# Implication of DNA methylation changes at chromosome 1q21.1 in the brain pathology of Primary Progressive Multiple Sclerosis

Majid Pahlevan Kakhki[1], Chiara Starvaggi Cucuzza[1], Alexandra Gyllenberg[1], Tejaswi Venkata S. Badam[1,2], Yun Liu[3], Sanjaykumar Boddul[4], Tojo James[1], Fredrik Wermeling[4], Mika Gustafsson[2], Patrizia Casaccia[5], Ingrid Kockum[1], Jan Hillert[1], Tomas Olsson[1], Lara Kular[1]*# and Maja Jagodic[1]*#.

[1]Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden.

[2]Department of Bioinformatics, Institute for Physics chemistry and Biology (IFM), Linköping university, Linköping, Sweden.

[3]MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences and Shanghai Xuhui Central Hospital, Fudan University, Shanghai, China.

[4]Department of Medicine, Solna, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden.

[5]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, USA.

*Contributed equally

#Corresponding author: Maja Jagodic and Lara Kular, Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. Email: maja.jagodic@ki.se and lara.kular@ki.se

## Abstract

Multiple Sclerosis (MS) is a highly heterogenous inflammatory and neurodegenerative disease of the central nervous system with unpredictable course towards progressive disability. Understanding and treating progressive forms of MS remains extremely challenging due to the limited knowledge of the underlying mechanisms. We examined the molecular changes that associate with primary progressive MS (PPMS) using a cross-tissue (blood and post-mortem brain) and multilayered data (genetic, epigenetic, transcriptomic) from independent cohorts. We first identified and replicated hypermethylation of an intergenic region within the chromosome 1q21.1 locus in the blood of PPMS patients compared to other MS patients and healthy individuals using Illumina 450K arrays and pyrosequencing. We next revealed that methylation is under the control of genetic variation within the extended locus both in the blood and brain. Several genetic variants also affected expression of proximal genes (*CHD1L*, *PRKAB2* and *FMO5*) in the brain and displayed evidence of the association with the risk of developing PPMS, suggesting a genetic-epigenetic-transcriptional interplay in PPMS pathogenesis. We next addressed the causal link between DNA methylation and gene expression using reporter systems and dCas9-TET1-induced demethylation of CpGs in the identified region, which resulted in upregulation of *CHD1L* and *PRKAB2* genes in SH-SY5Y neuron-like cells. Independent exploration using unbiased correlation network analysis, confirmed the putative implication of *CHD1L* and *PRKAB2* in brain processes in PPMS patients. We provide several lines of evidence suggesting distinct molecular changes in the 1q21.1 region, known to associate with brain development and disorders, associated with genetic predisposition to high methylation in PPSM patients that regulates the expression of genes within the extended locus.

**Keywords:** Multiple Sclerosis, progression, omics, DNA methylation, 1q21.1, CRISPR/dCas9, brain pathology, neurodegeneration.

## Background

Multiple Sclerosis (MS) is an inflammatory disease of the central nervous system (CNS) affecting young adults and leading to unpredictable and progressive physical and cognitive impairments. The pathological hallmarks of MS are represented by focal plaques of primary demyelination and diffuse neurodegeneration in the grey and white matter of the brain and spinal cord[1]. The most common form of disease, relapsing-remitting MS (RRMS), affects twice as many women as men and presents itself with neurological relapses followed by periods of partial or complete remission. In most cases, this inflammatory stage is followed by a secondary progressive phase of MS (SPMS) after several years. About 10-15% of patients, manifest a primary progressive form of MS disease (PPMS) with uninterrupted progression starting from disease onset, although relapses may occur[2]. Beyond epidemiological differences, both progressive forms of MS share similar features that differ from RRMS stage, notably a later onset (40 years vs. 30 years in RRMS) and a prevalence of CNS-intrinsic inflammatory and degenerative pathological processes as opposed to the predominant role of peripheral immune influence in RRMS stage[3]. However, while considerable progress has been achieved in deciphering the genetic variation predisposing for disease, the mechanisms underpinning disease progression and severity remain unresolved. Importantly, conventional immunomodulatory therapies are ineffective in progressive patients[4,5], reinforcing the need to better clarify processes underlying MS courses.

Disease trajectory likely relies on a complex interplay between genetic and non-genetic factors[6]. Epigenetic modifications such as DNA methylation, intersecting internal and external influences might provide novel opportunity to study the mechanisms underlying the different forms of MS disease. DNA methylation, which regulates gene expression without altering the genetic code, is the most studied epigenetic mark and its implications in MS pathogenesis have been reported in immune and nervous cells of MS patients[7-12]. We have shown that the major risk variant *HLA-DRB1*15:01* may mediate risk for MS via changes in *HLA-DRB1* DNA methylation and subsequent expression[11]. Recent comparative whole-genome DNA methylation studies shed further light on molecular mechanisms behind MS pathogenesis[9,13], the vast majority of them focusing on RRMS stage of disease.

In this study, we aimed to examine the molecular changes that associate with MS course. Using a multilayered (genetic, epigenetic, transcriptomic) and cross-tissue (blood, brain) approach in cohorts of MS patients and controls, we demonstrate an interplay between regional genetic variation, DNA methylation and gene expression in the chromosome 1q21.1 region, in PPMS patients specifically. We addressed the link between these three molecular layers using in-vitro reporter gene assays and

CRISPR/dCas9 epigenome editing. Our findings suggest that genetically controlled methylation at 1q21.1 locus might contribute to PPMS pathology though transcriptional regulation of proximal genes such as *CHD1L* and *PRKAB2*.

## Methods

### Cohorts

Details of the cohorts are described in the Supplementary Table 1. Briefly, for genome-wide DNA methylation and methylation quantitative trait locus (meQTL) analysis used in cohort 1, peripheral blood samples were collected form 140 MS, patients including 120 RRMS, 4 PPMS and 16 SPMS patients, and 139 healthy individuals (HC), as previously described[11]. An independent cohort, cohort 2, consisting of 48 RRMS and 36 PPMS patients (matched for age, sex, disease duration and Swedish descent) was used for pyrosequencing validation and locus-specific meQTL analysis. Cohort 3 used for genetic association study comprises 477 PPMS and 3871 RRMS patients. Gene expression data (RNA-sequencing) from bulk brain tissue samples of progressive MS patients (5 PPMS and 7 SPMS) and non-neurological controls (n = 10), previously described[9], were used for correlation network analysis. Additionally, we utilized publicly available databases from xQTL serve[14] (n = 543 bulk prefrontal cortex samples, http://mostafavilab.stat.ubc.ca/xQTLServe/ ) and GTEx (selecting all available nervous tissues, https://gtexportal.org/) platforms to address meQTL and eQTL effects in the CNS.

### Genome-wide DNA methylation and meQTL analyses (cohort 1)

*DNA methylation analysis.* The methylation data from Infinium HumanMethylation450 (450K) arrays was preprocessed as previously described[11] using the Illumina default procedure implemented in the Bioconductor minfi package[15]. Briefly, all samples were normalized together using the minfi preprocessQuantile function. Cell counts for the six major cell types in blood (granulocytes, B cells, $CD4^+$ T cells, $CD8^+$ T cells, monocytes, and NK cells) for each individual were estimated using the estimateCellCounts function in minfi package[15], which obtain sample-specific estimates of cell proportions based on reference information on cell-specific methylation signatures[16]. To identify differentially methylated regions (DMRs) associated with the PPMS phenotype, we used the bumphunter function in minfi package[15] with adjustment for confounders: age, sex, self-reported smoking status (ever smokers vs. never smokers), hybridization date, and the first two principle components of estimated differential

cell counts. Region that has a family wise error rate (FWER) less than 0.05 with 1000 resamples and contains at least 2 probes was identified as a trait-associated DMR.

*Methylation QTL analysis.* To identify potential genetic dependency, the PPMS-associated DMR was tested for association with genotype (594,262 SNPs) using an additive minor-allele dosage model. Genotype-DMR associations were corrected for multiple testing using a stringent Bonferroni-adjusted threshold of 0.05.

**Locus-specific DNA methylation and meQTL analyses (cohort 2)**

*DNA methylation analysis.* For validation of the identified PPMS-associated DMR, pyrosequencing analysis was performed using 500 ng of genomic DNA samples previously converted to bisulfite DNA (BS-DNA, EZ DNA methylation kit, ZYMO research) with PyroMark Q96 system (Qiagen). Primers and probes for three sequencing assays covering 7 CpG sites in the locus were designed by PyroMark Design software (Qiagen) (Supplementary Table 2, Supplementary Fig. 1). Around 10 ng of BS-DNA was amplified using PyroMark PCR kit (Qiagen) and the forward and 5'-biotinylated reverse primers. The entire PCR product, 4 pmol of the sequencing probe and streptavidin sepharose high-performance beads (GR Healthcare) were used for pyrosequencing on a PyroMark Q96 ID pyrosequencing instrument (Qiagen) using the PyroMark Gold 96 Reagent kit (Qiagen). Data analysis was done with PyroMark CpG software 1.0.11 (Qiagen). To test the differences in DNA methylation between PPMS and RRMS patients for each CpG site, Kruskal-Wallis test with Dunn's multiple comparison test was applied with GraphPad Prism software (PRISM 7.0; GraphPAD Software Inc., San Diego, CA, USA).

*Methylation QTL analysis*. Methylation data was RANK transformed in R using R Core team (Vienna, Austria, https://www.R-project.org/). Genotyping was carried out at deCODE (deCODE genetics/Amgen, Reykjavik, Iceland) using Illumina OmniExpress chip with 716,503 SNPs mapped to the Human Assembly Feb.2009 (GRCh37/hg19). Of 84 individuals, 83 where genotyped in deCODE and 82 of them passed QC. We performed meQTL analysis of chromosome 1 from bp 146500000 to bp 147000000, in PLINK[17] excluding SNPs with less than 98% genotyping rate and SNPs that were not in Hardy-Weinberg equilibrium ($p < 0.05$) and corrected for 5 population based (ancestral informative markers) principal component analysis covariates. After quality control, 123 SNPs remained in the region. Genotype-CpG associations were corrected for multiple testing using stringent Bonferroni-adjusted threshold of 0.05. Two markers passed Bonferroni correction of $p < 10^{-8}$, RS1969869 and RS4950357, for all seven CpG sites in the analysis and two others reached Bonferroni correction of $p < 10^{-5}$.

