



HAL
open science

Tag-SNP analysis of the GFI1-EVI5-RPL5-FAM69 risk locus for multiple sclerosis

Fuencisla Matesanz, Antonio Alcina, Oscar Fernández, Juan R González, Antonio Catalá-Rabasa, Maria Fedetz, Dorothy Ndagire, Laura Leyva, Miguel Guerrero, Carmen Arnal, et al.

► To cite this version:

Fuencisla Matesanz, Antonio Alcina, Oscar Fernández, Juan R González, Antonio Catalá-Rabasa, et al.. Tag-SNP analysis of the GFI1-EVI5-RPL5-FAM69 risk locus for multiple sclerosis. *European Journal of Human Genetics*, 2010, n/a (n/a), pp.n/a-n/a. 10.1038/ejhg.2009.240 . hal-00504138

HAL Id: hal-00504138

<https://hal.science/hal-00504138>

Submitted on 20 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 ***Tag-SNP analysis of the GFII-EVI5-RPL5-FAM69 risk***

2 ***locus for multiple sclerosis***

3 A Alcina¹, O Fernández², JR Gonzalez³, A Catalá-Rabasa¹, M Fedetz¹, D Ndagire¹, L
4 Leyva², M Guerrero², C Arnal⁴, C Delgado⁵, M Lucas⁶, G Izquierdo⁷, F Matesanz¹

5 **Authors' Affiliation:**

6 ¹Instituto de Parasitología y Biomedicina "López Neyra". Consejo Superior de
7 Investigaciones Científicas. Granada. Spain.

8 ²Servicio de Neurología. Instituto de Neurociencias Clínicas. Hospital Carlos Haya.
9 Málaga. Spain.

10 ³ Centre de Recerca en Epidemiologia Ambiental (CREAL); CIBERESP; Institut
11 Municipal d'Investigació Mèdica (IMIM).

12 ⁴ Servicio de Neurología. Hospital Virgen de las Nieves. Granada. Spain.

13 ⁵ Centro Regional de Transfusión Sanguínea Granada-Almería. Spain.

14 ⁶ Servicio de Biología Molecular. Hospital Virgen Macarena. Sevilla. Spain.

15 ⁷ Unidad de Esclerosis Múltiple. Hospital Virgen Macarena. Sevilla. Spain.

16

17 **Running Title:** The *GFII-EVI5-RPL5-FAM69* risk locus for MS

18 **Correspondence:**

19 Dr. Fuencisla Matesanz

20 Instituto de Parasitología y Biomedicina Lopez Neyra. CSIC. Parque Tecnológico de
21 Ciencias de la Salud. Avda. del Conocimiento s/n. 18100. Granada. Spain.

22 E-mail: pulgoso@ipb.csic.es.

23 Phone: +34-958-181668. Fax: +34-958-181632

24

25 **Abstract**

26

27 A recent genome-wide association study conducted by the International Multiple Sclerosis
28 Genetic Consortium (IMSGC) identified, among others, a number of putative multiple
29 sclerosis (MS) susceptibility variants at position 1p22. Twenty one SNPs positively
30 associated with MS were located at the *GFI1-EVI5-RPL5-FAM69A* locus. In this study,
31 we performed an analysis and fine mapping of this locus genotyping 8 Tag-SNPs up to
32 732 MS patients and 974 controls from Spain. We observed association with MS in 3 of 8
33 Tag-SNPs: rs11804321 (P=0.008. OR=1.29; 95% CI=1.08-1.54), rs11808092 (P=0.048.
34 OR=1.19; 95% CI=1.03-1.39) and rs6680578 (P=0.0082. OR=1.23; 95% CI=1.07-1.41).
35 After correcting for multiple comparisons and using logistic regression analysis to test the
36 addition of each SNP to the most associated SNPs, we observed that the rs11804321 alone
37 was sufficient to model the association. This Tag-SNP captures two SNPs in complete
38 linkage disequilibrium ($r^2 = 1$) both located within the 17th intron of *EVI5* gene. Our
39 findings agree with the corresponding data of the recent IMSGC study and present new
40 genetic evidence that point to *EVI5* as a factor of susceptibility to MS.

41

42 **Key words:** Multiple sclerosis (MS); Tag-SNP analysis; polymorphisms; genetics; *GFI1* -
43 *EVI5-RPL5-FAM69A* genes; association.

