This is a pre-copyedited, author-produced version of an article accepted for publication in **Human Molecular Genetics** following peer review. The version of record Fuencisla Matesanz, Victor Potenciano, Maria Fedetz, Priscila Ramos-Mozo, María del Mar Abad-Grau, Mohamad Karaky, Cristina Barrionuevo, Guillermo Izquierdo, Juan Luis Ruiz-Peña, María Isabel García-Sánchez, Miguel Lucas, Óscar Fernández, Laura Leyva, David Otaegui, Maider Muñoz-Culla, Javier Olascoaga, Koen Vandenbroeck, Iraide Alloza, Ianire Astobiza, Alfredo Antigüedad, Luisa María Villar, José Carlos Álvarez-Cermeño, Sunny Malhotra, Manuel Comabella, Xavier Montalban, Albert Saiz, Yolanda Blanco, Rafael Arroyo, Jezabel Varadé, Elena Urcelay, Antonio Alcina, *A functional variant that affects exon-skipping and protein expression of SP140 as genetic mechanism predisposing to multiple sclerosis*, H**uman Molecular Genetics**, Volume 24, Issue 19, 1 October 2015, Pages 5619–5627, is available online at: https://doi.org/10.1093/hmg/ddv256

A functional variant that affects exon-skipping and protein expression of *SP140* as genetic mechanism predisposing to multiple sclerosis

Fuencisla Matesanz[1,*,†], Victor Potenciano[1,2,†], Maria Fedetz[1,†], Priscila Ramos-Mozo[3,†], María del Mar Abad-Grau[2], Mohamad Karaky[1], Cristina Barrionuevo[1], Guillermo Izquierdo[4], Juan Luis Ruiz-Peña[4], María Isabel García-Sánchez[4], Miguel Lucas[5], Óscar Fernández[6], Laura Leyva[6], David Otaegui[7], Maider Muñoz[7], Javier Olascoaga[7], Koen Vandenbroeck[8,9,10], Iraide Alloza[8,9,10], Ianire Astobiza[8,9], Alfredo Antigüedad[11], Luisa-María Villar[12], José Carlos Álvarez-Cermeño[12], Sunny Malhotra[13], Manuel Comabella[13], Xavier Montalban[13], Albert Saiz[14], Yolanda Blanco[14], Rafael Arroyo[15], Jezabel Varadé[3], Elena Urcelay[3,†], Antonio Alcina[1,*,†]

[1]Department of Cell Biology and Immunology, Instituto de Parasitología y Biomedicina López Neyra (IPBLN), CSIC, Granada, Spain
[2]Department of Computer Languages and Systems-CITIC, Universidad de Granada, Granada, Spain
[3]Dept. of Immunology, Hospital Clínico San Carlos. Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain
[4]Unidad de Esclerosis Múltiple, Hospital Universitario Virgen Macarena, Sevilla, Spain
[5]Servicio de Biología Molecular, Hospital Virgen Macarena, Facultad de Medicina, Sevilla, Spain
[6]Unidad de Gestión Clínica de Neurociencias. Instituto de Biomedicina de Málaga (IBIMA). Hospital Regional Universitario de Málaga. Málaga, Spain
[7]Área de Neurociencias, Inst. Investigación Sanitaria Biodonostia, San Sebastián. Spain
[8]Neurogenomiks Group, Universidad del País Vasco (UPV/EHU), Leioa. Spain
[9]Achucarro Basque Center for Neuroscience, Zamudio, Spain
[10]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[11]Servicio de Neurología, Hospital de Basurto, Bilbao. Spain
[12]Hospital Ramon y Cajal, Departments of Immunology and Neurology. MS Unit. (IRYCIS), Madrid, Spain
[13]Servei de Neurologia-Neuroimmunologia, Centre d'Esclerosi Múltiple de Catalunya, Cemcat, Hospital Universitari Vall d´Hebron, Barcelona. Spain
[14]Neurology Service, Hospital Clinic and I. d´Investigació Biomèdica Pi iSunyer (IDIBAPS), Barcelona, Spain
[15]Multiple Sclerosis Unit, Hospital Clínico San Carlos. Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

[*]Corresponding authors: Dr. Fuencisla Matesanz. E-mail: lindo@ipb.csic.es. *Dr. Antonio Alcina. E-mail: pulgoso@ipb.csic.es, Instituto de Parasitología y Biomedicina López Neyra (IPBLN), Avda. Conocimiento SN, 18016 Granada, Spain, Tel.: +34 958 181668; Fax: +34 958 181632

[†]Contributed equally to this work

Abstract

Several variants in strong linkage disequilibrium (LD) at the *SP140* locus have been associated with multiple sclerosis (MS), Crohn's disease (CD) and chronic lymphocytic leukemia (CLL). To determine the causal polymorphism, we have integrated high density datasets of expression quantitative trait loci (eQTL), using GEUVADIS RNA sequences and 1000 Genomes genotypes, with MS-risk variants of the high density Immunochip array performed by the International Multiple Sclerosis Genetic Consortium (IMSGC). The variants most associated with MS were also correlated with a decreased expression of the full-length RNA isoform of *SP140* and an increase of an isoform lacking exon 7. By exon splicing assay, we have demonstrated that the rs28445040 variant was the causal factor for skipping of exon 7. Western blots of peripheral blood mononuclear cells from MS patients showed a significant allele-dependent reduction of the SP140 protein expression. To confirm the association of this functional variant with MS and to compare it with the best associated variant previously reported by GWAS (rs10201872), a case-control study including 4384 MS patients and 3197 controls was performed. Both variants, in strong LD (r^2=0.93), were found similarly associated with MS (P-values, odds ratios: 1.9 E-9, OR=1.35 [1.22-1.49] and 4.9 E-10, OR=1.37 [1.24-1.51], respectively). In conclusion, our data uncover the causal variant for the *SP140* locus and the molecular mechanism associated with MS risk. In addition, this study and others previously reported strongly suggest that this functional variant may be shared with other immune-mediated diseases as CD and CLL.

Introduction

The *SP140* locus has been implicated by Genome Wide Association Studies (GWAS) as risk factor in different autoimmune diseases, such as multiple sclerosis (MS) (1) and Crohn's disease (CD) (2, 3). Moreover, *SP140* has been associated with susceptibility to chronic lymphocytic leukemia (CLL) (4-7), reported as an autoantigen in primary biliary cirrhosis (8), and recurrent truncations of the gene have been observed in multiple myeloma (9).

The SP140 function is only partially known; however, its location in the nuclear bodies and its primary structure presenting several chromatin related modules suggest a role in chromatin-mediated regulation of gene expression (10). SP140 presents strong sequence homology with the autoimmune regulator (AIRE), a transcriptional activator governing the ectopic expression of peripheral tissue-specific antigens in the thymus (11). Similarly to AIRE, SP140 harbors a nuclear localization signal, a dimerization domain (HSR or CARD domain), a SAND domain, and a plant homeodomain (PHD) finger. In addition, SP140 has a bromodomain (BRD), which is frequently fused to PHD and both function in concert to read multivalent histone marks (12).

