
Antonio Alcina,1 Maria Fedetz,1 Óscar Fernández,2 Albert Saiz,3 Guillermo Izquierdo,4 Miguel Lucas,5 Laura Leyva,6 Juan-Antonio García-León,6 María del Mar Abad-Grau,7 Iraide Alloza,8 Alfredo Antigüedad,9 María J García-Barcina,10 Koen Vandenbroucke,8,11 Jezabel Varadé,12 Belén de la Hera,12 Rafael Arroyo,13 Manuel Comabella,14 Xavier Montalban,14 Natalia Petit-Martí,15 Arcadi Navarro,15,16,17 David Otaegui,18 Javier Olascoaga,19 Yolanda Blanco,3 Elena Urcelay,12 Fuencisla Matesanz1,4

ABSTRACT
Background and aim Several studies have highlighted the association of the 12q13.3–12q14.1 region with multiple sclerosis (MS); however, the causal variants underlying diseases are still unclear. The authors sought to identify the functional variant of this region associated with MS.

Methods Tag-single nucleotide polymorphism (SNP) analysis of the associated region encoding 15 genes was performed in 2876 MS patients and 2910 healthy Caucasian controls together with expression regulation analyses.

Results rs6581155, which tagged 18 variants within a region where 9 genes map, was sufficient to model the association. This SNP was in total linkage disequilibrium (LD) with other polymorphisms that associated with MS.

Conclusions This study determines a functional variant which alters the enhancer activity of a regulatory element in the locus affecting the expression of several genes and explains the association of the 12q13.3–12q14.1 region with MS.

INTRODUCTION
Multiple sclerosis (MS) is a chronic autoimmune disease with a complex pathogenesis in which demyelination and neurodegeneration are the main contributors to disability.1 Susceptibility to MS is thought to be conferred by a combination of genetic and environmental factors.2 The best characterised region implicated in predisposition to MS is the Major Histocompatibility Complex on chromosome 6p21; specifically the HLA-DRB1 *15:01 class II allele, but this accounts for less than 50% of MS heritability.1 New powerful methods as gene expression profiling and genome-wide association study (GWAS) have provided evidence for new susceptibility loci implicated in MS.10–12 These comprise over 40 loci including IL2RA, CBLB, IL7R, CLEC16A, TNFRSF1A, PTGER4 and IRF8.12–15 For many risk loci, the association signals do not directly implicate a single gene and the causative role for candidate genes in the region can only be speculated. A striking number of loci have demonstrated genome-wide evidence for association in multiple, distinct autoimmune disorders.10–11 One of these loci is 12q13–14 which has been associated in GWAS for rheumatoid arthritis (RA), coeliac disease (CD) and MS.16–18 However, different genes have been suggested in each study based on the main associated signal. A meta-analysis of two published GWAS totalling 3393 RA cases and 12 462 healthy controls identified an association at rs1678542 localised in the KIF5A intrinsic region.13 Another meta-analysis of two published GWAS on CD (4533 cases and 10 750 controls) and RA (5539 cases and 17 231 controls) described the association of both diseases at rs10876993 localised in the intergenic region between B4GALT1 and OS9 genes.14 Association at this locus was also described in a MS GWAS performed by the Australian and New Zealand Multiple Sclerosis Genetics Consortium in 1618 MS-cases. In this case, an associated single nucleotide polymorphism (SNP) (rs703842) was located at the 3’ untranslated region (3’ UTR) of the METTL1 gene.12 The last GWAS performed by the International Multiple Sclerosis Genetics Consortium with 10 000 MS patients also reported the association with this region at rs12368653 in theAGAP2 gene.8 In candidate gene studies, the KIF5A4 variant was demonstrated to be associated to MS15 and type 1 diabetes.16 Also rare variants in the CYP27B1 gene have been associated with MS.17 Other candidate-gene studies had

demonstrated association of variants located at the CYP27B1 gene with type 1 diabetes and MS.

