMSGC et al. (2011)		
			rs67297943	Intergenic	A	$1.8 \times 10^{-20}$	1.12	IMSGC et al. (2013)		
			rs1738074	5' UTR	G	$5.5 \times 10^{-13}$	1.11	IMSGC et al. (2013)		
			rs212405	Intergenic	T	$5.3 \times 10^{-11}$	1.13	IMSGC et al. (2011)		
			rs4410871	Intergenic	G	$8.0 \times 10^{-21}$	1.12	IMSGC et al. (2013)		
			rs67297943	Intergenic	A	$7.7 \times 10^{-9}$	1.11	IMSGC et al. (2011)		
			rs1738074	5' UTR	G	$4.3 \times 10^{-16}$	1.12	IMSGC et al. (2013)		
			rs212405	Intergenic	T	$5.2 \times 10^{-9}$	1.12	IMSGC et al. (2011)		
			rs4410871	Intergenic	G	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
8	8q24.21	<i>MYC, MIR1204, MIR1205, PVT1, MIR1208, MIR3686</i>	rs2019960	Intergenic	G	$4.3 \times 10^{-16}$	1.12	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.2 \times 10^{-9}$	1.12	IMSGC et al. (2011)		
			rs2019960	Intergenic	G	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
			rs2019960	Intergenic	G	$4.3 \times 10^{-16}$	1.12	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.2 \times 10^{-9}$	1.12	IMSGC et al. (2011)		
			rs2019960	Intergenic	G	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
			rs2019960	Intergenic	G	$4.3 \times 10^{-16}$	1.12	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.2 \times 10^{-9}$	1.12	IMSGC et al. (2011)		
			rs2019960	Intergenic	G	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		
			rs759648	Intergenic	C	$5.0 \times 10^{-10}$	1.08	IMSGC et al. (2013)		

Table 2 continued

Chromosome	Locus	Gene	rs ID	SNP localization	Risk allele	<i>p</i> value for combined datasets	OR	References
10	10p15.1	<i>IL2RA</i>	rs12722489	Intronic	C	$3.0 \times 10^{-8}$	1.25	IMSGC et al. (2007)
						$3.7 \times 10^{-8}$	1.23 <sup>a</sup>	Patsopoulos et al. (2011)
	10q22.3	<i>ZMIZ1</i>	rs2104286	Intronic	A	$9.0 \times 10^{-8}$	1.15	De Jager et al. (2009a, b, c)
						$7.4 \times 10^{-6}$	1.16	ANZgene et al. (2009)
			rs3118470	Intronic	G	$2.3 \times 10^{-47}$	1.22	IMSGC et al. (2013)
			rs7090512	Intergenic	G	$2.0 \times 10^{-9}$	1.12	IMSGC et al. (2011)
			rs1250540	Intronic	G	$2.6 \times 10^{-14}$	1.19	IMSGC et al. (2011)
			rs1250550	Intronic	A	$2.0 \times 10^{-6}$	1.12	De Jager et al. (2009a, b, c)
			rs1250542	Intronic	A	$6.3 \times 10^{-9}$	1.10	IMSGC et al. (2011)
			rs1782645	Intronic	A	$4.0 \times 10^{-7}$	1.15	Patsopoulos et al. (2011)
10q23.33	<i>HHEX</i>	rs7923837	Intergenic	G	$2.5 \times 10^{-15}$	1.10	IMSGC et al. (2013)	
		rs17824933	Intronic	G	$4.9 \times 10^{-9}$	1.10	IMSGC et al. (2011)	
		rs650258	Intronic	G	$4.3 \times 10^{-17}$	1.11	IMSGC et al. (2013)	
		rs34383631	Intergenic	G	$4.0 \times 10^{-9}$	1.18	De Jager et al. (2009a, b, c)	
11q12.2	<i>CD6</i>	rs630923	Intronic	G	$1.7 \times 10^{-9}$	1.12	IMSGC et al. (2011)	
		rs9736016	Intergenic	A	$3.7 \times 10^{-23}$	1.13	IMSGC et al. (2013)	
		rs4149584	Intergenic	C	$2.8 \times 10^{-7}$	1.12	IMSGC et al. (2011)	
		rs1800693	Intergenic	T	$3.0 \times 10^{-15}$	1.10	IMSGC et al. (2013)	
12p13.31	<i>TNFRSF1A</i>		Exonic	T	$5.0 \times 10^{-6}$	1.58	De Jager et al. (2009a, b, c)	
			Intronic	C	$2.0 \times 10^{-11}$	1.20	De Jager et al. (2009a, b, c)	
12q14.1	<i>METTL1, CYP27B1</i>	rs703842	3' UTR	A	$1.8 \times 10^{-10}$	1.12	IMSGC et al. (2011)	
					$6.7 \times 10^{-28}$	1.14	IMSGC et al. (2013)	
					$5.0 \times 10^{-11}$	1.23	ANZgene et al. (2009)	
14q24.1	<i>ZFP36L1</i>	rs12368653	Intronic	A	$2.0 \times 10^{-7}$	1.11	IMSGC et al. (2011)	
		rs4902647	Intergenic	G	$9.3 \times 10^{-12}$	1.11	IMSGC et al. (2011)	
		rs2236262	Intronic	A	$2.5 \times 10^{-12}$	1.08	IMSGC et al. (2013)	
14q31.3	<i>GALC</i>	rs2119704	Intronic	C	$2.2 \times 10^{-10}$	1.26	IMSGC et al. (2011)	
		rs74796499	Intronic	C	$2.4 \times 10^{-20}$	1.32	IMSGC et al. (2013)	