A study of expression quantitative trait loci (eQTL) in lymphoblastoid cell lines (LCL) has reported altered gene expression of *SP140* by cis-acting mechanisms (13). The obvious question is whether the association signals for the different diseases are consequence of the eQTLs localized in the same locus. If an altered gene expression is the driving force for disease susceptibility, then a genetic variant that affects gene expression should also explain or correlate precisely with disease association. In that case, provided that sample size and marker density are sufficient to determine the causal variant, both disease association and eQTL peaks should coincide (14). The best effort

toward the fine-mapping of the GWAS-loci associated with immune-related diseases came from the Immunochip Project (15). This work, based on the idea of shared etiopathogenic factors amongst 12 immune-related disorders, was designed for dense genotyping of 186 loci identified through previous GWAS. The variants reported from GWAS of MS, CD, and CLL in the *SP140* locus are different for each disease, albeit all of them located within the *SP140* gene.

To test whether variants affecting gene expression in the region are the causal origin of the association with MS and other *SP140*-associated diseases, we identified the eQTLs described in lymphoblastoid cell lines (LCLs) from 344 European out of the 1000 Genomes Project (16) by screening the RNA-Seq data registered by the GEUVADIS Project (17) and integrated this information with high density MS-associated dataset from the Immunochip Project (18) and other GWAS. Then, we identified the genetic mechanisms by which these variants affected the RNA-isoform expression levels and the protein levels of SP140 in blood cells from MS patients.

Results

Determination of eQTLs in the *SP140* locus

In order to determine whether the polymorphisms associated with MS, CD, and CLL in the *SP140* locus could be altering the expression of the genes in the region, we first identified the eQTLs in LCLs from European origin from the GEUVADIS Project (17). We processed the FASTQ files of mRNA sequences from four populations (Centre d' Etude du Polymorphisme Humain (CEPH) (CEU), Finns (FIN), British (GBR), Toscani (TSI)). A schematic explanation of the RNA-Seq analysis pipeline is depicted in

Supplementary Material, Fig. S1. The SNP genotypes from 344 LCLs were available from the 1000 Genomes Project phase 1 release v3, which were selected for the eQTL calculations.

Transcript quantification was performed by using Cufflinks and Ensembl gene annotation v.65, obtaining a total of 20,910 transcripts with levels of expression equal or higher than 3.0 fragments per kilobase of exon per million fragments mapped (FPKM) and present in more that 5% of the individuals. We mapped *cis*-eQTLs in 1MB fragments by Spearman correlation tests. Through this eQTL calculation process, the best correlated variant (best-eQTL) for each transcript was obtained, resulting in 1096 best-eQTLs with *P* values < 0.001.

The selection of the eQTLs at coordinates chr2:230856224-231357223, 250 Mb flanking each side of the SNP associated with MS, unrevealed 4 eQTLs, one associated with a transcript of the *SP100* gene and three with transcripts of the *SP140* gene (Table 1). The SNPs that best correlated with each of the four transcripts (best-eQTLs) were different ones, but those for the transcripts ENST00000343805 and ENST00000392045 of the *SP140* gene were in strong LD (r^2 =0.99) in the European population.

Colocalization of best-GWAS variants and best-eQTLs

To determine whether the GWAS-variants in the *SP140* locus colocalized with the eQTLs, we calculated the linkage disequilibrium (LD) between the best-eQTLs and the best-associated SNP for each disease (Table 1). The best-eQTLs for the *SP140* transcripts ENST00000343805 and ENST00000392045 were in almost total LD with the best GWAS-variants of MS, CD, and CLL (Fig. 1A and Table 1). These variants belong to a LD block including 18 variants with r^2 ranging between 0.94 and 1, as calculated from EUR populations of the 1000 Genomes Project. To verify the

colocalization between eQTLs and association signals, we examined data of the Immunochip project for MS (18). This dataset provides high genotyping density for this region in a large cohort of 14277 cases and 23605 controls. Given that both eQTLs and Immunochip have the 1000 Genomes as base of design, the data of MS association and the eQTLs were available for the same SNPs. Thus, we integrated both signals to determine whether they shared the causative variant (Fig. 1). Complete colocalization was observed between the best MS-associated SNPs and the best-eQTLs for the transcripts ENST00000343805 and ENST00000392045 of the *SP140* gene (Fig. 1B). However, there was no colocalization with the best-eQTLs for the other *SP140* transcript ENST00000350136 and the *SP100* transcripts. In the *locuszoom* graphs (Fig. 1C), we observed that the best associated variant in the Immunochip data for MS, rs9989735, is in total LD with a group of SNPs that are those top correlated with the transcription levels of ENST00000392045 and ENST00000343805, but not with transcript ENST00000350136.

To further support the colocalization results, we applied the Probabilistic Identification of Causal SNPs (PICS) described by Farh et al. (19) for each eQTL in the region, as well as for the SNPs from Immunochip data. We observed that the same SNPs from MS association and eQTLs for transcripts ENST00000392045 and ENST00000343805 had the best PICS score (Supplementary Material, Table S1).

Changes in the RNA isoform profile associated with disease

The expression levels of the two *SP140* RNA isoforms showed opposite correlations with the genotypes of the best MS-associated variant (Fig. 2A). The main difference between these two RNA isoforms was the alternative splicing of exon 7. In the LD block, rs28445040 was located in exon 7, at 5 bases downstream of the splicing

acceptor site. To explore whether the skipping of exon 7 is the functional cause of the association, we turned to eQTL exon-level analysis of the *SP140* for the European (EUR, n= 373) and African-Yoruba (YRI, n= 89) populations of the GEUVADIS Project (17). For YRI population, we observed that the best eQTL for *SP140* exon 7 was rs28445040 with a PICS score of 0.6429, much higher than the next one, rs13426106, with a PICS score of 0.0715 (Supplementary Material, Table S1). For the EUR population the results indicated that rs28445040 was an eQTL for all *SP140* exons, except for exons 24 and 25, albeit a significantly higher correlation coefficient was observed for exon 7 (Fig. 2B). It seemed that the reduction of the expression levels of the ENST00000392045 transcript was compensated by the increase in the ENST00000343805 isoform except for exon 7, not present in the latter.