**Genetic association analysis (cohort 3)**

Genotyping of Human OMNI (HOMNI) cohort (cohort 3) was conducted at deCODE (deCODE genetics/Amgen, Reykjavik, Iceland) using Illumina OmniExpress chip with 716,503 SNPs mapped to the Human Assembly Feb.2009 (GRCh37/hg19). SNP association analysis of chromosome 1 region from bp 146500000 to bp 147000000, was performed in PLINK[17]. A total of 477 PPMS and 3871 RRMS were used. After quality control, excluding SNPs with less than 98% genotyping rate and SNPs that were not in Hardy-Weinberg equilibrium ($p < 0.05$) and corrected for 5 population based (ancestral informative markers) principal component analysis covariates, 115 SNPs remained in the region, covering all locus-specific meQTL SNPs (cohort 2). Eight out of the 19 markers derived from genome-wide meQTL analysis (cohort 1), could be analyzed in deCODE HOMNI platform. Among the meQTL-SNPs, SNPs that could not be analyzed, two displayed HWE-issue and were removed and 11 were not present in HOMNI.

**In-vitro methylation assay**

To address the regulatory features of the identified DMR, we used in-vitro DNA methylation reporter assay. A 927 bp fragment encompassing the identified DMR was amplified using primers containing overhanging *SpeI* and *NsiI* restriction sites (Supplementary Table 2). We used blood genomic DNA from PPMS patients presenting with low (rs1969869: CC) and high (rs1969869:AA) methylation levels at the identified DMR. The amplified products in direct and reverse orientation were inserted into pCpG-free promoter vector (Invivogen) containing a Lucia luciferase reporter and into a pCpG-free basic vector (Invivogen) containing a murine secreted embryonic alkaline phosphatase (mSEAP) reporter gene for assessment of enhancer and promoter activity, respectively. As the body of these vectors is devoid of any CpGs, any impact of DNA methylation on reporter gene expression is restricted to the inserted fragment only. All the constructs were either completely methylated (57 CpGs) using *M.SssI* or partially methylated (7 CpGs residing in the GCGC sequence) by *HhaI* methyltransferases (New England BioLabs) using 1 ug of the vectors and 1 unit of the enzymes. The mock methylated control was treated equally but in absence of any methyltransferases and corresponds to unmethylated inserts. After the purification of the methylated, partially- and mock-methylated constructs (QIAquick PCR purification Kit, Qiagen), the efficiency of methylation was assessed using an EpiJET DNA Methylation analysis Kit (*MspI*/*HpaII*) (ThermoFisher Scientific), followed by gel electrophoresis (Supplementary Fig. 2). Original vectors treated by *M.SssI* and *HhaI* or mock-treated were used as controls. Human embryonic kidney HEK293T cells were

cultured in Dulbecco's Modified Eagle's medium in 96 well plates and co-transfected with 90 ng of the Lucia or SEAP constructs and 5 ng of the control vector pGL4-TK-hH Luc constitutively expressing Renilla luciferase, using Lipofectamine3000 Transfection Reagent (Qiagen). Approximately, 48 hours post transfection, Lucia, SEAP and Renilla activities were measured using QUANTI-Luc (Invivogen), the Phospha-Light System (Applied Biosystems) and the Dual-Glo Luciferase Assay System (Promega), respectively, according to manufacturer's instructions, on the GloMax 96 Microplate Luminometer (Promega). Both direct and reverse orientations of the sequence were tested. Lucia or SEAP signals were normalized against Renilla (triplicate) and experiments were replicated at least two times.

**CRISPR/dCas9-TET1 epigenome editing**

*dCas9-TET1 and gRNA generation*. Details and maps of the final constructs used in this study are presented in Supplementary Figure 3. Briefly, we engineered a P3-dCas9-Tet1-GFP-Puro construct as follows. First, to be able to express the gRNAs from the same vector, we mutated the *BbsI* sites in the TET1 sequence (without changing the protein sequence) synthesized by Eurofins (Eurofins MWG Operon Ebersberg, Germany). We then utilized the backbone of a pdCas9-DNMT3A-EGFP plasmid (Addgene #71666)[18] and replaced DNMT3A with TET1 sequence. This cassette was previously engineered to express the original EGFP sequence in a double reporter cassette containing EGFP-T2A-Puromycin under the control of an independent CMV promoter to allow sufficient expression of GFP signal for post-transfection cell sorting. We proceeded similarly with the TET1-IM construct which expresses a deactivated TET1 catalytic unit. The final plasmid expressed dCas9-TET1 (or dCas9-TET1-IM) and CMV-EGFP-T2A-Puromycin double marker unit. All gRNAs were designed by CRISPOR Version 4.98[19] both on the sense and antisense strands, with sequence and mapping presented in the Supplementary Table 2 and Supplementary Figure 2.

To address dCas9-TET1-mediated epigenome editing in SH-SY5Y cell line, we utilized the lentivirus version of the cassettes. We used Fuw-dCas9-Tet1CD (Addgene #84475) and Fuw-dCas9-Tet1CD_IM (Addgene #84479) plasmids and added the EGFP marker to facilitate sorting of positively transduced cells (new plasmids named as P3-Lenti-dCas9-Tet1-GFP and P3-Lenti-dCas9-Tet1IM-GFP). We used pKLV2-U6gRNA3(BbsI)-PGKpuro2ABFP (Addgene #67990) to express the gRNA and replaced the BFP with mCherry (named as P3-pKLV2-U6gRNA(BbsI)-PGKpuro2A-mCherry). All constructs were confirmed by Sanger sequencing (KiGene) and chromatograms were analyzed using SnapGene software 5.2.3 (GSLBiotech).

*dCas9-TET1 delivery to HEK293T and SH-SY5Y cells*. To test the efficiency of epigenome editing, we exploited HEK293T ease of transfection and performed gRNAs screen. Different gRNAs were transfected either individually or in combination based on the target site, using Lipofectamine 3000 (Invitrogen). For all experiments both on HEK293T and SH-SY5Y cells, DNA was extracted after 72 hours and bisulfite conversion was performed using 200 ng of the extracted DNA (BS-DNA, EZ DNA methylation kit, ZYMO research). DNA methylation was assessed using pyrosequencing, as described above. For all experiments conducted on SH-SY5Y cells, we delivered a mix of the two gRNAs that showed the highest efficiency in reducing methylation in HEK293T cells (gRNA #2 and #3).

For the generation of lentiviruses and subsequent transduction of SH-SY5Y cells, HEK293T cells were seeded (~40-50 % confluent) in 2 ml DMEM with 10% serum and 2 mM glutamine per well in 6-well plates. The next day, cells had reached 60-70 % confluency and were co-transfected with transfer plasmids as P3-Lenti-dCas9-Tet1-GFP or P3-Lenti-dCas9-Tet1IM-GFP, as well as P3-pKLV2-U6gRNA(BbsI)-PGKpuro2A-mCherry, psPAX2 (Addgene #12260) and pMD2.G (Addgene #12259) using Lipofectamine 3000 (Invitrogen) transfection reagent as per manufacturer instructions. After overnight incubation, medium was removed, and fresh medium added containing 5% serum and 1X Pen-strep. Cells were further incubated for 48h and 72h post-transfection and the virus containing supernatant collected and centrifuged for 10 mins at 500g to remove any cell debris. The supernatant was concentrated using Lentivirus Precipitation Solution (cat. #VC100; AlStem) according to the manufacturer's instructions and used to transduce SH-SY5Y cells. SH-SY5Y cell line was transduced with virus in complete medium (DMEM with 10% serum and 1X Pen-Strep) containing 8 µg/ml Polybrene (Sigma) with spin infection. Medium was changed 3-4 times to remove any virus residues and cells were used for further assays.

**qPCR analysis**

Total RNA and DNA were extracted using AllPrep DNA/RNA Kits (Qiagen) according to the manufacture instruction. RNA and DNA concentrations and quality were verified by QIAxpert (Qiagen). Reverse transcription of RNA was performed using the manufacturer's instructions of iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Inc., CA) with OligodT and Random Hexamer primers, generating cDNA for subsequent gene expression analysis. Real-time PCR was performed on a BioRad CFX384 Touch Real-Time PCR Detection System using iQ™ SYBR® Green Supermix (Bio-Rad Laboratories, Inc., CA) in a three step PCR: 95 °C:3 min, followed by 40 cycles of 95 °C:10 s, 60 °C:30 s and 72 °C:30 s. The relative expressions of the selected genes were normalized to the reference gene *GAPDH*. The specificity of real time PCR

reaction was verified by the melt curve analysis. The expression level of selected genes were analyzed using ΔΔCT method[20] and compared via independent t test. All statistical analyses were performed in GraphPad Prism 6 and 7 (GraphPad Software).

**Correlation network analysis in MS brain**

*Raw data analysis.* The fastq files corresponding to bulk gene expression (RNA-sequencing) data from brain tissue samples of progressive MS patients (n = 12) and non-neurological controls (n = 10)[9] were extracted from the RAW RNA sequence files and checked for quality control using multiqc software to make them ready for alignment[21]. After trimming using the trimgalore program[22], fastqc files were aligned and annotated using STAR aligner and Stringtie software[23] by applying human hg38 refseq information from UCSC. The analysis was performed on the extracted count matrix using bash and Python.

*Network analysis.* In order to utilize a brain-specific network module, we applied a previously established bioinformatic pipeline utilizing co-expression network analysis[24], as briefly described below. The count matrix was loaded into R(3.6.1) environment and quantile normalized using the glimma package[25]. Spearman correlation is applied on every gene pair and permuted for 10000 times to determine which interactions are significant (FDR < 0.05), which avoids biased filtering of the network based on correlation R value. The function also integrates hub connectivity significance for including only the interactions that have significant connectivity in the network. The produced network consisted of 5 million interactions among 27,059 genes, which limits the inherent resolution for defining modules which overlooks the multiscale organization of the network where compact clusters co-exist. In order to overcome this limitation, the correlation network was embedded on a spherical surface, thereby creating a planar maximally filtered network devoid of cross links. The final network consisted of 0.5 million interactions among 27,059 expressed genes from the RNAseq data. The planar maximally filtered graph is then clustered by implementing multiscale clustering algorithm (MCA) from the MEGENA package in R. MCA incorporates three distinct criteria to identify locally coherent clusters while maintaining a globally optimal partition. First, shortest path distances are utilized to optimize within-cluster compactness. Second, local path index is used to optimize local clustering structure. Third, overall modularity is employed to identify optimal partition. The final network clustered into 757 non-overlapping modules.