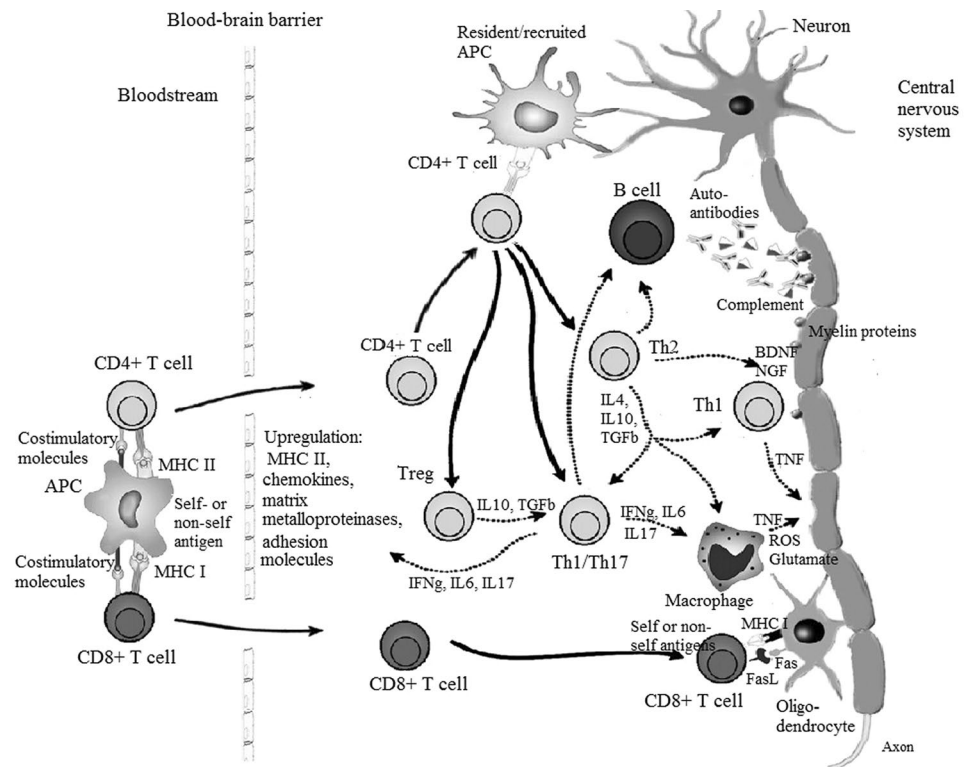
Table 2 continued

Chromosome	Locus	Gene	rs ID	SNP localization	Risk allele	<i>p</i> value for combined datasets	OR	References
16	16p13.13	<i>CLEC16A</i>	rs6498169	Intronic	G	$4.0 \times 10^{-6}$	1.14	IMSGC et al. (2007)
			rs11865121	Intronic	C	$2.0 \times 10^{-7}$	1.15	De Jager et al. (2009a, b, c)
	16q24.1	<i>IRF8</i>	rs7200786	Intronic	A	$6.3 \times 10^{-14}$	1.15	IMSGC et al. (2011)
			rs12927355	Intronic	G	$6.4 \times 10^{-46}$	1.20	IMSGC et al. (2013)
			rs17445836	Intergenic	G	$4.0 \times 10^{-9}$	1.25	De Jager et al. (2009a, b, c)
17	17q21.2	<i>STAT3</i>	rs1333054	Intergenic	A	$1.0 \times 10^{-8}$	1.12	IMSGC et al. (2011)
			rs35929052	Intergenic	G	$5.9 \times 10^{-12}$	1.15	IMSGC et al. (2013)
			rs744166	Intronic	G	$2.8 \times 10^{-10}$	1.15 <sup>a</sup>	Jakkula et al. (2010)
						$6.4 \times 10^{-6}$	1.13	Patsopoulos et al. (2011)
						$2.0 \times 10^{-10}$	1.10	IMSGC et al. (2011)
19	19p13.3	<i>TNFSF14</i>	rs9891119	Intronic	C	$4.1 \times 10^{-8}$	1.22 <sup>a</sup>	Patsopoulos et al. (2011)
			rs2293152	Intronic	C			
	19p13.2	<i>TYK2</i>	rs4796791	Intronic	A	$3.7 \times 10^{-20}$	1.12	IMSGC et al. (2013)
			rs1077667	Intronic	G	$9.4 \times 10^{-14}$	1.16	IMSGC et al. (2011)
			rs8112449	Intergenic	G	$1.7 \times 10^{-24}$	1.16	IMSGC et al. (2013)
19q13.33	<i>DKK1</i>	rs34536443	Exonic	C	$1.0 \times 10^{-6}$	1.10	IMSGC et al. (2011)	
		rs2303759	Exonic	C	$1.8 \times 10^{-14}$	1.29	IMSGC et al. (2013)	
20	20q13.2	<i>CYP24A1</i>	rs8107548	Intronic	G	$5.2 \times 10^{-9}$	1.11	IMSGC et al. (2011)
			rs2248359	Intergenic	G	$5.7 \times 10^{-15}$	1.11	IMSGC et al. (2013)
	20q13.12	<i>CD40</i>		Intergenic	G	$2.5 \times 10^{-11}$	1.12	IMSGC et al. (2011)
						$2.0 \times 10^{-13}$	1.09	IMSGC et al. (2013)
						$1.3 \times 10^{-7}$	1.20	ANZgene et al. (2009)
22	22q11.21	<i>MAPK1</i>	rs6074022	Intergenic	C			Patsopoulos et al. (2011)
						$4.9 \times 10^{-6}$	1.15	Patsopoulos et al. (2011)
			rs2425752	Intronic	A	$5.0 \times 10^{-10}$	1.10	IMSGC et al. (2011)
			rs4810485	Intronic	A	$7.7 \times 10^{-16}$	1.11	IMSGC et al. (2013)
			rs2283792	Intronic	C	$4.7 \times 10^{-9}$	1.09	IMSGC et al. (2011)
			$5.5 \times 10^{-16}$	1.10	IMSGC et al. (2013)			

The *p* values  $< 5 \times 10^{-8}$  are in bold

<sup>a</sup> OR was shown for the alternative (protective) allele in the respective original studies

**Fig. 2** The main stages of MS pathogenesis (see description in text). *APC* antigen-presenting cell, *BDNF* brain-derived neurotrophic factor, *HLA/MHC* human leukocyte antigen, *IFN $\gamma$*  interferon gamma, *GA* gliatramer acetate, *IL* interleukin, *NGF* nerve growth factor, *TNF* tumor necrosis factor, *TGF $\beta$*  transforming growth factor beta, *Th* T-helper cell, *Treg* T regulatory cell, *FasL* Fas-ligand



corresponds to loci identified in IMSGC study (IMSGC et al. 2011) and replicated by ImmunoChip study (IMSGC et al. 2013); several loci, including HLA, *IL2RA*, *IL7RA* genes, were identified in almost all studies and in many or all of them achieved  $p < 5 \times 10^{-8}$ . On the whole,  $p$  values as low as  $p < 1 \times 10^{-320}$  for HLA and  $2.3 \times 10^{-47}$  for non-HLA loci (*IL2RA* gene) were obtained (IMSGC et al. 2011, 2013).