44

45

46

47

48 **Introduction**

49 Multiple sclerosis (MS) is a complex disease presumed to be autoimmune and
50 characterized by inflammation and demyelination with axonal and neuronal degeneration.
51 MS is the most common central nervous system disease in young adults.¹ The prevalence
52 and incidence rates in Spain are around 77 /100000 habitants and 5.3/100000 habitants
53 per year respectively similar to what has been found in Britain.^{2,3} Susceptibility to MS is
54 thought to be conferred by the combination of genetic and environmental factors.^{1,4}

55 The best established region implicated in predisposition to MS is the major
56 histocompatibility complex (MHC) on chromosome 6p21, specifically the HLA-DRB1*
57 1501 class II allele; but this account for less than 50% of MS genetics.^{1,5} Recently, other
58 regions have been implicated in MS susceptibility and replicated in different independent
59 populations such as the interleukin 7 receptor alpha (*IL7RA*),⁶⁻⁸ the interferon regulatory
60 factor 5 (*IRF5*) gene⁹ and the interleukin-2 receptor alpha (*IL2RA*).^{8,10} However, except
61 for the *IL7RA*, the causal SNP of the new determined risk loci and the functional effect in
62 the encoded proteins are unknown.

63 A recent genome-wide association study conducted by the International Multiple
64 Sclerosis Genetics Consortium (IMSGC) has identified a number of putative MS
65 susceptibility genes,⁸ among them, 21 SNPs that are located in the locus containing the
66 growth factor independent 1 (*GFII*), ecotropic viral integration site 5 (*EVI5*), ribosomal
67 protein L5 (*RPL5*) and family with sequence similarity 69 (*FAM69*). The association of
68 two of these SNPs has been replicated in a posterior study with 1318 MS Canadian
69 patients.¹¹ In this study, our aim was to validate the association of the *GFII-EVI5-*
70 *RPL5-FAM69A* locus in a Caucasian Spanish population and try to identify the gene

71 responsible for the association with MS by a Tag-SNP strategy that captured 279
72 common variants at a pairwise $r^2 \geq 0.6$.

73

74 **Material and Methods**

75 **Study subjects**

76 Case samples comprised 732 patients with clinically defined MS according to Poser's
77 criteria.¹² They were obtained from four public hospitals: the Hospital Clínico in Granada
78 (n=126), the Hospital Virgen de las Nieves de Granada (n=165), the Hospital Carlos Haya in
79 Málaga (n=365) and the Hospital Virgen de la Macarena in Seville (n=76) all three cities
80 within a 200 km radius in the South of Spain. The mean age at the sample collection of the
81 cases was 36 years and mean age of controls at interview was 38 years. The percentage of
82 females was 68% for cases and 68 % for controls. All of them were classified as RR
83 (relapsing-remitting) or SP (secondary progressive) MS cases. Controls were 974 blood
84 donors with no history of inflammatory disease attending the blood banks of Granada
85 (n=823), Seville (n=71) and Málaga (n=80). The study was approved by the Ethics
86 Committees of each of the hospitals participating in the study and written informed consent
87 was obtained from all participants.

88 **Genotyping**

89 High-molecular-weight DNA was isolated from whole blood using the Flexigene Kit
90 (Qiagen. Hildren. Gemany) according to the manufacturer's protocol. The 8 SNPs were
91 genotyped by TaqMan technology under conditions recommended by the manufacturer
92 (Applied Biosystems, Foster City, CA. USA).

93 **Sample Power Calculation**

94 First, we computed power calculations using the Genetic Power Calculator
95 ([http://pngu.mgh.harvard.edu/~purcell\(gpc/\)](http://pngu.mgh.harvard.edu/~purcell/gpc/)). Thus, we determined that the case-control
96 sample had 84% power for detecting a risk allele with 20% frequency and a dominant
97 genotype relative risk of 1.3. To assess the power for detecting association due to LD with a
98 causal loci, we also carried out power calculations for an indirect association study that uses
99 tagSNPs. We estimated that our study was able to detect a susceptibility locus (80% power)
100 with an odds ratio (OR) of 1.4 if the MAF is 0.2 and an OR of 1.3 if the MAF is 0.3,
101 assuming a dominant effect at an unobserved locus, an alpha (α) value of 0.05, and $r^2 = 0.85$
102 for the ability of haplotypes to predict the allele count at the causal locus

103 **Statistical Analysis**

104 Departure from Hardy-Weinberg equilibrium (HWE) for all the biallelic SNP markers was
105 tested using an exact test.¹³ For individual SNP association analyses, genotype frequencies
106 were assessed by means of MAX-statistic. This test fits dominant, recessive and additive
107 models and consider the most significant among them. To avoid multiple comparisons by
108 fitting three different models, p-values are then computing using the exact distribution of the
109 MAX-statistic.¹⁴ Logistic regression models were used to estimate crude ORs and 95%
110 confidence intervals (95% CI). In order to determine which SNP is more associated with
111 MS, multiple logistic regression models were computed. By using likelihood ratio test, we
112 compared the effect of each SNP adjusted by the most promising SNP found in the crude
113 analysis. These analyses were performed using the SNPassoc R package¹⁵. To avoid false-
114 positive results due to multiple testing and to considering that the SNPs analyzed are not in
115 complete disequilibrium, we applied the Benjamini-Hochberg (BH) method¹⁶ that is robust
116 against positive dependence and controls the false discovery rate.¹⁷