The FAM119B gene has been pinpointed as the causative gene of the locus since its expression has been shown to be much lower in leukocytes of MS patients carrying the MS risk allele. On the other hand, since, in several studies, MS patients have lower levels of precursors of vitamin D in serum compared with controls, it has been suggested that CYP27B1, an enzyme implicated in Vitamin D activation, may be the causal gene.

In this work we performed a fine mapping of the 12q13.3–12q14.1 region by a Tag-SNP approach to identify the polymorphism leading the association with MS. We also performed functional studies to determine the involvement of the associated MS risk variants in gene expression.

MATERIAL AND METHODS
Study subjects and SNP genotyping

The entire data set studied consisted of 2876 MS patients diagnosed following Poser’s criteria and 2910 age and gender matched healthy controls, mostly blood donors and staff, matched by age, gender, ethnicity and place of recruitment. Patients and control subjects were included in the study based upon written informed consent. Demographic and clinical characteristics of the sample collections have been published before (a summary is in the online supplementary table S1). SNPs were genotyped using the iPLEX Sequenom MassARRAY platform in the Spanish National Genotyping Center’s facilities at Santiago de Compostela University (http://www.cegen.org). All DNAs were received in the Centro Nacional de Genotipado (CEGEN) centre for genotyping in 96-well plates; the concentration of double-stranded DNA was assessed using PicoGreen (Molecular Probes). As quality control, two trios of Center for Human Polymorphisms (CEPH) samples were genotyped (CEPH Na10830, Na10831, Na12147, Na10860, Na10861, Na11984). The genotypes of these samples corresponded with the ones deposited in HapMap with no detection of Mendelian inconsistencies. Additionally, a 10% of random samples were subjected to regenotyping. The resulting data were concordant with an average accuracy over 99.9%. We performed quality filtering of both samples and SNPs to ensure robust association testing.

Ethics statement

All participants provided a written informed consent to participate in this study, which was approved by the Institutional Review Board of Hospital Regional Universitario Carlos Haya of Malaga, Hospital Clinic of Barcelona, Hospital Virgen Macarena of Sevilla, Hospital de Basurto of Bilbao, Hospital Clínico San Carlos of Madrid, Hospital Universitari Vall d’Hebron of Barcelona and Hospital Donostia of San Sebastian.

Luciferase assays

Six fragments in both orientations with potential regulatory activity, as indicated by ENCODE, containing the tagged variants, 24 inserts in total, plus a plasmid without insert, were assayed for luciferase activity. Thus, fragments I–VI were PCR amplified (the coordinates and the primer used are indicated in online supplementary table S2) using DNA from homozygous individual for each allele of rs6581155. The fragments were previously cloned in Vaccinia DNA Topoisomerase I (TOPO) T/A vector (Invitrogen) and sequenced. Then, each clone was introduced in forward and reverse orientation in pGL4.25 (luc2CP/ mimP) vector which contains a minimal promoter upstream of the luciferase reporter gene luc2CP. Triplicates were transfected into human Raji cells (ATCC CCL86). The Raji cell line was maintained in Roswell Park Memorial Institute (RPMI) 1640 medium, supplemented with 10% fetal calf serum, 2.0 mM glutamine, 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 1 mM streptomycin, 100 U/ml penicillin, 0.05 mM 2-mercaptoethanol. Exponentially growing cells were transfected by nucleofection Amaxa Cell Line Nucleofector Kit V (Lonza) according to manufacturer protocols. Briefly, Raji cells (2×10⁶) were mixed with I–VI plasmid constructs or pGL4.25 control plasmid 2 μg of DNA and 0.2 μg of DNA of Renilla luciferase reporter gene plasmid pRL-TK(Promega) for normalising transfection efficiency. The transfected cells were cultured in 12 well-plates for 24 h. Cells were harvested and washed twice in Phosphate Buffered Saline (PBS) at 4°C. Luciferase activity was evaluated using Dual-Luciferase Assay System (Promega) according to manufacturers. For analysis, lysates were thawed on ice and the protein concentration was measured by the Bradford method. Then, 40 μg of protein from all supernatants were analysed for both firefly and Renilla luciferase activities in a luminometer F12 (Berthold Detection Systems). The following equation was used to determine the normalised fold changes in activity between tested groups:

\[
\text{Fold Activity} = \frac{(\text{Firefly}/\text{Renilla})_{\text{from construct}}}{(\text{Firefly}/\text{Renilla})_{\text{from empty plasmid}}}
\]

The normalised fold changes in activity from each of the three experiments are then averaged together.

