ELSEVIER

Journal of Neuroimmunology 143 (2003) 112–115

Journal of Neuroimmunology

www.elsevier.com/locate/jneuroim

A whole genome screen for association with multiple sclerosis in Portuguese patients

M. Santos[a,1], J. Pinto-Basto[a,b,1], M.E. Rio[c], M.J. Sá[c], A. Valença[d], A. Sá[e], J. Dinis[f], J. Figueiredo[g], L. Bigotte de Almeida[h], I. Coelho[i], S. Sawcer[j], E. Setakis[k], A. Compston[j], J. Sequeiros[a,l], P. Maciel[a,l,m,*]

[a] *UnIGENe-IBMC, University of Porto, Portugal*
[b] *Inst. Genét. Médica Jacinto de Magalhães, Porto, Portugal*
[c] *Serv. Neurol., Hosp. S. João, Porto, Portugal*
[d] *Serv. Neurol., Hosp. Força Aérea, Lisboa, Portugal*
[e] *Serv. Neurol., Hosp. Dist. Sto. André, Leiria, Portugal*
[f] *Serv. Neurol., Hosp. Dist. Sta. Luzia, Viana do Castelo, Portugal*
[g] *Serv. Neurol., Hosp. S. Marcos, Braga, Portugal*
[h] *Serv. Neurol., Hosp. Garcia de Orta, Almada, Portugal*
[i] *Serv. Neurol., Hosp. Dist. Sra. Oliveira, Guimarães, Portugal*
[j] *Neurology Unit, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK*
[k] *MRC Biostatistics Unit, Institute of Public Health, University of Forvie Site, Cambridge, UK*
[l] *Dept. Est. Pop., ICBAS, University of Porto, Portugal*
[m] *ICVS, Health Sciences School, University of Minho, Braga, Portugal*

Abstract

Multiple sclerosis (MS) is common in Europe affecting up to 1:500 people. In an effort to identify genes influencing susceptibility to the disease, we have performed a population-based whole genome screen for association. In this study, 6000 microsatellite markers were typed in separately pooled DNA samples from MS patients ($n = 188$) and matched controls ($n = 188$). Interpretable data was obtained from 4661 of these markers. Refining analysis of the most promising markers identified 10 showing potential evidence for association.

Keywords: Genome screen; Linkage disequilibrium; Portugal; Multiple sclerosis; DNA pooling

1. Introduction

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system with a wide range of clinical manifestations, which most commonly presents in the third or fourth decade. There are no published studies of prevalence from Portugal but a rate of 15 per 100,000 was observed by Callegaro et al. (2001) in Sao Paulo, Brazil amongst a population of mainly Portuguese descent. The importance of genetic factors in the aetiology of MS has been established by classical epidemiological studies such as those showing a higher concordance rate for MS in monozygotic (25%) compared with dizygotic twins (3%) (Ebers et al., 1996). Efforts to identify susceptibility genes have established association of the disease with MHC Class II alleles (Marrosu et al., 1988; Olerup and Hillert, 1991), but no other consistently supported genes have thus far been identified (reviewed in Compston et al., 1998).

For diseases not following simple Mendelian inheritance, such as MS, the absence of multigenerational families effectively precludes the use of parametric linkage analysis, at the same time non-parametric linkage analysis lacks power when the genetic effects attributable to individual loci are modest. In this setting, tests for association are optimal (Risch and Merikangas, 1996). The goal of popu-

* Corresponding author. Life and Health Sciences Research Institute, School of Health Sciences, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal. Tel.: +351-253604800; fax: +351-253604831.
E-mail address: pmaciel@ecsaude.uminho.pt (P. Maciel).
[1] Both authors contributed equally to this work.

0165-5728/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneuroim.2003.08.023

lation-based association studies is to identify genetic variants that differ in frequency between (unrelated) affected individuals and healthy controls (Kruglyak, 1999). This difference in frequency of a marker allele may reflect a direct causative effect of the allele in question, or may occur when the marker allele is in linkage disequilibrium (LD) with the biologically relevant genetic variant.