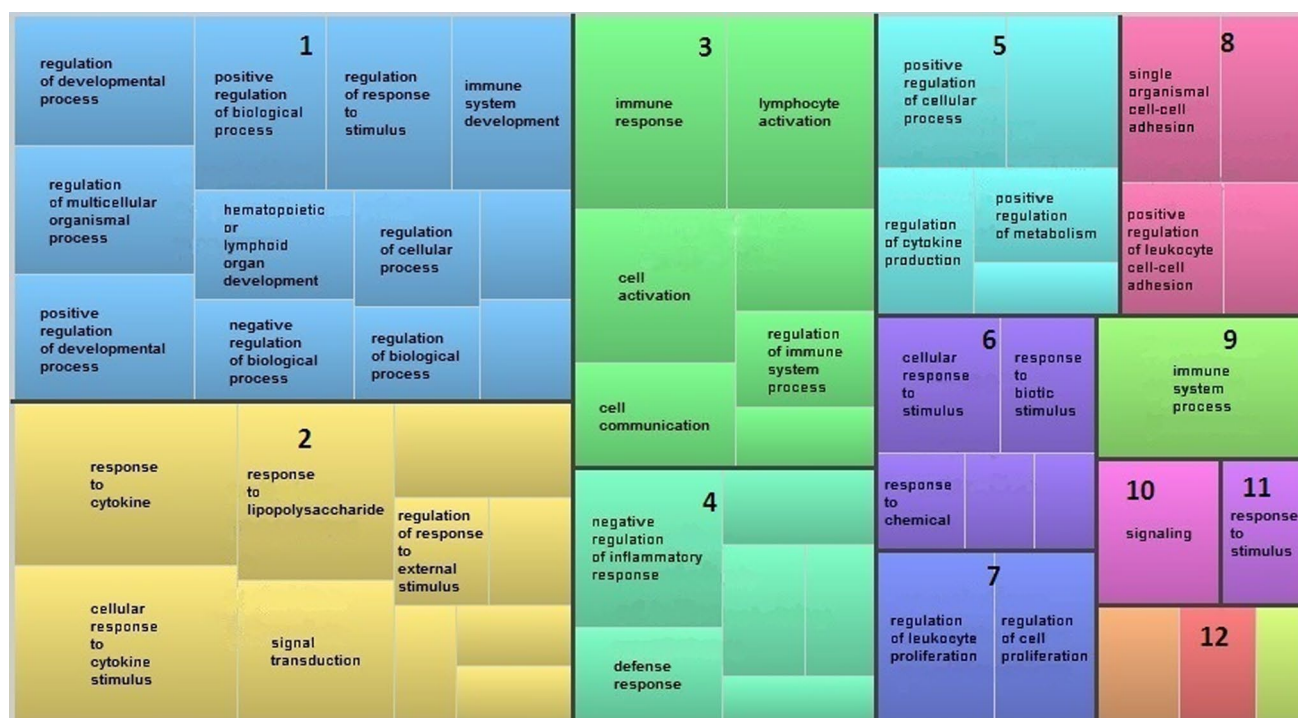
Odds ratio (OR) is a commonly applied measure of effect size. For the HLA locus, OR ranges from 2.05 to 3.3. For non-HLA genes, the MS risk alleles exert only a modest effect, odds ratio (OR) ranging from 1.1 to 1.3 with few exceptions (Table 2).

### Genes carrying gwas-identified polymorphisms: possible biological roles in pathogenesis of multiple sclerosis

Numerous studies report that MS development is mediated by autoimmune inflammation in CNS, which results in demyelination, oligodendrocyte destruction, axonal breakdown, gliosis and neurodegeneration leading to irreversible neurological dysfunction (Goverman 2009).

Figure 2 represents the main stages of MS pathogenesis. Peripheral anergic autoreactive lymphocytes (mainly T cells) undergo activation with microbial superantigens or

with self-antigens with enhanced immunogenicity, particularly due to chronic inflammation (Tauber et al. 2007). The next step is penetration of these cells through blood–brain barrier (BBB), a complex organization of cerebral endothelial cells, pericytes and their basal lamina, which are surrounded and supported by astrocytes and perivascular macrophages (Ortiz et al. 2014). Cytokine imbalance toward the increased production of pro-inflammatory cytokines by T-helpers 1 (Th1) and 17 (Th17) (Hedegaard et al. 2008) promotes the expression of adhesion molecules and HLA class II molecules on endothelial cells of BBB (Dore-Duffy et al. 1993). In addition, chemokines and pro-inflammatory cytokines with the assistance of matrix metalloproteinases may influence the integrity of tight junctions between endothelial cells, facilitating the migration of leukocytes into CNS (Holman et al. 2011). Activated myelin-specific T and B cells undergo the secondary reactivation in CNS by resident or recruited antigen-presenting cells (APCs). Activation of microglia and macrophages results in increased phagocytosis and production of cytotoxic agents such as oxidative radicals, NO, tumor necrosis factor (TNF), and glutamate, which contribute to myelin sheath damage (Lovett-Racke et al. 2011; Cunningham 2013). Activated B cells produce antibodies to myelin proteins and lipids, thus activating a complement system with the following formation of membrane attack complexes that directly damage myelin sheath (von Büdingen et al. 2011). Conversely,



**Fig. 3** The main biological processes, in which are involved genes from Table 2. These genes were analyzed with AmiGO website (<http://amigo.geneontology.org/amigo/landing>); after that we put all GO terms, which were significantly ( $p < 0.05$ ) involved in this gene

set, as well as their  $p$  values, in search window of REViGO web server (<http://revigo.irb.hr/>; Supek et al. 2011). The results are presented as Treemap; the size of squares reflects  $\log p$  values of these processes' association with the selected genes

FoxP3-expressing natural Treg cells (Tregs) and IL-10-producing T regulatory type 1 cells (Tr1) play a crucial role in restricting autoimmune neuroinflammation and establishing immunological tolerance (Kleinewietfeld and Hafler 2014). Activated CD8<sup>+</sup> T-lymphocytes participate in oligodendrocyte destruction and axonal breakdown via FAS/FAS-ligand-mediated cytolysis (Denic et al. 2013).