Statistical analysis

Departure from Hardy-Weinberg equilibrium for all the biallelic SNP markers was tested using an exact test. Raw data were analysed for comparison of allele and genotype counts using PLINK V1.05. To avoid false-positive results due to multiple testing and considering that the SNPs analysed are not in complete disequilibrium, we applied the Benjamini-Hochberg method that is robust against positive dependence and controls the false discovery rate. In order to determine the effect and independence of variant association with MS, multiple logistic regression models were computed. Linkage disequilibrium (LD) patterns between SNPs were analysed with Haploview 4.2. Gender was analysed for potential differential risk contribution. We used the Breslow-Day Test to evaluate the homogeneity between gender cohorts. The age of onset with respect to the genotypes was analysed by the Kruskal-Wallis test.

RESULTS

A tag-SNP association study of the region

We performed the fine mapping of the 12q13.3–12q14.1 region between coordinates 56227468 to 56469005 (B36) (figure 1). To analyse this region, 16 SNPs were chosen by pairwise tagging from the HapMap B36 Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) population which captured 89 markers with r²≥0.8 (mean r²=0.93) and a minor-allele frequency ≥0.1. These tag-SNPs captured the variants associated with RA, MS and CD previously reported in different GWAS (see online supplementary table S3).

Tag-SNPs were genotyped in 2895 MS patients and 2942 healthy controls, mostly blood donors, matched by age, gender, ethnicity and place of recruitment. The genotypic distribution between cases and controls is shown in table 1. The genotype frequencies were fitted to an additive model. The top associated SNP was rs6581155 (p=4.57E−7; OR=0.79; 95% CI 0.73 to 0.87). We performed a logistic regression analysis to determine
the effect of further independent variants, by testing the addition of each SNP to rs6581155 and by adding this polymorphism to each one of the other Tag-SNPs (table 2). The results showed that, in addition to generate the strongest signal, rs6581155 alone was sufficient to model the association in the locus. Stratification by gender and age at onset and genotypes of rs6581155 SNP showed absence of statistical significance.

Moreover, we genotyped the rare variant rs118204009 located in the CYP27B gene that has been recently associated with MS in a subgroup of 1006 patients and 1044 healthy

Table 1 Association study of the different Tag-SNPs with multiple sclerosis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Alleles</th>
<th>Multiple sclerosis</th>
<th>Healthy controls</th>
<th>Per allele OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>rs1678542</td>
<td>C</td>
<td>G</td>
<td>248</td>
<td>(9)</td>
</tr>
<tr>
<td>rs2888334</td>
<td>C</td>
<td>G</td>
<td>134</td>
<td>(4.7)</td>
</tr>
<tr>
<td>rs775249</td>
<td>T</td>
<td>C</td>
<td>181</td>
<td>(6.4)</td>
</tr>
<tr>
<td>rs1678536</td>
<td>G</td>
<td>C</td>
<td>129</td>
<td>(4.5)</td>
</tr>
<tr>
<td>rs7313599</td>
<td>G</td>
<td>A</td>
<td>752</td>
<td>(26.4)</td>
</tr>
<tr>
<td>rs1078109</td>
<td>A</td>
<td>G</td>
<td>6</td>
<td>(0.2)</td>
</tr>
<tr>
<td>rs7305391</td>
<td>A</td>
<td>G</td>
<td>128</td>
<td>(4.5)</td>
</tr>
<tr>
<td>rs1564374</td>
<td>G</td>
<td>A</td>
<td>477</td>
<td>(17.2)</td>
</tr>
<tr>
<td>rs1172290</td>
<td>G</td>
<td>A</td>
<td>58</td>
<td>(2)</td>
</tr>
<tr>
<td>rs3825078</td>
<td>A</td>
<td>T</td>
<td>176</td>
<td>(6.2)</td>
</tr>
<tr>
<td>rs701006</td>
<td>A</td>
<td>G</td>
<td>222</td>
<td>(7.8)</td>
</tr>
<tr>
<td>rs2301551</td>
<td>G</td>
<td>C</td>
<td>76</td>
<td>(2.7)</td>
</tr>
<tr>
<td>rs1230784</td>
<td>C</td>
<td>T</td>
<td>26</td>
<td>(0.9)</td>
</tr>
<tr>
<td>rs1236853</td>
<td>G</td>
<td>A</td>
<td>430</td>
<td>(15.1)</td>
</tr>
<tr>
<td>rs1048691</td>
<td>T</td>
<td>C</td>
<td>135</td>
<td>(4.8)</td>
</tr>
<tr>
<td>rs6581155</td>
<td>G</td>
<td>A</td>
<td>110</td>
<td>(4)</td>
</tr>
</tbody>
</table>