In order to screen the genome for LD in the Portuguese MS population, we have analysed a dense map of microsatellites markers in a case-control cohort using DNA pooling for efficiency (Barcellos et al., 1997; Sawcer et al., 2002).

2. Material and methods

2.1. Patients

We studied a sample of 188 unrelated patients (129 females: 59 males) with an average age of 37 (17–64) years. All were of Portuguese origin and all had a definite diagnosis of MS, according to the criteria of Poser et al. (1983). We also studied 188 unrelated healthy controls (HC), matched by sex, age ($\pm$ 2 years) and region of origin. All the participants were informed about the purposes of this study and gave written consent to genetic analysis.

2.2. DNA pooling

Genomic DNA from all subjects was isolated from peripheral blood, using a genomic DNA isolation kit (Puregene). DNA concentration from stocks was measured twice in a BECKMAN DU 650 spectrophotometer, and all samples were diluted to a final concentration of 60 ng/μl. DNA concentration of these dilutions was then measured with the PicoGreen system (Molecular Probes), according to the manufacturer's instructions, and again diluted to a final concentration of 25 ng/μl.

We used 100 μl (2.5 μg) of each sample, and pooled the 188 cases and 188 controls, separately.

2.3. Polymerase chain reaction and capillary electrophoresis

The GAMES collaborative provided the 6000 markers, which were analysed in the case and control pools. DNA from the two pools was amplified twice for each marker, in a final reaction volume of 15 μl using True Allele™ Premix (Applied Biosystems), according to manufacturer's instructions. Each set of PCR products was then electrophoresed twice, on a 3700 DNA analyser (Applied Biosystems). Semi-automated sizing was performed using the GENESCAN software (Applied Biosystems); 4661 markers provided electropherograms of sufficient quality to enable semi-automated analysis using the Applied Biosystems GENOTYPER (version 3.6) software.

2.4. Statistical analysis

Genotyping a microsatellite in a pooled DNA sample generates an allele image pattern (AIP) consisting of a series of fragments, the signal from each reflecting the frequency of the corresponding allele. For each marker, a weighted average AIP was calculated for each pool, using the available data from that pool (up to four AIPs for each). The peak height distribution thus generated was then normalised according to the number of alleles in the corresponding pool. The resulting allele count distributions were compared statistically using a Chi-square test, and the observed distribution of the Chi-square statistic was used to evaluate the significance of results empirically. This empirical approach was used because analysis by pooling introduces additional sources of variance, beyond the expected sampling variance, that are not normally distributed. A software package to perform this statistical analysis was specifically developed as part of the GAMES collaborative study (by ES). The statistical methods are described in detail in the accompanying paper from Setakis (2003).

The raw data from those markers with the most extreme empirical *p*-values were reassessed and the best were taken forward for a second analysis. In this second stage, we employed the adapting factors established by Yeo et al. (2003).

3. Results

We employed a two-stage approach in our search for genes influencing susceptibility to MS. In the first stage, we genotyped the 6000 microsatellite markers provided through the Genetic Analysis of Multiple sclerosis in EuropeanS (GAMES) collaborative in separately pooled DNA samples from 188 cases and 188 controls. Usable allele image patterns (AIPs) were obtained from 4661 markers. Seventy-eight of these markers gave empirical *p*-values of < 0.01. The raw AIPs from these 78 markers were re-inspected and assessed for data quality, homogeneity and genotyping errors. In the second stage, new AIPs were generated for

Table 1
Empirical *p*-values for the most extreme 10 markers

Marker	*p*-value
D10S1423	0.0085
D11S1333	0.0013
D11S1914	0.0034
D11S4046	0.0022
D4S2921[a]	0.0078
D4S426[a]	0.0052
D5S1464	0.0027
D6S2444	0.0038
TNFa	0.0015
D7S630	0.0062

[a] These two markers were not included in the 529 considered by Yeo et al.

the 34 most promising markers and the combined data set analysed employing the adapting factors suggested by Yeo et al. (2003). Ten markers retained empirical p-value <0.01 in this refining analysis (see Table 1).

