




# Identification of the genetic mechanism that associates L3MBTL3 to multiple sclerosis

Antonio Alcina <sup>1,\*</sup>, Maria Fedetz<sup>1</sup>, Isabel Vidal-Cobo<sup>1</sup>, Eduardo Andrés-León <sup>2</sup>, Maria-Isabel García-Sánchez<sup>3</sup>, Alicia Barroso-del-Jesus<sup>4</sup>, Sara Eichau<sup>5</sup>, Elia Gil-Varea<sup>6</sup>, Luisa-Maria Villar<sup>7</sup>, Albert Saiz<sup>8</sup>, Laura Leyva<sup>9</sup>, Koen Vandembroeck<sup>10,11</sup>, David Otaegui<sup>12</sup>, Guillermo Izquierdo<sup>13</sup>, Manuel Comabella<sup>6</sup>, Elena Urcelay<sup>14,†</sup> and Fuencisla Matesanz <sup>1,\*</sup>,†

<sup>1</sup>Department of Cell Biology and Immunology, Instituto de Parasitología y Biomedicina "López Neyra", Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), 18016 Granada, Spain

<sup>2</sup>Bioinformatic Unit, Instituto de Parasitología y Biomedicina López Neyra (IPBLN-CSIC), 18016 Granada, Spain

<sup>3</sup>UGC Neurología, Nodo Hospital Universitario Virgen Macarena, Biobanco del Sistema Sanitario Público de Andalucía, 41009 Sevilla, Spain

<sup>4</sup>Genomics Unit, Instituto de Parasitología y Biomedicina López Neyra (IPBLN-CSIC), 18016 Granada, Spain

<sup>5</sup>UGC Neurología, Hospital Universitario Virgen Macarena, 41009 Sevilla, Spain

<sup>6</sup>Servei de Neurologia-Neuroimmunologia, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Institut de Recerca Vall d'Hebron (VHIR), Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona 08035, Spain

<sup>7</sup>Departments of Immunology, Hospital Ramon y Cajal, (IRYCIS), 28034 Madrid, Spain

<sup>8</sup>Servicio de Neurología, Hospital Clinic and Institut d'Investigació Biomèdica Pi i Sunyer (IDIBAPS), Institut de Neurociències, Universitat de Barcelona, 08036 Barcelona, Spain

<sup>9</sup>Instituto de Investigación Biomédica de Málaga-IBIMA, UGC Neurología, Hospital Regional Universitario de Málaga, Málaga 29010, Spain

<sup>10</sup>Inflammation & Biomarkers Group, Biocruces Bizkaia Health Research Institute, Barakaldo 48903, Spain

<sup>11</sup>KERBASQUE, Basque Foundation for Science Bilbao 48013, Spain

<sup>12</sup>Neurosciences Area, Biodonostia Health Research Institute, 20014 San Sebastián, Spain

<sup>13</sup>Multiple Sclerosis Unit, Neurology Service, Vithas Nisa Hospital, Seville 41950, Spain

<sup>14</sup>Lab. of Genetics of Complex Diseases, Hospital Clinico San Carlos, Instituto de Investigacion Sanitaria San Carlos (IdISSC), Madrid 28040, Spain

\*To whom correspondence should be addressed at: Instituto de Parasitología y Biomedicina López Neyra (CSIC), Avda. Conocimiento 17, Granada 18016, Spain. Tel: +34 958181668; Fax: +34 958181632; E-mail: f.matesanz@csic.es; pulgoso@ipb.csic.es

†The authors wish it to be known that, in their opinion, the two last authors should be regarded as joint last authors

## Abstract

Multiple sclerosis (MS) is a complex and demyelinating disease of the central nervous system. One of the challenges of the post-genome-wide association studies (GWAS) era is to understand the molecular basis of statistical associations to reveal gene networks and potential therapeutic targets. The L3MBTL3 locus has been associated with MS risk by GWAS. To identify the causal variant of the locus, we performed fine mapping in a cohort of 3440 MS patients and 1688 healthy controls. The variant that best explained the association was rs6569648 ( $P = 4.13E-10$ , odds ratio = 0.71, 95% confidence interval (CI) = 0.64–0.79), which tagged rs7740107, located in intron 7 of L3MBTL3. The rs7740107 (A/T) variant has been reported to be the best expression and splice quantitative trait locus (eQTL and sQTL) of the region in up to 35 human genotype-tissue expression (GTEx) tissues. By sequencing RNA from blood of 17 MS patients and quantification by digital qPCR, we determined that this eQTL/sQTL originated from the expression of a novel short transcript starting in intron 7 near rs7740107. The short transcript was translated into three proteins starting at different translation initiation codons. These N-terminal truncated proteins lacked the region where L3MBTL3 interacts with the transcriptional regulator Recombination Signal Binding Protein for Immunoglobulin Kappa J Region which, in turn, regulates the Notch signalling pathway. Our data and other functional studies suggest that the genetic mechanism underlying the MS association of rs7740107 affects not only the expression of L3MBTL3 isoforms, but might also involve the Notch signalling pathway.

## Introduction

Multiple sclerosis (MS) is a common inflammatory and demyelinating disease of the central nervous system. Its aetiology is not fully understood, but evidence exists supporting the involvement of environmental and genetic factors that trigger the disease (1). In the last 15 years, several genome-wide association studies (GWAS) have contributed to a better understanding of these genetic factors involved in the disease (2). In addition to HLA, the first locus associated with the disease and the one

displaying the largest effect, over two hundred risk loci have been identified (3).

Nowadays, the definition of functional alterations underlying these genetic associations is key objective and faces great difficulties. Most of the variants associated with MS are not located in coding regions and their effects must be sought in the alteration of transcriptional regulation, which opens the possibility of influencing a large number of genes. Expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL) are

Received: July 29, 2021. Revised: December 19, 2021. Accepted: January 10, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

powerful tools to identify causal genes and functional effects in MS-associated variants and other diseases (4–8). They link trait-associated genetic variants to functional effects at the transcriptional level. International projects such as the genotype-tissue expression (GTEx) have built eQTL databases to describe the relationship between genetic variation and gene expression in human tissues (9). Although these databases are very useful, important questions remain to be solved regarding how each variant exerts its effect on the RNA levels of affected transcripts and how they are related to the disease.

