Title
Genetic burden in multiple sclerosis families.

Journal
Genes and Immunity, 14(7)

Authors
Isobe, N
Damotte, V
Re, V
et al.

Publication Date
2013-10-01

DOI
10.1038/gene.2013.37

Peer reviewed

Published in final edited form as:
Genes Immun. 2013 October ; 14(7): 434–440. doi:10.1038/gene.2013.37.

Genetic burden in multiple sclerosis families

N Isobe[#1], **V Damotte**[#1,2], **V Lo Re**[#3], **M Ban**[3], **D Pappas**[1], **L Guillot-Noel**[2], **I Rebeix**[2,4], **A Compston**[3], **T Mack**[5], **W Cozen**[5], **B Fontaine**[2,4], **SL Hauser**[1], **JR Oksenberg**[1], **S Sawcer**[3], and **P-A Gourraud**[1]

[1]Department of Neurology, University of California at San Francisco, San Francisco, CA, USA

[2]UPMC-INSERM-CNRS-ICM, UMR 975-7225, Institut du Cerveau et de la Moelle épinière (ICM), Hôpital Pitié-Salpêtrière, Paris, France

[3]Department of Clinical Neurosciences, Addenbrooke's, Hospital, University of Cambridge, Cambridge, UK

[4]Assistance Publique-Hôpitaux de Paris (AP-HP), Département de Neurologie, Hôpital Pitié-Salpêtrière, Paris, France

[5]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[#] These authors contributed equally to this work.

Abstract

A previous study using cumulative genetic risk estimations in multiple sclerosis (MS) successfully tracked the aggregation of susceptibility variants in multi-case and single-case families. It used a limited description of susceptibility loci available at the time (17 loci). Even though the full roster of MS risk genes remains unavailable, we estimated the genetic burden in MS families and assess its disease predictive power using up to 64 single-nucleotide polymorphism (SNP) markers according to the most recent literature. A total of 708 controls, 3251 MS patients and their relatives, as well as 117 twin pairs were genotyped. We validated the increased aggregation of genetic burden in multi-case compared with single-case families (P = 4.14e – 03) and confirm that these data offer little opportunity to accurately predict MS, even within sibships (area under receiver operating characteristic (AUROC) = 0.59 (0.55, 0.53)). Our results also suggest that the primary progressive and relapsing-type forms of MS share a common genetic architecture (P = 0.368; difference being limited to that corresponding to ±2 typical MS-associated SNPs). We have confirmed the properties of individual genetic risk score in MS. Comparing with previous reference point for MS genetics (17 SNPs), we underlined the corrective consequences of the integration of the new findings from GWAS and meta-analysis.

Correspondence: Dr P-A Gourraud, Department of Neurology, University of California at San Francisco, 675 Nelson Rising lane, Room NS235, San Francisco, CA 94158, USA. Pierreantoine.Gourraud@ucsf.edu.

CONFLICT OF INTEREST The authors declare no conflict of interest.

Supplementary Information accompanies this paper on Genes and Immunity website (http://www.nature.com/gene)

Keywords

multiple sclerosis; family study; genetic risk

Extensive epidemiological and laboratory data confirm that genetic variation is an important contributor to risk in multiple sclerosis (MS), a chronic and severe neuro-inflammatory disease that is seen most commonly in Whites of northern European descent.[1] As with many other multifactorial diseases, genome-wide association studies (GWAS) have been highly successful in identifying genomic regions associated with susceptibility. To date, multiple MS GWAS have been completed, and between them have identified nearly 60 susceptibility genomic regions. Given the limited power of this approach to identify rare variants, it is unsurprising that the majority of these associations relates to variants that are common in the general population.[2–13] The significant enrichment for genes with annotated immunological functions is supported by multiple statistical measures[4,14] with only a minority of the candidate genes implicated by these associations having bona fide neuronal functions independent of inflammation. The strongest signal consistently maps to the *HLA-DRB1* gene in the class II region of the major histocompatibility complex (MHC, 6p21.3). Extensive overlap with susceptibility markers for other autoimmune diseases has been well established,[15,16] and follow-up experiments provided important mechanistic insights linking associated variants in *IL7R*, *IL2RA*, *CD58*, *TNFRSF1A* and *TYK2*(refs 17–21) to their functional consequences. These results significantly broaden our understanding of etiology but the full potential of these data to model an individual's lifetime disease risk remains unknown.

A previous attempt to use cumulative genetic risk estimations in MS suggested a higher aggregation of susceptibility variants in multi-case compared with single-case MS families, but resulted in incomplete affectation status classification accuracy.[22–24] A number of factors could have contributed to the low sensitivity of the genomic burden metrics, including the limited description of susceptibility loci available at the time (17 loci). Even though the full roster of MS risk genes remains unavailable, we build on the most recent data sets[4,25] to estimate the genetic burden in MS families and assess its disease predictive power. We validate the increased aggregation of genetic burden in multi-case compared with single-case families and confirm that these data offer little opportunity to accurately predict MS, even within sibships. We also provide additional genetic evidence for the absence of any meaningful difference in the genetic burden between the primary progressive and relapsing-type forms of MS.