117 **Results**

118

119 We selected the region between SNPs rs11164607 and rs17380908 that covers 455413
120 base pairs containing the *GFII-EVI5-RPL5-FAM69A* risk locus which is associated with
121 MS in the IMSGC study.⁸ To analyze this region we chose eight SNPs by pairwise
122 tagging from the HapMap B35 CEU population which captured 279 markers with $r^2 \geq$
123 0.6 (mean $r^2 = 0.911$) and a minor-allele frequency (MAF) ≥ 0.1 .

124 We proceeded to genotype the Tag-SNPs in a Caucasian population of 732 MS
125 patients and 974 controls from the South of Spain. The results of genotypic and carrier
126 distribution between cases and controls are shown in Table 1. The genotype frequencies
127 assessed by MAX-statistic and the best fitted model are also shown in Table 1. Three
128 polymorphisms were associated with MS: rs11804321 (P=0.008. OR=1.29; 95%
129 CI=1.08-1.54), rs11808092 (P=0.048. OR=1.19; 95% CI=1.03-1.39) and rs6680578
130 (P=0.0082. OR=1.23; 95% CI=1.07-1.41) and from these, two survived BH correction,
131 rs11804321 and rs6680578. We performed a logistic regression analysis testing the
132 addition of each SNP to both SNPs (independently) to determine what the primarily
133 associated marker was (Table 2). We found that the rs11804321 alone was sufficient to
134 model the association with MS since as observed in the Table 2, the model that adds the
135 locus rs6680578 to rs11804321 was not statistically significant (p from likelihood ratio
136 test equal to 0.3074). This means that the effect found in the single analysis of
137 rs6680578 was confounded by the effect of the polymorphism rs11804321. In addition,
138 we also observed that the variant rs11804321 remained statistically significant after
139 being adjusted for the others SNPs (Table 2, second analysis). Therefore, the
140 rs11804321 polymorphism, among the ones analyzed with minor allele frequency ≥ 0.1

141 and $r^2 \geq 0.6$ with rs11804321, showed the strongest association in the *GFII-EVI5-RPL5-*
142 *FAM69* locus and independent effect on MS susceptibility.

143 The IMSGC⁸ analyzed 34 SNPs in this region, 32 of them were represented in the
144 HapMap B35. Based on the linkage disequilibrium (LD) for these 32 SNPs from the
145 HapMap, we grouped them in 6 Tag-SNPs (SNP2, 4, 5, 6, 7, 8) as shown in Table 3. The
146 other two Tag-SNPs were not represented by any IMSGC marker. Thirteen out of 21
147 IMSGC associated SNPs were tagged by SNP6 and 4 by SNP4. The SNP7 and SNP8
148 were not associated in our cohort. Twelve SNPs not associated in the IMSGC work were
149 tagged by SNP2 and SNP5 which also resulted negative for association with MS in this
150 study. The 87 HapMap SNPs tagged by SNP6 were localized in the four genes and
151 intergenic regions of the 1p22 locus while the 23 tagged by SNP4 were located in *EVI5*
152 gene. The 2 SNPs tagged by SNP1 are in the 17th intron of the *EVI5* gene. The frequency
153 as well as the LD map of the 8 Tag-SNPs in our population was similar to the ones
154 obtained from the HapMap CEU genotypes (Fig.1).

155 SNPs rs6680578 and rs11164838 have been analyzed in previous studies with
156 discordant results. To address this issue, we performed a meta-analysis with the data of
157 the IMSGC,⁸ Canadian,¹¹ Australian,¹⁸ and Spanish (this work) cohorts, as shown in
158 Figure 2. The summary effect estimate for the rs6680578 is significant with a combined
159 OR = 1.12, 95% CI [1.07- 1.17] and an estimated heterogeneity variance no significant
160 (P= 0.395). On the other hand, the rs11164838 showed heterogeneity among results (P=
161 0.025) with a combined effect estimate no significant (OR= 1.04, 95% CI [0.93-1.16]).
162

163 **Discussion**

164 In the present study we have analyzed the *GFI-EVI5-RPL5-FAM69A* locus in order to
165 test, in a Spanish MS cohort, the association found in a genome-wide association study
166 by the IMSGC.⁸ A fine mapping of this genomic region was performed by a Tag-SNP
167 approach that points to the *EVI5* as the most probable risk gene for MS. Our results
168 replicate those obtained by the IMSGC with the exception for rs7514280 and
169 rs11164838 that were not associated with MS in our cohort.

