Human Molecular Genetics, 2014, Vol. 23, No. 25 **6746–6761**
doi:10.1093/hmg/ddu392
Advance Access published on July 30, 2014

Functional variations modulating *PRKCA* expression and alternative splicing predispose to multiple sclerosis

Elvezia M. Paraboschi[1], Valeria Rimoldi[1], Giulia Soldà[1], Tommaso Tabaglio[1], Claudia Dall'Osso[1], Elena Saba[1], Marco Vigliano[2], Alessandro Salviati[3], Maurizio Leone[4,5], Maria D. Benedetti[7], Diego Fornasari[1], Janna Saarela[8], Philip L. De Jager[9,10,11,12], Nikolaos A. Patsopoulos[9,10,11,12], Sandra D'Alfonso[5,6], Donato Gemmati[2], Stefano Duga[1] and Rosanna Asselta[1,*]

[1]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milano 20133, Italy, [2]Center Hemostasis and Thrombosis, Hematology Section, Department of Medical Sciences, University of Ferrara, Ferrara 44124, Italy, [3]Unit of Neurological Genetics, Department of Neurological and Movement Sciences, University of Verona, Verona 37134, Italy, [4]SCDU Neurology, AOU "Maggiore della Carità", Novara 28100, Italy, [5]Interdisciplinary Research Center of Autoimmune Diseases IRCAD and , [6]Department of Health Sciences, University of Eastern Piedmont, Novara 28100, Italy, [7]Department of Neurosciences, University Hospital, Verona 37134, Italy, [8]Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland, [9]Department of Neurology, Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Brigham & Women's Hospital, Boston, MA, USA, [10]Division of Genetics, Department of Medicine, Brigham & Women's Hospital and, [11]Harvard Medical School, Boston, MA, USA and [12]Broad Institute of Harvard and MIT, Cambridge, MA, USA

Received April 30, 2014; Revised July 25, 2014; Accepted July 27, 2014

The protein kinase C alpha (*PRKCA*) gene, encoding a Th17-cell-selective kinase, was repeatedly associated with multiple sclerosis (MS), but the underlying pathogenic mechanism remains unknown. We replicated the association in Italians (409 cases, 723 controls), identifying a protective signal in the *PRKCA* promoter ($P = 0.033$), and a risk haplotype in intron 3 ($P = 7.7 \times 10^{-4}$; meta-analysis with previously published data: $P = 4.01 \times 10^{-8}$). Expression experiments demonstrated that the protective signal is associated with alleles conferring higher *PRKCA* expression levels, well fitting our observation that MS patients have significantly lower *PRKCA* mRNA levels in blood. The risk haplotype was shown to be driven by a GGTG ins/del polymorphism influencing the heterogeneous nuclear ribonucleoprotein H-dependent inclusion/skipping of a *PRKCA* alternative exon 3*. Indeed, exon 3* can be present in two different versions in *PRKCA* mRNAs (out-of-frame 61 bp or in-frame 66 bp long), and is preferentially included in transcripts generated through a premature polyadenylation event. The GGTG insertion downregulates 3* inclusion and shifts splicing towards the 66 bp isoform. Both events reduce the nonsense-mediated mRNA-decay-induced degradation of exon 3*-containing mRNAs. Since we demonstrated that the protein isoform produced through premature polyadenylation aberrantly localizes to the plasma membrane and/or in cytoplasmic clusters, dysregulated *PRKCA* 3* inclusion may represent an additional mechanism relevant to MS susceptibility.

INTRODUCTION

Multiple sclerosis (MS) (OMIM #126200) is an autoimmune disease characterized by demyelination in the central nervous system and chronic inflammation (1). According to the presentation and severity of symptoms, MS subtypes are classified as relapsing remitting (RR, the commonest form), primary progressive (PP) or secondary progressive (SP) (2). Current models of MS

*To whom correspondence should be addressed at: Department of Medical Biotechnologies and Translational Medicine, University of Milan, Via Viotti 3/5, 20133 Milan, Italy. Tel: +39 0250315853; Fax: +39 0250315864; Email: rosanna.asselta@unimi.it; rosanna.asselta@gmail.com

pathogenesis support the contribution of two interlaced processes: inflammation and neurodegeneration. Bursts of focal inflammation are indicated as responsible for the episodic RR phase, whereas axonal loss and neurodegeneration underlie progressive symptoms, which are the predominant cause of disability (3).

MS has a complex genetic architecture: besides environmental factors certainly playing a role (4), it is thought that multiple genes contribute, each with a modest effect, to the disease pathogenesis (5,6). To date, the human leukocyte antigen (*HLA*) gene cluster still remains the strongest susceptibility *locus* associated with MS (7). As for non-*HLA loci*, genome-wide association studies (GWAS) identified ~100 common variants contributing to disease pathogenesis with mild effects on risk (odds ratio, OR < 1.3), altogether pointing to a primary role for cell-mediated immune mechanisms but explaining only a modest fraction of MS heritability (8,9). Moreover, even when association signals were identified, the functional characterization of the underlying variants was often missing (for a list of the few fully characterized polymorphisms) (10).