RESULTS

Distribution of MS genetic burden (MSGB) scores in multi-case and single-case MS families

The genetic risk captured by the MSGB score was higher in the probands and parents from the multi-case families compared with those with unaffected first-degree relatives (*P*-values for probands = 4.14e – 03, *P* for mothers = 1.16e – 02 and *P* for fathers = 3.13e – 02) (Figure 1 and Table 1). These results are consistent with those we reported previously using

just 17 markers[23] and in line with the increased genetic information conferred by the inclusion of additional MS-associated single-nucleotide polymorphisms (SNPs), which affords more statistical significance when compared with controls. Using the maximum number of MS risk variants genotyped in the University of California San Francisco (UCSF) samples, we did not observe any indication of balancing aggregation (Supplementary Figure 1), meaning that neither the presence of the homozygous major risk factor (*HLA-DRB1*15:01*) nor female status results in any trend toward lower accumulation of the other non-MHC genetic risk variants. However, power computations suggest that we would need ~2.5 times as many samples to reach 80% power for detecting an one-sided statistically significant difference of half the one observed between the group of *HLA-DRB1*15:01* homozygous female and the group of *HLA-DRB1*15:01*-negative males. So this negative observation is not unexpected.

To further assess the effect of the additional information content on the individual risk, we plotted $MSGB_{17SNPs}$ (previous reference point for MS genetics) vs the new part of $MSGB_{64SNPs}$ (most up-to-date MS Genetics $MSGB_{64SNPs}$ minus previously used $MSGB_{17SNPs}$) using UCSF samples typed for SNPs required for preparation of both MSGB scores (Figure 2). Surprisingly, this analysis yielded a statistically significant negative correlation in both, cases and controls (rho = −0.16 in cases P = 2.82e – 10 and rho = −0.31 in controls P = 9.81e – 17). It shows that our knowledge of MS genetic not only grows in number of regions identified but also refines the association signal in previously identified regions. Several factors are likely to contribute to this correlation. First, the lack of confirmation for *CD226* and *GPC5* in the recent GWAS[4,25] suggest that the previous analyses may have been confounded by the inclusion of unassociated variants. In line with the winners curse, the effect sizes were reduced for 13 SNPs but increased for only 2 (*HLA-DRB1* and *EVI5*). In two regions (*IL2RA* and the MHC), we included some degree of genetic heterogeneity with multiple SNPs capturing second signals. Furthermore, updating the MSGB model with the latest results obtained with independent samples from these family collections not only resulted in the addition of lower effect-size common genetic variants but also tended to replace the low-frequency variants with more common ones at the previously identified genomic regions because the SNP with the lowest *P*-value was chosen. For example, in the *TYK2* locus, the MS risk SNPs rs34536443 (NP_003322.3:p.Pro1104Ala – MAF = 0.0285) has been replaced by rs8112449 (NT_011295.11:g.1782866G>A – MAF = 0.305 in dbSNP). Several changes resulting from the redefinition of SNPs considered to represent the most up-to-date definition of the MSGB probably explain the negative correlation in both cases and controls. Supplementary Figure 2 presents the MSGB score distribution in both cases and controls when increasing the number of SNPs. While the difference between cases and controls increases, there is also a noteworthy increase in the variance of the scores in both cases and controls (0.53 vs 0.74, P = 2.09e – 04 in cases, 0.37 vs 0.55, P = 2.19e – 05 in controls; Supplementary Table 3). $MSGB_{64SNPs}$ shows larger variance than $MSGB_{17SNPs}$ in both cases and controls and the skewness of the distribution is reduced in $MSGB_{64SNPs}$ (from 0.40 to 0.10 in cases, and from 0.73 to 0.25 in controls) showing that the combination of 62 non-MHC SNPs with high frequencies and modest effect sizes in the scores smoothen the distorted distribution observed in $MSGB_{17SNPs}$.

Distribution of MSGB scores in twin and sib pairs

We explored the properties of the MSGB statistics (including HLA and gender) in three additional familial data sets, including 117 twin pairs (Supplementary Table 4, Figure 3). In this analysis, data from 57 markers were available (Supplementary Table 1). As expected, MS twins have higher MSGB compared with controls (P = 1.14e – 09 vs MZ twins pairs and P = 1.92e – 06 for DZ pairs). Given the limited power of this data set, it is unsurprising that we saw no statistically significant difference in MSGB score when comparing concordant with discordant monozygotic twins (P = 0.81, estimated power = 30%), nor between discordant dizygotic twins (P = 0.24, estimated power = 61%).

In siblings of the same sibship, having a greater or equal $MSGB_{56\ SNPS}$ than the proband is significantly associated with MS with an odds ratio (OR) = 2.10 (1.4, 3.1) (conditional logistic regression P = 1e – 04, 348 informative pedigrees, 804 individuals) (Figure 4a). However, this statistically significant association had very little predictive value in the sibship. Figure 4b displays the receiver operating characteristic (ROC) curves corresponding to the MSGB2 sib-proband contrast and confirms that contrasting MSGB in MS sibships is statistically significant but not suitable for prediction: The AUROC (area under ROC) areas is 0.58 (0.55, 0.62), whereas the contrasting MSGB in sibling results only in AUROC areas between 0.5 and 0.6. (AUROC 0.59 (0.55, 0.53) for $MSGB_{56SNPs}$ including HLA and gender, and 0.58 (0.54, 0.62) for non-MHC SNPs only). In siblings, the full MSGB risk score does not appear to do better than *HLA-DRB1*15:01*-tagging SNP and gender alone (P = 0.97 Hanley test (not shown)). In the general population, on the other hand, the full $MSGB_{56SNPs}$ does better than HLA and gender only (AUROC = 0.75 (0.73, 0.77) vs 0.72 (0.69, 0.73) P = 1e – 4)). However, the scores fail to reach the high specificity values that would be required for the prediction of a relatively rare condition such as MS in the general population.