*Cluster Trait association analysis.* Principal component analysis (PCA) is first performed for each cluster. Next, correlation between the first principal component and each trait was computed as cluster relevance

to the trait. The 757 clusters identified from the correlation network were evaluated for the relevance to PPMS, SPMS and control phenotypes. Three clustered passed FDR P-value < 0.05.

## Results

### Hypermethylation of an intergenic region in the 1q21.1 locus of PPMS patients

We profiled genome-wide DNA methylation in the blood of RRMS, SPMS and rare PPMS patients and matched HC (cohort 1, n = 279, Fig. 1, Supplementary Table 1), using the Illumina Infinium HumanMethylation450 BeadChip (450K array). After adjustment for confounders, we identified one differentially methylated region (DMR) in PPMS patients in comparison with HC ($P = 6.16 \times 10^{-6}$), RRMS ($P = 2.51 \times 10^{-6}$) and SPMS ($P = 7.33 \times 10^{-6}$) patients (FWER < 0.05, Fig. 2a, Supplementary Table 3). The DMR, located in an intergenic CG-rich region on chromosome 1 (q21.1): 146549909–146551201 (GRCh37/hg19), consists of 8 consecutive CpG probes displaying hypermethylation in PPMS patients compared to RRMS, SPMS as well as HC (Fig. 2b). Considering the low number of PPMS patients (n = 4), we selected an independent cohort for replication including 36 PPMS and 48 RRMS patients, matched for gender, age and disease duration (cohort 2, Fig. 1, Supplementary Table 1). We designed pyrosequencing assays covering seven CpG sites, including the 450K probes exhibiting the largest changes (cg21263664, cg03526459, cg16814483, cg02487331) and three additional adjacent CpG sites (Fig. 2a). A total of six out of these seven CpGs were found significantly hypermethylated in PPMS compared to RRMS patients (Fig. 2c). These data imply that PPMS patients display hypermethylation of an intergenic region within the 1q21.1 locus, suggesting a potential involvement in the PPMS pathogenesis.

### Methylation of the PPMS-associated 1q21.1 locus is under strong genetic control

Previous studies have revealed a complex interplay between genetic and epigenetic variations by reporting that part of the epigenome is under genetic control in general[26, 14], and more particularly, that genetically dependent DNA methylation can mediate risk to develop MS[11]. Interestingly, our data revealed that methylation clustered within three methylation levels (i.e. low, medium and high) in both cohorts (Fig. 2b-c), a pattern typically observed for genetically controlled methylation. To test for potential genetic dependence, we performed genome-wide methylation Quantitative Trait Loci (meQTL) analysis in cohort

1 and found 19 SNPs, located in the extended 1q21.1 locus (-50 kb to +400 kb), associating with methylation levels of the DMR, the strongest effect coming from rs1969869 (genotype *vs.* methylation estimate = 0.19, $p < 10^{-13}$) (Fig. 3a, Supplementary Table 4). To further validate this finding, we conducted locus-specific meQTL analysis in cohort 2 by testing all SNPs covering ± 500 kb window encompassing the DMR against each pyrosequenced CpG. Of note, only eight out of the 19 meQTL-SNPs identified in cohort 1 could be assessed in this cohort. Out of the 123 SNPs tested in cohort 2, two variants showed significant (rs1969869, rs4950357, $p < 10^{-8}$) and two suggestive (rs647596, rs10900384, $p < 10^{-5}$) associations with CpGs methylation (Fig. 3a). An example of genetic control of methylation at CpG 4 is illustrated in Figure 3b. The two strongest SNPs, rs1969869 and rs4950357, displayed positive association between the minor allele and CpG methylation levels (Fig. 3a, Supplementary Table 5). In addition to these four SNPs, 66 SNPs demonstrated nominal association with methylation ($p < 0.05$) (Supplementary Table 5). Overall, all eight overlapping SNPs identified in cohort 1 displayed significant and similar direction of changes also in the cohort 2 (Fig. 3a, Supplementary Tables 4 and 5). Many of the SNPs with similar effect, i.e. minor allele associating with either high or low methylation, were in strong linkage disequilibrium (LD > 0.8) and were therefore grouped into clusters (Fig. 3a). Altogether, these findings strongly suggest that methylation at CpGs in the 1q21.1 DMR are under genetic control.

We next asked whether the identified methylation-controlling SNPs associate with the risk to develop PPMS phenotype by testing association of meQTL-SNPs with PPMS (n = 477) in comparison to RRMS patients (n = 3,870) in a Swedish cohort (cohort 3, Fig. 1, Supplementary Table 1). Six and 115 SNPs from the genome-wide (cohort 1) and locus-specific (cohort 2) meQTL analyses, respectively, could be analyzed in this cohort. Despite the low number of PPMS patients, 17 SNPs exhibited nominal association ($p < 0.05$) with PPMS (Fig. 3a, Table 1, Supplementary Table 6). The strong LD (LD > 0.8) between some of the SNPs further suggested that these 17 SNPs exert influence from 6 variants, the minor alleles of two and four of them conferring risk (SNP cluster 11 and 14) and protection (SNP clusters 2, 8, 9, 10), respectively, for PPMS (Fig. 3a). The two variants conferring risk for PPMS predispose for higher levels of methylation at 1q21.1 DMR while all but one protective variants associate with low methylation at 1q21.1 DMR in the blood. These findings suggest a functional link, albeit modest, between high methylation at 1q21.1 DMR and the genetic predisposition to develop PPMS.

**Variants that control methylation of the PPMS locus associate with methylation and gene expression in the brain**

Our results implying a link between genetic variation and blood DNA methylation at the 1q21.1 region, a locus reported to be crucial for brain size[27,28] and neuropsychiatric disorders[29-31], pose the question whether such effect might occur in the CNS as well. To address this, we examined putative genetic influences on the methylation at 1q21.1 DMR in the brain using publicly available meQTL data from brain tissue of 543 individuals[14]. Results showed that 15 SNPs corresponding to 12 out of 14 SNP clusters demonstrated strong meQTL associations in the CNS as well, with similar direction of the effect as observed in the blood (Fig. 3c, Supplementary Table 7). As exemplified for cg02487331 (Fig. 3d), the strongest genetic associations to methylation levels at this locus arose from two variants, SNP cluster 5 tagged by rs1969869 and SNP cluster 6 tagged by rs21327 and rs4950357.

We next assessed whether the identified methylation-controlling SNPs could have an impact on gene expression in the brain using GTEx database and found that a large fraction of the meQTL-SNPs conditioned transcript levels of neighboring genes within the locus, *PRKAB2*, *CHD1L* and *FMO5* as well as long non-coding RNAs, in the CNS (Fig. 3d). Overall, SNPs predisposing for high DMR methylation and PPMS risk associated with lower *CHD1L* expression while SNPs conferring low DMR methylation and PPMS protection associated with higher *CHD1L and FMO5* expression and lower *PRKAB2* transcript levels. These data connecting the DNA sequence, methylation and transcription in the extended 1q21.1 region are consistent with genomic annotations of the DMR underscoring putative regulatory features as well as short-range physical interaction with proximal chromatin segments harboring eQTL genes (Supplementary Fig. 5). Thus, the variants controlling DNA methylation at 1q21.1 locus in PPMS patients likely have a functional impact on the expression of proximal genes, particularly *CHD1L, FMO5* and *PRKAB2*, in the brain.

**Methylation at the 1q21.1 locus regulates proximal gene expression in neuron-like cells**

In order to functionally address the causal relationship between epigenetic and transcriptional changes, we first tested whether methylation at the 1q21.1 DMR can exhibit regulatory properties on transcription, as suggested by genomic annotations of the locus (Fig. 2a, Supplementary Fig. 5), by examining methylation-sensitive enhancer and promoter activity using in-vitro methylation assays. Two distinct 1 kb fragments harboring the DMR sequence, isolated from PPMS patients with low (rs1969869: CC) and high (rs1969869: AA) methylation levels at the identified DMR, were inserted in CpG-free vector-based reporter system. The sequences were subsequently methylated using two different methyltransferases, *M.SssI* enzyme methylating all CpG sites (57 CpGs) and *HhaI* enzyme targeting only internal cytosine

residues from the consensus sequence GCGC (7 CpGs). The 3'-5' oriented DNA sequence of both genotypes exhibited potent constitutive promoter activity on the reporter gene (Mock condition), this basal activity being nonetheless halved in carriers of the minor allele compared to AA-homozygotes (Fig. 4a). The DMR manifested enhancer activity as well, although to a lesser extent, with no clear contribution of each allele or sequence orientation (Fig. 4b). Comparison of the unmethylated and methylated sequences revealed a significant reduction of the promoter and enhancer activity when the insert was fully methylated (Fig. 4a-b). Thus, controlling the level of methylation in the DMR region might have the potential regulatory effect on gene expression.