170 The rs6680578 and rs10735781, two SNPs located in the *EVI5* gene in nearly
171 complete LD ($D'=0.98$), have been previously tested in an independent set of 756
172 Canadian families containing 1318 MS patients.¹¹ Both SNPs had a weak but significant
173 contribution in this population. However, a replication study of 4 SNPs in the *GFI-EVI5-*
174 *RPL5-FAM69A* locus with 1134 Australian MS patients¹⁸ showed not association. The
175 meta-analysis we performed for the rs6680578 using data of the IMSGC,⁸ Canadian,¹¹
176 Australian¹⁸ and Spanish (this work) cohorts resulted significant with a combined odds
177 ratio 1.12 (95% CI [1.07- 1.17]). On the other hand, the rs11164838, that has been
178 associated with MS in the IMSGC genome-wide study,⁸ resulted not associated in both
179 Spanish and Australian¹⁸ cohorts. This SNP is in low LD ($D'=0.27$, $r^2=0.014$) with
180 rs11804321, the most associated Tag-SNPs in this work, therefore, the power to capture
181 the association with MS due to rs11804321 with rs11164838 would be very low for the
182 Australian and Spanish cohorts. Thus, the lack of power seems to be the cause of the
183 different results among cohorts. In fact, a recent genome-wide study performed by the
184 Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene)¹⁹ in
185 1,618 cases and 3,413 controls have found association with MS of the rs6604026 in the
186 *EVI5-RPL5* locus. This polymorphism is in high LD with rs11804321 ($D'=0.72$). All

187 together, these data support the association of the locus with MS in all the populations
188 studied.

189 The logistic regression analysis to determine the primarily associated marker showed
190 that the association within this region could best be explained by one disease locus
191 tagged by rs11804321. The two SNPs tagged from the HapMap database by rs11804321
192 are located in the *EVI5* gene. We note, however, that until a more complete set of
193 polymorphisms is identified and genotyped in a large collection of cases and control
194 subjects, we cannot exclude another variant in LD with SNP rs11804321 being the
195 causal variant. Future resequencing of the locus may provide as-yet-undiscovered
196 variants that will need to be assessed for diseases susceptibility.

197 The EVI-5 protein contains a TBC (TRE2/BUB/CDC16 homology) motif located in
198 the N-terminal 200 amino acid region (EVI-5N) that binds and activates guanosine
199 triphosphate-bound form of Rab11 (GTPase-activating protein. GAP).²⁰ This
200 heterodimer complex may coordinate vesicular trafficking, cytokinesis and cell cycle
201 control independent of GTPase-activating protein function.²¹ EVI-5 also is as a nuclear
202 zinc finger protein that functions as a transcriptional repressor possibly involved in
203 centrosome stability and dynamics.²² Interestingly, SNP4 located in exon 15 of the *EVI5*
204 gene is a non-synonymous polymorphism (612 His/Gln) but the potential effect on the
205 EVI5N domain is unknown. Statistically, SNP4 did not survive BH correction. SNP5 at
206 exon 7 of *EVI5* gene produces a 316 Val-Ile change but it did not show association with
207 MS. None of the SNPs analyzed seems to interfere with any splice site or transcription
208 factor recognition site at the promoter of the genes in the locus.

209 In conclusion these findings suggest that *EVI5* is the most likely candidate in the
210 *GFII-EVI5-RPL5-FAM69A* locus as a risk gene for MS in a cohort of Caucasian from

211 the south of Spain. We confirm the data of the GWAS of IMSGC⁸ remaining open the
212 question of the determination of the causal polymorphism.

213

214 **Acknowledgments**

215

216 We thank the patients with multiple sclerosis and persons acted as controls for making
217 this study possible. Financial support for the study was provided by the Ministerio de
218 Ciencia e Innovación-Fondos Feder (grant SAF2009-11491) and Junta de Andalucía
219 (P07-CVI-02551) to A. Alcina and Fondo de Investigación Sanitaria (PI081636) to F.
220 Matesanz. María Fedetz is a holder of a fellowship from Fundación IMABIS. Dorothy
221 Ndagire is a holder of AECI-Ministerio de Asuntos Exteriores fellowship.