To confirm these data experimentally, we analyzed *SP140* RNA levels by a reverse transcriptase (RT)-PCR in LCLs and PBMCs from individuals carrying the different genotypes of rs28445040 (Fig. 2C), using primers that hybridized in the flanking exon 6 and exon 8 and visualized by acrylamide-gel electrophoresis. The samples from TT and TC carries showed a band corresponding in size to a fragment lacking exon 7, and with a T-allele dose effect that was absent in the CC carriers. In order to quantify the expression levels of the two spliced variants and to validate the data obtained from the RNA-Seq from GEUVADIS, we performed real time qPCR in RNA from 59 LCLs (32 CC, 22 CT and 5 TT for rs28445040) using a bridge primer that hybridized between exons 6 and 8 for the transcript with skipped-exon 7 and another qPCR with a bridge primer between exon 7 and 8 for the full-length transcript, as shown in Fig. 2D. After Sperman's rho correlation test of expression levels respect to the rs28445040 genotypes, we observed that the expression of the exon 7-skipped transcript was highly correlated

with the T-allele dose while the full-length transcript, containing exon 7, was inversely correlated (Fig. 2D).

rs28445040 as a functional variant affecting the exon 7-skipped RNA isoform

To confirm that rs28445040 is the causal variant of the splicing alteration observed in the *SP140* transcript profile, we used an alternative splicing strategy by cloning the exon 7 and its flanking sequences carrying the two alleles into the pSPL3 plasmid (Fig. 3A). After transfection in HEK cells, RNA purification, RT-PCR amplification, analysis of the RNA products by agarose-gel electrophoresis and sequencing, we determined that the exon 7- T allele was spliced in about 60% of the molecules and the C allele was spliced in < 10% of the molecules (Fig. 3B). These data were in agreement with the results shown in Fig. 2C, confirming that the rs28445040*T allele produced splicing alterations by exon 7-skipping.

Analysis of SP140 protein in blood cells

To evaluate how these allele-dependent RNA isoforms translated into protein expression, semi quantitative immunoblots of genotyped PBMCs from MS patients (n=28 [8 CC, 10 CT, 10 TT]) and controls (n=25 [10 CC, 8 CT, 7 TT]) were performed (Fig. 4). Using a polyclonal antibody against SP140 protein, we observed a unique band of the expected size for the translation of the full-length transcript isoform (estimated size 98 kDa, apparent size in SDS-PAGE < 92 kDa), and there was a significant allele-dependent reduction of expression levels in the TT-carrier samples (P-value=0.007) (Fig. 4A). Whether the antibody used did not react with the possible short-size band corresponding with the exon 7-deleted SP140 isoform (estimated size-difference with

the full-length isoform approx. 4 kDa) or this RNA isoform or the corresponding product were not stable and degraded is unknown. Quantification of SP140 protein levels and comparison between MS and controls samples with similar proportion of each allele did not show significant differences (Fig. 4B).

Confirmation of the genetic association by a case-control study

We focused our attention on rs28445040 as the causal variant due to its location in the *SP140* exon 7 and its high LD with the variant best associated with MS risk by GWAS (Supplementary Material, Table S2). Initially, the LD between these two SNPs was calculated from the CEU population of the 1000 Genomes Project (r^2=0.96); however, when the LD was analyzed in other populations we observed that holds great variability. In African populations the LD between these two variants is lower than in the CEU population, ranging from 0.6 to 0.64, and in Asiatic populations both variants are missing. The highest LD, though with differences, was observed in the Amerindian and European populations. Therefore, we considered important to estimate the LD and to discern the primary association signal in our Caucasian Spanish population. Thus, we performed a case-control study with 4384 patients and 3197 controls to confirm the association of this functional variant (rs28445040) with MS and to compare it with the best associated variant from GWAS (rs10201872). Results shown in Table 2 indicated that both variants were in strong LD (r^2=0.93) and evidenced similar MS risk association P-values (MAF (T allele) P-values, odds ratios: 1.9 E-9, OR=1.35 [1.22-1.49] and 4.9 E-10, OR=1.37 [1.24-1.51], respectively). After logistic regression analyses, we found that the dominant model was the one that best fitted the data for both SNPs.

Discussion

The aim of this study was to identify the causal variant and the functional mechanism behind the association of the *SP140* locus with MS susceptibility. This was performed by integration of the high density map of SNPs associated with MS-risk available from the Immunochip Project (15), and the high density map of eQTLs generated in our study using RNA sequences from the GEUVADIS Project (17), which in turn have been obtained from the lymphoblastoid cell lines of the 1000 Genomes Project (16). This strategy has pointed to rs28445040 as the causative variant of the association. We have demonstrated that this variant interferes with the splicing of the exon 7 of the *SP140* gene, resulting in a decrease of the full length transcript and, as a consequence, the reduction of the produced protein in blood cells.

Evidences suggest rs28445040 as the causal variant of the *SP140* association with MS, which is in strong LD with variants associated with CLL and CD obtained by different GWAS (1). Common susceptibility loci have been widely reported among autoimmune diseases, indicating shared pathological pathways (20), so it is not surprising that MS and CD showed association with the same risk variants. Moreover, the variants selected in the locus by CLL GWAS were also in strong LD with the *SP140* variant described in this study. Although not an autoimmune disease, some of the susceptibility loci for CLL have been associated with autoimmune diseases as it is the case for the *IRF4* locus (21), which has also been associated with rheumatoid arthritis (RA) (22), or the *IRF8* locus associated with MS (1), RA (23), systemic lupus erythematosus (SLE) (24), inflammatory bowel disease (IBD) (3) and systemic scleroderma (SS) (24). Apparently, CLL and autoimmune diseases seemed to have different pathogenic mechanisms; however, the immunological tolerance failure characteristic of CLL (25) could be a common background for both pathologies.

The eQTL studies, recently published in tissues and cell lines, have revealed their great value to identify the functional variants implicated in disease pathogenesis (26). However, due to the different density of markers used in GWAS and eQTLs studies, the colocalization of both signals in a locus does not always indicate a common origin of effects (27). The use of high density datasets, as the ones mentioned here, allows to accurately selecting the common variants underlying both effects: transcript levels and association to disease. However, in many cases, LD between variants hampers the identification of a unique SNP as the causal variant of eQTL and risk association. This is the case for the *SP140* locus in which 18 SNPs, with r^2 ranging between 0.965 and 1, are potential causal variants, and therefore, the identification of the ultimate causal one has required functional studies. The opposite effects of the eQTLs on the levels of expression of two *SP140* transcripts, differing in the presence or absence of exon 7, and the localization of one of these 18 SNPs in this exon, were suggestive of rs28445040 as the causal SNP. The reproduction by alternative splicing construct experiments of the splicing effect observed in the RT-PCR data demonstrates that this variant is the responsible for the alteration of the splicing of exon 7. However, this new RNA isoform with the exon 7 deleted does not seem to give rise to the expected shorter protein, since it could not be detected by western blot experiments with a polyclonal antibody. There are several potential explanations for the failure in finding the expected shorter product in immunoblots. The polyclonal antibody used for detecting SP140 expression was generated against a 30-amino acids peptide located in the exon 8. It is possible that a shorter protein lacking the amino acid sequence encoded in the exon 7 would hampered the antibody's reactivity in the adjacent exon. It is also possible that exon 7 spliced RNA or the corresponding amino acid sequence be degraded and therefore not detected. Nonetheless, we could quantify the exon 7-skipped RNA isoform and there was a T-

allele dependent reduction in full-length protein expression. Therefore, the ultimate effect of the exon-skipping seems to be the reduction of the SP140 protein.