To formally address the impact of DNA methylation on endogenous gene expression, we exploited the CRISPR/dCas9-based epigenome editing platform which allows editing of methylation in a sequence-specific manner in combination with tailored single guide RNA (gRNA). We took advantage of the fact that several cell lines display hypermethylation at the 1q21.1 locus (as annotated by ENCODE Roadmap), mimicking methylation pattern in PPMS patients, and developed constructs expressing deactivated Cas9 fused to the catalytic domain of the demethylating enzyme TET1 (dCas9-TET1) to remove methylation from specific CpGs (see Methods for further description). We could validate the editing efficiency of dCas9-TET1 constructs by targeting 5 CpGs composing *MGAT3* gene promoter, with 10-50% de-methylation as previously reported[32] (Supplementary Fig. 6). To remove methylation at CpGs of the PPMS-associated DMR, we first designed and functionally screened 9 gRNAs targeting the two 450K probes exhibiting the largest changes (cg21263664, cg03526459) and adjacent CpGs in HEK293T cells (Fig. 4c). Negative controls consisted of cells transfected with a non-targeting (nt) gRNA, the catalytically inactive form of TET1 (TET1-IM) or non-transfected cells. DNA methylation analysis, using pyrosequencing, showed different degrees of de-methylation ranging from 10-50% reduction in methylation at single or multiple CpG sites depending on the gRNA (Fig. 4d). Delivery of dCas9-TET1 together with gRNA2 and gRNA3 could achieve robust demethylation at four sites, including cg03526459, in HEK293T cells (Fig. 4d). We next adapted this strategy to neuroblastoma SH-SY5Y cells using transduction with dCas9-TET1 + gRNA2 + gRNA3 lentiviruses (Fig. 4e) and could confirm the editing efficiency in these cells with 15-35% decrease of DNA methylation spanning throughout CpGs 1-3 (Fig. 4f). We then assessed by qPCR the expression of genes located in the vicinity of the DMR. These include the three coding genes found regulated by the same meQTL-SNPs in the normal brain (Fig. 3d), namely *CHD1L* gene encoding a DNA helicase protein involved in DNA repair, *PRKAB2* gene encoding the non-catalytic subunit of the AMP-activated protein kinase (AMPK), which serves as an energy sensor protein kinase regulating cellular energy metabolism and *FMO5* gene encoding a Baeyer-Villiger monooxygenase implicated in liver function and metabolic

ageing[33]. The induced de-methylation, even moderate, of CpGs in the 1q21.1 DMR in SH-SY5Y cells resulted in increased expression of *CHD1L* and *PRKAB2* genes while the expression of *FMO5* and other proximal transcripts, *PDIA3P1* and *PFN1P5*, did not vary significantly (Fig. 4g). Of note, we did not observe any significant gene expression changes following dCas9-TET1 co-delivery with single or combined gRNAs in HEK293T cells (Supplementary Fig. 7).

Collectively, these data suggest that the intergenic PPMS-associated DNA methylation at 1q21.1 region exerts regulatory properties on transcriptional processes. The methylation-mediated regulation of gene expression affects *CHD1L* and *PRKAB2* genes in neuron-like cells.

**Network analysis implicates *CHD1L* and *PRKAB2* genes in PPMS brain pathology**

To further delineate the putative relevance of the genes included in the 1q21.1 locus in PPMS brain pathology, we constructed an unbiased correlation network analysis using RNA-sequencing-derived gene expression matrix from brain tissue samples of progressive MS patients (n = 12) and non-neurological controls (n = 10)[9] (Fig. 5). After gene-pair permutation and planar filtering, the network consisted on 0.5 million interactions among 27,059 genes (FDR < 0.05) (Fig. 5a). Multiscale clustering analysis further clustered these interactions into 757 non-overlapping modules (Fig. 5b). We then evaluated the relevance of these modules for PPMS (n = 5) and SPMS (n = 7) phenotypic traits by applying cluster trait association analysis (Fig. 5b). Three modules survived statistical testing, one of them (module #1) being correlated to SPMS trait while the two others (modules #2 and #3) showing high significance to PPMS phenotype (Fig. 4b, Supplementary Table 9). Strikingly, these modules were found centered on genes from the extended 1q21.1 locus. Indeed, exploration of the biggest module (module #2), consisting of 380 nodes and 1079 interactions, revealed *CHD1L* as a central node within this network, closely interacting with six genes of a sub-module (Fig. 5c). Relative centrality could also be observed for the other two modules, which include *FMO5* among 40 nodes (module #1) and *PRKAB2* among 54 nodes (module #3) (Fig. 5c). Of note, *CHD1L* and *PRKAB2* were the only genes from the extended 1q21.1 locus to partake in the regulatory gene network underlying PPMS brain pathology in our analysis. Gene ontology analysis, using Ingenuity Pathway Analysis, underscored implication of all modules in nervous processes. This is particularly the case for module #2 with terms linked to neuronal functions, while additional *Biological Functions* were found associated to module #1 (e.g. metabolic processes) and module #3 (e.g. cell cycle) (Fig. 5c). Thus, findings from unbiased *in-silico* approach support a potential role of genotype-methylation-expression

regulatory network affecting genes of chromosome 1q21.1 extended locus, such as *CHD1L* and *PRKAB2,* in CNS processes in PPMS patients.

## Discussion

MS is a multifactorial and heterogeneous neurological disease with unpredictable trajectories affecting young individuals. Treatment of progressive forms remains the greatest challenge in the care of MS patients, which is likely due to our still partial knowledge of the pathological processes underlying these forms of disease. In this study, we utilized a multi-omics (genetic, epigenetic, transcriptomic) approach in several cohorts, in combination with in-vitro functional experiments and network analysis, to gain more insights into the primary progressive form of MS. We identified a genetically controlled hypermethylated region in the chromosome 1q21.1 locus which might affect the expression of proximal genes, particularly *CHD1L* and *PRKAB2*, in the brain of PPMS patients.

Attempts to identify genetic variants that predispose for PPMS have not yielded reproducible significant genome-wide associations with the exception of the known RRMS-risk *HLA-DRB1*15:01* variant[34,35]. One of the challenges of the genetic studies in PPMS is the limited availability of PPMS cases that comprise around 10% of all MS cases. We and others have shown that integrating additional layers of gene regulation, such as DNA methylation, can reveal genetic associations that are difficult to identify using conventional genetic studies[11,36-38]. In this study, we combined methylome profiling with locus-specific genetic association analyses in independent cohorts and identified a genetically controlled DMR in the 1q21.1 locus that displayed hypermethylation in PPMS patients specifically. The same genetic variants that control methylation and gene expression in the locus yielded nominally significant association with PPMS risk suggesting a functional link between genetic, epigenetic, and transcriptional modalities in 1q21.1 in PPMS patients. Undoubtedly, despite having nearly 500 PPMS cases, such nominal associations require replication. Our findings should also be considered in the light of the remarkably complex genomic architecture of the 1q21.1 locus displaying considerable copy-number variations. Indeed, the extended 1q21.1 genomic locus covering the identified DMR is characterized by the largest copy number expansion between non-human and human primate lineage, such variation correlating with evolutionary brain size in a dose-dependent manner[27,28]. Copy-number variations in this region have also been associated with brain disorders such cognitive and motor deficits, neuroblastoma, autistic spectrum disorder and schizophrenia[29-31]. These associations with brain size and neurodevelopmental disorders further support

the idea that inherent brain tissue vulnerability influenced by the genes in the locus may predispose for rapid and progressive MS disease. Whether the changes identified in the 1q21.1 locus in our study are related to structural variations of the genome warrants further investigation.

The observed dual meQTL and eQTL effects in the CNS tissue, underscore the possible contribution of genetic-epigenetic-transcriptional regulation in PPMS brain pathology. In this context, we found that genetic variation that strongly associated with elevated DNA methylation levels at 1q21.1 and modestly associated with increased PPMS risk is linked to lower expression of *CHD1L* transcript in the CNS. We next sought to formally explore the possibility that DNA methylation may directly influence gene expression. We first demonstrated intrinsic methylation-sensitive promoter and enhancer activity of PPMS-associated DMR sequence using in-vitro methylation assays. We then utilized CRISPR/dCas9-TET1 epigenome editing approach to achieve effective and locus-specific demethylation at CpGs of the identified DMR and found that manipulating DNA methylation at the 1q21.1 DMR affects the expression of the proximal *CHD1L* and *PRKAB2* genes in neuron-like cells, an effect we could not observe in HEK293T cells. These data, together with the indisputable involvement of 1q21.1 region in neuronal circuitry and brain integrity[27-31], implicate genes of the locus in neuropathological processes in PPMS. Undeniably, given the occurrence of meQTL effects in the blood, one cannot exclude a contribution of 1q21.1 regulatory network in the peripheral immune compartment as well. Similar trans-tissue effects with shared pattern between the immune and nervous compartments have previously observed in the context of progressive MS disease[39]. Collectively, these findings portray DNA methylation as a putative intermediary of genetic influence on gene expression, particularly in the CNS of PPMS patients.

Our unbiased correlation network analysis of published transcriptomic data obtained from post-mortem brain tissue from PPMS, SPMS patients and controls[9], identified *CHD1L* and *PRKAB2* transcripts as implicated in PPMS specific regulatory gene networks while *FMO5* gene was found significantly connected to SPMS phenotype. *FMO5* encodes for an atypical flavin containing monooxygenase protein with the ability to oxidize a wide range of substrates and xenobiotic, whose disruption has been shown to affect metabolic ageing in rodents[40]. *CHD1L* encodes a DNA helicase implicated in chromatin-remodelling during DNA repair, with both depletion and overexpression of CHD1L leading to genomic sensitivity to DNA-damaging agents[41]. *PRKAB2* encodes the beta subunit of AMP-activated kinase, a neuroprotective energy-sensing kinase exerting pivotal roles in the cellular energy metabolism required for maintenance of neuronal integrity[42]. Accordingly, previous studies on neuron-specific loss of PRKAB2 function fruit flies resulted in progressive fragmentation of dendritic arborisation, cognitive impairment as well as reduced

stress resistance and lifespan[42]. Interestingly, bulk and single cell transcriptomics in brain tissue have found dysregulation of several genes located within the identified 1q21.1 locus, including *CHD1L*, *PRKAB2*, *FMO5* transcripts along with several members of the *NBPF* family of genes, in progressive MS patients compared to control individuals[9,43]. Markedly, *CHD1L* gene has been identified in the stress-related signature of excitatory CUX2-expressing projection neurons in upper-cortical layers, which are particularly susceptible to degeneration in MS patients[43]. Additionally, a large fraction of *CHD1L* (3/6) closest network neighbours identified in our study were found significantly upregulated in these stressed excitatory neurons as well. It is therefore reasonable to speculate on the involvement of *CHD1L* and *PRKAB2* genes in the compensatory mechanisms counteracting the cellular burden, e.g. genomic and metabolic stress, observed in these neurons[43]. More specifically, genetic predisposition to high methylation and low expression of *CHD1L* gene might influence susceptibility to PPMS by limiting the inherent repair capacity in the brain.

## Conclusions

Our study suggests that PPMS patients display distinct molecular changes compared to SPMS, RRMS and healthy individuals. The data further support the hypothesis of a causal genetic-epigenetic-transcriptional interplay within the extended 1q21.1 locus, with functional evidence of contribution from *CHD1L* and *PRKAB2* in progressive MS pathology. Given the reversible nature of DNA methylation, our findings open new avenues for development of therapeutic strategies, such as the targeted epigenetic therapy, in the treatment of progressive MS.

# Declarations

### Ethics approval and consent to participate

All experiments on human subjects were approved by the Regional Ethical Review Board in Stockholm and carried out in accordance with institutional guidelines. Written informed consent was obtained from all study participants.