Counts (percentages in brackets) are represented for each genotype.
controls.\(^\text{17}\) We did not detect this polymorphism in any individual of the MS or control groups (data not shown).

### Association of rs6581155-tagged SNPs with gene transcription levels

To determine if selected Tag-SNPs could have a functional effect on the transcription levels of the genes located in the 12q13.3–12q141 region, we first used expression quantitative trait loci (eQTL) data from Zeller et al.\(^\text{27}\) obtained from monocyte samples of 1490 German individuals (http://genecanvas.ecgene.net/news.php?readmore=45). There were 23 eQTLs for FAM119B, 6 for AVIL, 18 for XRCC6BP1, 18 for TSFM and 20 for TSPAN31 genes in the whole region. The variants best correlated with expression of FAM119B, AVIL, TSFM and TSPAN31 were also SNPs tagged by rs6581155 and, therefore, in high LD with it \((r^2>0.8)\) (table 3). The eQTLs for the XRCC6BP gene were in low LD with rs6581155 \((r^2=0.18)\) indicating that they were not involved in MS risk. Interestingly, the minor alleles for the rs6581155-tagged SNPs, which were protective for MS, correlated with high expression of FAM119B and AVIL but low expression of TSFM and TSPAN31 (see online supplementary figure S1) although there were large differences between the correlation coefficients obtained for each gene. FAM119B was the gene most strongly correlated with the rs6581155-tagged SNPs in monocytes.

The eQTL data from Dixon et al.\(^\text{28}\) obtained in lymphoblastic cell lines (LCL) from 206 families of British descent confirmed the eQTLs previously observed for FAM119B and TSFM, although the correlation coefficients were similar for both genes in contrast to the data from Zeller et al.\(^\text{27}\). The eQTLs for FAM119B were also reported in three additional data sets carried out in different cell types, as shown in table 3.

The data from Montgomery et al.\(^\text{29}\) obtained by RNA-Seq from LCL of 60 HapMap CEU individuals, have reported a rs6581155-tagged SNP to be eQTLs for FAM119B and CYP27B1, in which the minor allele inversely correlated with high expression of these genes (correlation coefficient \(r\) for CYP27B1 = –0.472, and for FAM119B = –0.474).

### Enhancer activity

Since all the rs6581155-tagged polymorphisms were located in intronic and intergenic non-coding regions, another strategy to search for functionality was to use annotations of gene regulatory regions from ENCODE (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=173506627&c=chr8&g=wgEncodeReg). Six out of 18 captured SNPs were located in remarkable regions associated with histone methylation, binding sites of transcription factors or in DNAse I hypersensitive site as shown in figure 2A.

To test their potential activities as enhancers or repressors of gene expression, we cloned the fragments containing these six variants upstream of the basic promoter of a luciferase reporter system and transfected into Raji lymphoblastic B cells.