Interestingly, using this updated version of genetic risk score, we confirmed the lack of differences in MSGB between progressive (PP (primary progressive) + PR (progressive relapsing), n = 182) and relapsing-type (CIS (clinically isolated syndrome) + RR (relapsing remitting) + SP (secondary progressive), n = 1914) MS (Figure 5) (P = 0.368). However, classical statistical methods are primarily developed to establish the statistical significance of a difference, the non-significant P-value of P = 0.368 do not provide correct statistical evidence to claim for similarity of MSGB scores. We therefore used equivalence testing and found that primary progressive MS and relapsing-type MS have similar scores. This difference being limited to that corresponding to $\pm_2$ typical MS-associated SNPs (equivalence test P = 5.1e – 03). It means that if different, the MSGB scores in PP MS and RR MS differs of less than the contribution of two risk alleles. Taken together, these results confirm that primary progressive and relapsing-type MS share the same underlying genetic architecture.

DISCUSSION

MS is a prototypic multifactorial disease in which susceptibility is determined by polygenic inheritance. Over the last 5 years, large and multi-center DNA collections have been successfully established and the application of the hypothesis-free GWAS approach has

resulted in remarkable progress in identifying the non-*HLA* genetic components of MS genetic risk. The most recent multi-center collaborative GWAS involving 9772 cases of European descent collected by 23 research groups working in 15 different countries replicated nearly all of the previously GWAS-suggested associations together with the identification of 29 novel susceptibility loci.[4] Understanding how these associated alleles exert their effects on risk constitutes a priority in MS genetics. We used the most updated genetic data to build a genetic profile associated with the cumulative genetic risk measured by the probability of an individual having MS.

We have confirmed previously obtained results (17 MS-associated SNPs) with the updated version of the MSGB scores using up to 64 MS markers. Our observations are as follows: (1) individuals from multi-case families have a greater MS genetic load than members of single-case families; (2) association of MSGB with disease co-occurrence in sibships is highly significant and yet, of limited power in terms of disease-prediction; (3) we confirmed the similarity of genetic load in primary progressive MS and relapsing-type MS. While the significance of any comparison between MS family members and healthy controls increases with the addition of the recently identified SNPs in the MSGB score, the current results further highlight the limited discriminative power of common alleles when patients and their relatives are being compared. Whereas a greater MSGB in siblings of MS patients was associated with an increased risk of MS, the ROC curves of MSGB differences between probands and sibs show that case–control status prediction using 56 rather than 17 markers did not really increase the predictive value of genetic risk score. The predictive power of genetic burden scores in the population is challenged by the false positive signal generated by common alleles, which by definition, are neither specific of MS patients nor they are for MS families. Interestingly, as suggested by the striking similarity of the distribution of MSGB score when comparing the relapse type of MS (CIS, RR, SP) to the progressive forms of MS (PR, PP), the normality test significance in patients with $MSGB_{64SNPs}$ do not suggest that genetic risk score would support the existence of etiologically heterogeneity in MS. Similarly, the analysis of the residuals in the negative correlation reveals no evidence of new subgroups patients that could account specifically for the newest genetic risk discovered. In addition, no multi-modal distribution of the score supports a high predictive power of genetic risk score for a small proportion of strongly genetically predisposed patients.

Modeling of the potential multiple non-independent contributions of variants within several regions will certainly require more elaborated genetic risk score computations than the log-additive model presented in this study. This particularly applies to current knowledge in the MHC region, underlying the need of full HLA characterization. Our current findings are based on a relatively simplistic modeling of the MHC contribution to MS risk. Much information may be lost by using just two SNPs partially tagging for *HLA-DRB1*15:01* and *HLA-B*44*. It underlines that the log-additive model inherited from GWAS has limited power to describe complex genotype risk hierarchies, including dominant and recessive models of association. Finally, allelic heterogeneity is an established feature of polygenic inheritance[26] and it seems likely that many of the MS-associated loci will ultimately prove to contain additional risk allele associations.

The information of the 56-marker model is barely any greater than that explained by the 17 SNP model. Unremarkably, most of the classification accuracy comes from the variables with larger effects included in the model (gender and *HLA-DRB1*15:01*). These variables with large effects account for most of the information, and the number of associated SNPs that would need to be added to significantly increase the information captured and thereby improve the classification accuracy is substantial. Unfortunately, while adding variants to the model increases the mean difference in the MSGB score between cases and controls, it also increases the variance in this score, that is, the extent to which the distribution of the scores may overlap. Given the absence of significant disease linkage signals outside of the MHC, it is unlikely that large-effect rare variants (OR>2) remains undetected after high-density GWAS.

With several large-scale replication and genome-wide scans currently going on, one must expect the list of MS-associated genomic regions to continue its exponential growth. Our results on the correction and enhancement of the $MSGB_{17SNPs}$ (2010–2011) to account for novel discoveries underline the risk for excess fitting of the data. In this study, we used historic familial data sets that were used in the first generation of GWAS, which could suggest that previously reported MSGB associations were slightly inflated. The accumulation of new variants will slowly increase the power of genetic risk score for the genetic study beyond susceptibility and will most likely require large multi-center efforts. However, the need for summarizing the wealth of information represented by hundreds of disease-associated regions reinforce the need to further develop genetic risk score models with parameters emerging from the most replicated results of the literature. These will be of great importance when studying quantitative phenotypic traits such as age of onset, progression or response to treatment.