### Consent for publication

Not applicable.

### Availability of data and materials

The 450K array methylome data generated from whole blood (cohort 1) are available in the Gene Expression Omnibus (GEO) database under accession number GSE106648. Results from the genetic association analysis (cohort 3) are available in supplementary material, the raw data containing sensitive information can be made available from the corresponding author upon request and signature of data transfer agreement.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This study was supported by grants from the Swedish Research Council, the Swedish Association for Persons with Neurological Disabilities, the Swedish Brain Foundation, the Swedish MS Foundation, the Stockholm County Council - ALF project, the European Union's Horizon 2020 research, innovation programme (grant agreement No 733161) and the European Research Council (ERC, grant agreement No 818170), the Knut and Alice Wallenberg Foundation, Hedlund Foundation, Åke Wiberg Foundation and Karolinska Institute's funds. LK is supported by a fellowship from the Margaretha af Ugglas Foundation. MPK is supported by McDonald Fellowship from Multiple Sclerosis International Federation (MSIF) and the EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking (EUbOPEN

grant agreement No 875510). CSC is supported by the Blanceflor Boncompagni Ludovisi, née Bildt Foundation. PC is supported by NIH R35-NS111604. The funders of the study had no role in study design, sample acquisition, data collection, data analysis, data interpretation or writing of the manuscript.

**Authors' contributions**

MJ conceptualized, designed and supervised the study. MPK conducted the clonings, BS-pyrosequencing, in-vitro DNA methylation assay and dCas9-TET1 experiments with the assistance of CSC. SB and FW assisted in lentiviral production and in-vitro experiments. YL performed genome-wide DNA methylation and meQTL analyses (cohort 1). AG carried out locus-specific meQTL and genetic association analyses (cohort 2). TVSB performed correlation network analysis and MG supervised network analysis. TJ aided in data acquisition. LK supervised the study, performed analyses, interpreted the data and wrote the manuscript with contribution from all co-authors. IK provided genetic data. TO, JH and PC contributed to the collection of patient samples. All authors read and approved the final manuscript.

**Acknowledgments**

We are grateful to all volunteers for contributing to the study. We thank Maria Kalomoiri, Jessica Aguilar and Jacqueline Hammer for their help in experimental setup and Dr. Alexander Espinosa for kindly providing plasmid construct. We acknowledge KIgene facility at the Center for Molecular Medicine and Karolinska Institutet for Sanger sequencing service. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

**Table 1.** Genetic association of methylation-controlling variants (p < 0.05) with PPMS (n = 477) in comparison to RRMS (n = 3871).

| SNP | BP | CHR | A1 | N | ORX | SE | L95 | U95 | *P* |
|---|---|---|---|---|---|---|---|---|---|
| RS1123333 | 146694571 | 1 | G | 4342 | 0.83 | 0.07 | 0.72 | 0.95 | 0.0089 |
| RS1036851 | 146633750 | 1 | G | 4343 | 0.83 | 0.07 | 0.72 | 0.95 | 0.0091 |
| RS1470261 | 146694922 | 1 | G | 4345 | 0.83 | 0.07 | 0.72 | 0.95 | 0.0092 |
| RS930495 | 146634514 | 1 | C | 4344 | 0.83 | 0.07 | 0.72 | 0.96 | 0.0097 |
| RS6937 | 146626922 | 1 | G | 4347 | 0.75 | 0.11 | 0.61 | 0.94 | 0.0103 |
| RS12141187 | 146982763 | 1 | A | 4345 | 1.18 | 0.07 | 1.03 | 1.36 | 0.0198 |
| RS11240099 | 146506626 | 1 | G | 4313 | 0.71 | 0.15 | 0.53 | 0.96 | 0.0261 |
| RS1348316 | 146643555 | 1 | G | 4346 | 0.85 | 0.07 | 0.74 | 0.98 | 0.0268 |
| RS7553584 | 146693429 | 1 | G | 4334 | 1.32 | 0.13 | 1.02 | 1.71 | 0.0382 |
| RS6703847 | 146669444 | 1 | G | 4348 | 1.31 | 0.13 | 1.01 | 1.71 | 0.0408 |
| RS4950385 | 146656292 | 1 | A | 4347 | 1.31 | 0.13 | 1.01 | 1.71 | 0.0411 |
| RS894469 | 146672906 | 1 | G | 4346 | 1.31 | 0.13 | 1.01 | 1.70 | 0.0412 |
| RS894471 | 146687913 | 1 | G | 4346 | 1.31 | 0.13 | 1.01 | 1.70 | 0.0412 |
| RS1530336 | 146685121 | 1 | G | 4344 | 1.31 | 0.13 | 1.01 | 1.70 | 0.0416 |
| RS750467 | 146589758 | 1 | A | 4345 | 0.81 | 0.10 | 0.67 | 0.99 | 0.0422 |
| RS7523812 | 146673489 | 1 | G | 4336 | 1.31 | 0.13 | 1.01 | 1.70 | 0.0432 |
| RS11239951 | 146709784 | 1 | A | 4347 | 0.87 | 0.07 | 0.76 | 1.00 | 0.0470 |

MS, multiple sclerosis; PPMS, primary progressive MS; RRMS, relapsing-remitting MS; CHR, chromosome; N, total number; ORX, odds ratio; SE, standard error; L95 and U95, 95% confidence interval for odds ratio, lower bound and upper bound, respectively. Complete data is listed in Supplementary Table 6.

## Figures legends

**Figure 1.** Workflow of the study.

MS, multiple sclerosis; PPMS, primary progressive MS; SPMS, secondary progressive MS; RRMS, relapsing-remitting MS; SNP, single nucleotide polymorphism, meQTL and eQTL, methylation and expression quantitative trait loci.

**Figure 2.** Hypermethylation of the 1q21.1 locus in primary progressive multiple sclerosis patients.

**a**. Genomic annotation of the identified differentially methylated region (DMR), including 450K and pyrosequenced CpG probes and regulatory features, i.e. CpG island (CGI) and chromatin state segmentation by hidden Markov model (ChromHMM) from ENCODE/Broad. **b**. Dot plot and violin plot of DNA methylation differences at the identified DMR on chromosome 1: 146549909-146551201 (hg19) in PPMS (n = 4) compared to RRMS (n = 120), SPMS (n= 16) and HC (n = 140), obtained using Illumina 450K in cohort 1. **c**. Replication of methylation at 7 CpGs in an independent cohort (RRMS n=48, PPMS n=36) using pyrosequencing. * and ** correspond to P-value < 0.5 and 0.01, respectively, using Kruskal-Wallis test with Dunn's multiple comparison test. MS, multiple sclerosis; PPMS, primary progressive MS; RRMS, relapsing-remitting MS; SPMS, secondary progressive MS; HC, healthy controls.

**Figure 3.** Genetic control of methylation at the 1q21.1 locus DMR.

**a**. Genomic and functional annotation of the methylation-controlling variants (SNPs). Linkage disequilibrium (LD) is indicated by shades of red (increasing $R^2$ from light to dark red) and clustering of SNPs is based on high LD between SNPs ($R^2 > 0.8$). The effects of SNPs on DMR methylation and PPMS risk are depicted using colored circles: the size and color of circles represent significance (–log10 P-value) and direction of the effect, respectively, conferred by the minor allele. Gene location and regulatory features, i.e. CpG island (CGI) and chromatin state segmentation by hidden Markov model (ChromHMM), from ENCODE/Broad. **b**. Association between genetic variation at the extended locus (upper panel) and DNA methylation in blood at CpG4 of the 1q21.1 DMR, obtained using meQTL analysis in cohort 2 (n=82). Significance is represented as –log10 (P-value). Association of the two strongest variants rs1969869 and rs4950357 with CpG 4 are depicted in the lower panel. **c**. Effects of SNPs within each SNP cluster on DNA methylation at 1q21.1 DMR (using xQTL serve platform[14]) and gene expression (using GTEx database) in

the brain. Only significant associations are displayed in colors, blue and red colors representing negative and positive effect of the minor allele, respectively. **d**. Association between genetic variation in the extended locus and DNA methylation at cg02487331 of the identified 1q21.1 DMR in the brain, retrieved from xQTL serve platform[14]. Significance is represented as –log10 (P-value). *indicates that the SNPs are in strong LD with rs21327 (LD = 0.7). The data shown in this figure are available in Supplementary Tables 2-7.

**Figure 4.** Functional impact of methylation at the 1q21.1 DMR on gene expression.

**a-b**. Promoter (**a**) and enhancer (**b**) activity of the DMR, using CpG-free promoter-free (SEAP) and promoter-containing (Lucia) reporter gene vectors, respectively. Constructs in direct or inverted orientation of DNA segments derived from individual varying according to the genotype at rs1969869 were partially or fully methylated using *HhaI* and *M.SssI* enzymatic treatment, respectively. Results show relative activity (Lucia or SEAP normalized against Renilla) using three replicates in a representative experiment performed at least two times (2-way ANOVA test with Bonferroni correction for multiple comparisons (± SD)). **c**. Schematic representation of the experimental design for gRNAs screen in HEK293T cells including the features of the constructs and gRNA locations. **d**. Heatmap of the DNA methylation levels in successfully transfected (GFP positive) HEK293T cells three days following co-transfection of dCas9-TET1 with single or combined gRNAs in comparison to control conditions, deactivated TET1 (TET1-IM), non-targeting gRNA (ntgRNA) and non-transfected cells. **e**. Schematic representation of the experimental design for functional investigation in SH-SY5Y cells. **f**. DNA methylation levels in SH-SY5Y cells following delivery of dCas9-TET1-, gRNA2- and gRNA3-containing lentiviruses. Methylation percentages represent the average methylation levels of three experiments. **g**. Experiment showing expression of *CHD1L*, *PRKAB2*, *PDIA3P1* and *FMO5* genes relative to *GAPDH* transcript levels, quantified using RT-qPCR. The expression levels represent the average of three experiments (Two-tailed Student's t test (± SD)). ****=p<0,0001, n.s = non-significant.