Autoimmune mechanisms are mostly intense during disease exacerbations in RRMS. The chronic MS forms are characterized by more intensive myelin degeneration and oligodendrocyte and axonal loss (Lassmann et al. 2007). However, axonal loss was shown both in acute and chronically active lesions and in normal-appearing white matter in brain tissue of MS patients (Nylander and Hafler 2012). Many different immunological mechanisms may lead to axon injury including destruction by specific T cells, activated microglia, invading macrophages, natural killer cells and auto-antibodies against specific antigens (Iwanowski and Losy 2015). In addition, several non-immunologic mechanisms, such as neurotrophic factors imbalance (Hohlfeld 2008) and glutamate excitotoxicity (Stys 2005), can contribute to axon injury and neurodegeneration.

SNPs identified and replicated in GWASs of MS (Table 2) are located mostly in or near protein-coding genes and relatively rarely in known RNA-coding genes.

The majority of MS-associated protein-coding genes are directly involved in immune-related functions. Relevant references are available in the GeneCards human gene database (<http://www.genecards.org>), a US National Center for Biotechnology Information (NCBI) resource (<http://www.ncbi.nlm.nih.gov/gene>), and in other sources. Here, we performed a brief GO search to overview the functions, which are enriched in this gene set. We searched for biological processes of genes from Table 2 in AmiGO database (<http://amigo.geneontology.org/amigo/landing>) using Bonferroni correction, provided by the website. Fifty-two GO terms significantly associated with this gene set were found ( $p < 0.05$ ); for genes *MIR1204*, *MIR1205*, *MIR1208*, *MIR3686*, and *PVT1*, the search failed. To get rid of redundant GO terms, we put all found GO terms and their  $p$  values in the search window of REViGO web server (<http://revigo.irb.hr/>; Supek et al. 2011); one of the convenient graphical representations, which is available at the server, is shown at Fig. 3. It is seen that the majority of genes are involved in lymphocyte activation, cytokine response, cell–cell adhesion, regulation of immune response and other immune-related functions, as well as development regulation.

It is well known that HLA locus is an essential component of immune response and immune system

development. After 10 years of GWAS, the MHC region still represents roughly half of the MS genetic risk. The fact that HLA locus has a highly significant association with MS is consistent with the autoimmune nature of MS. A genotypic gradient of risk with ascending hierarchy from *DRB1\*11:01/DRB1\*11:01* up to *DRB1\*15:01/DRB1\*15:01* has been described (Oksenberg et al. 2008). Multiple epistatic interactions occur in the class II region, where certain genes nullify the risk conferred by *DRB1* genotypes, and other genes act as modifiers of disease severity. It is likely that independent signals in the telomeric class I region of the HLA locus confer protection against MS (Oksenberg 2013). Further studies with denser coverage of the class I and class III HLA gene regions are necessary for identifying other MS-associated genes of the locus.

As for the non-HLA genes associated with MS, an essential part of those presented in Table 2 are directly involved in the T cell function; that may indicate the leading role of T cell immunity in MS development. We will consider briefly the functions known for the protein products of these genes. T cell differentiation and survival of immature CD4 + CD8+ thymocytes depends on the TCF7 transcription factor, whose activation is mediated by the prostaglandin E receptor 4 (PTGER4). Cytokines and cytokine receptors (IL12, IL7R, and IL2R), as well as the costimulatory molecules CD58 and CD86, also play a substantial role in T cell proliferation and differentiation. T cell activation depends on the regulatory protein encoded by *TAGAP*, and activation maintenance depends on expression of the CD6 receptor on the T cell surface. In turn, the product of *NDFIP1* prevents T-helper-mediated inflammation from passing into a chronic phase. TYK2 kinase is involved in T-helper polarization; i.e., a decrease in TYK2 activity is associated with a Th2 bias of T-helper differentiation.

The products of *CD40*, *RGS1*, and *CXCR5* are directly involved in the B cell function. A large group of genes codes for proteins involved in the NF- $\kappa$ B signaling pathway and inflammatory response. Genes of the group belong to the superfamily of the tumor necrosis factor (*TNFSF14* and *TNFAIP3*) or its receptors (*TNFRSF1A*), cytokine receptors (*IL22RA2*), and adhesion molecules (*VCAM1*). The protein products of *STAT3*, *IRF8*, *TYK2*, and *ZMIZ1* are involved in the JAK/STAT signaling pathways that regulate cytokine expression and affect the immune response development.

Some of the genes presented in Table 2 are expressed in the nervous system to a higher or lower level; these are *EOMES*, *OLIG3*, *GALC*, *AH11*, *RPL5*, and *MMEL1*. The transcriptional activator eomesodermin, which is encoded by *EOMES*, plays an essential role in early differentiation of the embryo and the brain and is involved in CD8+ T cell differentiation. The *GALC* product acts as a galactosylceramidase and plays a role in lysosomal catabolism of

galactosylceramide, which is a major glycolipid of the axon myelin sheath.

The functional role the other GWAS-identified genes play in MS susceptibility is not that clear. These genes encode transcription factors (e.g., the repressor HHEX, which is involved in hematopoietic cell differentiation, and protooncogene MYC, an important regulator of the cell cycle), cytochrome family proteins (CYP24A1 and CYP27B1), important components of various signaling pathways (e.g., MAPK1 protein kinase, regulator of G protein signaling RGS14, transcription factor BACH2), and many enzymes (among them methyltransferase METTL1, which is involved in tRNA methylation, and dimethylamino-hydroxylase DDAH1, which plays a role in nitric oxide generation). The function remains unknown for some of the genes (e.g., *CLEC16A*, although there is evidence that the gene is expressed almost exclusively in cells of the immune system).

Several microRNA genes—*MIR1204*, *MIR1205*, *MIR1208*, and *MIR3686*—map to MS-associated locus 8q24.21. This locus also harbors *PVT1*, which encodes a regulatory RNA, attributed to the long noncoding RNA class. As the majority of identified SNPs are located in introns or intergenic regions (Table 2) of protein-coding genes, it can be suggested that the list of MS-associated RNA-coding genes will be extended. These findings and considerations may indicate that the epigenetic regulation plays an important role in MS pathogenesis.