The association of the rs4364506 variant, located at the intron 12 of the *L3MBTL3* gene, with MS susceptibility was identified in a GWAS by Andlauer et al. (10). The association of this locus was corroborated by the last meta-analysis performed by the International Multiple Sclerosis Genetic Consortium, which identified 233 statistically independent associations with MS susceptibility using data from 47 429 MS cases and 68 374 controls (3). To date, no studies have established whether the MS-associated variant in the locus is functionally altering the *L3MBTL3* gene or another gene in the vicinity.

In the present study, we performed a fine-mapping of the region to define the causal variant responsible for the association with MS susceptibility, and to identify the effect of this variant on the splicing, transcription and expression of the *L3MBTL3* gene. We could correlate the causal variant with a new transcription initiation site within intron 7, that originates a previously undescribed transcript, which codified a truncated protein lacking the N-terminal region, involved in interacting with other proteins such as the transcriptional regulator Recombination Signal Binding Protein for Immunoglobulin Kappa J Region (RBPJ) that influences in the Notch signalling pathway (11).

## Results

### Identification of the causal variant by a tag-SNP approach

First, we addressed the validation of the association with MS of the *L3MBTL3* locus by genotyping the rs4364506 variant (10) in a Caucasian cohort of 3440 MS patients and 1688 healthy controls from Spain. The cohort was described in Gil-Varea et al. (12). Our results corroborated the association of the variant with MS [ $P=4.80E-5$ , odds ratio (OR) = 0.81 (0.73–0.9)].

Second, we performed fine-mapping of the region on chromosome 6 with coordinates from 129949823 to 130205324 (hg38). SNPs were selected by pairwise tagging of the EUR population from The 1000 Genome Project (13). The selection criteria captured all variants in the region with linkage disequilibrium  $r^2 > 0.4$  with the rs4364506 variant, initially associated with MS by GWAS. We selected 6 SNPs that captured 86 markers with  $r^2 \geq 0.7$  (mean  $r^2 = 0.94$ ) and a minor allele frequency  $\geq 0.1$  (Supplementary Material, Table S1 and Supplementary Material, Fig. S1).

The six tag-SNPs were genotyped in the same cohort used to replicate the rs4364506 association. Genotype frequencies were fitted to an additive model. The best-associated SNP was rs6569648 [ $P$ -value = 4.13E-10, OR = 0.71 (0.64–0.79)] (Table 1). We performed a logistic regression analysis to establish the effect of other independent variants, testing the addition of each SNP to rs6569648, and adding this polymorphism to each one of the other tag-SNPs. The results showed that, in addition to generating the strongest signal, rs6569648 alone was sufficient to model the association of the locus.

### The MS-associated variant is an eQTL and sQTL for the *L3MBTL3* gene

To ascertain whether rs6569648 could correlate with the expression of any gene in the region, we explored the GTEx database (9). We observed that this variant is an eQTL for the *L3MBTL3* gene in 35 human tissues, showing the best correlation in spleen [normalized effect size (NES) = -1.1;  $P$ -value =  $1.7 \times 10^{-41}$ ]. We also observed that this variant is a sQTL for the *L3MBTL3* gene in 26 tissues, showing the highest correlation in pituitary (NES = 1.4;  $P$ -value =  $10^{-42}$ ) and spleen (NES = 1.3;  $P$ -value =  $9.3 \times 10^{-34}$ ).

To evaluate whether rs6569648 was the causal variant driving these expression effects, we scrutinized the GTEx results and found that the best eQTL and sQTL was rs7740107 in all tissues analysed. This variant was tagged by rs6569648 in the fine-mapping study (Supplementary Material, Table S1). The LD between these two variants was  $r^2 = 0.79$  and  $D' = 0.99$  for the European population. Interestingly, the pattern of variants associated with *L3MBTL3* expression and altered splicing was almost identical (Fig. 1), which led us to think that both effects were related.

### The *L3MBTL3* locus has also been associated with other traits and diseases

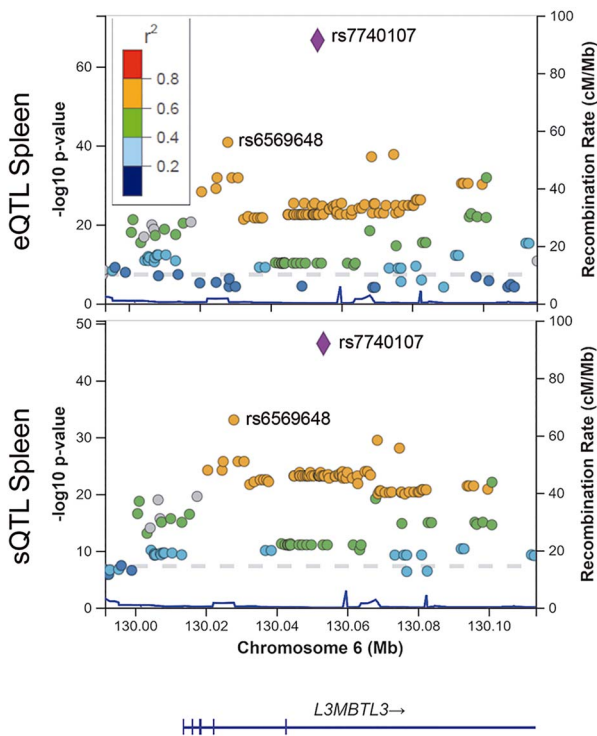
In the present work, we analysed other traits that have been reported associated with this locus. Sixteen associations with traits such as white blood cell count, monocyte count, lymphocyte count, height, hair colour, body mass index, breast cancer, or hip circumference have been published for the rs6569648 variant in 11 well-powered GWAS studies (14–24).