MATERIALS AND METHODS

Subjects

Subject recruitment occurred at four centers, two in the United States (University of California San Francisco and University of South Carolina (USC)), one in France (French Network for MS Genetics) and one in the United Kingdom (University of Cambridge) (Table 2). The study was approved by the corresponding Institutional Review Board or ethics committee, and written informed consent was obtained from all study participants.

DNA samples from 3251 MS patients and some of their relatives were available in each data sets. A group of 708 controls DNA (including 253 familial controls and spouses of probands) from UCSF were also included using stringent inclusion and exclusion criteria as previously described.[23] We first used the full UCSF cohort familial data set, consisting of 422 multi-case families in which at least one first-degree relative of the affected proband also had clinically definite MS, and 807 ‘single-case’ families in which the affected individual reported no known history of MS in any family member. The study also included 551 samples from a French familial data set (134 cases with reported first-degree relative with MS and 417 without), and 1042 samples from University of Cambridge (253 with reported family history of MS and 789 without). Families or cases with ambiguous records of co-occurrence were omitted from the study. Diagnostic criteria and ascertainment

protocols of patients were similar for all data sets and are summarized elsewhere.[27,28] In addition to sib pairs from the familial data sets, a collection of 117 twin pairs from USC and UCSF were also studied (24 concordant monozygotic twins, 47 discordant monozygotic twins, 10 concordant dizygotic twins, 36 discordant dizygotic twins; Supplementary Table 4).

SNP genotyping

Sixty four MS-SNPs were genotyped using either individual TaqMan assays or the TaqMan OpenArray genotyping technology (Life Technologies, Carlsbad, CA, USA). MS-SNPs were selected from the latest GWAS[4] and meta-analysis.[25] Samples were loaded into customized TaqMan OpenArray genotyping plates with the OpenArray Autoloader and amplified in a Dual Flat Block GeneAmp PCR System 9700, as recommended by the manufacturer. The OpenArray Genotyping Analysis Software and the Taqman Genotyper Software were used for assigning genotypes.

Quality control

SNPs had to meet several criteria to be included in the analysis: <10% missing genotypes, Hardy–Weinberg proportion test $P>0.001$ in controls, and no significant SNP call rate differences between patients and controls ($P>0.001$) (Table 2).

MSGB statistics

The MSGB score was computed based on a weighted scoring algorithm using independent 64 MS-SNPs, typically one in each genomic region of interest. This method extends the log-additive models used in previous analyses[23] with weights given to each SNP based on its effect size as reported by odds ratios in the largest GWAS and meta-analysis[3,4,25] (Supplementary Table 1 for markers and equation). Briefly, the computation of the MSGB cumulative genetic risk scores follows a log-additive model. It corresponds to the trend test typically used in the literature to identify SNP associated with MS susceptibility. In the weighted genetic risk score, the log of odds ratio associated to the presence of a single dose of risk allele is added to the score of each risk allele carried by the subject. For example, on the presence of one allele of HLA-DRB1*15:01 would add 1.089 (log(OR = 2.97)) to the MSGB score; or, for a typical MS-associated SNP such as the IL7R SNP rs6897932 in IL7R, one risk allele would add 0.104 (log(1.11)). Where possible, missing values are substituted by proxy SNPs that tag the primarily identified risk SNPs (Supplementary Table 2, $R^2>0.8$ in samples of European ancestry from the 1000 genome project[29]) or if not available, compensated with average frequency of risk allele dose in controls (Supplementary Table 1). When SNPs were not included in one of the centers, the risk SNP was excluded; therefore MSGB score do not account for this risk component in the analysis and the exact number of SNPs used for computation is indicated in subscript (for example, $MSGB_{64SNPs}$, $MSGB_{57SNPs}$ and $MSGB_{56SNPs}$; details are in Supplementary Table 1). The minimum number of SNPs genotyped in common is 56. Mean and s.d. of the score are given in Table 2. To facilitate comparison with previous model, gender was optionally included in the score and assigned an odds ratio (OR) of 1.6 as a lower bound of sex ratio observed in epidemiological longitudinal studies.[23] To identify the specific effects on familial

aggregation of the MHC component of MS risk, we also computed a non-MHC version of the score without contribution from *HLA-DRB1*15:01* and *HLA-B*44* (see primary and associated proxy SNPs in Supplementary Tables 1 and 2). Statistical test used are indicated in the legends of the tables and the figures. ROC curves were computed to assess the predictive power of MSGB scores in identifying affectation status in patients and controls, which represent the various false positive and false negative rate of quantitative metric for all possible cutoff thresholds. s.e. were computed using the Hanley method. When power computation are presented, expected difference between the group of patients of the comparison is set to half of the gap observed between all the cases and controls.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the National Institutes of Health (grant RO1NS26799 RO1NS19142), the National Multiple Sclerosis Society (grant RG2901) and the Cambridge NIHR Biomedical Research Centre, the Institut National de la Santé et de la Recherche Médicale (INSERM), the Fondation d'Aide pour la Recherche sur la Sclérose En Plaques (ARSEP), the Association Française contre les Myopathies (AFM) and GISIBISA. The research leading to these results has received funding from the program 'Investissements d'avenir' ANR-10-IAIHU-06. We acknowledge use of the cohort of the REFGENSEP and thank ICM, CIC Pitié-Salpêtrière, Généthon and REFGENSEP's members for their help and support. NI is a postdoctoral fellow supported by Japan Society for Promotion of Science. VD received a travel grant from ARSEP. P-AG is a recipient of the Nancy Davis Foundation young investigator award. We acknowledge the assistance of Stacy Caillier, Adam Santaniello, Vinod Bakthavachalam Jinyang Wang and Matthew Chan in sample and data management.