**Figure 5.** Implication of *CHD1L* and *PRKAB2* genes in PPMS brain pathology.

**a**. Schematic representation of the correlation network analysis using bulk MS brain transcriptome[9]. The different steps are depicted from left to right: correlation matrix containing gene-gene pairwise Spearman correlations; illustration of the permutation test filtered correlation network surviving an FDR P-value <

0.05 and the planar maximally filtering to convert the scale free network to be able to overlay on a plane spherical surface, reducing the network to 0.5 million interactions among 27509 genes and, finally; representation of the multiscale clustering of the network resulting into 757 non-overlapping clusters. **b**. Heatmap of the correlation coefficients of all 757 non-overlapping modules (left) and the three modules surviving correlation FDR P-value < 0.05 (right) to each tested phenotypic trait, with red gradient colors representing negative (not significant) to positive (significant) correlation. **c**. Representation of the three modules (left), the closest neighbors to candidate genes in each module (middle) and Gene Ontology findings, i.e. *Disease and Biological functions,* obtained using Ingenuity Pathway Analysis. MS, multiple sclerosis; PPMS, primary progressive MS; SPMS, secondary progressive MS; NNC, non-neurological controls.

## Supplementary Materials

**Supplementary Table 1:** Description of cohorts

**Supplementary Table 2.** Primers sequences

**Supplementary Table 3.** Genome-wide methylation analysis (cohort 1)

**Supplementary Table 4.** Genome-wide meQTL analysis (cohort 1)

**Supplementary Table 5.** Locus-specific-meQTL analysis (cohort 2)

**Supplementary Table 6.** Genetic association (cohort 3)

**Supplementary Table 7.** Brain meQTL and eQTL data

**Supplementary Table 8.** Brain correlation network analysis (RNAseq from Huynh et al., 2014)

**Supplementary Table 9.** Gene ontology analysis (using Ingenuity Pathway Analysis)

**Supplementary Figure 1.** Bisulfite amplicons for the pyrosequencing assays.

**Supplementary Figure 2.** Methylation of 1q21.1 DMR vector used for in-vitro DNA methylation assay.

**Supplementary Figure 3.** Sequence and map of the plasmids used for DNA methylation editing.

**Supplementary Figure 4.** Map of the gRNAs designed to target the 1q21.1 DMR.

**Supplementary Figure 5.** Chromatin chromosome interaction (Hi-C) annotation of the locus.

**Supplementary Figure 6.** Validation of the editing efficiency of dCas9-TET1 constructs by targeting 5 CpGs composing *MGAT3* gene promoter.

**Supplementary Figure 7.** Expression of the selected genes in the 1q21.1 locus in HEK293T cells.

## References


1. Kutzelnigg, A. *et al.* Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* **128**, 2705-2712 (2005).
2. Ransohoff, R.M., Hafler, D.A. & Lucchinetti, C.F. Multiple sclerosis—a quiet revolution. *Nature Reviews Neurology* **11**, 134-142 (2015).
3. Miller, D.H. & Leary, S.M. Primary-progressive multiple sclerosis. *The Lancet Neurology* **6**, 903-912 (2007).
4. Mahad, D.H., Trapp, B.D. & Lassmann, H. Pathological mechanisms in progressive multiple sclerosis. *Lancet Neurol* **14**, 183-93 (2015).
5. Thompson, A.J. Challenge of progressive multiple sclerosis therapy. *Curr Opin Neurol* **30**, 237-240 (2017).
6. Alfredsson, L. & Olsson, T. Lifestyle and environmental factors in multiple sclerosis. *Cold Spring Harbor perspectives in medicine* **9**, a028944 (2019).
7. Ewing, E. *et al.* Combining evidence from four immune cell types identifies DNA methylation patterns that implicate functionally distinct pathways during Multiple Sclerosis progression. *EBioMedicine* **43**, 411-423 (2019).
8. Graves, M.C. *et al.* Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. *Mult Scler* **20**, 1033-41 (2014).
9. Huynh, J.L. *et al.* Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci* **17**, 121-30 (2014).
10. Kular, L. *et al.* Neuronal methylome reveals CREB-associated neuro-axonal impairment in multiple sclerosis. *Clin Epigenetics* **11**, 86 (2019).
11. Kular, L. *et al.* DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun* **9**, 2397 (2018).
12. Kular, L. *et al.* DNA methylation changes in glial cells of the normal-appearing white matter in Multiple Sclerosis patients. *Epigenetics*, 1-20 (2022).
13. Bos, S.D. *et al.* Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. *PloS one* **10**, e0117403 (2015).
14. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* **20**, 1418-1426 (2017).
15. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
16. Houseman, E.A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
17. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
18. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* **44**, 5615-28 (2016).
19. Concordet, J.P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**, W242-W245 (2018).
20. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *methods* **25**, 402-408 (2001).

21. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-8 (2016).
22. Wu, Z., Wang, X. & Zhang, X. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics* **27**, 502-8 (2011).
23. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-5 (2015).
24. Song, W.-M. & Zhang, B. Multiscale embedded gene co-expression network analysis. *PLoS computational biology* **11**, e1004574 (2015).
25. Su, S. *et al.* Glimma: interactive graphics for gene expression analysis. *Bioinformatics* **33**, 2050-2052 (2017).
26. Chen, L. *et al.* Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398-1414. e24 (2016).
27. Dumas, L.J. *et al.* DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am J Hum Genet* **91**, 444-54 (2012).
28. Zimmer, F. & Montgomery, S.H. Phylogenetic Analysis Supports a Link between DUF1220 Domain Number and Primate Brain Expansion. *Genome Biol Evol* **7**, 2083-8 (2015).
29. Diskin, S.J. *et al.* Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459**, 987-91 (2009).
30. Bernier, R. *et al.* Clinical phenotype of the recurrent 1q21.1 copy-number variant. *Genet Med* **18**, 341-9 (2016).
31. Mehta, D. *et al.* Comprehensive survey of CNVs influencing gene expression in the human brain and its implications for pathophysiology. *Neurosci Res* **79**, 22-33 (2014).
32. Josipovic, G. *et al.* Antagonistic and synergistic epigenetic modulation using orthologous CRISPR/dCas9-based modular system. *Nucleic Acids Res* **47**, 9637-9657 (2019).
33. Malagon, S.G.G. *et al.* The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochemical pharmacology* **96**, 267-277 (2015).
34. International Multiple Sclerosis Genetics, C. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-9 (2011).
35. Martinelli-Boneschi, F. *et al.* A genome-wide association study in progressive multiple sclerosis. *Mult Scler* **18**, 1384-94 (2012).
36. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature biotechnology* **31**, 142-147 (2013).
37. Olsson, A.H. *et al.* Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS genetics* **10**, e1004735 (2014).
38. Hong, X. *et al.* Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nature communications* **6**, 1-12 (2015).
39. Kular, L. & Jagodic, M. Epigenetic insights into multiple sclerosis disease progression. *J Intern Med* **288**, 82-102 (2020).
40. Gonzalez Malagon, S.G. *et al.* The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochem Pharmacol* **96**, 267-77 (2015).
41. Ahel, D. *et al.* Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240-3 (2009).
42. Nagy, S. *et al.* AMPK signaling linked to the schizophrenia-associated 1q21.1 deletion is required for neuronal and sleep maintenance. *PLoS Genet* **14**, e1007623 (2018).

43. Schirmer, L. *et al.* Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75-82 (2019).

# Figure 1

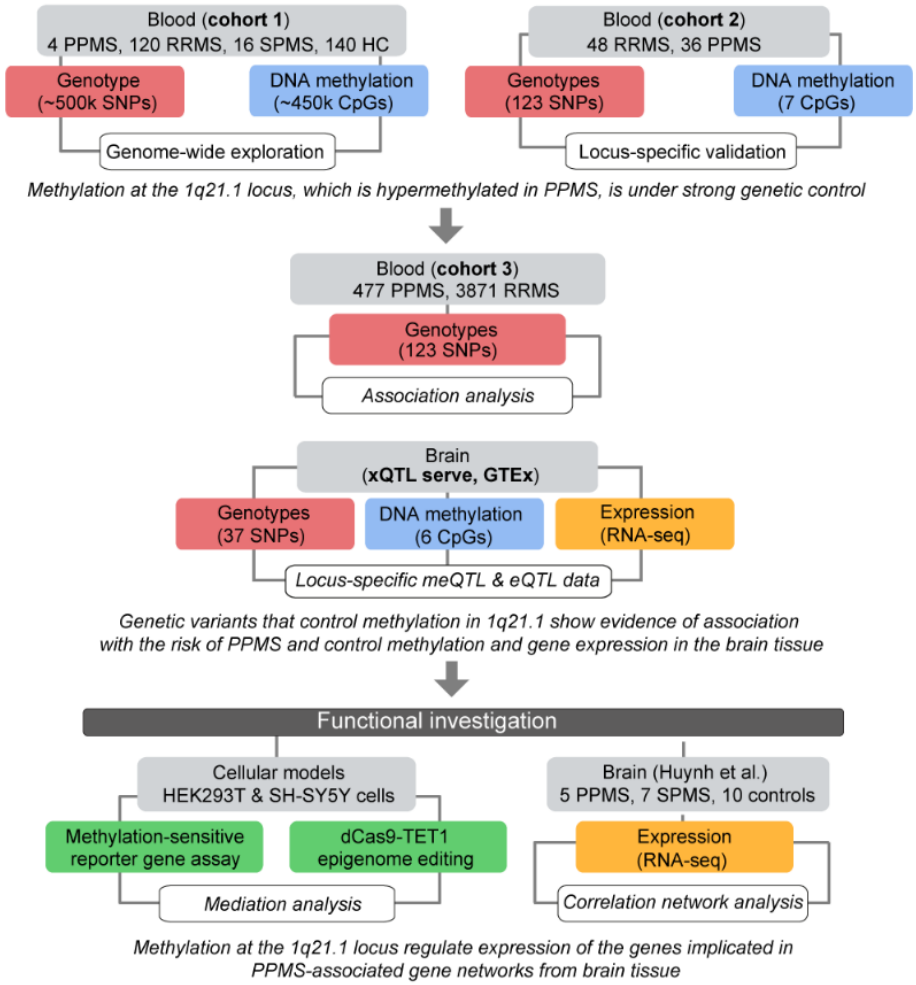


**Figure 1.** Workflow of the study.

MS, multiple sclerosis; PPMS, primary progressive MS; SPMS, secondary progressive MS; RRMS, relapsing-remitting MS; SNP, single nucleotide polymorphism, meQTL and eQTL, methylation and expression quantitative trait loci.