In general, the results of GWASs, both classical and hypothesis-driven, taken in the context of biological functions of identified genes, are in line with the modern concept of MS as an autoimmune disease affecting CNS, and can even expand our knowledge about MS pathogenesis. It should be mentioned that about one-third of MS-associated genes were shown to be associated with other autoimmune disorders in the ImmunoChip study (Oksenberg 2013).

## Main conclusions and future perspectives

GWASs identified the multiple loci and genes involved in MS development, thus leading, according to the figurative expression (Oksenberg 2013), to creation of a new genetic atlas of MS. The currently available list of genes has already provided a better understanding of the potential molecular mechanisms involved in MS pathogenesis. However, we apparently are at the initial stage of this process. Recent estimations based on sibling studies (Sadee et al. 2014) suggest that the MS risk loci identified to date explain only about 27 % of its total heritability (Lill 2014). At that, HLA-*DRB1\*15:01* alone explains about 20 % of heritability (Lill 2014). There are several explanations of this “missing heritability” (Manolio et al. 2009; Sadee

et al. 2014) in complex diseases, of which MS is a typical example.

One of them is a poor reproducibility of the GWAS-identified MS loci in other GWASs or subsequent candidate gene studies. As a matter of fact, the stringent requirements of data significance fail to completely prevent false-positive findings and simultaneously can prevent detection of some true associations because of insufficient statistical power of the studies.

Another possible reason of “missing heritability” is that the GWAS design is based on the common disease-common variant (CDCV) hypothesis, which states that weak effects of common alleles (MAF more than 1 %) provide a genetic basis for common diseases. The CDCV hypothesis is applicable to the polygenic diseases including MS (Bush and Moore 2012), but omits a possible effect of rare alleles (with a frequency of 0.01–0.1 %), while these alleles may contribute substantially to the disease (Visscher et al. 2012). Discovery of rare polymorphisms potentially conferring high risk in distinct individuals requires full-genome sequencing, which is still too expensive to provide in every individual.

“Missing heritability” may arise also as a result of the traditional locus-by-locus association analysis (Zuk et al. 2012). Meanwhile, epistasis between loci or pathways can take place in genetic susceptibility to MS (Lvovs et al. 2012; Cotsapas and Hafler 2013). The bioinformatics analysis currently used in GWASs does not report the risk factors that are determined by nonlinear (epistatic) interactions between alleles (including both rare and common ones) in an individual allele set (Lvovs et al. 2012), as well as interactions of genotype with nongenetic factors.

It is useful to keep in mind when interpreting GWAS results that many adaptive frequent variations, remaining in the noise of GWAS results, are in gene–gene interactions and can turn deleterious under unfavorable conditions (Sadec et al. 2014). Recently, novel approaches to GWAS analysis have been proposed. These strategies focus on the additive effects of multiple loci, acknowledging that each may make a small contribution to the overall phenotype, potentially providing valuable insights into the genetic basis of common disease (IMSGC et al. 2007; ISC et al. 2009). Rather than focusing on individual markers, network-based analysis methods take into account multiple loci in the context of molecular networks. Due to this critical feature, these methods can afford to use sub-genome-wide (nominal) statistical significance and increase the power to detect new associations and functional relationships between genes in complex traits (Wang et al. 2015). By applying network-based analysis of (Patsopoulos et al. 2011) and (IMSGC et al. 2011) results, in the study (IMSGC 2013), several plausible candidate genes were suggested. Of them, *BCL10* and *TRAF3* were associated

at the genome-wide significance level in the ImmunoChip study (IMSGC et al. 2013). Analyzing nominal gene-level significance and studying genes in the context of biological networks seem a reasonable approach for future GWAS analysis (Matsushita et al. 2015).

One more approach considers association signal from haplotypes (SNP-strings) instead that of SNPs (Khankhanian et al. 2015). In this wise, there is search for clusters of SNPs, which are jointly associated with the disease and which presumably belong to a particular disease-associated haplotype. In this study, 32 regions were shown to be more significantly associated at SNP-haplotype model than at single-SNP model in the individuals studied at (IMSGC et al. 2011, 2013). Upon such kind of analysis, the area under curve characterizing overall heritability explained by haplotypes was higher than that of distinct SNPs (Khankhanian et al. 2015). Of note, odds ratios of associations of haplotypes are considerably higher than ORs of SNPs by themselves (Goodin and Khankhanian 2014).

Another aspect of the problem became clear in 2012 after completion of the ENCODE project, which was aimed at deciphering the functional part of the genome. A report of its completion (ENCODE Project Consortium et al. 2012) stated, in particular, that 80.4 % of the genome is functional, being involved in at least one biological process in at least one cell type. This conclusion indicates that, to understand the nature of MS, it may be insufficient to analyze the associations with the allelic variants that occur in protein-coding regions, which account for approximately 1.5 % of the total genome in humans, or in their vicinity. Other genome regions are important to decipher as well. Indeed, most SNPs associated with MS according to GWASs are located in introns and intergenic regions (see Tables S1 and 2 for example). There is growing evidence of vast involvement of noncoding DNA regions via eQTLs, noncoding regulatory RNAs and other mechanisms in gene expression regulation, and some data on MS risk loci suggest abundance of eQTLs among them (Sawcer et al. 2014). Considering this, it is reasonable to expect the appearance of novel GWAS SNP panels with more in-depth coverage of RNA-coding regions. In addition, in the currently existing GWAS design, copy number variations are underexplored (Lin et al. 2012).

The above considerations give grounds to expect that next generation sequencing (NGS) techniques, based on high-throughput parallel DNA sequencing, will be employed in GWASs of the genome variants associated with particular phenotypes, the more so as the NGS techniques continuously improve and grow cheaper (Pavlopoulos et al. 2013). The use of these techniques in no way alters the basic ideology of the GWASs, but rather provides for its further development.



New targets for MS therapy can be identified in new studies. For instance, GWAS data should prompt a search for target proteins among the NF- $\kappa$ B signaling pathway components. Apart from shedding light on the molecular pathogenesis of MS, which is required for developing new MS therapies, GWASs aim at identifying, via genotyping, the genetic risk factors suitable for predicting the individual susceptibility to MS. In other words, the most important result of studying the genetic architecture of MS would be formulation of a rule to predict the disease in a particular individual with high probability by carriage of certain alleles and allelic combinations. Studies estimating the individual MS risk on the basis of GWAS data alone (Wang et al. 2011) or in combination with findings from candidate gene studies (Jafari et al. 2011; Gourraud et al. 2011) have led to some encouraging results, though not clinically applicable yet. For example, individuals carrying more MS risk variants (estimated using the weighted Genetic Risk Score) have increased odds of MS development (De Jager et al. 2009c). Patients with severe disease, higher oligoclonal bands, or earlier age at onset typically have a higher genetic load (i.e., number of genetic markers of MS, assessed using Multiple Sclerosis Genetic Burden score (Gourraud et al. 2011). Patients with clinically isolated syndrome having a higher genetic load tend to develop MS more rapidly (Gourraud et al. 2012).