REFERENCES

1. Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. Nat Rev Genet. 2008; 9:516–526. [PubMed: 18542080]
2. Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA, et al. Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. Am J Hum Genet. 2010; 86:621–625. [PubMed: 20362272]
3. Gourraud PA, Harbo HF, Hauser SL, Baranzini SE. The genetics of multiple sclerosis: an up-to-date review. Immunol Rev. 2012; 248:87–103. [PubMed: 22725956]
4. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–219. [PubMed: 21833088]
5. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet. 2007; 39:1329–1337. [PubMed: 17952073]
6. 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; 357:851–862. [PubMed: 17660530]
7. Comabella M, Craig DW, Camina-Tato M, Morcillo C, Lopez C, Navarro A, et al. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. PLoS One. 2008; 3:e3490. [PubMed: 18941528]
8. ANZGENE-consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009; 41:824–828. [PubMed: 19525955]
9. 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; 18:767–778. [PubMed: 19010793]

10. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009; 41:776–782. [PubMed: 19525953]
11. Jakkula E, Leppa 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; 86:285–291. [PubMed: 20159113]
12. Nischwitz S, Cepok S, Kroner A, Wolf C, Knop M, Muller-Sarnowski F, et al. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. J Neuroimmunol. 2010; 227:162–166. [PubMed: 20598377]
13. 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; 42:495–497. [PubMed: 20453840]
14. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet. 2009; 18:2078–2090. [PubMed: 19286671]
15. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011; 7:e1002254. [PubMed: 21852963]
16. Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T, et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proc Natl Acad Sci USA. 2009; 106:18680–18685. [PubMed: 19846760]
17. Couturier N, Bucciarelli F, Nurtdinov RN, Debouverie M, Lebrun-Frenay C, Defer G, et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. Brain. 2011; 134(Pt 3):693–703. [PubMed: 21354972]
18. 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 USA. 2009; 106:5264–5269. [PubMed: 19237575]
19. Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, Butter F, et al. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. Nature. 2012; 488:508–511. [PubMed: 22801493]
20. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet. 2007; 39:1083–1091. [PubMed: 17660817]
21. 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:e1000322. [PubMed: 19119414]
22. De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009; 8:1111–1119. [PubMed: 19879194]
23. Gourraud PA, McElroy JP, Caillier SJ, Johnson BA, Santaniello A, Hauser SL, et al. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. Ann Neurol. 2011; 69:65–74. [PubMed: 21280076]
24. Sawcer S, Ban M, Wason J, Dudbridge F. What role for genetics in the prediction of multiple sclerosis? Ann Neurol. 2010; 67:3–10. [PubMed: 20186855]
25. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol. 2011; 70:897–912. [PubMed: 22190364]
26. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. [PubMed: 20881960]
27. Barcellos LF, Oksenberg JR, Green AJ, Bucher P, Rimmler JB, Schmidt S, et al. Genetic basis for clinical expression in multiple sclerosis. Brain. 2002; 125(Pt 1):150–158. [PubMed: 11834600]
28. Goodkin DE, Doolittle TH, Hauser SS, Ransohoff RM, Roses AD, Rudick RA. Diagnostic criteria for multiple sclerosis research involving multiply affected families. Arch Neurol. 1991; 48:805–807. [PubMed: 1898254]

29. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24:2938–2939. [PubMed: 18974171]

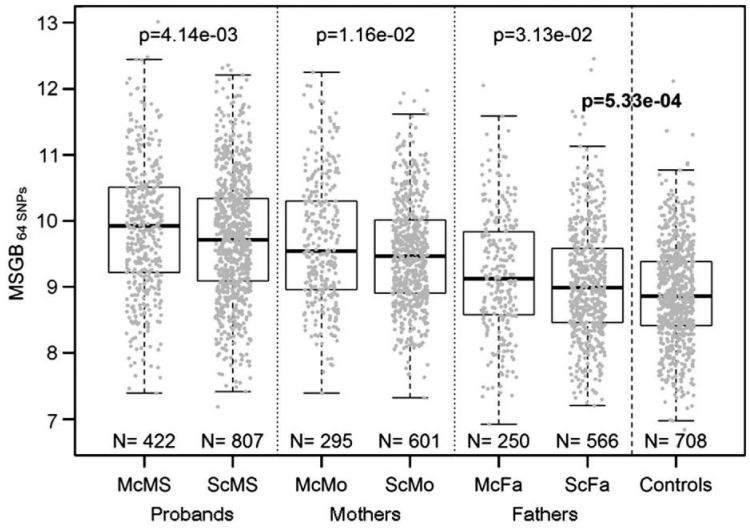

Figure 1.