Eleven associations of rs7740107 have been described with immature reticulocyte fraction, red blood cell count, appendicular lean mass, estimated glomerular filtration rate, chronic kidney disease, white blood cell count, height, body mass index and haematocrit (14,15,17,21,25–28). Other variants, in LD with those reported here and associated with MS at the locus, have been also associated with other traits or diseases (Supplementary Material, Fig. S2), most of them related to blood cell counts and biometric features.

**Table 1.** MS-association of the Tag-SNPs in the *L3MBTL3* locus

SNP	Position Chr 6 (hg38)	Allele effect	P-value	OR (95% CI)	P effect conditional
rs6569648	130 027 974	C	4.13E-10	0.7143 (0.64–0.79)	NA
rs9388768	130 052 957	C	0.0007179	0.8531 (0.78–0.94)	0.43
rs4364506	130 068 795	A	4.80E-05	0.8123 (0.73–0.90)	0.51
rs6569650	130 082 671	A	0.0001449	0.8421 (0.77–0.92)	0.36
rs13211683	130 096 688	C	2.98E-05	0.8147 (0.74–0.90)	0.44
rs7740188	130 024 960	G	0.001014	0.8525 (0.78–0.94)	0.06

OR, odds ratio; CI, confidence interval.



**Figure 1.** The best MS-associated variant rs6569648 is in high LD ( $r^2 = 0.79$ ) with the best eQTLs and sQTLs for *L3MBTL3* gene as shown in the LocusZoom plot from spleen GTEx data. Colours refers to the LD with rs7740107 variant using the European population of The 1000 Genomes project as reference, coordinates are in hg38.

### rs7740107 is associated with transcription arrest and a new transcription initiation at *L3MBTL3* intron 7

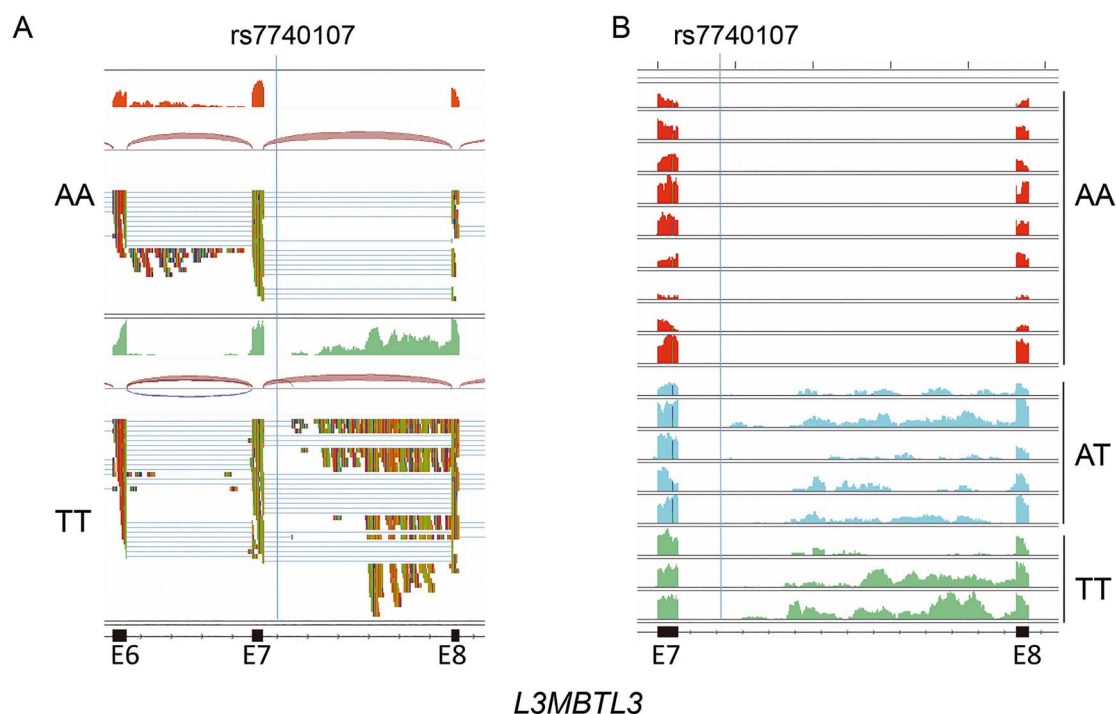
Since the sQTL and eQTL of the locus appeared to be related (Fig. 1), we further examined the region in which GTEx described the sQTL, defined as the difference in intron-excision ratio among genotypes. This region is located at position 130 052 991:130055171 (hg 38) on chromosome 6, overlapping intron 7 of the *L3MBTL3* gene. Interestingly, rs7740107 was also located in intron 7, at 325 bp of the splice donor. The RNA-Seq from two MS patient blood samples, carriers of TT or AA genotypes of the rs7740107 variant, showed multiple reads aligned to intron 7 in the TT but not in the AA sample (Fig. 2A). However, no new joints could be observed between these aligned sequences and the exons, ruling out novel alternative splicing. To confirm these results, the number of

MS patients analysed was expanded to 17 carriers of TT, AT or AA genotypes for the rs7740107 variant (Fig. 2B). Reads in intron 7 were not observed in any of the 9 AA samples, whereas they were observed in the samples of the AT and TT carriers. These data proved that there were transcript initiations in intron 7 of the *L3MBTL3* gene with the T allele, the minor allele associated with MS protection.