The genetic MS atlas will certainly be improved in the future by adding new findings and excluding the data that lack support. The heterogeneity of the MS patient samples examined so far makes it possible to expect that several partly different atlases will be obtained for different ethnic groups and for different MS forms as data replication studies will be performed in homogenous samples.

**Acknowledgments** This work was supported by the Russian Foundation for Basic Research (projects 13-04-40281-H, 13-04-40279-H, and 15-04-04866-A).

#### Compliance with ethical standards

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV et al (2008) Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. *Nat Genet* 40:1402–1403. doi:10.1038/ng.251
- Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene), Bahlo M et al (2009) Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 41:824–828. doi:10.1038/ng.396
- Bahreini SA, Jabalameli MR, Saadatnia M, Zahednasab H (2010) The role of non-HLA single nucleotide polymorphisms in multiple sclerosis susceptibility. *J Neuroimmunol* 229:5–15
- Ban M, Stewart GJ, Bennetts BH, Heard R, Simmons R, Maranian M, Compston A, Sawcer SJ (2002) A genome screen for linkage in Australian sibling-pairs with multiple sclerosis. *Genes Immun* 3:464–469
- 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. doi:10.1093/hmg/ddn388
- Bomprezzi R, Kovanen PE, Martin R (2003) New approaches to investigating heterogeneity in complex traits. *J Med Genet* 40:553–559. doi:10.1136/jmg.40.8.553
- Bradl M, Lassmann H (2009) Progressive multiple sclerosis. *Semin Immunopathol* 31:455–465
- Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, Thompson AJ (2014) Atlas of Multiple Sclerosis 2013: a growing global problem with widespread inequity. *Neurology* 83:1022–1024
- Burdett T, Hall PN, Hasting E, Hindorf LA, Junkins HA, Klemm AK, MacArthur J, Manolio TA, Morales J, Parkinson H, Welter D (2015) TheNHGRI-EBI Catalog of published genome-wide association studies. <http://www.ebi.ac.uk/gwas>. Accessed 17 April 2015
- Bush WS, Moore JH (2012) Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 8:e1002822. doi:10.1371/journal.pcbi.1002822
- Comabella M, Craig DW, Carmiña-Tato M et al (2008) Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. *PLoS One* 3:e3490. doi:10.1371/journal.pone.0003490
- Compston DA, Batchelor JR, McDonald WI (1976) B-lymphocyte alloantigens associated with multiple sclerosis. *Lancet* 2:1261–1265. doi:10.1016/S0140-6736(76)92027-4
- Cotsapas C, Hafler DA (2013) Immune-mediated disease genetics: the shared basis of pathogenesis. *Trends Immunol* 34:22–26
- Cree BA (2014) Multiple sclerosis genetics. *Handb Clin Neurol* 122:193–209
- Cunningham C (2013) Microglia and neurodegeneration: the role of systemic inflammation. *Glia* 61:71–90. doi:10.1002/glia.22350
- De Jager PL, Jia X, Wang J et al (2009a) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 41:776–782. doi:10.1038/ng.401.Meta-analysis
- De Jager PL, Baecher-Allan C, Maier LM et al (2009b) The role of the CD58 locus in multiple sclerosis. *Proc Natl Acad Sci USA* 106:5264–5269. doi:10.1073/pnas.0813310106
- De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S et al (2009c) Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* 8:1111–1119. doi:10.1016/S1474-4422(09)70275-3
- Denic A, Wootla B, Rodriguez M (2013) CD8(+) T cells in multiple sclerosis. *Expert Opin Ther Targets* 17:1053–1066
- Dore-Duffy P, Washington R, Dragovic L (1993) Expression of endothelial cell activation antigens in microvessels from patients with multiple sclerosis. *Adv Exp Med Biol* 331:243–248
- Ebers GC, Kukay K, Bulman DE et al (1996) A full genome search in multiple sclerosis. *Nat Genet* 13:472–476. doi:10.1038/ng0896-472
- Favorov AV, Andreewski TV, Sudomoina MA, Favorova OO, Parmigiani G, Ochs MF (2005) A Markov chain Monte Carlo technique for identification of combinations of allelic variants underlying complex diseases in humans. *Genetics* 171:2113–2121. doi:10.1534/genetics.105.048090