MSGB score differentiates multi-case from single-case MS in UCSF families. The distribution of MSGB is presented using box plots. MSGB is computed using components derived from gender, MHC and non-MHC SNPs. Gray dots represent the MSGB of an individual subject. Groups separated by dotted lines (probands, mothers of probands, fathers of probands and unrelated controls) are divided into multi-case and single-case samples. Spouses of patients were considered genetically unrelated controls. Sample sizes are indicated at the bottom of each box plot. *P*-values in each of the three left panels indicate the significance of Wilcoxon's tests of the null hypothesis that MSGB of members of multi-case families are greater than those of members of single-case MS families (*P*-values without affected mothers and fathers: McMo vs ScMo = 6.63e – 02 and McFa vs ScFa = 3.47e – 02). The *P*-value in the right panel corresponds to the test that the MSGB of fathers of single-case MS patients is greater than unrelated controls (Wilcoxon's test). Spouses as healthy unrelated individuals are taken as controls. Fa, father; Mc, multi-case; Mo, mother; Sc, single-case.

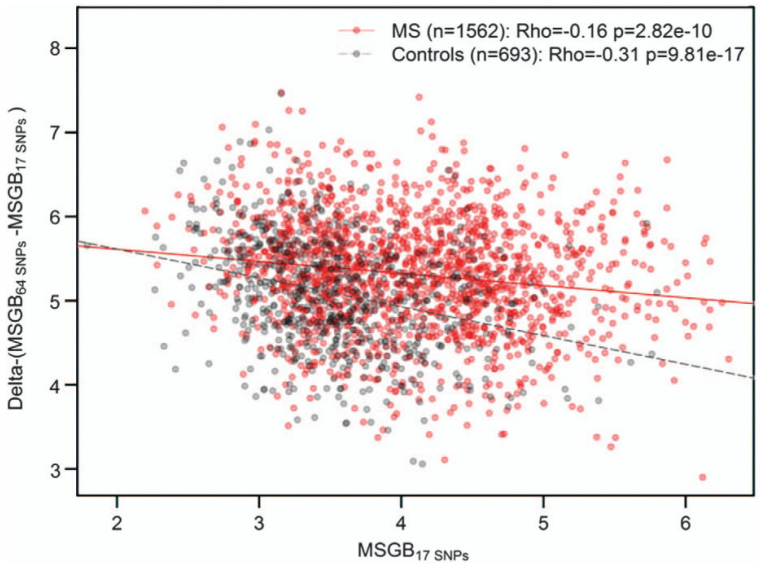

Figure 2.

Characteristics of the new part of MSGB scores compared with the previous MSGB scores with 17SNPs. *y* axis represents the new and eventually corrective part of MSGB scores calculated by subtracting $MSGB_{17SNPs}$ from $MSGB_{64SNPs}$. Similar correlation are obtained by plotting $MSGB_{17SNPs}$ vs $MSGB_{64SNPs}$ (data not shown). UCSF samples with available MSGB scores for both $MSGB_{17SNPs}$ and $MSGB_{64SNPs}$ are included. For MS patients, probands of the family data set and cases in the case–control data set are enrolled. Negative correlation between $MSGB_{17SNPs}$ and new part of MSGB ($MSGB_{64SNPs} - MSGB_{17SNPs}$) were seen for both MS patients and controls but controls had significantly stronger negative correlation than patients.

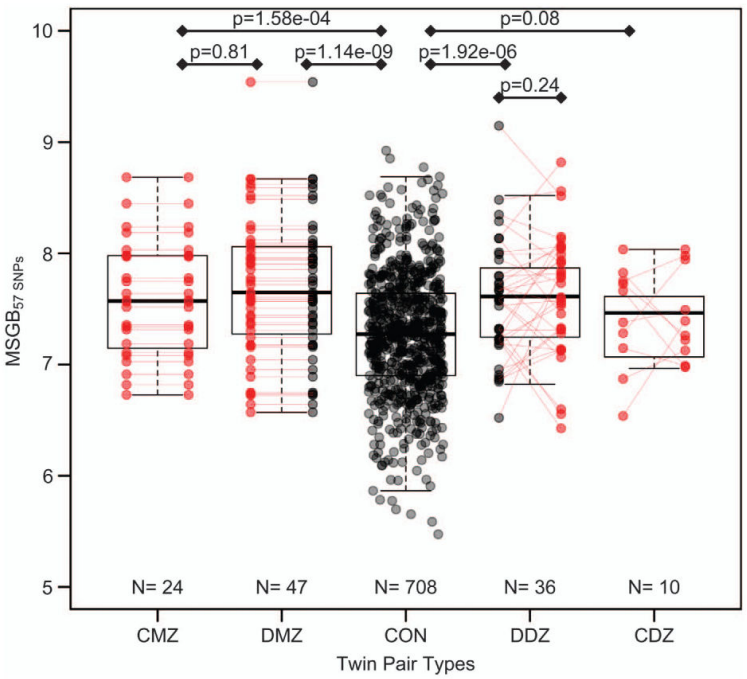

Figure 3.

Comparison of MSGB scores in monozygotic and dizygotic twin pairs. The distribution of MSGB is represented using box plots. The calculated MSGB does not account for gender or the presence of *HLA-DRB1*15:01*. Black dots represent the MSGB of unaffected twins or controls; red dots represent the MSGB of affected twins. Individual twin pairs are connected with a red line. *P*-values between adjacent groups or indicated groups (black horizontal line) represent the significance from a Wilcoxon's test. CDZ, concordant dizygotic; CMZ, concordant monozygotic; DDZ, discordant dizygotic; DMZ, discordant monozygotic.