4. Discussion

Given the importance of gene–environment interactions in MS, it is possible that different genetic factors are relevant to its pathogenesis in different populations. Therefore, it may be helpful to consider various populations, and those where the risk of the disease is less striking may be of use precisely because the genetic risk factors are less common, and easier to pinpoint.

We have performed a population-based whole-genome screen for association with MS in the Portuguese population using polymorphic microsatellites to identify chromosomal regions containing genes potentially involved in susceptibility to this disease. The study is dependant upon the hypothesis that linkage disequilibrium will result in associations between neutral polymorphisms and functional risk-conferring variants of relevant genes, so that certain marker alleles will be over represented in patients compared with the normal population (associated alleles).

In association studies, the sample size and the choice of the control population are of crucial importance (Cardon and Bell, 2001). Selection of a control group that really represents the population, avoiding artefacts that may either mask real effects or introduce population stratification (and generate false positive results) is therefore extremely important. It is also important to have a group of patients that is large enough to provide the statistical strength needed for detection of genes with subtle effects. In this study, we individually paired every one of the 188 patients with a control individual from the same region of the country (sharing a similar environment and genetic background), of the same gender and with approximately the same age (± 2 years).

In order to overcome the barrier provided by the large number of genotypes needing to be completed, a DNA pooling strategy was used, as proposed by Barcellos et al. (1997) and previously employed by Sawcer et al. (2002) and Shaw et al. (1998). This strategy allowed the genotyping of thousands of markers for large case ($n=188$) and control ($n=188$) groups, and then comparison of allele frequencies in these groups, reflected in their respective allele image patterns (AIP) (Daniels et al., 1998). Identification of differences in allele frequency distribution between cases and controls was undertaken using a statistical program specifically developed (by ES) for the purpose of the collaborative study. We selected 34 markers from the screening phase that were studied a second time during the refining stage. Of these, 10 retained an empirical p-value of $p<0.01$ after refined analysis. Selection of markers was performed using systematic criteria involving an evaluation of the data quality (intensity of fluorescence) and the homogeneity of results between duplicate electropherograms. We also considered the consistency between the original screening AIPs and those observed in the refining phase. These markers correspond to seven chromosomal regions: 4q, 5q, 6p, 7q, 10p, 11p and 11q. Of these, regions 5q, 6p have already been referred as having a positive lod score in other whole genome linkage studies (Sawcer et al., 1996; Haines et al., 1996; Ebers et al., 1996), as well as the region on chromosome 7q identified in the Canadian study (Ebers et al., 1996). More recently, whole genome LD studies suggested regions that we could also identify, namely, chromosomal regions 6p and 11p (Goedde et al., 2002) and regions 6p and 4q (Sawcer et al., 2002). Besides these five regions, we suggest two additional ones (10p and11q), that might represent new regions associated with MS: to our knowledge, these regions have not previously been referred as relevant in MS. The replication of previous findings of association with markers in the HLA region suggests that this methodology is appropriate for the detection of *loci* associated to MS.

Future work will involve individually genotyping the best markers for association with MS in the individuals used to construct the pool, replication of the results in a different sample of MS patients, and refinement of the regions of interest using additional markers.

Acknowledgements

We thank all patients and healthy individuals who contributed samples to this study.

Special acknowledgement to nurses M[a]. Fátima Cunha and Joaquim Martins.

This work was supported by SERONO (Portugal) and British Council/ICCTI. Santos M supported by FCT, grant SFRH/BD/9111/2002. This study represents one component of the GAMES project, designed and co-ordinated by DASC and SJS (Cambridge, UK), with funding from the Wellcome Trust (grant 057097), supplemented by the Multiple Sclerosis Societies of the United States and Great Britain, and the Multiple Sclerosis International Federation.

References

Barcellos, L., Klitz, W., Field, L., Tobias, R., Bowcock, A., Wilson, R., Nelson, M., Nagatomi, J., Thomson, G., 1997. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am. J. Hum. Genet. 61, 734–747.