**Figure 2**

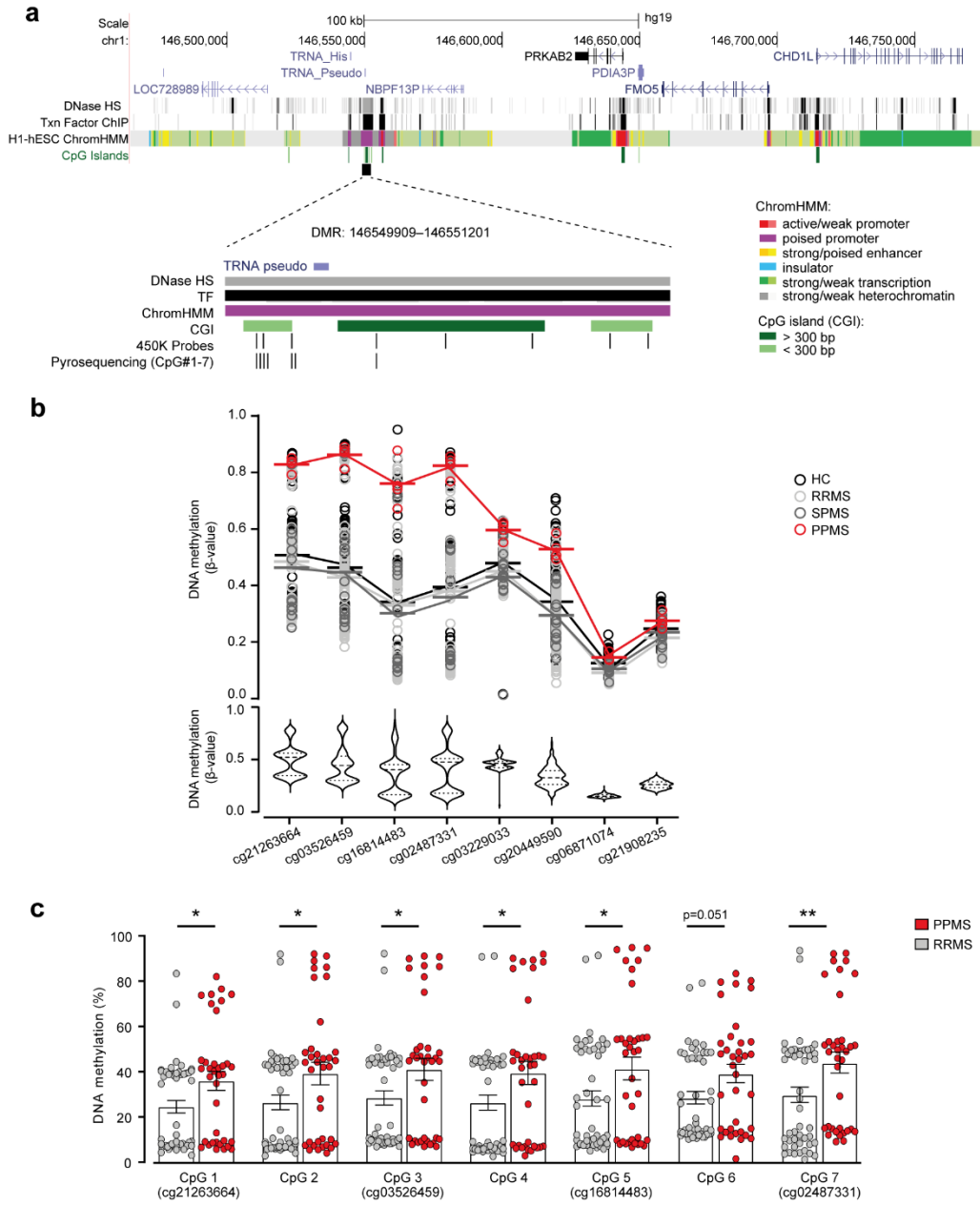


**Figure 2.** Hypermethylation of the 1q21.1 locus in primary progressive multiple sclerosis patients.

**a**. Genomic annotation of the identified differentially methylated region (DMR), including 450K and pyrosequenced CpG probes and regulatory features, i.e. CpG island (CGI) and chromatin state segmentation by hidden Markov model (ChromHMM) from ENCODE/Broad. **b**. Dot plot and violin plot of DNA methylation differences at the identified DMR on chromosome 1: 146549909-146551201 (hg19) in PPMS (n = 4) compared to RRMS (n = 120), SPMS (n= 16) and HC (n = 140), obtained using Illumina 450K in cohort 1. **c**. Replication of methylation at 7 CpGs in an independent cohort (RRMS n=48, PPMS n=36) using pyrosequencing. * and ** correspond to P-value < 0.5 and 0.01, respectively, using Kruskal-Wallis test with Dunn's multiple comparison test. MS, multiple sclerosis; PPMS, primary progressive MS; RRMS, relapsing-remitting MS; SPMS, secondary progressive MS; HC, healthy controls.

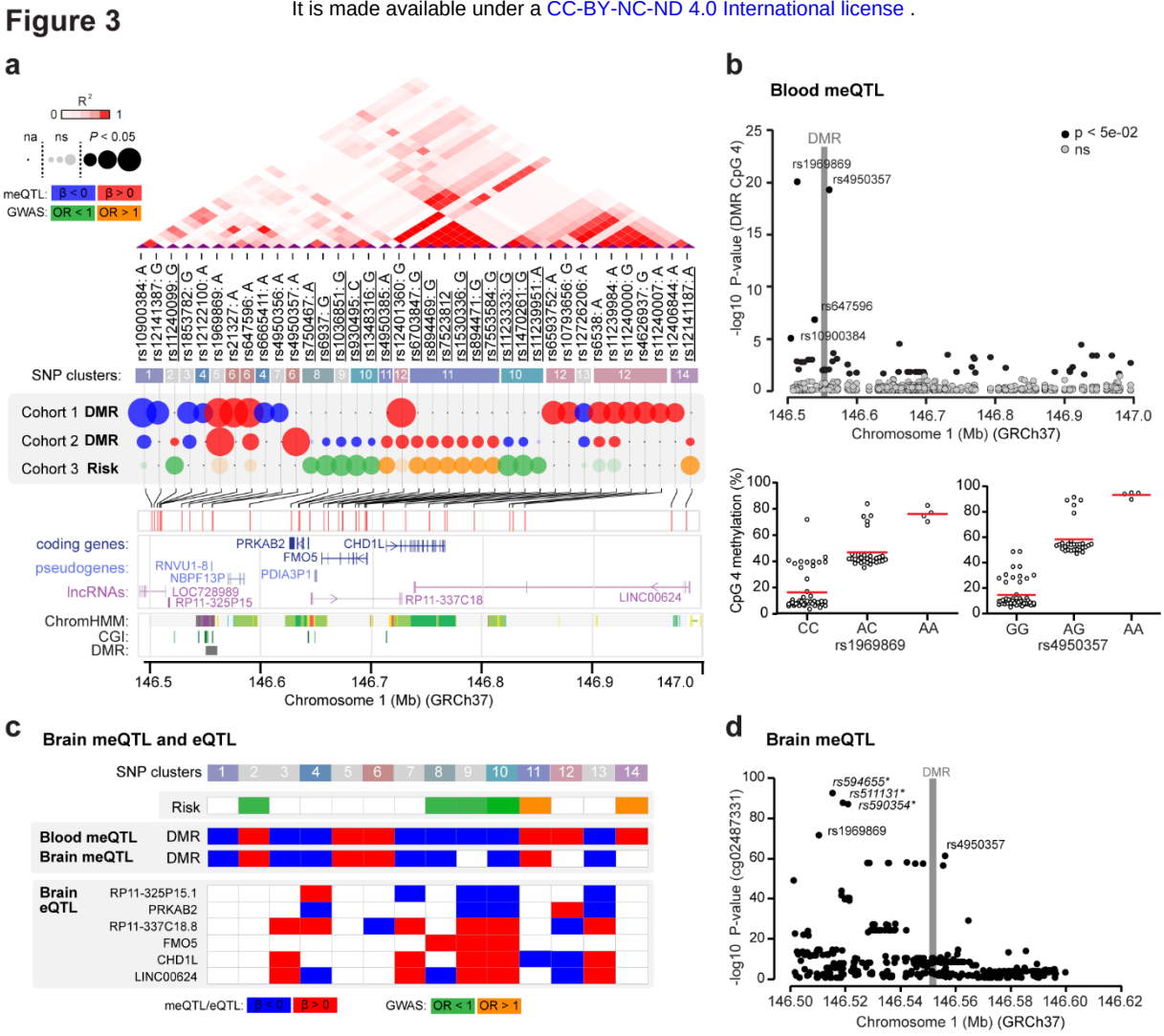


**Figure 3.** Genetic control of methylation at the 1q21.1 locus DMR.

**a**. Genomic and functional annotation of the methylation-controlling variants (SNPs). Linkage disequilibrium (LD) is indicated by shades of red (increasing $R^2$ from light to dark red) and clustering of SNPs is based on high LD between SNPs ($R^2 > 0.8$). The effects of SNPs on DMR methylation and PPMS risk are depicted using colored circles: the size and color of circles represent significance (–log10 P-value) and direction of the effect, respectively, conferred by the minor allele. Gene location and regulatory features, i.e. CpG island (CGI) and chromatin state segmentation by hidden Markov model (ChromHMM), from ENCODE/Broad. **b**. Association between genetic variation at the extended locus (upper panel) and DNA methylation in blood at CpG4 of the 1q21.1 DMR, obtained using meQTL analysis in cohort 2 (n=82). Significance is represented as –log10 (P-value). Association of the two strongest variants rs1969869 and rs4950357 with CpG 4 are depicted in the lower panel. **c**. Effects of SNPs within each SNP cluster on DNA methylation at 1q21.1 DMR (using xQTL serve platform[14]) and

gene expression (using GTEx database) in the brain. Only significant associations are displayed in colors, blue and red colors representing negative and positive effect of the minor allele, respectively. **d**. Association between genetic variation in the extended locus and DNA methylation at cg02487331 of the identified 1q21.1 DMR in the brain, retrieved from xQTL serve platform[14]. Significance is represented as –log10 (P-value). *indicates that the SNPs are in strong LD with rs21327 (LD = 0.7). The data shown in this figure are available in Supplementary Tables 2-7.