The eQTL for rs7740107 described by GTEx shows that *L3MBTL3* mRNA levels were higher for TT carriers in whole blood cells (Fig. 3), as well as in other tissues. To validate eQTL in blood samples from MS patients, and also to examine whether the transcription initiation observed from RNA-Seq data affects gene expression, we designed PCR primers at different positions of the gene and performed reverse transcription (RT) and digital qPCR in samples from 36 individuals: 14 AA, 11 AT and 11 TT genotype carriers. We did not detect significant differences in expression levels with amplification primers localized between exon 2 and exon 3 or exon 4 and exon 6 (Fig. 3a and b). However, we detected lower expression in TT carriers when we amplified between exon 7 and 8, flanking the intron where transcription initiation was observed (Fig. 3c). Opposite results were observed when we quantified *L3MBTL3* expression between intron 7 and exon 9 (Fig. 3d). The forward primer was located in intron 7, 35 bp upstream of the exon 8 acceptor splice site. These primers could only amplify transcripts originating from the intron, and not those with intron 7 spliced. Higher expression was also observed with primers located between exon 21 and exon 23 in the AT and TT samples compared with the AA samples (Fig. 3e). However, in the latter amplification, we cannot distinguish between transcripts originating from the 5' end of the gene or from intron 7.

### Translation of *L3MBTL3* transcripts into proteins

To demonstrate whether the transcript originating from intron 7 (short transcript) can be translated into protein, we cloned both the full-length and short transcripts obtained from a carrier of the TT genotype (Fig. 4). For cloning, the forward primer was placed in the 5' UTR region in exon 2 for the full-length transcript, and at the end of intron 7 for the short transcript. The reverse primer was placed in the last 20 nt of the coding mRNA of exon 23. Both cDNAs were cloned into the plasmid pcDNA6.2/V5/GW/D-TOPO to produce



**Figure 2.** Transcription initiation at intron 7 of *L3MBTL3*. **(A)** RNA-Seq data from blood of two MS patients carrying TT or AA genotypes for the rs7740107 variant, represented by Integrative Genomics Viewer (IGV). The region depicted corresponds to chr6:130,051,116–130,055,425 (hg19), showing exons (E) 6, 7 and 8. From bottom to top, for each sample, are represented: alignment track, joints track and reads coverage track. **(B)** Reads coverage track from RNA-Seq data of 17 MS patients grouped by rs7740107 genotypes.

the *L3MBTL3* proteins fused to the V5 peptide at the carboxyl terminus. The plasmids were transfected into HEK293T cells and protein production was detected by western blot using an anti-V5 antibody. As expected, cells transfected with the full-length isoform produced the complete protein fused to the V5 fragment. The short isoform originated 3 protein bands with sizes compatible with initiation at the first, second or third, and fourth AUG translation start codon, respectively (Fig. 4). This was also supported by sequence analysis around the translation initiation codons looking for the presence of Kozak consensus sequences (data not shown) (29).

## Discussion

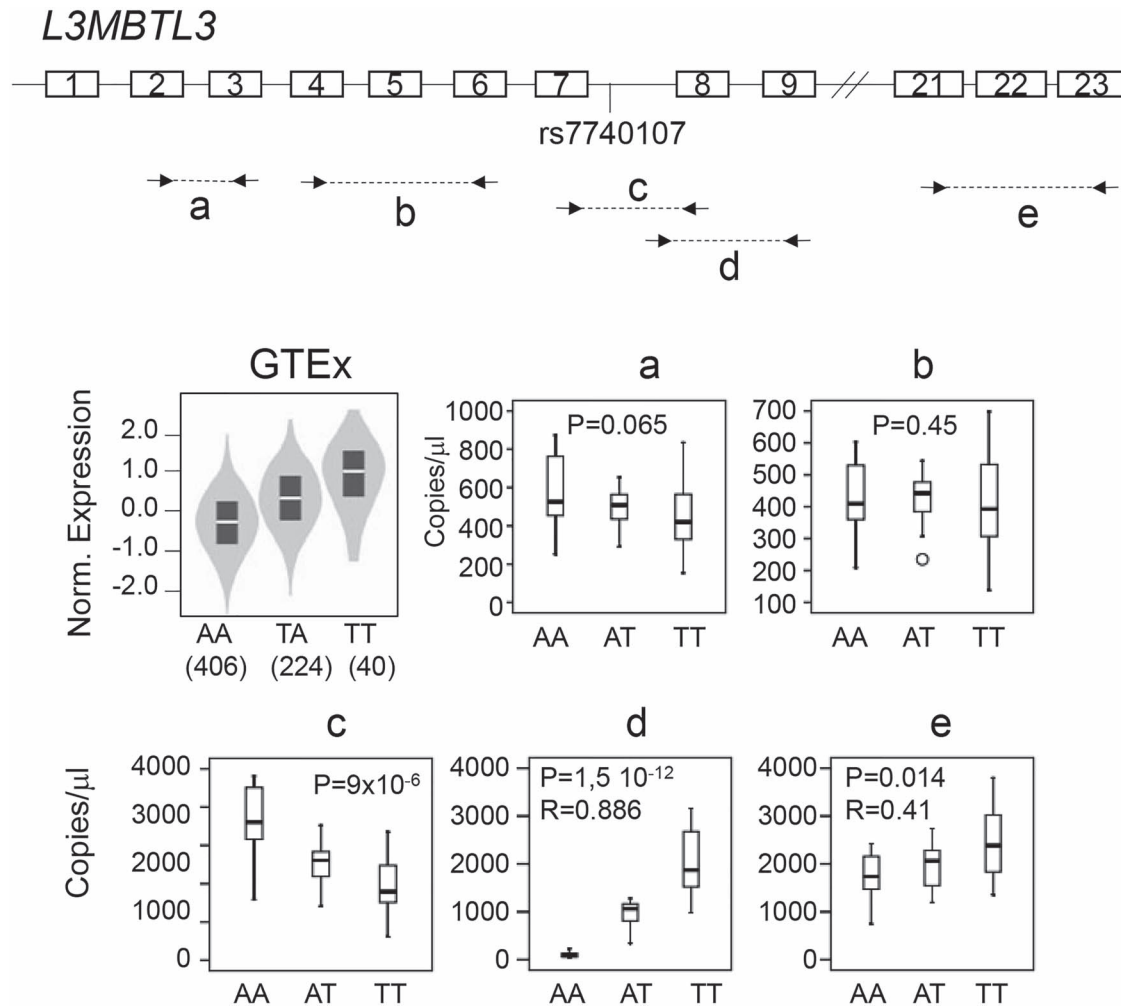
In the present study, we searched for the causal variant of the MS-associated *L3MBTL3* locus found by Andlauer et al. (10) in a GWAS. Our main finding revealed that in a fine-mapping of 3450 MS patients and 1688 healthy controls, the variant that best explained the association with MS was rs6569648. This polymorphism is a proxy for rs7740107, located in intron 7 of the *L3MBTL3* gene. We showed that this variant correlates with decreased expression of the full-length transcript of the *L3MBTL3* gene and with the initiation, in intron 7, of a new shortened isoform, which translates into three N-terminal truncated proteins.