222

223 **References**

- 224 1 Compston A, Coles A : Multiple sclerosis. *Lancet* 2002; **359**: 1221-1231.
- 225 2 Fernández O, Fernández V, Martínez-Cabrera V *et al*: Multiple sclerosis in
226 Gypsies from southern Spain: prevalence, mitochondrial DNA haplogroups and
227 HLA class II association. *Tissue Antigens* 2008; **71**: 426-433.
- 228 3 Alonso A, Jick SS, Olek MJ, Hernán MA: Incidence of multiple sclerosis in the
229 United Kingdom: findings from a population-based cohort. *J Neurol* 2007; **254**:
230 1736-1741.
- 231 4 Lyndsay JW: Familiar recurrence rates and genetic models of multiple sclerosis.
232 *Am J Med Genet* 2005; **135**: 53-58.
- 233 5 McElroy JP, Oksenberg JR: Multiple sclerosis genetics. *Curr Top Microbiol*
234 *Immunol* 2008; **318**: 45-72.
- 235 6 Lundmark F, Duvefelt K, Iacobaeus E *et al*: Variation in interleukin 7 receptor
236 alpha chain (IL7R) influences risk of multiple sclerosis. *Nat Genet* 2007; **39**:
237 1108-1113.
- 238 7 Gregory SG, Schmidt S, Seth P *et al*: Interleukin 7 receptor alpha chain (IL7R)
239 shows allelic and functional association with multiple sclerosis. *Nat Genet* 2007;
240 **39**: 1083-1091.
- 241 8 The International Multiple Sclerosis Genetics Consortium *et al*: Risk alleles for
242 multiple sclerosis identified by a genome wide study. *N Engl J Med* 2007; **357**:
243 851-862.
- 244 9 Kristjansdottir G, Sandling JK, Bonetti A *et al*: Interferon Regulatory Factor 5
245 (*IRF5*) Gene Variants are Associated with Multiple Sclerosis in Three Distinct
246 Populations. *J Med Genet* 2008; **45**: 362-369.

- 247 10 Matesanz F, Caro-Maldonado A, Fedetz M *et al*: IL2RA/CD25 polymorphisms
248 contribute to multiple sclerosis susceptibility. *J Neurol* 2007; **254**: 682-684.
- 249 11 Hoppenbrouwers IA, Aulchenko YS, Ebers GC *et al*: EVI5 is a risk gene for
250 multiple sclerosis. *Genes Immun* 2008; **9**: 334-337.
- 251 12 Poser CM, Paty DW, Scheinberg L *et al*: New diagnostic criteria for multiple
252 sclerosis: guidelines for research protocols. *Ann Neurol* 1983; **13**: 227-231.
- 253 13 Wigginton JE, Cutler DJ, Abecasis GR: A Note on Exact Tests of Hardy-
254 Weinberg Equilibrium. *Am J Hum Genet* 2005; **76**: 887-893
- 255 14 González JR, Carrasco JL, Dudbridge F, Armengol L, Estivill X, Moreno V:
256 Maximizing association statistics over genetic models. *Genet Epidemiol.* 2008;
257 **32**: 246-54.
- 258 15 González JR, Armengol L, Solé X *et al*: SNPAssoc: an R package to perform
259 whole genome association studies. *Bioinformatics.* 2007; **23**: 644-645.
- 260 16 Benjamini Y, Hochberg, Y: Controlling the false discovery rate: A practical and
261 powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B.* 1995; **57**: 289-300.
- 262 17 Sarkar SK: False discovery and false nondiscovery rates in single-step multiple
263 testing procedures. *The Annals of Statistics.* 2006; **34**:394-415.
- 264 18 Rubio JP, Stankovich J, Field J *et al*: Replication of KIAA0350, IL2RA, RPL5
265 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes Immun*
266 2008; **9**: 624-630.
- 267 19 Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene):
268 Genome-wide association study identifies new multiple sclerosis susceptibility
269 loci on chromosomes 12 and 20. *Nat Genet* 2009; **41**: 824-828.

270 20 Westlake CJ, Junutula JR, Simon GC *et al*: Identification of Rab11 as a small
271 GTPase binding protein for the Evi5 oncogene. *Proc Natl Acad Sci U S A* 2007;
272 **104**: 1236-1241.

273 21 Richardson PM, Zon LI: Molecular cloning of a cDNA with a novel domain
274 present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and
275 cdc16. *Oncogene* 1995; **11**: 1139-1148.

276 22 Faitar SL, Dabbekeh JT, Ranalli TA, Cowell JK: EVI5 is a novel centrosomal
277 protein that binds to alpha- and gamma-tubulin. *Genomics* 2005; **86**: 594-605.