## Figure 4

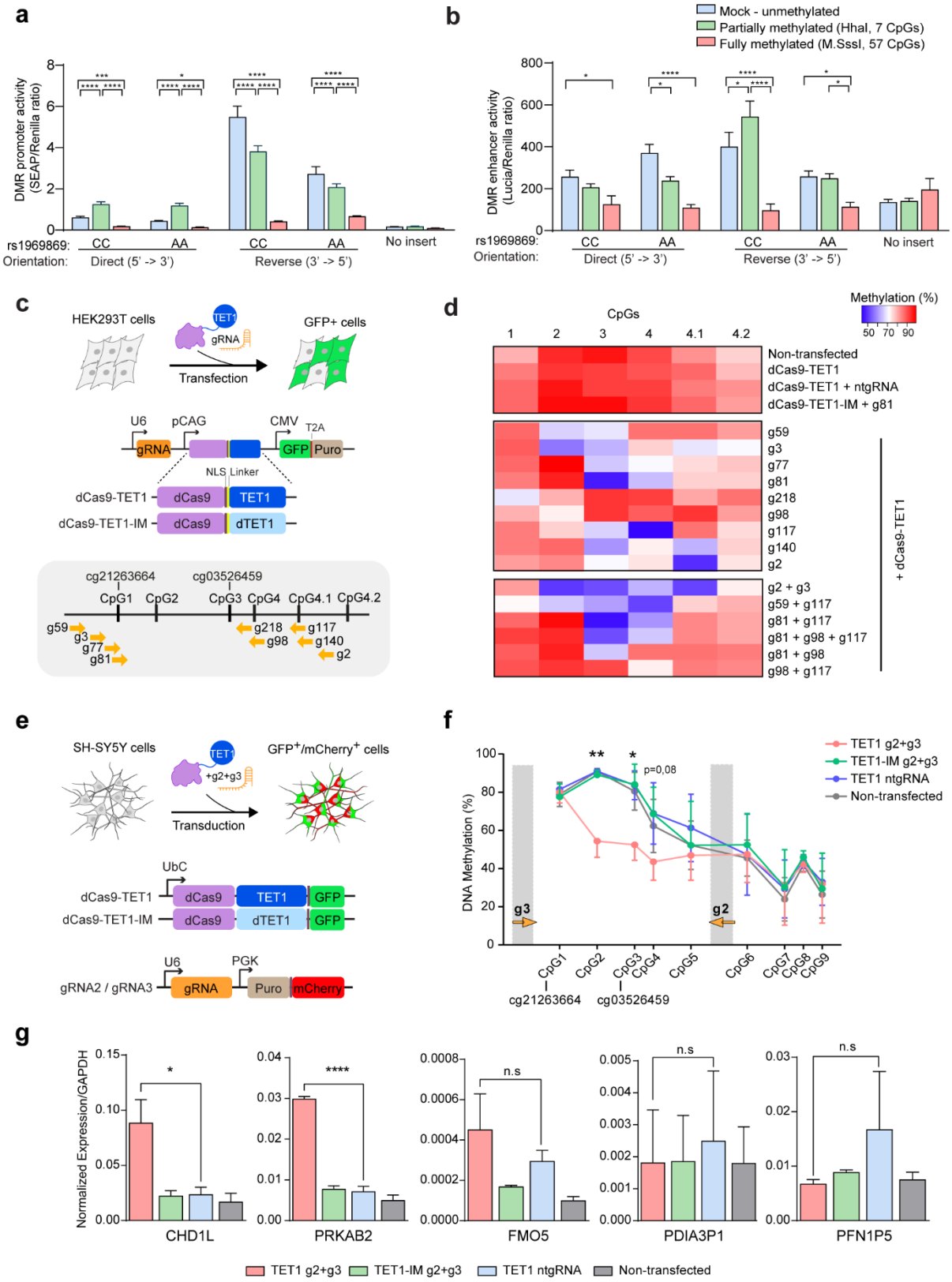


**Figure 4.** Functional impact of methylation at the 1q21.1 DMR on gene expression.

**a-b**. Promoter (**a**) and enhancer (**b**) activity of the DMR, using CpG-free promoter-free (SEAP) and promoter-containing (Lucia) reporter gene vectors, respectively. Constructs in direct or inverted

orientation of DNA segments derived from individual varying according to the genotype at rs1969869 were partially or fully methylated using *HhaI* and *M.SssI* enzymatic treatment, respectively. Results show relative activity (Lucia or SEAP normalized against Renilla) using three replicates in a representative experiment performed at least two times (2-way ANOVA test with Bonferroni correction for multiple comparisons (± SD)). **c**. Schematic representation of the experimental design for gRNAs screen in HEK293T cells including the features of the constructs and gRNA locations. **d**. Heatmap of the DNA methylation levels in successfully transfected (GFP positive) HEK293T cells three days following co-transfection of dCas9-TET1 with single or combined gRNAs in comparison to control conditions, deactivated TET1 (TET1-IM), non-targeting gRNA (ntgRNA) and non-transfected cells. **e**. Schematic representation of the experimental design for functional investigation in SH-SY5Y cells. **f**. DNA methylation levels in SH-SY5Y cells following delivery of dCas9-TET1-, gRNA2- and gRNA3-containing lentiviruses. Methylation percentages represent the average methylation levels of three experiments. **g**. Experiment showing expression of *CHD1L*, *PRKAB2*, *PDIA3P1* and *FMO5* genes relative to *GAPDH* transcript levels, quantified using RT-qPCR. The expression levels represent the average of three experiments (Two-tailed Student's t test (± SD)). ****=p<0,0001, n.s = non-significant.

# Figure 5

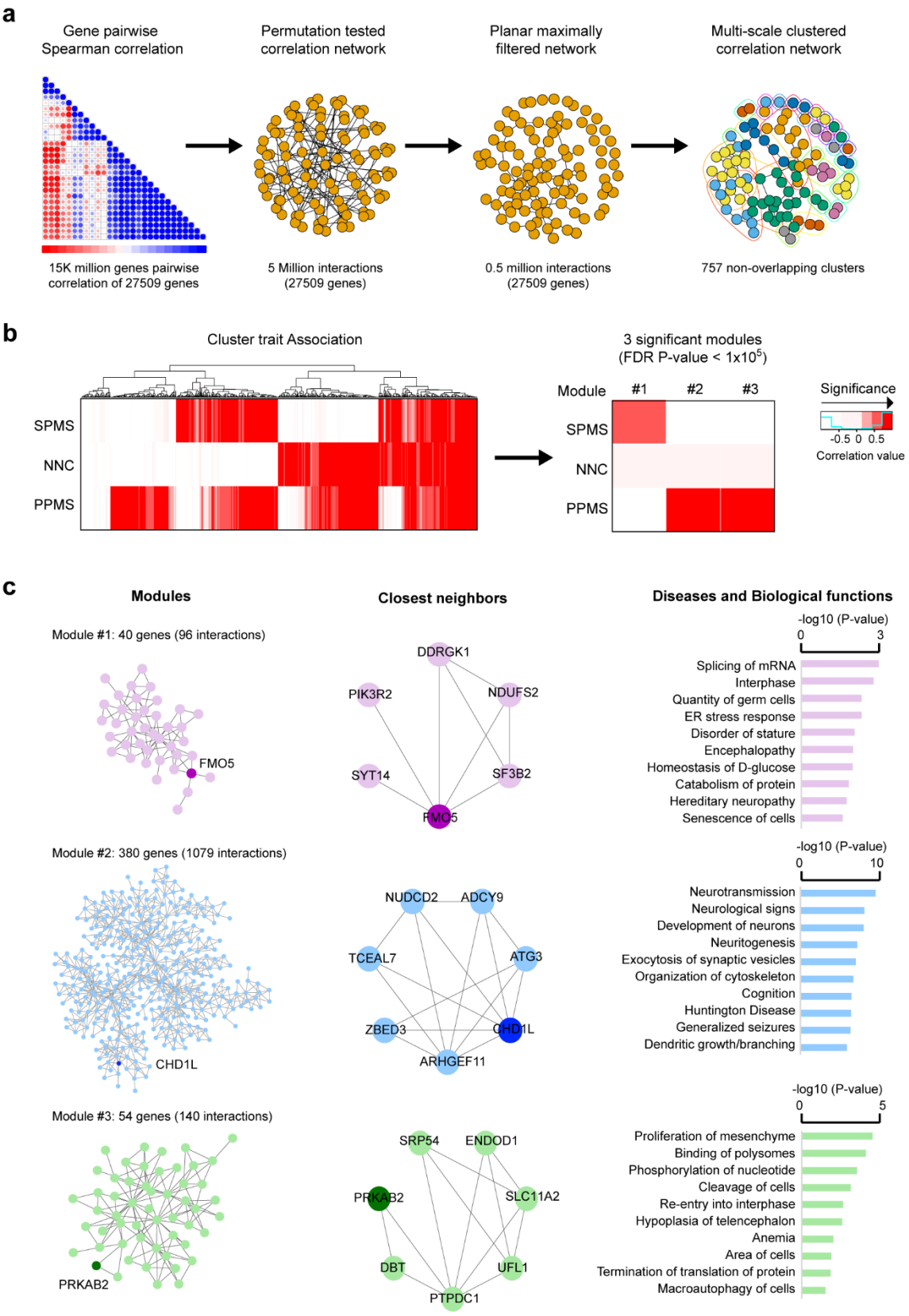

**Figure 5.** Implication of *CHD1L* and *PRKAB2* genes in PPMS brain pathology.

**a**. Schematic representation of the correlation network analysis using bulk MS brain transcriptome[5]. The different steps are depicted from left to right: correlation matrix containing gene-gene pairwise Spearman correlations; illustration of the permutation test filtered correlation network surviving an FDR P-value < 0.05 and the planar maximally filtering to convert the scale free network to be able to overlay on a plane spherical surface, reducing the network to 0.5 million interactions among 27509 genes and, finally; representation of the multiscale clustering of the network resulting into 757 non-overlapping clusters. **b**. Heatmap of the correlation coefficients of all 757 non-overlapping modules (left) and the three modules surviving correlation FDR P-value < 0.05 (right) to each tested phenotypic trait, with red gradient colors representing negative (not significant) to positive (significant) correlation. **c**. Representation of the three modules (left), the closest neighbors to candidate genes in each module (middle) and Gene Ontology findings, i.e. *Disease and Biological functions,* obtained using Ingenuity Pathway Analysis. MS, multiple sclerosis; PPMS, primary progressive MS; SPMS, secondary progressive MS; NNC, non-neurological controls.