The association assay performed in an Spanish cohort with the best MS variant from the GWAS (17) and the rs28445040 did not allow distinguishing which one was the primary signal of the association due to the strong LD between them (28). Nevertheless, we have confirmed the association of the locus in the Spanish cohort, showing that the T carriers, producing lower expression of the protein, had a higher MS risk. The use of eQTL data from an African population, having different LD pattern in the *SP140* locus respect to the EUR population, resulted in an important help to narrow down the causal variant. Thus, data of YRI eQTLs, obtained from the GEUVADIS Project (17), pointed to rs28445040 as the most likely functional variant affecting the splicing of *SP140* exon 7.

The aetiology of multiple sclerosis is considered complex, with genetic and environmental factors contributing to the onset of disease. Given the limited knowledge of the functional activity of SP140, it is difficult to envisage the pathogenic relevance of the reduction of SP140 protein expression in any of the associated diseases. Two plausible, non exclusive, hypotheses are considered here. First, due to its strong sequence homology with the autoimmune regulator AIRE, a transcriptional activator governing the ectopic expression of peripheral tissue-specific antigens in the thymus (29), it is tempting to speculate that the implication of SP140 in MS and other immune-mediated diseases could be related with the process of immune self-tolerance acquisition, potentially contributing to the autoimmune component of MS, CLL and CD. The second hypothesis is based on the potential role of SP140 as an antiviral component of nuclear bodies induced by interferons (IFNs). SP140 is shown to

physically interact with the viral infectivity factor of the human immunodeficiency virus type 1 (HIV-1) as part of the antiviral response of non permissive cells (30, 31). One of the putative risk factors for MS is infection with Epstein-Barr virus (EBV) (32) with evidences suggesting the collaboration of neuropathogenic HERV-W/MSRV endogenous retroviruses (33, 34). The "low producer" of SP140, rs28445040*T carriers, could have lower effective antiviral response against viruses potentially implicated in MS and in the other SP140-associated diseases. Given the implications of these results, further functional studies of SP140 would be required to finally dissect its pathogenic role in these diseases and potential interventions.

Material and Methods

Study subjects

The entire cohort comprised 4384 MS patients meeting established diagnostic criteria (35) and 3197 healthy controls (mostly blood donors and staff). Patients and controls were recruited from 11 different Spanish hospitals: 733 MS patients from Hospital Virgen Macarena of Sevilla; 154 from Hospital Virgen de la Nieves of Granada; 105 MS patients from Hospital San Cecilio of Granada; 450 healthy controls from Blood Bank of Andalucía; 612 MS patients and 599 healthy controls from Hospital Carlos Haya of Málaga ; 1076 MS patients and 829 healthy controls from Hospital Clínico San Carlos of Madrid; 373 MS patients from Hospital Ramón y Cajal of Madrid; 245 patients from Hospital Clinic of Barcelona; 189 MS patients and 371 healthy controls from Hospital Universitari Vall d´Hebron of Barcelona; 259 MS patients and 267 healthy controls from Hospital Donostia of San Sebastian; 479 MS patients and 494 healthy controls from Hospital de Basurto of Bilbao; 160 MS patients and 187 healthy

controls from the Spanish Biobank of DNA. Patients provided informed consent and the institutional ethics committees of these centers approved the study. PBMCs from 28 relapsing-remitting MS patients and 25 healthy controls, all Caucasians, were used for the RT-PCR amplifications and Western blot studies. LCLs from the 1000 Genomes Project were obtained from Coriell Biorepository.

Gene expression analysis

For this study we used the raw data from GEUVADIS RNA sequencing project (http://www.geuvadis.org/web/genvadis) obtained from LCLs of the 1000 Genomes Project (http://www.1000genomes.org/) (17). We selected the FASTQ files of 344 subjects belonging to the CEU, GBR, FIN and TSI populations. A QC phase was performed. We launched the FastQC tool on randomly selected samples of the dataset and obtained that all of them had the first 14 bases repeated in every read. We used fastx-toolkit to get a clean read set. First, a read trimming process was performed removing the first 14 bases of every read. Next, reads shorter than 50 bp and those with quality less than 20 PHRED across 66% of the sequence were discarded.

For Gene Expression Profiling analysis, reads were mapped to the human reference genome (assembly GRCh37.68) using Tophat v1.4 (36). This application mapped the reads using a splicing-aware algorithm. Transcript abundance was estimated using Cufflinks v1.3 (37) and Ensembl GRCh37.68. We forced Cufflinks to estimate only isoform expression compatible with reference annotation, no novel transcripts were reconstructed. More detailed information is provided in Supplementary Material, Fig. S1.

Cis-eQTL calculation

We conducted Spearman's rank correlation analyses with SNPs from the 1000 Genomes Project phase1 release v3 data set with MAF (minor-frequent allele) > 0.05. For each data set, we performed associations in 1-megabase windows overlapping 100 kb. A *P*-value of false discovery rate (FDR) was indicated in Table 1.

RNA isoform profile analysis

RNA from LCLs and PBMCs was extracted with RNeasy system (Qiagen). Primers for PCR amplification were designed with Primer 3 (v. 0.4.0) software (38): Forward primer at exon 6, 5'-*CAGTTAGCTCTCCCAAAGGC-3′* and Reverse primer at exon 8, 5'-*CTGTGCTGTATGTCCTTGCC-3′*. The cDNA was synthesized using a total of 50 ng mRNA of each sample, reverse transcribed using the Superscript III reverse transcription reagents (Invitrogen S.A., Invitrogene Ltd., UK) and then subjected to PCR amplification. Electrophoresis was carried out in 10% PAGE gel (40% Acrylamide/Bis solution 37, 5:1 (BioRad)) with running buffer (90mM Tris-Borate and 2mM EDTA). PCR products visualized with GelRed (Biotium Inc) and detected by UV light, were evaluated by densitometry (Quantity One; BioRad Laboratories).

Reverse Transcriptase and Quantification by qPCR

cDNA was generated using iScript™ Reverse Transcription Supermix (BioRad). Then, isoforms quantification was performed by real-time qPCR using GoTaq qPCR Master Mix (Promega) and normalized to *UBE2D2* mRNA levels using 2E (Ct sample-Ct reference) method (39). The primer sequences were designed using Primer3 browser (FORWARD -F; REVERSE-R, 5'-3' direction): *UBE2D2* F-CAATTCCGAAGAGAATCCACAAGGAATTG, *UBE2D2*-R-

GTGTTCCAACAGGACCTGCTGAACAC. In order to amplify transcripts with exon 7 spliced, we used a forward bridge primer between exon 6 and 8: *SP140*B F-CTACCAGGTGGGGGAGTTCT and reverse primer at exon 8: *SP140*B R-TTCCCTCTGGACTCTCTTGG. In order to amplify the full transcript, with exon 7 included, we used a reverse bridge primer between exons 8 and 7: *SP140*I R-CCGTTGCTTTCTAGAACTTC and forward primer at exon 6: *SP140*I F-TGGTGGAGGAGATGCTGAAG.