Our results were in apparent contradiction with the eQTL described by GTEx, as they showed higher

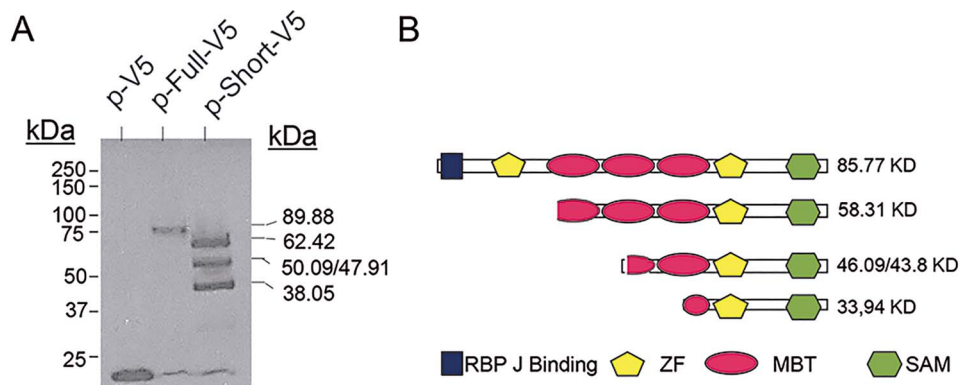
expression of *L3MBTL3* in TT carriers, whereas we observed differences in expression levels depending on the primers used for RT-digital qPCR quantification. These differences are due to the fact that GTEx performs expression quantification at the gene level, using an isoform collapsing procedure to a single transcript model for each gene. In our case, the primers were designed to capture the expression of the different isoforms, which allowed us to detect the new transcript originating from intron 7. This transcript has not been previously described in Ensembl or GENCODE databases. The GTEx results would have led us to conclude that the causal effect associated with MS protection would be the increased expression of *L3MBTL3*. Although the number of RNA-Seq reads mapped in the *L3MBTL3* gene is higher in T allele compared with A allele carrier, they are distributed in different transcripts. In addition to the eQTL, GTEx also described a sQTL in intron 7. They used LeafCutter (30), a system based on finding introns that share splice sites to detect split reads and uncover alternative intron-excision options. The sQTL from GTEx in intron 7 allowed us to design RT-digital qPCR assays around the variant to reveal the effect on *L3MBTL3* transcription. Our data suggested that the functional effect underlying the MS-association might lie in the reduction of full-length protein expression, and/or in the appearance of N-terminal truncated proteins that might have distinct activity.

As our data suggest, the truncated *L3MBTL3* proteins lost the amino-terminal region, which carries several





**Figure 3.** A novel transcript originating from intron 7 of *L3MBTL3* occurs in carriers of the rs7740107 T allele. (Top) Schematic of the *L3MBTL3* gene showing the location of the primers used for reverse transcriptase and quantification by digital qPCR. The exon number and position of rs7740107 are indicated. (Bottom) The graphs represent *L3MBTL3* expression levels with respect to rs7740107 genotypes: (GTEx) violin plot representing RNA-Seq of blood cells from GTEx; (a) to (e) box plots representing quantification by RT-digital qPCR of *L3MBTL3* expression levels using the different primers depicted at the top. RNA samples were purified from blood of 36 individuals: 14 AA, 11 AT and 11 TT carriers.



**Figure 4.** Expression of recombinant *L3MBTL3* full-length and short transcripts obtained from blood of a patient bearing rs7740107 TT alleles. (A) Western blot of extracts from HEK293T cells transfected with constructs carrying the full-length (p-Full-V5) and short (p-Short-V5) cDNA inserts, cloned into pcDNA<sup>TM</sup>6.2/V5 directional expression plasmid (p-V5). The proteins were detected with an anti-V5 antibody. Molecular weight markers are at the left of the blot. Expected molecular weight of the full-length and short protein-V5 bands are indicated at the right side of the blot. (B) Scheme of the *L3MBTL3* protein (XP\_006715641.1) with the functional domains as described by Xu et al. (11): binding site for RBPJ (recombination signal binding protein for immunoglobulin kappa J region, blue); two ZF (C2C2 zinc finger, yellow); three MBT (malignant brain tumour domain, red); one SAM (sterile alpha motif, green). The expected molecular weight of the produced proteins, according with the possible starting ATG translation sites are indicated.

functional domains. L3MBTL3 is a regulator of Notch: down-regulation of L3MBTL3 leads to up-regulation of Notch target genes, while its overexpression produces the opposite effect, i.e. repression of gene expression (11). This balanced regulation is mediated by RBPJ, which interacts with amino acids 1–64 in the N-terminus of L3MBTL3. In addition, the truncated protein has a dominant effect over the full-length protein. It is tentative to speculate that the alteration of Notch signalling in an opposite manner between the full-length and truncated proteins could be the functional reason of the locus association with MS. Notch has an important role in adult neurogenesis and oligodendrocyte differentiation, and it is also essential for T cell development (31–33).