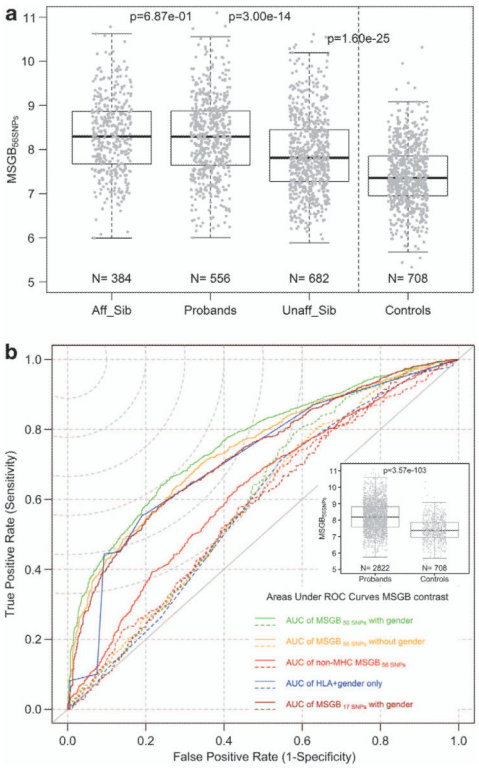

Figure 4.

(**a**) Distribution of MSGB in siblings of MS families. Computations have been done with the 56 SNPs typed in common with all data sets. Distribution of MSGB in siblings of UCSF and French MS multi-case families using box plots. MSGB is computed using components derived from gender, MHC and non-MHC SNPs. Gray dots correspond to the MSGBs of individual subjects. The three left box plots correspond to subjects' status in sibship (Aff_Sib = affected sibs; Unaff_Sib = unaffected sibs). Sample sizes are indicated at the bottom of each box plot. The first *P*-value corresponds to the test that MSGBs of affected sibs are different from MSGBs of probands (Wilcoxon's test). The second *P*-value corresponds to the test that the MSGBs of multi-case probands is greater than MSGBs of unaffected sibs of probands (Wilcoxon's test). The *P*-value overlaying the dotted line indicates the significance of Wilcoxon tests of the null hypothesis that MSGBs of unaffected siblings of probands are greater than the MSGBs of the controls. (**b**) ROC curves for MS prediction comparing achievement of various MSGBs sib-proband contrasts in sibships compared with the direct use of MSGB scores as predictors in general population. Computations have been done with up to 56 SNPs typed in common with all data sets. ROCs corresponding to: the prediction of MS status of the sibs of the probands from UCSF and French multi-case families based on the contrast between sib's and proband's MSGB (dotted lines); only UCSF multi-case families are used for the brown line; the prediction of MS status of the general population based on the contrast between unrelated UCSF controls and UCSF, French and Cambridge multi-case and single-case families probands (full lines); only UCSF samples are used for the brown line. In green, MSGB contrasts are computed using the gender, MHC and non-MHC SNPs components. In orange, MSGB contrasts are computed using the MHC and non-MHC SNPs components. In red, MSGB contrasts are

computed using the only the non-MHC SNPs components. In blue, MSGB contrasts are computed using the gender and the MHC components. In brown, MSGB contrasts are computed using the MSGB values of the previously published study (Gourraud *et al.*[23]) only for UCSF samples of families and case–control data set. The inset corresponds to the distribution of MSGB in probands from UCSF, French and Cambridge multi-case and single-case families and in unrelated UCSF controls, using box plots. Sample sizes are indicated at the bottom of each box plot. The *P*-value corresponds to the test that the MSGBs of probands are greater than MSGBs of controls (Wilcoxon's test). AUC, area under the curve; HLA, human leukocyte antigen.

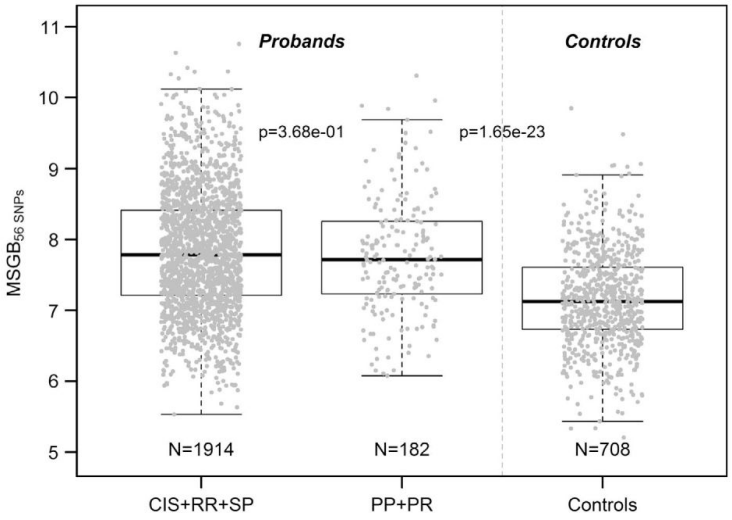

Figure 5.