- Favorova OO, Andreewski TV, Boiko AN et al (2002) The chemokine receptor CCR5 deletion mutation is associated with MS in HLA-DR4-positive Russians. *Neurology* 59:1652–1655. doi:10.1212/01.WNL.0000035626.92372.0A
- Favorova OO, Favorov AV, Boiko AN, Andreewski TV, Sudomoina MA, Alekseenkov AD, Kulakova OG, Gusev EI, Parmigiani G, Ochs MF (2006) Three allele combinations associated with multiple sclerosis. *BMC Med Genet* 7:63. doi:10.1186/1471-2350-7-63
- GAMES, Transatlantic Multiple Sclerosis Genetics Cooperative (2003) A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* 143:39–46
- Goodin DS, Khankhanian P (2014) Single nucleotide polymorphism (SNP)-strings: an alternative method for assessing genetic associations. *PLoS One* 9:e90034
- Gourraud PA, International Multiple Sclerosis Genetics Consortium (IMSGC) (2011) When is the absence of evidence, evidence of absence? Use of equivalence-based analyses in genetic epidemiology and a conclusion for the KIF1B rs10492972\*C allelic association in multiple sclerosis. *Genet Epidemiol* 35:568–571. doi:10.1002/gepi.20592
- Gourraud PA, McElroy JP, Caillier SJ, Johnson BA, Santaniello A, Hauser SL, Oksenberg JR (2011) Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. *Ann Neurol* 69:65–74
- Gourraud PA, Harbo HF, Hauser SL, Baranzini SE (2012) The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev* 248:87–103. doi:10.1111/j.1600-065X.2012.01134.x
- Gourraud PA, Sdika M, Khankhanian P, Henry RG, Beheshtian A, Matthews PM, Hauser SL, Oksenberg JR, Pelletier D, Baranzini SE (2013) A genome-wide association study of brain lesion distribution in multiple sclerosis. *Brain* 136:1012–1024
- Goverman J (2009) Autoimmune T cell responses in the central nervous system. *Nat Rev Immunol* 9:393–407. doi:10.1038/nri2550
- 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. doi:10.1038/ng2103
- Gregory AP, Dendrou CA, Attfield KE et al (2012) TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 488:508–511. doi:10.1038/nature11307
- Guerini FR, Ferrante P, Losciale L, Caputo D, Lombardi ML, Pirozzi G, Luongo V, Sudomoina MA, Andreewski TV, Alekseenkov AD, Boiko AN, Gusev EI, Favorova OO (2003) Myelin basic protein gene is associated with MS in DR4- and DR5-positive Italians and Russians. *Neurology* 61:520–526
- Hedegaard CJ, Krakauer M, Bendtzen K et al (2008) T helper cell type 1 (Th1), Th2 and Th17 responses to myelin basic protein and disease activity in multiple sclerosis. *Immunology* 125:161–169. doi:10.1111/j.1365-2567.2008.02837.x
- Hermanowski J, Bouzigon E, Forabosco P, Ng MY, Fisher SA, Lewis CM (2007) Meta-analysis of genome-wide linkage studies for multiple sclerosis, using an extended GSMA method. *Eur J Hum Genet* 15:703–710. doi:10.1038/sj.ejhg.5201818
- Hindorf LA, Sethupathy P, Junkins HA et al (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106:9362–9367. doi:10.1073/pnas.0903103106
- Hohlfeld R (2008) Neurotrophic cross-talk between the nervous and immune systems: relevance for repair strategies in multiple sclerosis? *J Neurol Sci* 265:93–96. doi:10.1016/j.jns.2007.03.012
- Holman DW, Klein RS, Ransohoff RM (2011) The blood–brain barrier, chemokines and multiple sclerosis. *Biochim Biophys Acta Mol Basis Dis* 1812:220–230  
<http://www.genecards.org>. Accessed 20 April 2015  
<http://www.ncbi.nlm.nih.gov/gene>. Accessed 20 April 2015
- <http://amigo.geneontology.org/amigo/landing>. Accessed 25 April 2015
- <http://revigo.irb.hr/>. Accessed 25 April 2015
- International Multiple Sclerosis Genetics Consortium (2013) Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am J Hum Genet* 92:854–865. doi:10.1016/j.ajhg.2013.04.019
- International Multiple Sclerosis Genetics Consortium (IMSGC), Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA, Haines JL (2010) Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 86:621–625. doi:10.1016/j.ajhg.2010.02.027
- International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, Xifara DK et al (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 45:1353–1360. doi:10.1038/ng.2770
- International Multiple Sclerosis Genetics Consortium, David A, Hafler MD, Compston A et al (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357:2373–2383. doi:10.1056/NEJMoa1407764
- International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium, Sawcer S et al (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476:214–219. doi:10.1038/nature10251
- International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL et al (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752. doi:10.1038/nature08185
- Isobe N, Madireddy L, Khankhanian P, Matsushita T, Caillier SJ, Moré JM, Gourraud PA, McCauley JL, Beecham AH, International Multiple Sclerosis Genetics Consortium, Piccio L, Herbert J, Khan O, Cohen J, Stone L, Santaniello A, Cree BA, Onengut-Gumuscu S, Rich SS, Hauser SL, Sawcer S, Oksenberg JR (2015) An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain* 138:1518–1530
- Iwanowski P, Losy J (2015) Immunological differences between classical phenotypes of multiple sclerosis. *J Neurol Sci* 349:10–14
- Jafari N, Broer L, van Duijn CM, Janssens AC, Hintzen RQ (2011) Perspectives on the use of multiple sclerosis risk genes for prediction. *PLoS One* 6:e26493. doi:10.1371/journal.pone.0026493
- 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. doi:10.1016/j.ajhg.2010.01.017
- Kaufman DW, Reshef S, Golub HL, Peucker M, Corwin MJ, Goodin DS, Knappertz V, Pleimes D, Cutter G (2014) Survival in commercially insured multiple sclerosis patients and comparator subjects in the US. *Mult Scler Relat Disord* 3:364–371. doi:10.1016/j.msard.2013.12.003
- Khankhanian P, Gourraud PA, Lizee A, Goodin DS (2015) Haplotype-based approach to known MS-associated regions increases the amount of explained risk. *J Med Genet* 52:587–594
- Kilpinen H, Barrett JC (2013) How next-generation sequencing is transforming complex disease genetics. *Trends Genet* 29:23–30
- Kleinewietfeld M, Hafler DA (2014) Regulatory T cells in autoimmune neuroinflammation. *Immunol Rev* 259:231–244. doi:10.1111/imr.12169
- Kuhle J, Disanto G, Dobson R, Adiatori R et al (2015) Conversion from clinically isolated syndrome to multiple sclerosis: a large multicentre study. *Mult Scler* 21:1013–1024. doi:10.1177/1352458514568827