278
279
280

281 **Table 1** Tag-SNP mapping of the *GFII-EVI5-RPL5-FAM69* locus

282

<i>SNP</i>	<i>Alleles</i>		<i>Genotypes</i>								<i>p-value</i>	<i>additive</i>	
	<i>I</i>	<i>2</i>	<i>Multiple sclerosis</i>			<i>Healthy controls</i>							
			<i>N</i>	<i>11</i>	<i>12</i>	<i>22</i>	<i>N</i>	<i>11</i>	<i>12</i>	<i>22</i>	<i>MAX-statistic</i> <i>(best genetic model)</i>	<i>Benjamini-Hochberg</i>	<i>OR [95% CI]</i>
1. rs11804321	<u>C</u>	T	681	0.05	0.36	0.59	871	0.03	0.3	0.66	0.008 (add)	0.034	1.29 (1.08,1.54)
2. rs6603984	A	G	724	0.04	0.32	0.64	911	0.04	0.31	0.65	0.751 (dom)	ns	1.00 (0.84,1.19)
3. rs17380378	C	T	698	0.01	0.18	0.8	881	0.01	0.18	0.81	0.952 (dom)	ns	1.03 (0.82,1.29)
4. rs11808092	<u>A</u>	C	726	0.1	0.44	0.45	889	0.08	0.41	0.51	0.048 (add)	ns	1.19 (1.03,1.39)
5. rs2391199	T	C	700	0.01	0.15	0.83	884	0.01	0.16	0.83	0.909 (rec)	ns	0.96 (0.76,1.22)
6. rs6680578	<u>T</u>	A	732	0.22	0.49	0.29	907	0.17	0.49	0.35	0.008 (add)	0.034	1.23 (1.07,1.41)
7. rs7514280	T	C	728	0.09	0.45	0.46	908	0.09	0.42	0.49	0.346 (rec)	ns	1.07 (0.92,1.25)
8. rs11164838	C	T	730	0.17	0.46	0.37	978	0.19	0.47	0.34	0.335 (add)	ns	0.91 (0.80,1.04)

283 Abbreviations: N. number of samples; NS. Nonsignificant; 95% CI, 95% confidence interval; OR, odds ratio; dom, dominant; rec, recessive; add, additive

284 Risk allele is bolt and underlined.

285

286

287 **Table 2** Regression analyses of 8 Tag SNPs in the *GFII-EVIS-RPL5-FAM69* region

288 genotyped in up to 732 affected individuals and 974 control samples.

289

SNP	A1	¹ Add locus to rs11804321		² Add rs11804321 to locus	
		P	OR (95% c.i.)	P	OR (95% c.i.)
rs11804321	C	NA	NA	NA	NA
rs6603984	A	0.7164	1.03 (0.85,1.23)	0.0092	1.28 (1.06,1.54)
rs17380378	C	0.4577	1.09 (0.86,1.38)	0.0048	1.30 (1.08,1.55)
rs11808092	A	0.8394	0.98 (0.78,1.22)	0.0208	1.35 (1.05,1.75)
rs2391199	T	0.9669	1.01 (0.79,1.28)	0.0044	1.30 (1.08,1.56)
rs6680578	T	0.3074	1.09 (0.92,1.29)	0.0306	1.26 (1.02,1.55)
rs7514280	T	0.6027	0.95 (0.79,1.14)	0.0064	1.34 (1.09,1.65)
rs11164838	C	0.2561	0.92 (0.80,1.06)	0.0055	1.29 (1.08,1.55)

290 Results for models assuming additive effects. OR, odds ratio, NA

291 1 denotes results adjusted for rs11804321.

292 2 denotes results adjusted for locus tested.

293

294

295

296

297

298

299

300 **Table 3** Tagged SNPs from the GWAS and location

301

302

303

304

305

306

307

308

SNP	IMSGC-GWAS					Gene localization
	SNPs $r^2 \geq 0.6$	Risk allele	MAF	Allelic P-value	OR 95% C.I.	
1. rs11804321						EVI5
2. rs6603984	rs6692187*		0.2	0.9		EVI5-GFI1
	rs4970712*		0.2	0.7		
	rs4970705*		0.2	0.9		
	rs1556562*		0.2	0.8		
3. rs17380378						EVI5-FAM69
4. rs11808092						EVI5
	rs12743520	A	0.28	0.000047	1.14 [1.07-1.21]	
	rs11800848	C	0.28	0.00021	1.12 [1.06-1.2]	
	rs1415296	C	0.29	0.00037	1.12 [1.05-1.19]	
	rs11578004*	C	0.28	0.0005	1.24 [1.09-1.4]	
5. rs2391199						EVI5-FAM69
	rs2107521*		0.09	0.1		
	rs4847280*		0.08	0.2		
	rs6604015*		0.09	0.1		
	rs2391164*		0.09	0.1		
	rs10747448*		0.09	0.1		
	rs2027060*		0.09	0.06		
	rs6659942*		0.1	0.2		
	rs7543964*		0.09	0.07		
6. rs6680578	rs6680578	T	0.38	0.0005	1.11 [1.04-1.17]	GFI1-EVI5-
	rs2811600	C	0.37	0.0022	1.09 [1.03-1.16]	RPL5-FAM69
	rs4847267	C	0.38	0.0036	1.09 [1.03-1.15]	
	rs17380908	T	0.36	0.00026	1.11 [1.05-1.18]	
	rs2255723	G	0.37	0.0094	1.08 [1.02-1.14]	
	rs10874746	T	0.35	0.0002	1.11 [1.05-1.18]	
	rs7536563	A	0.38	0.000009	1.12 [1.06-1.18]	
	rs11164607	A	0.38	0.011	1.08 [1.02-1.14]	
	rs9651257	C	0.37	0.00049	1.11 [1.04-1.17]	
	rs12745968	G	0.37	0.00065	1.1 [1.04-1.17]	
	rs10735781	G	0.38	0.00033	1.11 [1.05-1.18]	
	rs11164835	A	0.37	0.00097	1.10 [1.05-1.24]	
	rs4970700	A	0.37	0.018	1.07 [1.01-1.14]	
7. rs7514280	rs7514280	T	0.28	0.00007	1.13 [1.06-1.2]	FAM69
	rs6604026	C	0.29	0.000008	1.15 [1.08-1.22]	
8. rs11164838	rs11164838	C	0.43	0.00019	1.11 [1.05-1.18]	FAM69-EVI5