On the other hand, the truncated proteins translated from the short transcript also lack malignant brain tumour (MBT) interaction domains. It has been demonstrated that the L3MBTL3 MBT domains can directly recognize SOX-2 methylated Lys-42 (34) and also the K142-methylated DNMT1 (35) proteins and regulates their degradation. This opens the possibility that the cells of individuals carrying rs7740107 T allele could have altered the homeostasis of SOX2 and DNMT1 proteins. SOX2, is a master transcriptional factor involved in the regulation of embryonic development, determination of cell fate and maintenance of stem-cells in the central nervous system (36,37). Moreover, DNMT1 is a major DNA methyltransferase that methylates cytosine residues in the CpG dinucleotides of the genome and is the major enzyme responsible for maintaining methylation patterns following DNA replication to preserve epigenetic inheritance (38).

As mentioned, several other traits or diseases are associated with the same polymorphism or are in very high LD with the MS variant. Among these traits, no other autoimmune diseases were found, as commonly happens in other MS-risk polymorphisms (39). These traits do not seem to have much in common, but the possible effect of the variant on Notch signalling pathway, maintenance of CpG DNA methylation patterns by DNMT1 or regulation of the self-renewal of stem cells by SOX2 are mechanisms key on the growth, development and differentiation of multiple tissues, which could explain the implication of the genetic variant in so different traits.

This work evidences the complexity of transcending the original GWAS signals towards the identification of the actual etiologic polymorphisms, the genes involved and their products and/or regulatory roles. Even though the genomic approaches open a valuable panorama, a close look also helps in their correct interpretation.

## Material and Methods

### MS patients and healthy donors

Validation and fine mapping of the L3MBTL3 locus was performed with DNA samples from a cohort of 3450 MS patients and 1688 age- and sex-matched healthy donors

of European origin, recruited from 8 Spanish MS centres. The demographic characteristics of the cohort are described in Gil-Varea et al. (12). RNA-Seq was performed on blood from 16 MS donors and 1 healthy donor. Of the MS donors, 9 had relapsing–remitting MS (RRMS) and 9 secondary progressive MS. None of the donors had received any treatment in the 3 months prior to blood collection. For RT-digital qPCR, in order to have balance number of samples of each genotype and since the frequency of TT genotype is around 6% on Caucasian population, we preselected 14 AA, 11AT and 11TT carriers out of 100 patients and 100 healthy controls. Out of the 32 samples, 12 samples were from MS patients and 20 from healthy controls. There was not significant L3MBTL3 expression difference associated within gender or disease, therefore we performed the eQTL calculation using data from MS and healthy controls combined. In [Supplementary Material, Table S3](#), we show the demographic characteristics of the cohort used for expression analysis. All participants signed the informed consent form and the study was approved by the Andalusian Ethics Committee.

### SNPs genotyping

Genotyping of rs4364506 was performed using iPLEX® Gold chemistry at the National Genotyping Center (CeGen-PRB2, USC node). The remaining SNPs genotyped for fine mapping were performed in the same patient and healthy donor cohorts using pre-designed TaqMan assays by Thermo Fisher Scientific (Thermo Scientific, IL, USA). The assays are listed in [Supplementary Material, Table S2](#). Random samples were genotyped in duplicate to check for reproducibility. Genotyping data were analysed with PLINK V.1.07 (40). Logistic regression models were calculated to determine the independence of effects. Genotypes of rs7740107 from blood donor were obtained from GWAS data performed with Infinium Global Screening Array, GSA (Illumina, CA, USA). Deviation from Hardy–Weinberg equilibrium for all SNPs was tested by exact test.

### RNA-Seq from blood

The RNA-Seq data were obtained from seventeen RNA samples extracted from stabilized whole blood using the PAXgene Blood RNA Kit (QIAGEN GmbH, Hilden, Germany) and quantified using Nanodrop One® and Qubit® (Thermo Scientific, IL, USA). Library preparation and Illumina sequencing were carried out at the IPBLN Genomics Facility (CSIC, Granada, Spain). Total RNA quality was verified by Bioanalyzer RNA 6000 Nano chip electrophoresis (Agilent Technologies, CA, USA). All RNA samples showed a RIN value above 8. In order to increase detection sensitivity, the alpha and beta globin mRNA was depleted using hybridization technology. Briefly, for each sample, 400 ng of total RNA were treated with the GLOBInclear™-Human Kit (Thermo Scientific, IL, USA), with a globin mRNA depletion efficiency >95% (tested by qPCR in a representative set of samples). The totality of recovered

RNA was used as input for RNA-seq library construction performed with Truseq® stranded mRNA kit (Illumina, CA, USA) and IDT for Illumina® Truseq® UD RNA indexes kit. Quality and size distribution of PCR-enriched libraries was validated through Bioanalyzer High-Sensitivity DNA assay and the concentration was measured on the Qubit® fluorometer. Final libraries were pooled in an equimolecular manner, and then diluted and denatured as recommended by Illumina NextSeq 500 library preparation guide. The 75 × 2 nt paired-end sequencing was conducted on a NextSeq 500 sequencer, in two consecutive runs, using the high output configuration and producing 41.591.000 raw paired reads sample on average.

### RNA-Seq data analysis

To analyse transcriptomic samples, miARma-Seq pipeline was used (41). This workflow performs all steps from raw data to differential expressed genes (DEG) calculation. Firstly, raw data were evaluated using FastQC software to examine the quality of the reads (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). No adapter accumulation or bad quality reads ( $q < 30$ ) were found. In the second step, miARma-Seq aligns all sequences using HISAT2 (42), resulting in a 95.98% of properly aligned reads. With this aim, *Homo sapiens* Gencode v26 genome-build: GRCh37 was used as reference genome. Data were deposited at BioProject database with accession number PRJNA732389.