- Lambert CA, Tishkoff SA (2009) Genetic structure in African populations: implications for human demographic history. *Cold Spring Harb Symp Quant Biol* 74:395–402
- Lassmann H, Brück W, Lucchinetti CF (2007) The immunopathology of multiple sclerosis: an overview. *Brain Pathol* 17:210–218
- Leray E, Vukusic S, Debouverie M, Clanet M et al (2015) Excess mortality in patients with multiple sclerosis starts at 20 years from clinical onset: data from a large-scale French observational study. *PLoS One* 10:e0132033. doi:10.1371/journal.pone.0132033
- Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. *Annu Rev Genomics Hum Genet* 10:387–406. doi:10.1146/annurev.genom.9.081307.164242
- Lill CM (2014) Recent advances and future challenges in the genetics of multiple sclerosis. *Front Neurol* 5:130
- Lin R, Charlesworth J, van der Mei I, Taylor BV (2012) The genetics of multiple sclerosis. *Pract Neurol* 12:279–288
- Lovett-Racke AE, Yang Y, Racke MK (2011) Th1 versus Th17: are T cell cytokines relevant in multiple sclerosis? *Biochim Biophys Acta Mol Basis Dis* 1812:246–251. doi:10.1016/j.bbdis.2010.05.012
- Lu YF, Goldstein DB, Angrist M, Cavalleri G (2014) Personalized medicine and human genetic diversity. *Cold Spring Harb Perspect Med* 4:a008581
- Lundmark F, Duvefelt K, Iacobaeus E et al (2007) Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat Genet* 39:1108–1113. doi:10.1038/ng2106
- Lvovs D, Favorova OO, Favorov AV (2012) A polygenic approach to the study of polygenic diseases. *Acta Naturae* 4:59–71
- Maier LM, Anderson DE, Severson CA, Baecher-Allan C, Healy B, Liu DV, Wittrup KD, De Jager PL, Hafler DA (2009) Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *J Immunol* 182:1541–1547
- Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753. doi:10.1038/nature08494
- Marian AJ (2012) Molecular genetic studies of complex phenotypes. *Transl Res* 159:64–79. doi:10.1016/j.trsl.2011.08.001
- Marigorta UM, Lao O, Casals F et al (2011) Recent human evolution has shaped geographical differences in susceptibility to disease. *BMC Genom* 12:55. doi:10.1186/1471-2164-12-55
- Marrie RA, Elliott L, Marriott J, Cossoy M, Blanchard J, Leung S, Yu N (2015) Effect of comorbidity on mortality in multiple sclerosis. *Neurology* 85:240–247. doi:10.1212/WNL.0000000000001718
- Martinelli-Boneschi F, Esposito F, Brambilla P et al (2012) A genome-wide association study in progressive multiple sclerosis. *Mult Scler* 18:1384–1394. doi:10.1177/1352458512439118
- Matesanz F, González-Pérez A, Lucas M, et al. (2012) Genome-wide association study of multiple sclerosis confirms a novel locus at 5p13.1. *PLoS One* 7:e36140. doi:10.1371/journal.pone.0036140
- Matsushita T, Madireddy L, Sprenger T, Khankhanian P, Magon S, Naegelin Y, Caverzasi E, Lindberg RL, Kappos L, Hauser SL, Oksenberg JR, Henry R, Pelletier D, Baranzini SE (2015) Genetic associations with brain cortical thickness in multiple sclerosis. *Genes Brain Behav* 14:217–227
- Naito S, Namerow N, Mickey MR, Terasaki PI (1972) Multiple sclerosis: association with HL-A3. *Tissue Antigens* 2:1–4
- 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. doi:10.1016/j.jneuroim.2010.06.003
- Nylander A, Hafler DA (2012) Multiple sclerosis. *J Clin Invest* 122:1180–1188
- Oksenberg JR (2013) Decoding multiple sclerosis: an update on genomics and future directions. *Expert Rev Neurother* 13:11–19. doi:10.1586/14737175.2013.865867
- Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL (2008) The genetics of multiple sclerosis: sNPs to pathways to pathogenesis. *Nat Rev Genet* 9:516–526
- Ortiz GG, Pacheco-Moisés FP, Macías-Islas MÁ, Flores-Alvarado LJ, Mireles-Ramírez MA, González-Renovato ED, Hernández-Navarro VE, Sánchez-López AL, Alatorre-Jiménez MA (2014) Role of the blood–brain barrier in multiple sclerosis. *Arch Med Res* 45:687–697
- Parkes M, Cortes A, van Heel DA, Brown MA (2013) Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet* 14:661–673. doi:10.1038/nrg3502
- Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFN[ $\beta$ ]-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. doi:10.1002/ana.22609
- Pavlopoulos GA, Oulas A, Iacucci E et al (2013) Unraveling genomic variation from next generation sequencing data. *BioData Min* 6:13. doi:10.1186/1756-0381-6-13
- Pèer I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 32:381–385. doi:10.1002/gepi.20303
- Ramagopalan SV, Ebers GC (2009) Multiple sclerosis: major histocompatibility complexity and antigen presentation. *Genome Med* 1:105. doi:10.1186/gm105
- Rasmussen HB, Kelly MA, Clausen J (2001) Genetic susceptibility to multiple sclerosis: detection of polymorphic nucleotides and an intron in the 3' untranslated region of the major histocompatibility complex class II transactivator gene. *Hum Immunol* 62:371–377. doi:10.1016/S0198-8859(01)00215-4
- Sadee W, Hartmann K, Seweryn M, Pietrzak M, Handelman SK, Rempala GA (2014) Missing heritability of common diseases and treatments outside the protein-coding exome. *Hum Genet* 133:1199–1215
- Sanna S, Pitzalis M, Zoledziwska M et al (2010) Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet* 42:495–497. doi:10.1038/ng.584
- 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. doi:10.1038/ng0896-464
- 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. doi:10.1086/444547
- Sawcer S, Franklin RJ, Ban M (2014) Multiple sclerosis genetics. *Lancet Neurol* 13:700–709
- Spencer CC, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* 5:e1000477. doi:10.1371/journal.pgen.1000477
- Stys PK (2005) General mechanisms of axonal damage and its prevention. *J Neurol Sci* 233:3–13
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi:10.1371/journal.pone.0021800
- Tauber SC, Nau R, Gerber J (2007) Systemic infections in multiple sclerosis and experimental autoimmune encephalomyelitis. *Arch Physiol Biochem* 113:124–130
- The ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M (2012) An integrated

- encyclopedia of DNA elements in the human genome. *Nature* 489:57–74. doi:[10.1038/nature11247](https://doi.org/10.1038/nature11247)
- Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90:7–24
- von Büdingen H-C, Bar-Or A, Zamvil SS (2011) B cells in multiple sclerosis: connecting the dots. *Curr Opin Immunol* 23:713–720
- Wang JH, Pappas D, De Jager PL et al (2011) Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data. *Genome Med* 3:3. doi:[10.1186/gm217](https://doi.org/10.1186/gm217)
- Wang L, Mousavi P, Baranzini SE (2015) iPINBPA: an integrative network-based functional module discovery tool for genome-wide association studies. *Pac Symp Biocomput* 2015:255–266
- Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR et al (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 39:1329–1337. doi:[10.1038/ng.2007.17](https://doi.org/10.1038/ng.2007.17)
- Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 109:1193–1198. doi:[10.1073/pnas.1119675109](https://doi.org/10.1073/pnas.1119675109)