347

348 Abbreviations: CI. confidence interval; EVI5. ecotropic viral integration site 5; FAM69. family with sequence
 349 similarity 69; GFI1. growth factor independent 1; IMSGC-GWAS. International multiple sclerosis genetics
 350 consortium-genome wide association study; OR. odds ratio; RPL5. ribosomal protein L5; SNP single-
 351 nucleotide polymorphism.
 352 Data from IMSGC study⁸ and (supplementary material) combining the analysis of 1540 family trios. 2322
 353 cases and 5418 control subjects.

354 **Figure 1**

355 LD structure of the Tag-SNPs at the *GFII-EVI5-RPL5-FAM69A* locus and schematic
356 illustration of the genes in the locus with the positions in chromosome 1 (Ch1). Pair wise r^2
357 values are shown for Spanish population.

358

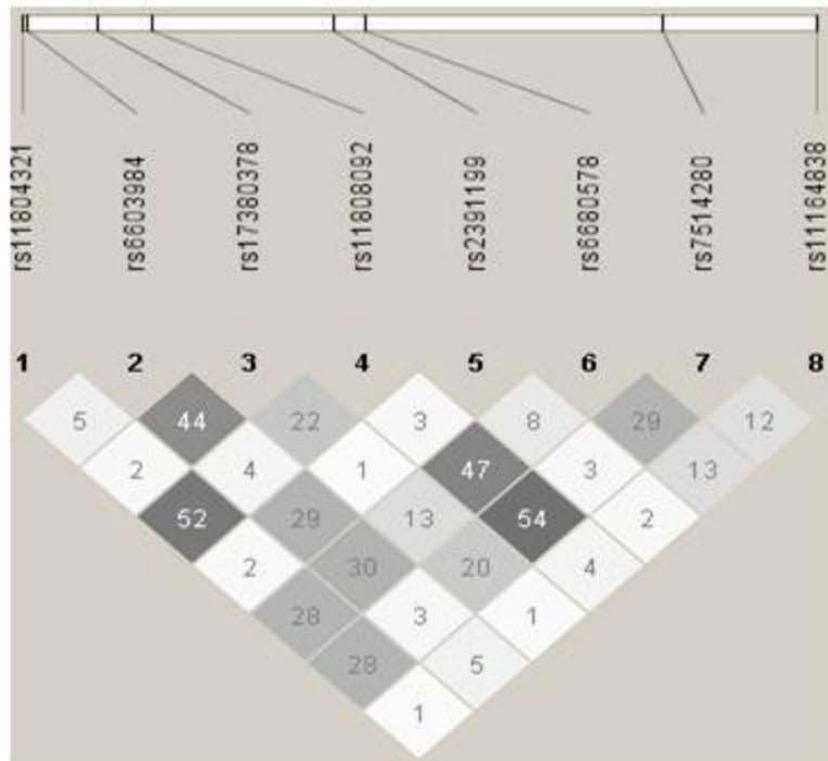
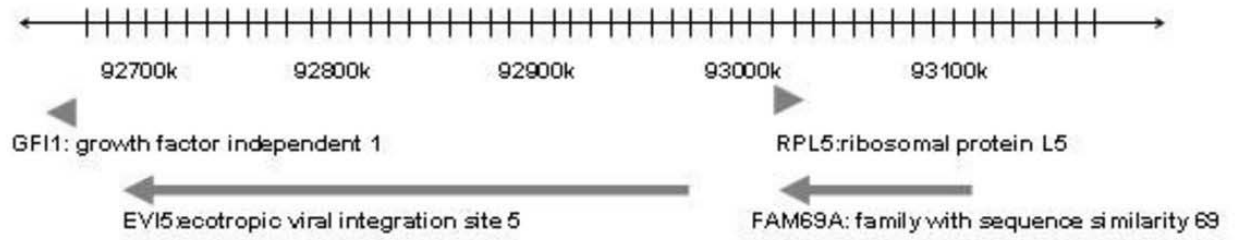
359 **Figure 2**