### RNA quantification by digital qPCR

To quantitate the expression of the *L3MBTL3* gene by RT-digital qPCR, 1500 ng of RNA were reverse transcribed into cDNA with oligodT and random primers, according to Superscript IV First-Strand Synthesis SuperMix protocol (Thermo Scientific, IL, USA). We performed digital qPCR in duplicates using QX200 Droplet Digital PCR EvaGreen Supermix (BioRad, USA). Data was analysed with QuantaSoft (BioRad, CA USA), which counts the fluorescent positive and negative droplets to calculate target DNA concentration. Primers for digital qPCR amplification were designed with Primer 3 V1.07 software, and their sequences are shown in [Supplementary Material, Table S2](#).

### Amplification of the *L3MBTL3* full and short isoforms

Blood cells cDNA from a carrier of rs7740107\*TT was amplified by PCR using the Q5® High-Fidelity DNA Polymerase (New England BioLab, MA, USA). The full isoform of *L3MBTL3* was amplified using a forward primer located in the 5' UTR, position chr6:130022261–130022280 hg38. The short *L3MBTL3* isoform was amplified using a forward primer located in intron 7 of the *L3MBTL3* gene at position chr6:130055117–130055136. These two primers have 4 nt with sequence 5'-CACC for directional insertion into the plasmid. For both isoforms, the same reverse primer located at the last 20 coding nucleotides of the gene, position chr6:130012271–130012270, was used.

PCR conditions were: 1 cycle of 94°C for 3 min; 34 cycles of 94°C for 30 s; 60°C for 30 s; 72°C for 2 min.

### Cloning and expression of the *L3MBTL3* isoforms in HEK293T

PCR products were cloned into the expression plasmid pcDNA6.2/V5/GW/D-TOPO (Thermo Scientific, IL, USA). Positive colonies were validated by Sanger sequencing. The full isoform cloned corresponded to the ENST00000533560 transcript, and the short isoform to the sequence of 35 bp of intron 7 near the splice donor plus exons 8 to 23. The recombinant plasmids for transfection were purified with the EndoFree Plasmid Maxi Kit (QIAGEN GmbH, Hilden, Germany). The HEK293T (human embryonic kidney [epithelial]) cell line was cultured in DMEM high glucose medium supplemented with 10% foetal calf serum and 100 U/ml penicillin/streptomycin (all from Gibco, Thermo Scientific, IL, USA). Following the manufacturer's instructions,  $2 \times 10^5$  cells/well were seeded in a six-well plate and transfected the day after with the different recombinant plasmids and the control plasmid pcDNA™6.2/V5/GW-CAT using LTX lipofectamine reagent (Thermo Scientific, IL, USA). After 24 h of culture, cell extracts were dissolved in Laemmli buffer and processed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) at 10% concentration and transferred to PVDF-P transfer membranes (Thermo Scientific, IL, USA). The blot membrane was incubated with an antibody against the V5 label (ab27671, Abcam, Cambridge, United Kingdom) and an HRP-conjugated mouse anti-IgG secondary antibody (H + L) (#62–6520, ThermoScientific, IL, USA). Protein bands were revealed with Pierce ECL Western Blotting substrate (Thermo Scientific, IL, USA). The molecular weight marker used as a guide was Precision Plus Protein™ All Blue Prestained Protein Standards (1 610 373, Bio-Rad, CA, USA).

### Supplementary Material

[Supplementary Material](#) is available at HMG online.

### Acknowledgements

We want to particularly acknowledge the patients and the Nodo Biobanco HU Virgen Macarena [PT20/00069] (Biobanco del Sistema Sanitario Público de Andalucía) integrated in the Spanish National Biobanks Network and supported by Instituto de Salud Carlos III-FEDER for its collaboration.

*Conflict of Interest statement.* None declared.

### Funding

Agencia Española de Investigación-Fondo Europeo de Desarrollo Regional (Feder) (PID2019-110487R-C21, PID2019-110487R-C22); Junta de Andalucía (P18-RT-2623); Instituto de Salud Carlos III and co-funded by the

European Regional Development Fund (ERDF) 'A way to make Europe' (PI20/01259, REEM RD16/0015/0016, REEM RD16/0015/0002); Fundacion Esclerosis Múltiple España.