We cloned four inserts for each of the six regions carrying both alleles in forward and reverse orientation. Only region III containing variant SNP rs10877013 had allele-dependent and orientation-dependent enhancer activity. Thus, allele A versus allele G in forward orientation showed 9.5-fold higher ratio of luciferase activity (figure 2B). We observed a 2.7 times higher enhancer activity of the T allele in the reverse fragment.

### DISCUSSION

The 12q13.3–12q14.1 region has been associated to several autoimmune diseases\(^\text{8}–\text{12,16}\) but no causal variant has been found so far for any of them exception made for a rare variant...
### Table 3: rs6581155 tagged SNPs that have been reported to be eQTLs in the region

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>1,490 German individuals</td>
<td>206 families of British descent</td>
<td>60 CEU individuals of the HapMap collection</td>
<td>60 individuals of the HapMap collection</td>
<td>80 individuals of Western European origin</td>
</tr>
<tr>
<td>Gene</td>
<td>AVIL</td>
<td>XRPCC6BP1</td>
<td>CYP27B1</td>
<td>FAM119B</td>
<td>FAM119B</td>
</tr>
<tr>
<td>Cell type</td>
<td>Monocytes</td>
<td>LCL</td>
<td>LCL</td>
<td>LCL</td>
<td>LCL</td>
</tr>
<tr>
<td>Probe/Transcript</td>
<td>ILMN_20148</td>
<td>ILMN_16915</td>
<td>ENST00000228606</td>
<td>GI_24308058-S</td>
<td>ILMN_1723846</td>
</tr>
<tr>
<td></td>
<td>ILMN_27946</td>
<td>ILMN_12426</td>
<td>ENST0000030209</td>
<td>ILMN_1723846</td>
<td>ILMN_1723846</td>
</tr>
<tr>
<td></td>
<td>ILMN_11441</td>
<td>p Value/Effect</td>
<td>p Value/Effect</td>
<td>p Value</td>
<td>p Value</td>
</tr>
<tr>
<td>rs6581155</td>
<td></td>
<td></td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10774783</td>
<td>5.00E−21/0.731</td>
<td>9.00E−130/5.64</td>
<td></td>
<td>7.5E−11/0.58</td>
<td></td>
</tr>
<tr>
<td>rs923829</td>
<td>5.00E−21/0.731</td>
<td>9.00E−130/5.64</td>
<td></td>
<td>6E−11/0.58</td>
<td></td>
</tr>
<tr>
<td>rs10431552</td>
<td>2.80E−21/0.745</td>
<td>1.00E−120/5.69</td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs2072052</td>
<td>7.70E−19/0.689</td>
<td>1.00E−130/5.88</td>
<td></td>
<td>5E−10/0.56</td>
<td></td>
</tr>
<tr>
<td>rs2291617</td>
<td>4.10E−11/0.62</td>
<td>7.60E−110/6.26</td>
<td></td>
<td>6E−11/0.58</td>
<td></td>
</tr>
<tr>
<td>rs4646536</td>
<td>1.7E−07/0.021</td>
<td>5E−380.11</td>
<td>7.1E−190.055</td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs274834</td>
<td>5.00E−21/0.731</td>
<td>9.00E−130/5.64</td>
<td></td>
<td>7.5E−11/0.58</td>
<td></td>
</tr>
<tr>
<td>rs2069502</td>
<td>2.80E−21/0.745</td>
<td>1.00E−120/5.69</td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10877013</td>
<td>7.7E−19/0.689</td>
<td>1.40E−130/5.83</td>
<td></td>
<td>5E−10/0.56</td>
<td></td>
</tr>
<tr>
<td>rs10877011</td>
<td>8.80E−20/0.711</td>
<td>1.80E−120/5.69</td>
<td></td>
<td>6E−11/0.58</td>
<td></td>
</tr>
<tr>
<td>rs1070008</td>
<td>7.7E−19/0.689</td>
<td>1.40E−130/5.83</td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs11172335</td>
<td>8.90E−14/0.612</td>
<td>9.00E−130/6.</td>
<td>0.0001457E−0.472</td>
<td>4.1E−10/0.56</td>
<td>0.000001/−0.571</td>
</tr>
<tr>
<td>rs11172333</td>
<td>2.30E−12/0.563</td>
<td>2.30E−110/5.5</td>
<td></td>
<td>9.5E−11/0.57</td>
<td></td>
</tr>
<tr>
<td>rs11172335</td>
<td>1.6E−8/−0.51</td>
<td></td>
<td></td>
<td>1.6E−8/−0.51</td>
<td></td>
</tr>
<tr>
<td>rs8181644</td>
<td>1.2E−07/0.022</td>
<td>3.8E−380.11</td>
<td>1.4E−190.057</td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10877019</td>
<td>4.2E−07/0.023</td>
<td>6.3E−380.11</td>
<td>1.02E−190.058</td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10431552</td>
<td>2.42E−370.11</td>
<td>1.42E−180.054</td>
<td>7.5E−235/0.52</td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10877015</td>
<td>7.80E−17/0.694</td>
<td>2.50E−110.568</td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs11172333</td>
<td>1.1E−07/0.023</td>
<td>7.3E−380.11</td>
<td>1.74E−190.057</td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
<tr>
<td>rs10877019</td>
<td>2.07E−23/−0.51</td>
<td>5.63E−120.035</td>
<td></td>
<td>2.1E−10/0.57</td>
<td></td>
</tr>
</tbody>
</table>