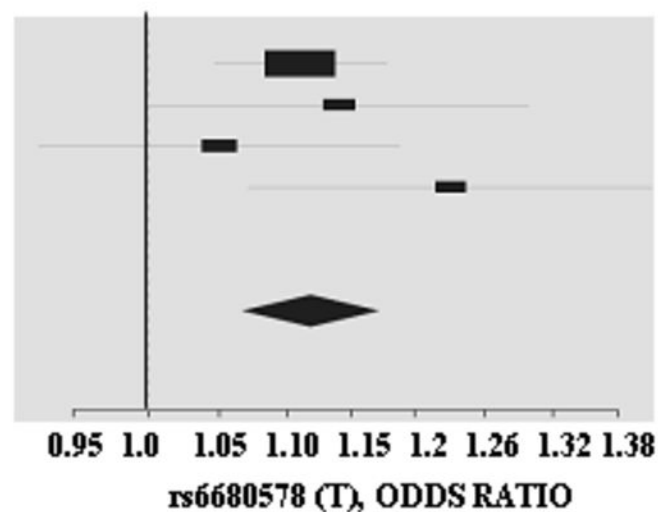
360 Summary effect estimate for the SNPs rs6680578 (upper panel) and rs11164838 (lower
361 panel) using information from 4 different studies: IMSGC⁸ (combined analysis of 1540
362 family trios and replicated in 2322 cases and 5418 controls), CANADIAN¹¹ (Canadian
363 Collaborative Project on the Genetic Susceptibility to MS performed on 756 families which
364 included 1318 individuals with definite MS and 1507 of their unaffected first-degree
365 relatives), AUSTRALIAN¹⁸ (cohort of 1134 cases and 1265 controls), SPANISH, this
366 work (732 patients with clinically defined MS and 974 healthy controls). Individual effect
367 estimates reported by the authors and their confidence intervals have been used. A random
368 effects model was used when heterogeneity was observed ([http://cran.r-](http://cran.r-project.org/web/packages/rmeta/index.html)
369 [project.org/web/packages/rmeta/index.html](http://cran.r-project.org/web/packages/rmeta/index.html)). The odds ratio (OR) estimate of each study is
370 marked with a square, which is also indicated in the figure. The size of the square
371 represents the weight that the corresponding study exerts in the meta-analysis. Estimated
372 heterogeneity variance for rs6680578 = 0 (P= 0.395), and for rs11164838: 0.0071 (P=
373 0.025).

374

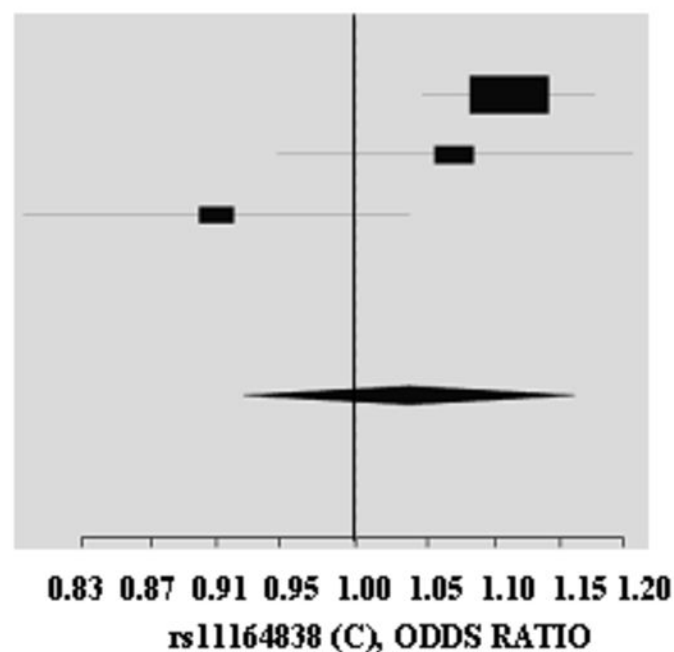
375

Ch1





IMSGC	1.11 [1.04-1.17]
CANADIAN	1.14 [1.01-1.31]
AUSTRALIAN	1.07 [0.95-1.21]
SPANISH	1.23 [1.07-1.41]
COMBINED	1.12 [1.07, 1.17]



IMSGC	1.11 [1.05-1.18]
AUSTRALIAN	1.07 [0.95-1.21]
SPANISH	0.91 [0.80-1.04]
COMBINED	1.04 [0.93-1.16]