Callegaro, D., Goldbaum, M., Morais, L., Tilbery, C.P., Moreira, M.A., Gabbai, A.A., Scaff, M., 2001. The prevalence of multiple sclerosis in the city of Sao Paulo, 1997, Brazil. Acta Neurol. Scand. 104, 208–213.

Cardon, L.R., Bell, J.I., 2001. Association study designs for complex diseases. Nat. Rev., Genet. 2, 91–99.

Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews, B.,

Wekerle, H., 1998. Genetic Susceptibility to Multiple Sclerosis. McAlpine's Multiple Sclerosis. Churchill Livingstone, London, UK.

Daniels, J., Holmans, P., Williams, N., Turic, D., McGuffin, P., Plomin, R., Owen, M., 1998. A simple method for analysing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. Am. J. Hum. Genet. 62, 1189–1197.

Ebers, G.C., Kukay, K., Bulman, D., Sadovnick, A., Rice, G., Anderson, C., Armstrong, H., Cousin, K., Bell, R., Hader, W., Paty, D., Hashimoto, S., Oger, J., Duquette, P., Warren, S., Gray, T., O'Connor, P., Nath, A., Auty, A., Metz, L., Francis, G., Paulseth, J., Murray, T., Pryse-Phillips, W., Nelson, R., Freedman, M., Brunet, D., Bouchard, J.P., Hinds, D., Risch, N., 1996. A full genome search in multiple sclerosis. Nat. Genet. 13, 472–476.

Goedde, R., Sawcer, S., Boehringer, S., Miterski, B., Sindern, E., Haupts, M., Schimrigk, S., Compston, A., Epplen, J.T., 2002. A genome screen for linkage disequilibrium in HLA-DRB1*15-positive Germans with multiple sclerosis based on 4666 microsatellite markers. Hum. Genet. 111, 270–277.

Haines, J., Pericak-Vance, M., Seboun, E., Hauser, S., 1996. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. Nat. Genet. 13, 469–471.

Kruglyak, L., 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. 22, 139–144.

Marrosu, M.G., Muntoni, F., Murru, M.R., Spinicci, G., Pischedda, M.P., Goddi, F., Cossu, P., Pirastu, M., 1988. Sardinian multiple sclerosis is associated with HLA-DR4: a serologic and molecular analysis. Neurology 38, 1749–1753.

Olerup, O., Hillert, J., 1991. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens 38, 1–15.

Poser, C.M., Paty, D.W., Scheinberg, L., McDonald, W.I., Davis, F.A., Ebers, G.C., Johnson, K.P., Sibley, W.A., Silberberg, D.H., Tourtellotte, W.W., 1983. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann. Neurol. 13, 227–231.

Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science 13, 1516–1517.

Sawcer, S., Jones, H.B., Feakes, R., Gray, J., Smaldon, N., Chataway, J., Robertson, N., Clayton, D., Goodfellow, P., Compston, A., 1996. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nat. Genet. 13, 464–468.

Sawcer, S., Maranian, M., Setakis, E., Curwen, V., Akesson, E., Hensiek, A., Coraddu, F., Roxburgh, R., Sawcer, D., Gray, J., Deans, J., Goodfellow, P.N., Walker, N., Clayton, D., Compston, A., 2002. A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. Brain 125, 1337–1347.

Setakis, E., 2003. Statistical analysis of the GAMES studies. J. Neuroimmunol. 143, 47–52 (this issue).

Shaw, S., Carrasquillo, M., Kashuk, C., Puffenberger, E., Chakravarti, A., 1998. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Res. 8, 111–123.

Yeo, T.W., Roxburgh, R., Maranian, M., Singlehurst, S., Gray, J., Hensiek, A., Setakis, E., Compston, A., Sawcer, S., 2003. Refining the analysis of a whole genome linkage disequilibrium association map: the United Kingdom results. J. Neuroimmunol. 143, 53–59 (this issue).