Western Blotting and densitometry

PBMCs were obtained from sodium heparine venous blood by centrifugation in CPT tubes (Becton Dickinson Vacutainer). PBMCs cell extracts were sonicated and resuspended in lysis buffer (150 mM NaCl, 2 mM EDTA, 2 mM EGTA, 0.2% Triton X-100, 0.3% NP40 and 50 mM Tris-HCl). Protein concentrations were measured using NanoDrop 1000 (Thermo Scientific). Equal amounts of proteins (40 μg) separated on denaturing SDS/12% (w/v) polyacrylamide gels were then blotted onto PVDF membranes (BioRad). The blots were blocked with 10% (w/v) non-fat dry milk in TBST (10mM Tris (pH 7.7), 100mM NaCl and 0.1% Tween 20). Membranes were incubated with polyclonal antibodies against either SP140 (1:500, Abcam) and β-actin (1:5000, Santa Cruz). Then, they were incubated with HRP (horseradish peroxidase)-conjugated anti-(rabbit or mouse IgG) antibodies at a dilution of 1:2500. Proteins were detected by enhanced chemiluminescence (ECL Western Blotting Detection Reagents, BioRad) and evaluated by densitometry (Quantity One; BioRad Laboratories). Pre-stained protein markers (Precision Plus Protein standards, BioRad) were used for molecular mass determinations.

Alternative splicing constructs

Alternative exon 7 splicing analysis was performed as described by Desviat et al. (40). Briefly: PCR amplification of a DNA fragment containing the exon 7 of *SP140* gene (78bp) and the flanking intronic sequences (103bp at the 5' and 75bp at the 3' of the exon). The oligonucleotides for this amplification were: forward 5'-*CCCGAATATTAGAGCTCAGCA*-3' and reverse 5'- *TGGGAAGGGAGATGAAAGAG*-3'. The amplification of the DNA fragment was performed from the LCL *NA12383* which is heterozygous for the rs28445040 variant. The PCR product was cloned in TOPO-TA vector and sequenced to identify the clones carrying each allele and to discard potential PCR errors. An EcoRI fragment from each TOPO vectors was subcloned in the EcoRI site of the pSPL3 minigene plasmid and orientation checked by sequencing. The T and C pSPL3 (pC, pT) plasmids and plasmid without insert (p) were transfected in HEK cells and harvested 24 h after transfection. RNA was extracted and amplified by RT-PCR using SD6 and SA2 primers. The PCR products were visualized in 2% agarose gel electrophoresis and sequenced to confirm the DNA origin.

Acknowledgments

We thank patients with multiple sclerosis and control subjects for making this study feasible. We also thank to the GEUVADIS Consortium for the RNA-Seq data and to the IMSGC (International Multiple Sclerosis Genetics Consortium) and WTCCC (Wellcome Trust Case Control Consortium) for the Immunochip Project data. We also acknowledge the University Hospital Virgen Macarena Biobank (Andalusian Public Health System Biobank), integrated in the Spanish National Biobank Network and to the Banco Nacional de ADN for its help and support with the clinical samples used in this work. We thank Dr. L.R. Desviat for kindly provide us the pSPL3 plasmid. This

work was supported by Fondo de Investigación Sanitaria (FIS)-Instituto de Salud Carlos III(ISCIII)-Fondos Europeos de Desarrollo Regional (FEDER), Unión Europea [grant numbers P12/00555, PI13/01527, PI13/02714, PI13/01466, PI13/00879, RETICS-Red Española de Esclerosis Múltiple (REEM) [grant numbers RD12/0032/0002-RD12/0032/0015-RD12/0032/0005-RD12/0032/0006- RD12/0032/0009, PT13/0010/0041] and Junta de Andalucía (JA)- Fondos Europeos de Desarrollo Regional (FEDER) [grant number CTS2704].

Conflict of interest statement

None declared

Reference

1. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature,* **476,** 214-9.
2. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R. *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet,* **42,** 1118-25.
3. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. *et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature,* **491,** 119-24.
4. Speedy, H.E., Di Bernardo, M.C., Sava, G.P., Dyer, M.J., Holroyd, A., Wang, Y., Sunter, N.J., Mansouri, L., Juliusson, G., Smedby, K.E. *et al.* (2014) A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet,* **46,** 56-60.
5. Berndt, S.I., Skibola, C.F., Joseph, V., Camp, N.J., Nieters, A., Wang, Z., Cozen, W., Monnereau, A., Wang, S.S., Kelly, R.S. *et al.* (2013) Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet,* **45,** 868-76.
6. Slager, S.L., Skibola, C.F., Di Bernardo, M.C., Conde, L., Broderick, P., McDonnell, S.K., Goldin, L.R., Croft, N., Holroyd, A., Harris, S. *et al.* (2012) Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood,* **120,** 843-6.
7. Di Bernardo, M.C., Crowther-Swanepoel, D., Broderick, P., Webb, E., Sellick, G., Wild, R., Sullivan, K., Vijayakrishnan, J., Wang, Y., Pittman, A.M. *et al.*

(2008) A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet,* **40,** 1204-10.

8. Granito, A., Yang, W.H., Muratori, L., Lim, M.J., Nakajima, A., Ferri, S., Pappas, G., Quarneti, C., Bianchi, F.B., Bloch, D.B. *et al.* (2010) PML nuclear body component Sp140 is a novel autoantigen in primary biliary cirrhosis. *Am J Gastroenterol,* **105,** 125-31.
9. Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R. *et al.* (2014) Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun,* **5,** 2997.
10. Zucchelli, C., Tamburri, S., Quilici, G., Palagano, E., Berardi, A., Saare, M., Peterson, P., Bachi, A. and Musco, G. (2014) Structure of human Sp140 PHD finger: an atypical fold interacting with Pin1. *FEBS J,* **281,** 216-31.
11. Kisand, K. and Peterson, P. (2011) Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy: known and novel aspects of the syndrome. *Ann N Y Acad Sci,* **1246,** 77-91.
12. Park, S., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2013) The CH2 domain of CBP/p300 is a novel zinc finger. *FEBS Lett,* **587,** 2506-11.
13. Sille, F.C., Thomas, R., Smith, M.T., Conde, L. and Skibola, C.F. (2012) Post-GWAS functional characterization of susceptibility variants for chronic lymphocytic leukemia. *PLoS One,* **7,** e29632.
14. Plagnol, V., Smyth, D.J., Todd, J.A. and Clayton, D.G. (2009) Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics,* **10,** 327-34.
15. Parkes, M., Cortes, A., van Heel, D.A. and Brown, M.A. (2013) Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet,* **14,** 661-73.
16. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature,* **491,** 56-65.
17. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature,* **501,** 506-11.
18. Beecham, A.H. and Patsopoulos, N.A. and Xifara, D.K. and Davis, M.F. and Kemppinen, A. and Cotsapas, C. and Shah, T.S. and Spencer, C. and Booth, D. and Goris, A. *et al.* (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet,* **45,** 1353-60.
19. Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A. *et al.* (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature,* **518,** 337-43.
20. Voight, B.F. and Cotsapas, C. (2012) Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity. *Curr Opin Immunol,* **24,** 552-7.
21. Slager, S.L., Rabe, K.G., Achenbach, S.J., Vachon, C.M., Goldin, L.R., Strom, S.S., Lanasa, M.C., Spector, L.G., Rassenti, L.Z., Leis, J.F. *et al.* (2011) Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood,* **117,** 1911-6.

22. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature,* **506,** 376-81.
23. Okada, Y., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Kawaguchi, T., Stahl, E.A., Kurreeman, F.A., Nishida, N. *et al.* (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet,* **44,** 511-6.
24. Martin, J.E., Assassi, S., Diaz-Gallo, L.M., Broen, J.C., Simeon, C.P., Castellvi, I., Vicente-Rabaneda, E., Fonollosa, V., Ortego-Centeno, N., Gonzalez-Gay, M.A. *et al.* (2013) A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet,* **22,** 4021-9.
25. Garcia-Munoz, R., Feliu, J. and Llorente, L. (2015) The top ten clues to understand the origin of chronic lymphocytic leukemia (CLL). *J Autoimmun,* **56,** 81-6.
26. Pai, A.A., Pritchard, J.K. and Gilad, Y. (2015) The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLoS Genet,* **11,** e1004857.
27. Battle, A. and Montgomery, S.B. (2014) Determining causality and consequence of expression quantitative trait loci. *Hum Genet,* **133,** 727-35.
28. Malo, N., Libiger, O. and Schork, N.J. (2008) Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. *Am J Hum Genet,* **82,** 375-85.
29. Metzger, T.C. and Anderson, M.S. (2011) Control of central and peripheral tolerance by Aire. *Immunol Rev,* **241,** 89-103.
30. Madani, N., Millette, R., Platt, E.J., Marin, M., Kozak, S.L., Bloch, D.B. and Kabat, D. (2002) Implication of the lymphocyte-specific nuclear body protein Sp140 in an innate response to human immunodeficiency virus type 1. *J Virol,* **76,** 11133-8.
31. Guo, Y., Dong, L., Qiu, X., Wang, Y., Zhang, B., Liu, H., Yu, Y., Zang, Y., Yang, M. and Huang, Z. (2014) Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature,* **505,** 229-33.
32. Tzartos, J.S., Khan, G., Vossenkamper, A., Cruz-Sadaba, M., Lonardi, S., Sefia, E., Meager, A., Elia, A., Middeldorp, J.M., Clemens, M. *et al.* (2012) Association of innate immune activation with latent Epstein-Barr virus in active MS lesions. *Neurology,* **78,** 15-23.
33. Mameli, G., Madeddu, G., Mei, A., Uleri, E., Poddighe, L., Delogu, L.G., Maida, I., Babudieri, S., Serra, C., Manetti, R. *et al.* (2013) Activation of MSRV-type endogenous retroviruses during infectious mononucleosis and Epstein-Barr virus latency: the missing link with multiple sclerosis? *PLoS One,* **8,** e78474.
34. Garcia-Montojo, M., de la Hera, B., Varade, J., de la Encarnacion, A., Camacho, I., Dominguez-Mozo, M., Arias-Leal, A., Garcia-Martinez, A., Casanova, I., Izquierdo, G. *et al.* (2014) HERV-W polymorphism in chromosome X is associated with multiple sclerosis risk and with differential expression of MSRV. *Retrovirology,* **11,** 2.
35. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L. *et al.* (2010) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol,* **69,** 292-302.

36. Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* **25,** 1105-11.
37. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2011) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol,* **28,** 511-5.
38. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res,* **40,** e115.
39. Catala-Rabasa, A., Ndagire, D., Sabio, J.M., Fedetz, M., Matesanz, F. and Alcina, A. (2011) High ACSL5 transcript levels associate with systemic lupus erythematosus and apoptosis in Jurkat T lymphocytes and peripheral blood cells. *PLoS One,* **6,** e28591.
40. Desviat, L.R., Perez, B. and Ugarte, M. (2012) Minigenes to confirm exon skipping mutations. *Methods Mol Biol,* **867,** 37-47.

Legends to Figures

Figure 1. Colocalization of eQTLs and best MS-associated SNPs at the *SP140* locus. (A) Scheme of the region analyzed in the chr2:231,045,388-231,276,234 (hg19) locus showing, from top to bottom: the 18 best-eQTLs for ENST00000392045 and ENST00000343805 transcripts of the *SP140* gene, which correlated with the best MS-associated variants in the Immunochip (red); SNPs associated with different diseases identified by GWAS (green). (B) Scatter plots representing the expression correlation coefficient (absolute value of Spearman's rho coefficient) for each indicated transcript versus the MS-association values (-log *P*). Determination of eQTLs in the region was performed in this work using the RNA sequencing data from GEUVADIS Project, together with the genotype information from the 1000 Genomes Project. In each plot, the best MS-associated SNP and the best-eQTL are indicated. (C) LocusZoom plots showing the expression-correlation levels of variants in the region. The best MS-associated SNP in the locus from the Immunochip dataset is in purple and indicated with an arrow. Colours scale represents the linkage disequilibrium (r^2 values) respect to this variant obtained from the 1000 Genomes EUR population.

Figure 2. Inverse correlation of two *SP140* RNA isoforms respect to rs28445040 genotypes. (A) Box plots representing the expression levels from RNA-Seq (GEUVADIS Project) of the indicated transcripts in 344 LCLs versus the rs28445040 genotypes, and the Spearman's correlation index (*rho*) and *P*-values inside. (B) Spearman's correlation index (*rho*-values) between each *SP140* exons and the rs28445040 genotypes. (C) Polyacrylamide gel electrophoresis (PAGE) showing the results of the RT-PCR amplification of RNA from LCLs (cell lines: TT, NA20518; CT, NA20766; CC, NA12004) and PBMCs carrying the different genotypes for the

rs28445040 and the scheme of primers' positions in exon 6 and 8 used for amplification and variant's position in exon 7. (D) Box plots of relative expression levels of exon 7-skipped and full-length transcripts respect to rs28445040 genotypes measured by real time qPCR from 59 LCLs with bridge primers as depicted in the figure. Spearman's correlation index (*rho*) and *P*-values are indicated inside the plots.