## References

- Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B. and Ciccarelli, O. (2018) Multiple sclerosis. *Lancet*, **391**, 1622–1636.
- Patsopoulos, N.A. (2018) Genetics of multiple sclerosis: an overview and new directions. *Cold Spring Harb. Perspect. Med.*, **8**, a028951.
- International Multiple Sclerosis Genetics Consortium (2019) Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*, **365**, eaav7188.
- Nica, A.C. and Dermitzakis, E.T. (2013) Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.*, **368**, 20120362.
- Matesanz, F., Potenciano, V., Fedetz, M., Ramos-Mozo, P., Abad-Grau, M., Del, M., Karaky, M., Barrionuevo, C., Izquierdo, G., Ruiz-Peña, J.L., García-Sánchez, M.I. et al. (2015) A functional variant that affects exon-skipping and protein expression of SP140 as genetic mechanism predisposing to multiple sclerosis. *Hum. Mol. Genet.*, **24**, 5619–5627.
- Matesanz, F., Fedetz, M., Barrionuevo, C., Karaky, M., Catalá-Rabasa, A., Potenciano, V., Bello-Morales, R., López-Guerrero, J.A. and Alcina, A. (2016) A splice variant in the ACSL5 gene relates migraine with fatty acid activation in mitochondria. *Eur. J. Hum. Genet.*, **24**, 1572–1577.
- Pala, M., Zappala, Z., Marongiu, M., Li, X., Davis, J.R., Cusano, R., Crobu, F., Kukurba, K.R., Gloudemans, M.J., Reinier, F. et al. (2017) Population- and individual-specific regulatory variation in Sardinia. *Nat. Genet.*, **49**, 700–707.
- Pérez-Núñez, I., Karaky, M., Fedetz, M., Barrionuevo, C., Izquierdo, G., Matesanz, F. and Alcina, A. (2019) Splice-site variant in ACSL5: a marker promoting opposing effect on cell viability and protein expression. *Eur. J. Hum. Genet.*, **27**, 1836–1844.
- GTEX Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
- Andlauer, T.F., Buck, D., Antony, G., Bayas, A., Bechmann, L., Berthele, A., Chan, A., Gasperi, C., Gold, R., Graetz, C. et al. (2016) Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci. Adv.*, **2**, e1501678.
- Xu, T., Park, S.S., Giaimo, B.D., Hall, D., Ferrante, F., Ho, D.M., Hori, K., Anhezini, L., Ertl, I., Bartkuhn, M. et al. (2017) RBPJ/CBF1 interacts with L3MBTL3/MBT1 to promote repression of notch signaling via histone demethylase KDM1A/LSD1. *EMBO J.*, **36**, 3232–3249.
- Gil-Varea, E., Fedetz, M., Eixarch, H., Spataro, N., Villar, L.M., Urcey, E., Saiz, A., Fernández, Ó., Leyva, L., Ramió-Torrentà, L. et al. (2020) New risk variant for multiple sclerosis at 11q23.3 locus is associated with expansion of CXCR5+ circulating regulatory T cells. *J. Clin. Med.*, **9**, 625.
- Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., Tassé, A.M. and Flicek, P. (2017) The international genome sample resource (IGSR): a worldwide collection of genome variation incorporating the 1000 genomes project data. *Nucleic Acids Res.*, **45**, D854–D859.
- Vuckovic, D., Bao, E.L., Akbari, P., Lareau, C.A., Mousas, A., Jiang, T., Chen, M.H., Raffield, L.M., Tardaguila, M., Huffman, J.E. et al. (2020) The polygenic and monogenic basis of blood traits and diseases. *Cell*, **182**, 1214–1231.e11.
- Chen, M.H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D. et al. (2020) Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell*, **182**, 1198–1213.e14.
- Akiyama, M., Ishigaki, K., Sakaue, S., Momozawa, Y., Horikoshi, M., Hirata, M., Matsuda, K., Ikegawa, S., Takahashi, A., Kanai, M. et al. (2019) Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.*, **10**, 4393.
- Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B. and Price, A.L. (2019) Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.*, **104**, 65–75.
- Hoffmann, T.J., Choquet, H., Yin, J., Banda, Y., Kvale, M.N., Glymour, M., Schaefer, C., Risch, N. and Jorgenson, E. (2018) A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics*, **210**, 499–515.
- Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A. et al. (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature*, **551**, 92–94.
- Milne, R.L., Kuchenbaecker, K.B., Michailidou, K., Beesley, J., Kar, S., Lindström, S., Hui, S., Lemaçon, A., Soucy, P., Dennis, J. et al. (2017) Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat. Genet.*, **49**, 1767–1778.
- Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklika, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A. et al. (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*, **167**, 1415, e19–1429.
- Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.
- Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E. et al. (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, **518**, 187–196.
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
- Hernandez Cordero, A.I., Gonzales, N.M., Parker, C.C., Sokolof, G., Vandenberg, D.J., Cheng, R., Abney, M., Sko, A., Douglas, A., Palmer, A.A. et al. (2019) Genome-wide associations reveal human-mouse genetic convergence and modifiers of Myogenesis, CPNE1 and STC2. *Am. J. Hum. Genet.*, **105**, 1222–1236.
- Wuttke, M., Li, Y., Li, M., Sieber, K.B., Feitosa, M.F., Gorski, M., Tin, A., Wang, L., Chu, A.Y., Hoppmann, A. et al. (2019) A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.*, **5**, 957–972.
- Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y. et al. (2019) Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *Hum. Mol. Genet.*, **28**, 166–174.
- Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z. et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125–8148.



30. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K. and Pritchard, J.K. (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.*, **50**, 151–158.
31. Ho, D.M., Artavanis-Tsakonas, S. and Louvi, A. (2020) The notch pathway in CNS homeostasis and neurodegeneration. *Wiley Interdiscip. Rev. Dev. Biol.*, **9**, e358.
32. Li, C., Xie, Z., Xing, Z., Zhu, H., Zhou, W., Xie, S., Zhang, Z. and Li, M.H. (2021) The notch Signaling pathway regulates differentiation of NG2 cells into oligodendrocytes in demyelinating diseases. *Cell. Mol. Neurobiol.* [10.1007/s10571-021-01089-0](https://doi.org/10.1007/s10571-021-01089-0).
33. Yuan, J.S., Kousis, P.C., Suliman, S., Visan, I. and Guidos, C.J. (2010) Functions of notch signaling in the immune system: consensus and controversies. *Annu. Rev. Immunol.*, **28**, 343–365.
34. Zhang, C., Leng, F., Saxena, L., Hoang, N., Yu, J., Alejo, S., Lee, L., Qi, D., Lu, F., Sun, H. et al. (2019) Proteolysis of methylated SOX2 protein is regulated by L3MBTL3 and CRL4(DCAF5) ubiquitin ligase. *J. Biol. Chem.*, **294**, 476–489.
35. Leng, F., Yu, J., Zhang, C., Alejo, S., Hoang, N., Sun, H., Lu, F. and Zhang, H. (2018) Methylated DNMT1 and E2F1 are targeted for proteolysis by L3MBTL3 and CRL4 (DCAF5) ubiquitin ligase. *Nat. Commun.*, **9**, 1641.
36. Rizzino, A. (2013) Concise review: the Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells*, **31**, 1033–1039.
37. Pevny, L.H. and Nicolis, S.K. (2010) Sox2 roles in neural stem cells. *Int. J. Biochem. Cell Biol.*, **4**, 421–424.
38. Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T. et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908–912.
39. International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kempainen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D. et al. (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.*, **45**, 1353–1360.
40. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
41. Andres-Leon, E., Nunez-Torres, R. and Rojas, A.M. (2016) miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Sci. Rep.*, **6**, 25749.
42. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.