

EDITORIAL



Old Suspects Found Guilty — The First Genome Profile of Multiple Sclerosis

Leena Peltonen, M.D., Ph.D.

High-resolution genomewide association studies using panels of 300,000 to 1 million single-nucleotide polymorphisms (SNPs) aim to define genetic risk profiles of common diseases. These studies herald a fundamentally new opportunity to explore human biology and medicine, since they are unbiased by previous hypotheses or assumptions about the nature of genes that influence complex diseases. Underscoring the importance of this approach is the fact that many genetic variants identified as risk factors in type 2 diabetes and Crohn's disease by such studies have been localized to previously unsuspected pathways, to genes without a known function, or to noncoding regions of genes.

Since 1868, when multiple sclerosis was first described as a clinical entity by Charcot, scientists have tried to identify the underlying causal factors. Studies of twins, adopted children, and the epidemiology of multiple sclerosis indicate a complex set of causative factors — both genetic predisposition and largely unknown environmental factors are required to cause the disease. The importance of the genetic background is obvious from the concordance rate of multiple sclerosis in monozygotic twins of nearly 30%; siblings or dizygotic twins have a risk of 2%, still higher than the risk of 0.1% in the average Northern European population. A biologically significant genetic component is also implied by familial aggregation of patients and by incidence rates that vary among different ethnic groups, independently of geographic location. The prevalence of multiple sclerosis is high among Northern Europeans but low in African, Chinese, Japanese, and Saami populations.

As early as 1972, the association between mul-

tle sclerosis and the HLA region of the genome was established. This association has since been narrowed down to the *HLA-DRB1* gene on chromosome 6p21. Indeed, this locus is the single strongest genetic factor influencing susceptibility to multiple sclerosis. Multiple linkage-based genome scans performed with the use of 300 to 400 multiallelic markers in exceptional multiplex families or hundreds of sibling pairs have confirmed a major role for this HLA locus but also have revealed several other putative susceptibility loci.¹⁻⁴ Meta-analyses of the combined data sets of these linkage studies have failed to identify major loci outside the HLA locus but have suggested minor loci on 5p, 17q, and 19q, which were replicated in multiple samples. One can predict from the accumulated genetic information, including data from twin and family studies, that the susceptibility profile for multiple sclerosis consists of many, probably interacting, risk alleles, each contributing a relatively small effect to the overall risk.

In this issue of the *Journal*, Hafler et al.⁵ describe the outcome of the first effort of the International Multiple Sclerosis Genetics Consortium to define the genetic profile underlying a predisposition to multiple sclerosis. This large-scale association study, a joint analysis of an impressive data set of more than 12,000 subjects, supports the prediction of multiple risk alleles.

The consortium's initial genomewide scan analyzed the association of multiple sclerosis with more than 330,000 SNP markers in 931 family trios (consisting of an affected child and both parents) by monitoring the overtransmission of any SNP to the affected child with the use of transmission disequilibrium testing. The statis-

tical significance of potentially overtransmitted SNPs was substantiated by the inclusion of genotypes from additional control samples; the researchers had access to genomewide association data from nearly 2000 control subjects, produced by the Wellcome Trust Case Control Consortium and by a project funded by the National Institute of Mental Health. No SNP outside the HLA region provided a significant association with multiple sclerosis at the genomewide level. However, the initial screen yielded promising signals, which were pursued by genotyping of 110 associated SNPs in an independent sample of 2322 case subjects with multiple sclerosis and 2987 control subjects. A combined analysis of all the data from more than 12,000 samples was then used to confirm associations between SNPs and multiple sclerosis and to estimate the size of the effect.

The data strongly support the dominance of the HLA locus in the genetic background of patients with multiple sclerosis but also indicate the involvement of two interesting genes: *IL2RA*, which encodes the alpha subunit of the interleukin-2 receptor (also known as CD25) on chromosome 10p15, and *IL7RA*, which encodes the alpha chain of the interleukin-7 receptor on chromosome 5p13. Both of the genes scored as next-best signals in the final data analysis.

These two interleukin-receptor genes and corresponding proteins are important in T-cell-mediated immunity. Interleukin-2 receptor is critical for the regulation of T-cell responses, and interleukin-7 is crucial for the homeostasis of the memory-T-cell pool and may also be important in the generation of autoreactive T cells in multiple sclerosis. Other data support the relevance of these genes in the pathogenesis of multiple sclerosis.^{6,7} The *IL2RA* gene has also been associated with type 1 diabetes and Graves' disease, which implies that it plays a more general role in autoimmunity.