Zeller et al Genome-Wide Expression analysis performed with Illumina HT-12 v3 BeadChip. Association significance was obtained by ANalysis Of VAriance (ANOVA).

Dixon et al Genome-Wide Expression analysis performed by U133 Plus 2.0 GeneChips (Affymetrix). Association analysis was obtained by fitting a simple regression model to each trait and used a variance component approach to account for correlation between different observed phenotypes within each family.

Montgomery et al Genome-Wide Expression analysis performed by RNA sequencing in an Illumina GAII. Association analysis and multiple testing corrections were conducted by spearman rank correlation and permuting the expression phenotype 10 000 times and summarising the extreme p value distribution for each particular exon, transcript and probe.

Stranger et al Genome-Wide Expression analysis performed with Sentrix Human-6 Expression BeadChip V1.1, Illumina. Association significance was obtained by spearman rank correlation with 10 000 permutations of the expression phenotypes.

Dimas et al Gene expression was performed by commercial whole-genome expression array (WG-6 V3 Expression BeadChip, Illumina). Association analysis and multiple testing corrections were conducted by spearman rank correlation. Thresholds for each gene were assigned after 10 000 permutations of expression values relative to genotypes where the haplotypic (LD) structure was maintained in each interaction.

eQTL, expression quantitative trait loci; LCL, lymphoblastic cell lines.
in the coding region of CYP27B1 gene associated with MS.17

In the present work, by integrating several strategies, we have
uncovered the variant most strongly associated with MS in the
region. This variant also affects the activity of an enhancer in an
allele-dependent and orientation-dependent manner and corre-
lates with the expression of five genes of the locus. Interestingly,
the major allele of this SNP (rs10877013) correlates with high
luciferase activity and at the same time with high expression of
TSFM and TSPAN31 and low expression of FAM119B,
CYP27B1 and AVIL. Although it could seem contradictory that
one enhancer may differentially affect the expression of differ-
ent genes, parallel situations have been described extensively in
two scenarios: first, the same transcription factor is capable of
activating or repressing depending on the context;32 33 second,
a single DNA sequence can be bound by different factors with
opposite activities.34 On the other hand, we cannot forget the
important role of H3K4me preventing aberrant gene expression
or modulating transcriptional response. H3K4me turns to mark
for both positive and negative regulation. The role of H3K4me2
in histones deacetylation could serve as a fine-tuning regulatory
mechanism.35