Absence of different MSGB scores calculated with 56 SNPs between patients with relapsing-type MS vs patients with primary progressive MS. The distribution of $MSGB_{56SNPs}$ scores are shown on *y* axis. The *P*-value in probands indicates the significance of Wilcoxon tests of the null hypothesis that $MSGB_{56SNPs}$ of relapsing-type MS (CIS + RR + SP) patients are not different from those of primary progressive (PP + PR) patients. The *P*-value in the right part of the figure corresponds to the Wilcoxon test for the null hypothesis that the $MSGB_{56SNPs}$ of PP + PR MS patients are not different from those of unrelated controls. CIS, clinically isolated syndrome; PP, primary progressive; PR, progressive relapsing; RR, relapsing remitting; SP, secondary progressive.

Table 1

MSGB differenciates multi-case families from single-case families

MSGB components median, (p25–p75)	***Multi-case probands*** **n = *422***	***Single-case probands*** **n = *807***	***Multi-case parents*** **n = *545***	***Single-case parents*** **n = *1167***	***Controls*** **n = *708***	***Cuzick*** **z P**
Gender, HLA, non-MHC SNPs	9.93 (9.22–10.51)	9.71 (9.09–10.34)	9.35 (8.77–10.10)	9.23 (8.70–9.83)	8.86 (8.42–9.38)	$z = -23.07$, $P<1e-04$
HLA, non-MHC SNPs	9.53 (8.92–10.09)	9.33 (8.76–9.94)	9.09 (8.55–9.83)	8.99 (8.45–9.56)	8.61 (8.14–9.10)	$z = -17.73$, $P<1e-04$
Gender, non-MHC SNPs	8.84 (8.45–9.29)	8.89 (8.45–9.28)	8.64 (8.24–9.07)	8.64 (8.21–9.06)	8.48 (8.04–8.86)	$z = -17.73$, $P<1e-04$
Non-MHC SNPs	8.50 (8.09–8.90)	8.54 (8.10–8.90)	8.42 (7.98–8.79)	8.36 (8.00–8.79)	8.22 (7.80–8.62)	$z = -8.93$, $P<1e-04$

Abbreviations: HLA, human leukocyte antigen; MSGB, Multiple Sclerosis Genetic Burden ; SNP, single-nucleotide polymorphism.

Table 2

Demographic features of the enrolled individuals

	Multi-case families[a]			Single-case families			UCSF case/control	UCSF additional
	UCSF	French	Cambridge	UCSF	French	Cambridge		PP MS[b]
Probands								
Number	422	134	253	807	417	789	380	49
M/F/unknown (*n*)	103/319/0	47/87/0	60/193/0	192/615/0	134/283/0	202/587/0	123/257/0	24/25/0
Age of onset (yo)	30.15 ± 8.72	23.62 ± 9.09	26.67 ± 7.80	30.79 ± 9.07	26.13 ± 7.09	27.21 ± 6.87	35.01 ± 9.73	37.18 ± 10.99
Disease duration (years)	11.98 ± 8.98	14.51 ± 9.41	10.81 ± 8.11	10.23 ± 8.35	9.02 ± 7.15	11.16 ± 7.22	9.32 ± 9.31	NA
Disease course (CIS + RR + SP/PP + PR/unknown) (*n*)	387/28/7	68/2/64	231/18/4	738/56/13	365/31/21	725/60/4	356/16/8	0/49/0
MSSS (mean ± s.d.)	4.84 ± 2.51	3.29 ± 3.03	5.82 ± 2.74	4.59 ± 2.69	4.87 ± 2.78	5.98 ± 2.54	4.13 ± 2.18	6.80 ± 1.38
Sample call rates (mean ± s.d.) (%)	97.26 ± 5.48	99.30 ± 1.33	99.30 ± 1.75	98.42 ± 2.79	99.32 ± 1.20	99.29 ± 1.65	99.23 ± 1.56	92.38 ± 4.99
MSGB (*n* = 56) (mean ± s.d.)	8.54 ± 0.88	8.37 ± 0.82	8.01 ± 0.95	8.4 ± 0.88	8.39 ± 0.84	7.85 ± 0.82	8.24 ± 0.84	8.25 ± 0.77
*Relatives (*n)								
Mother (affected/ unaffected/unknown)	22/272/1	11/94/0						
Father (affected/ unaffected/unknown)	11/239/0	3/69/0						
Siblings (affected/ unaffected/unknown)	258/570/16	126/112/0	0/0/0	0/368/0	0/0/0	0/0/0		
Unrelated controls								
Number	253						455	
M/F/unknown (*n*)	189/64/0						146/309/0	
Sample call rates (mean ± s.d.) (%)	97.53 ± 4.01						99.34 ± 1.43	
MSGB (*n* = 56) (mean ± s.d.)	7.76 ± 0.66						7.66 ± 0.72	

Abbreviations: CIS, clinically isolated syndrome; F, female; M, Male; MS, multiple sclerosis; MSGB, multiple sclerosis genetic burden; MSSS, multiple sclerosis severity score; NA, not applicable; PP, primary progressive; PR, progressive relapsing; RR, relapsing-remitting; SP, secondary progressive.

[a] Definition of multicase family is as follows: UCSF and French group defined it as families with at least one affected first-degree relative of the affected proband, while Cambridge defined it as families with more than one affected individuals with no degree-related limitation.

[b] In the analysis of the comparison of the MSGB scores according to disease course shown in Figure 5, UCSF PPMS patients were included in addition to the patients in the multi-case and single-case families. In total, 1914 patients with relapsing-type MS (CIS + RR + SP) and 182 primary progressive MS patients (PP + PR) from UCSF and French group are enrolled in the analysis.