The variant of *IL7RA* with the best association with multiple sclerosis has a mutation (T244I) in the alternatively spliced exon that codes for a transmembrane domain of the polypeptide. This particular variant of *IL7RA* was also found to be associated with multiple sclerosis in two other independent candidate-gene studies involving 1800 Scandinavian patients with multiple sclerosis⁷ and in the combined data set of more than 2500 patients from families of European origin with multiple sclerosis, as well as in an indepen-

dent data set of more than 1100 case subjects and 3000 control subjects⁶ (much of which was also reported in the study by Hafler et al.). Initial functional data for the *IL7RA* variant indicate an effect on gene expression with a change in the ratio of soluble to cell-bound interleukin-7 receptor, at least in vitro. If further genetic and functional studies confirm the role of these immunologically relevant genes in the genetic profile of multiple sclerosis, they would support the hypothesis of abnormal regulation of T cells in multiple sclerosis and again point to an autoimmune basis for the disease.

It is important to recognize that the increased risk contributed by *IL2RA* and *IL7RA* is very low and that these two alleles explain only a very small proportion of the variance (0.2%) in the risk of multiple sclerosis. Since the associated SNPs for these two genes are also very common in the population (with a frequency of up to 70%), their presence cannot be a major risk factor for genetic disease. The HLA region retains its unique position as the only known major risk gene for multiple sclerosis.

What have we learned from this "hypothesis-free approach" about the genetic risk profile behind multiple sclerosis? First, the critical and repeated lesson from all the complex disease association studies is the need for a massive number of study samples and the collaboration of large consortia because of the small effect of common alleles. Second, a genomewide approach, even in the large study sample used by Hafler et al., did not reveal major genes with an effect on the risk of multiple sclerosis that is even remotely similar to that of the HLA region. Third, no clues to new pathways emerged from among the top-scoring associations, in contrast to the results of some genomewide studies involving other complex diseases, such as type 2 diabetes. Also, one of the study's best hits, *IL7RA*, is positioned under one of the previously reported linkage peaks on 5p.^{1,8} This finding also differs somewhat from the results of other genomewide scans, which typically do not arrive at loci found in previous linkage studies. Furthermore, the best-associated variants in the combined data analysis were actually not among the best hits in the initial screen — both of them showed only a very marginal association. This result should signal caution to groups that are hastily progressing to replication studies with only a handful of best-associated SNPs.

Finally, we should remember that, by definition, genomewide association studies that rely on common SNPs monitor only common alleles, but there is so much more in the genetic risk profiles behind common diseases such as multiple sclerosis. We need to define the full allelic diversity of the “suspicious genes” that are initially identified by genomewide studies. By sequencing the potential risk alleles in large study samples, we will probably encounter rare, high-impact alleles with critical importance for disease risk in some families or patients. The lessons learned from studies of genetic risk variants like *BRCA1* and *BRCA2*, which are rare alleles with a high impact but which explain only a small fraction of breast cancers, will be instructive for our thinking about other complex diseases. The somewhat disappointing outcome with respect to the attributable risk conferred by the SNPs sampled in *IL2RA* and *IL7RA* indicates that other types of genome variants should be sought. Without a doubt, the multiple sclerosis community will soon be informed about the systematic scans for copy-number variations or other more complex changes in the genomic architecture of risk alleles for multiple sclerosis.

No potential conflict of interest relevant to this article was reported.

This article (10.1056/NEJMe078147) was published at www.nejm.org on July 29, 2007.

From the University of Helsinki and the National Public Health Institute, Biomedicum, Helsinki, Finland; and the Broad Institute, Massachusetts Institute of Technology, Cambridge, MA.

1. Ebers GC, Kukay K, Bulman DE, et al. A full genome search in multiple sclerosis. *Nat Genet* 1996;13:472-6.
2. Haines JL, Ter-Minassian M, Bazyk A, et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. *Nat Genet* 1996;13:469-71.
3. Sawcer S, Jones HB, Feakes R, et al. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 1996;13:464-8.
4. Kuokkanen S, Gschwend M, Rioux JD, et al. Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am J Hum Genet* 1997;61:1379-87.
5. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;357. DOI: 10.1056/NEJMoa073493.
6. Gregory SG, Schmidt S, Seth P, et al. Interleukin 7 receptor alpha chain shows allelic and functional association with multiple sclerosis. *Nat Genet* (in press).
7. Lundmark F, Duvefeldt K, Jacobaeus E, et al. Variation in interleukin-7 receptor alpha chain influences risk of multiple sclerosis. *Nat Genet* (in press).
8. Kuokkanen S, Sundvall M, Terwilliger JD, et al. A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus *Eae2*. *Nat Genet* 1996;13:477-80.

Copyright © 2007 Massachusetts Medical Society.