Figure 3. Alternative splicing of SP140 exon 7 carrying the rs28445040 C/T alleles analyzed by alternative splicing construct assay. (A) Genomic DNA construct showing the cloned sequence within the pSPL3 vector in the multi-cloning site position (MCS). The scheme represents the size in bp of the different exons corresponding to the vector, containing a cryptic exon, and the recombinant fragment of the SP140 exon 7 with intron flanking sequences (white). The position of the rs28445040 variant (C/T) and the primers for RT-PCR amplification are also indicated. (B) Agarose gel electrophoresis of RT-PCR from HEK cells RNA transfected with the constructs carrying the C or T alleles of the rs28445040 variant (pC, pT) and control plasmid (p). Lane m is the molecular weight marker.

Figure 4. Western blot analysis of PBMCs from MS patients and healthy controls. (A) Western blot of PBMCs from MS patients showing the decrease of SP140 full-length protein in TT carriers. (B) Box plot representing the quantification of the SP140 bands by densitometry of Western blots from 28 MS patients according to the genotypes of the rs28445040 variant. (C) Box plots representing the quantification of SP140 protein band in PBMCs from 28 MS patients (CC=8, CT=10, TT=10) and 25 healthy controls (CC=10, CT=8, TT=7).

Table 1. Linkage disequilibrium (LD) between the eQTLs at the chr2:230856224-231357223 locus and the GWAS associated variants for different diseases in the region.

						LD between eQTLs and associated SNP (r^2)(2)			
						CLL	MS	CD	CD
Gene	**Transcript**	**rho (1)**	**P value**	**FDR_P value**	**eQTL**	**rs13397985**	**rs10201872**	**rs6716753**	**rs7423615**
SP100	ENST00000452345	0.32	1.37E-07	5.974E-05	rs1649884	0.009	0.005	0.013	0.009
SP140	ENST00000350136	-0.35	2.87E-09	1.54E-06	rs10498245	0.492	0.449	0.509	0.449
SP140	ENST00000343805	-0.49	3.26E-17	6.18E-14	rs13426106	0.991	0.899	0.959	1
SP140	ENST00000392045	0.44	5.81E-14	7.603E-11	rs13397985	1	0.891	0.967	0.991

(1) Spearman's rho correlations between RNA expression levels and genotypes; (2) the LD has been calculated with the EUR population of 1000 Genomes Project; MS, multiple sclerosis; CD Crohn's disease; CLL, chronic lymphocytic leukemia.

Table 2. MS-association of the best GWAS-MS variant (rs10201872) and the functional variant described in this work (rs28445040) by logistic regression analysis.

SNPs	**MS**			**Control**			**Dominant model**		**Conditioning on rs10201872**		**Conditioning on rs28445040**	
	TT	**CT**	**CC**	**TT**	**CT**	**CC**	**P**	**OR (CI 0.95)**	**P**	**OR (CI 0.95)**	**P**	**OR (CI 0.95)**
rs10201872	158 (3.6)	1390 (31.7)	2836 (64.7)	88 (2.7)	824 (25.8)	2285 (71.5)	4.9 E-10	1.37 (1.24-1.51)	0.96	1 (0.70-1.45)	NA	NA
rs28445040	167 (3.8)	1434 (32.7)	2783 (63.5)	99 (3.1)	857 (26.8)	2241 (70.1)	1.9 E-9	1.35 (1.22-1.49)	NA	NA	0.1	1.35 (0.94-1.96)

Genotype distributions are shown as the number (%); odds ratio (OR), 95% confidence interval (CI), and P-values were determined by logistic regression analysis with dominant model.