It has been described that interacting promoters are coex-
pressed in a tissue-specific manner. Based on the Chromatin
Interaction Analysis with Paired-End-Tag sequencing results
mapping long-range chromatin interactions associated with
RNAPII and H3K4me2,36 we observed that the genes in the
analysed region are engaged through promoter–promoter, pro-
moter–enhancer interactions forming a multigene complex
that could provide a structural framework for cotranscription
(figure 3). In this sense, the effect of a polymorphism on one
enhancer embedded in this multigene complex could have diver-
ging effects depending on the interaction of this enhancer with
the different promoters. Additionally, the interactions could be
tissue-type or cell-type specific, as has been shown by Li
et al.36 Therefore, it is not evident which of the genes in the region
may be responsible for the genetic association with MS, since
they seem to be coordinately regulated, and the enhancer con-
taining the MS-associated variant may be affecting either of
them; in fact, it acts as an eQTL for all of them. We observed
that the Spearman’s correlation coefficients (r) between the
expression levels of the different genes in this cluster and the
rs6581155-tagged SNPs were different in the different assays
performed. For instance, in monocytes the best correlation is
observed with FAM119B (r = −0.52), however in data obtained
from LCL by RNA-seq a medium-high correlation is found for both
CYP27B1 (r = 0.472) and FAM119B (r = −0.474), and with microarrays the correlation coefficient is

Figure 2
Enhancer-suppressor activity of the rs6581155-tagged variants. (A) Schematic illustration to show the localisation of the six tagged SNPs
(in red) in potential regulatory regions (in black, I–VI) as indicated by ENCODE for the GM12878 lymphoblastoid cell line and the genes present in
the region: Enhancer H3K4Me1 track shows where modification of histone proteins is suggestive of enhancer; Promoter H3K4Me3 track shows a
histone mark associated with promoters; Layered H3K4Me3 track shows histone mark associated with promoters that are active or poised to be
activated; DNase Clusters track shows regions where chromatin is hypersensitive to DNase I enzyme; Tfn Factor ChIP track shows DNA regions
where transcription factors bind to DNA as assayed by chromatin immunoprecipitation (ChIP) with antibodies specific to the transcription factor
followed by sequencing of the precipitated DNA (ChIP-Seq). (B) Δ Fold luciferase activity respect to the empty vector of the different constructs
containing the rs6581155 tagged polymorphisms transfected into Raji B
cells. Four clones for each region bearing the different alleles (allele m, minor and M, major) and in both orientations (Forward and Reverse) from
three independent transfection experiments are represented. Luciferase activity levels are referred to the level of the control plasmid containing only
the basic promoter and the Renilla activity.

Very similar for TSFM and FAM119B. Based on data from LCL cells, TSFM, FAM119B or CYP27B1 would be candidates, while FAM119B appears to emerge from monocytes. In fact, Gandhi et al reported that FAM119B expression level was much lower in MS patients carrying the susceptibility haplotype. We observed that this correlation occurs also in samples from healthy donors as they are the CEU HapMap samples. However, we do not know the precise conditions (cell type, stage of disease, etc) under which expression of these genes could be determinant for the development of MS. For the same reason the variant-specific enhancer activity of fragment III, reported in this work, could vary in different cell-types. Here we have used in the activity assays the same cell-type used in the eQTL analysis to reproduce the effect observed in the expression of these genes. The Raji cells are lymphoblast-like cells derived from B-lymphocyte as the LCL from HapMap. The effect of the enhancer on the gene expression of proximal tubule cells of the kidney and the disease-activated macrophages, which are the major source of CYP27B1, could shed light on the involvement of this variant in MS.

Relying on published data from other groups on the concentration of precursors of vitamin D in serum, two independent groups reported a correlation between the major alleles of rs703842, tagged by rs6581155, and low 1,25(OH)2D serum levels. The major allele of the MS-associated SNPs, in total LD with the previously mentioned SNPs, correlated with low expression of the CYP27B1 gene. This enzyme catalyses the conversion of 25(OH)D to 1,25(OH)2D, thus a low expression of this enzyme, could reconcile the above seemingly conflicting data, via accumulation of 25(OH)D and a low 1,25(OH)2D product. These data support the CYP27B1 gene as a very plausible candidate gene for association of the region with MS.

Vitamin D, besides having well-known control functions of calcium and phosphorus metabolism, bone formation and mineralisation, also has a complex role in the maintenance of immune homeostasis. The administration of vitamin D in animal models leads to improvement of immune-mediated symptoms. The correlation between the MS-associated variants and circulating levels of vitamin D supports the important role of vitamin D in susceptibility to MS. Thus, if this is the case in human MS, both the pharmacogenomics and the expression analysis of all the genes affecting the causing factor can reveal the kind of intervention necessary for neutralising a pathogenic situation.

In this work we localise a group of variants in almost total LD that explain the association of the 12q13.3–12q14.1 region with MS. One of these variants (rs10877013) strongly affects the activity of one enhancer in the region, in an allele-dependent and orientation-dependent way, that could be the cause of the alteration of TSFM, TSPAN31, FAM119B, and...
CYP27B1 and AVIL gene expression, due to the promoter multigene interactions observed in the region. Some of the tagged variants have been described to be associated with the vitamin D serum concentration. Our results point to one of these as the casual gene for MS association in the locus.

Acknowledgements We thank patients with multiple sclerosis and control subjects for making this study feasible. SNP genotyping services were provided by the Spanish Centro Nacional de Genotipado CEGEN-USC (http://www.cegen.org).

Contributors AA, EU, KV and FM designed the study, performed the main statistical analyses and co-wrote the manuscript. MF carried part of the experiments. OA, AS, GI, ML, LL, JAGL, IA, AA, MIGOB, JF, DH, BA, MC, XM, MD, JO, YB coordinated sample collection. MAA, NPM, AN carried out some bioinformatics analysis.

Funding Financial support for the study was provided by Ministerio de Economía y Competitividad (MINECO)-Fondos Europeos de Desarrollo Regional (FEDER) (grant numbers SAF2011-11491; Fondo de Investigación Sanitaria (FIS) (grant numbers RETICS-REEM RD07/0060, CP 10/00526, P10/1985, P09/02105); Junta de Andalucía- Fondos Europeos de Desarrollo Regional (FEDER) (grant numbers P07-CVI-02551, PO09-CTS-5218); Ikerbasque; the Basque Foundation for Science (Bilbao) and Fundación Ilundain.

Competing interests None.

Patient consent Obtained.

Ethics approval This study was approved by the Institutional Review Boards of Hospital Regional Universitario Carlos Haya of Malaga, Hospital Clinic of Barcelona, Hospital Virgen Macarena of Sevilla, Hospital Basurto of Bilbao, Hospital Clínico San Carlos of Madrid, Hospital Universitari Vall d’Hebron of Barcelona and Hospital Donostia of San Sebastian.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data presented in the manuscript are available on request.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/.

REFERENCES


Antonio Alcina, Maria Fedetz, Óscar Fernández, Albert Saiz, Guillermo Izquierdo, Miguel Lucas, Laura Leyva, Juan-Antonio García-León, María del Mar Abad-Grau, Iraide Alloza, Alfredo Antiguédad, María J García-Barcina, Koen Vandenbroeck, Jezabel Varadé, Belén de la Hera, Rafael Arroyo, Manuel Comabella, Xavier Montalban, Natalia Petit-Marty, Arcadi Navarro, David Otaegui, Javier Olascoaga, Yolanda Blanco, Elena Urcelay and Fuencisla Matesanz

J Med Genet 2013 50: 25-33 originally published online November 17, 2012
doi: 10.1136/jmedgenet-2012-101085
Open access (124)
Diabetes (101)
Immunology (including allergy) (557)
Multiple sclerosis (18)
Genetic screening / counselling (823)
Coeliac disease (11)
Connective tissue disease (90)
Metabolic disorders (307)
Rheumatoid arthritis (13)
Clinical genetics (240)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/