Abbreviations

SP140: SP140 nuclear body protein

RT-PCR: Reverse transcription polymerase chain reaction

eQTL: Expression quantitative trait loci

PBMC: Peripheral blood mononuclear cell

GWAS: Genome-wide association study

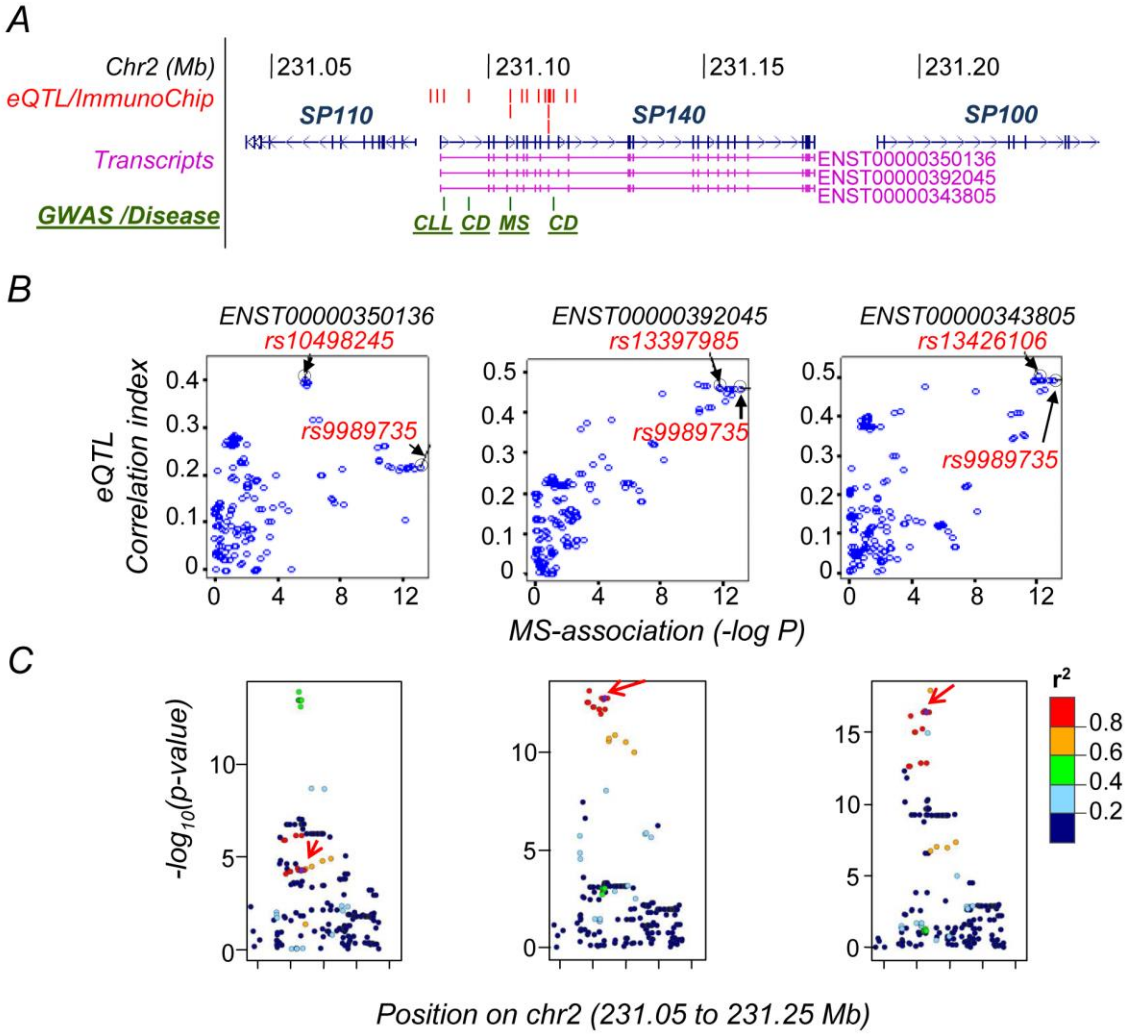
MS: Multiple sclerosis

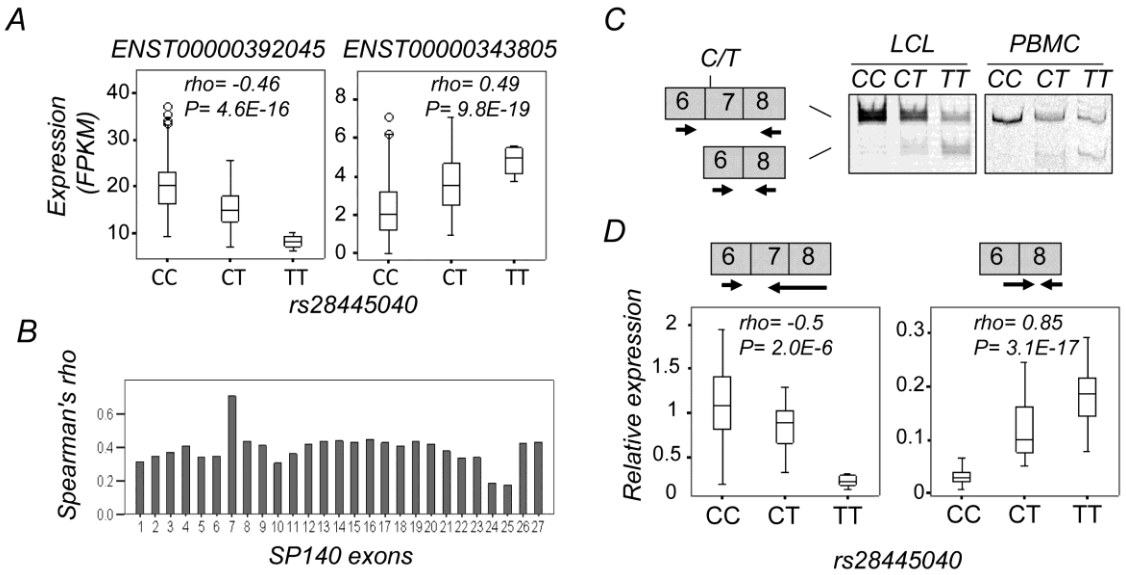
LD: Linkage disequilibrium

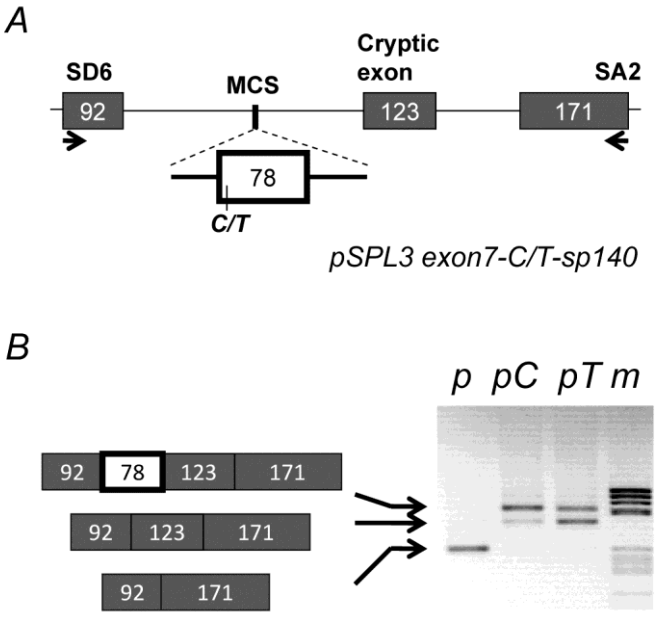
CD: Crohn's disease

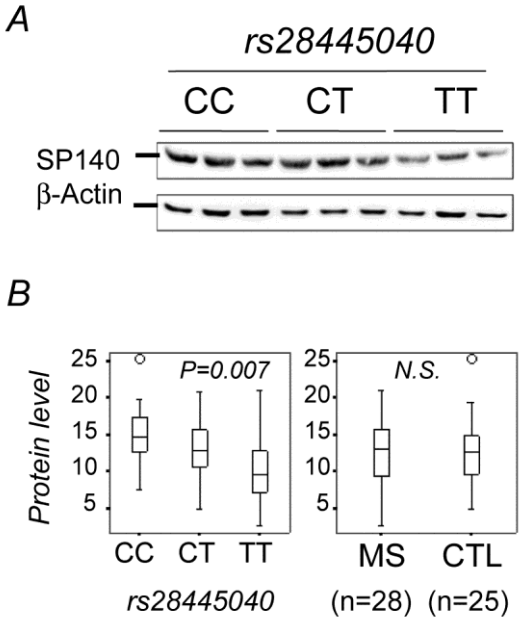
CLL: Chronic lymphocytic leukemia

IMSGC: International Multiple Sclerosis Genetic Consortium


A
Chr2 (Mb)
231.05
231.10
231.15
231.20
eQTL/ImmunoChip
SP110
SP140
SP100
Transcripts
ENST00000350136
ENST00000392045
ENST00000343805
GWAS /Disease
CLL
CD
MS
CD
B
ENST00000350136
rs10498245
rs9989735
ENST00000392045
rs13397985
rs9989735
ENST00000343805
rs13426106
rs9989735
eQTL
Correlation index
MS-association (-log P)
C
-log10(p-value)
r2
0.8
0.6
0.4
0.2
Position on chr2 (231.05 to 231.25 Mb)


A
ENST00000392045
ENST00000343805
rho= -0.46
P= 4.6E-16
rho= 0.49
P= 9.8E-19
Expression (FPKM)
CC CT TT
rs28445040
B
Spearman's rho
SP140 exons
C
C/T
LCL
PBMC
CC CT TT
D
rho= -0.5
P= 2.0E-6
rho= 0.85
P= 3.1E-17
Relative expression
rs28445040


A
SD6
MCS
Cryptic exon
SA2
92
123
171
78
C/T
pSPL3 exon7-C/T-sp140
B
p
pC
pT
m
92
78
123
171
92
123
171
92
171


A
rs28445040
CC
CT
TT
SP140
β-Actin
B
Protein level
25
20
15
10
5
P=0.007
N.S.
CC
CT
TT
MS
CTL
rs28445040
(n=28)
(n=25)