The protein kinase C alpha (*PRKCA*) gene is one of the few *loci* that have been related to MS by both linkage analysis and association studies in different populations (Brits, Finns and Canadians), in which specific risk haplotypes were identified (11,12). The *PRKCA* protein (PKCα, 672 amino acids and 76.8 kDa) is a member of the conventional PKC family, which comprises three closely related serine/threonine isozymes (α, β and γ). These require Ca^{2+} and diacylglycerol (DAG) to become activated, and are involved in many signal transduction pathways connected with proliferation, apoptosis and differentiation (13). The PKCα domain structure consists of an N-terminal pseudosubstrate motif that provides kinase autoinhibition, a couple of DAG-binding C1 domains, a single plasma membrane-targeting C2 domain (specifically recognizing the inner leaflet lipids phosphatidylserine, PS and phosphatidylinositol-4,5-bisphosphate, PIP2) and a C-terminal kinase domain (14).

In the mouse, PKCα was shown to be part of a signalling pathway that is necessary both for full antigen receptor-mediated T-cell activation and for proper interferon-γ production, even though loss of PKCα was shown to have no apparent effects on T-cell development and selection in the thymus (15). On the other hand, in thymocytes of mice overexpressing PKCα, the kinase translocates to the cell membrane acting as a second messenger of the T-cell receptor-CD3 complex-delivered signal for both proliferation and interleukin (IL)-2 production (16). In human HepG2 cells, PKCα has been noted to inhibit inflammation by negatively regulating nuclear factor-κB-induced cytokine expression, in particular that of IL-1 (17). All these findings, together with the recent description of PKCα as a Th17-cell-selective kinase upregulating IL-17A during Th17-cell-mediated immune responses in mouse (18), make the *PRKCA* gene an excellent functional candidate for predisposition to MS.

In this work, we present data from a replication case–control study revealing a double-faced association between the *PRKCA* gene and MS. We characterized the protective/predisposing functional variants underlying the identified associations, demonstrating that MS-associated genetic variants act by modulating PRKCA levels and alternative splicing (AS) pattern.

RESULTS

PRKCA is associated with MS in the Italian population

To investigate the role of *PRKCA* as susceptibility gene for MS in the Italian population, we performed an association analysis on a cohort of 409 cases and 723 controls by genotyping three microsatellite markers located in the 5′ regulatory region (15xGCC), in intron 2 (20xAC) and 94 kb downstream of the 3′ end of the gene (20xAC-2) (Fig. 1A; Supplementary Material, Table S1). The 15xGCC and 20xAC microsatellites showed a robust association in the allelic analyses, being all the 4 χ^2 statistics T1–T4 significant (Supplementary Material, Table S1). The best associated were the 11-repeat-long allele ($P = 0.033$; OR = 0.12, 95% confidence interval, CI = 0.015–0.94) and the 9-repeat-long allele ($P = 0.013$; OR = 0.63, 95% CI = 0.40–0.98) for 15xGCC and 20xAC, respectively; both showed a protective effect. Genotypic analyses confirmed the association for both markers (Supplementary Material, Table S1).

To further dissect this association, 27 single-nucleotide polymorphisms (SNPs) were selected for genotyping (Fig. 1; Supplementary Material, Table S2). Among these, we chose all the 18 SNPs previously associated (nominally or within a haplotype) with MS (11,12), and 9 SNPs because of their position within potentially interesting regions of the gene (i.e. the promoter and the *hsa*-miR-634 gene, located within the *PRKCA* gene). None of the 20 informative SNPs showed any significant allele/genotype frequency difference between MS patients and controls (Supplementary Material, Table S3).

Haplotype analysis was performed based on the linkage disequilibrium (LD) structure of the gene (Fig. 1B) and revealed that haplotype CCGCAGG (Block 2) was overrepresented in MS cases (20.7%) respect to controls (14.4%), thus resulting strongly associated with the disease status ($P = 7.7 \times 10^{-4}$) (Supplementary Material, Table S4). This haplotype confers a 1.57-fold increased genetic risk of developing MS (95% CI = 1.24–1.99), and partially overlaps with those observed associated with MS in the Finnish and Canadian populations (Fig. 1A). Taking advantage of GWAS data derived from four independent case–control studies, as well as genotypes imputed in the same cohorts using the 1000Genome Project data, we tried to get a higher level of confidence of our association signal using a double approach. First, we performed a haplotype-based meta-analysis, including both the previously reported Finnish and Canadian haplotypes (Fig. 1A) and our CCGCAGG haplotype inferred in the four GWAS cohorts, for a total of 5181 cases and 9605 controls (Supplementary Material, Table S5). This resulted in an overall OR of 1.23 (95% CI = 1.14–1.32) with a *P*-value of 2.31×10^{-6} (Cochran *Q* test for heterogeneity, $P = 2.25 \times 10^{-5}$, mainly due to the major contribution of the Finnish cohort of the IMSGC study group). Second, we performed a single-marker association meta-analysis in the four GWAS cohorts using 1284 SNPs distributed along the *PRKCA* gene (all having a minor allele frequency >5%). This analysis pointed out 24 significantly associated markers ($P < 0.01$), all clustered in a 50 kb-long region within *PRKCA* intron 3 (Supplementary Material, Fig. S1). Again, this region overlaps with the CCGCAGG haplotype.

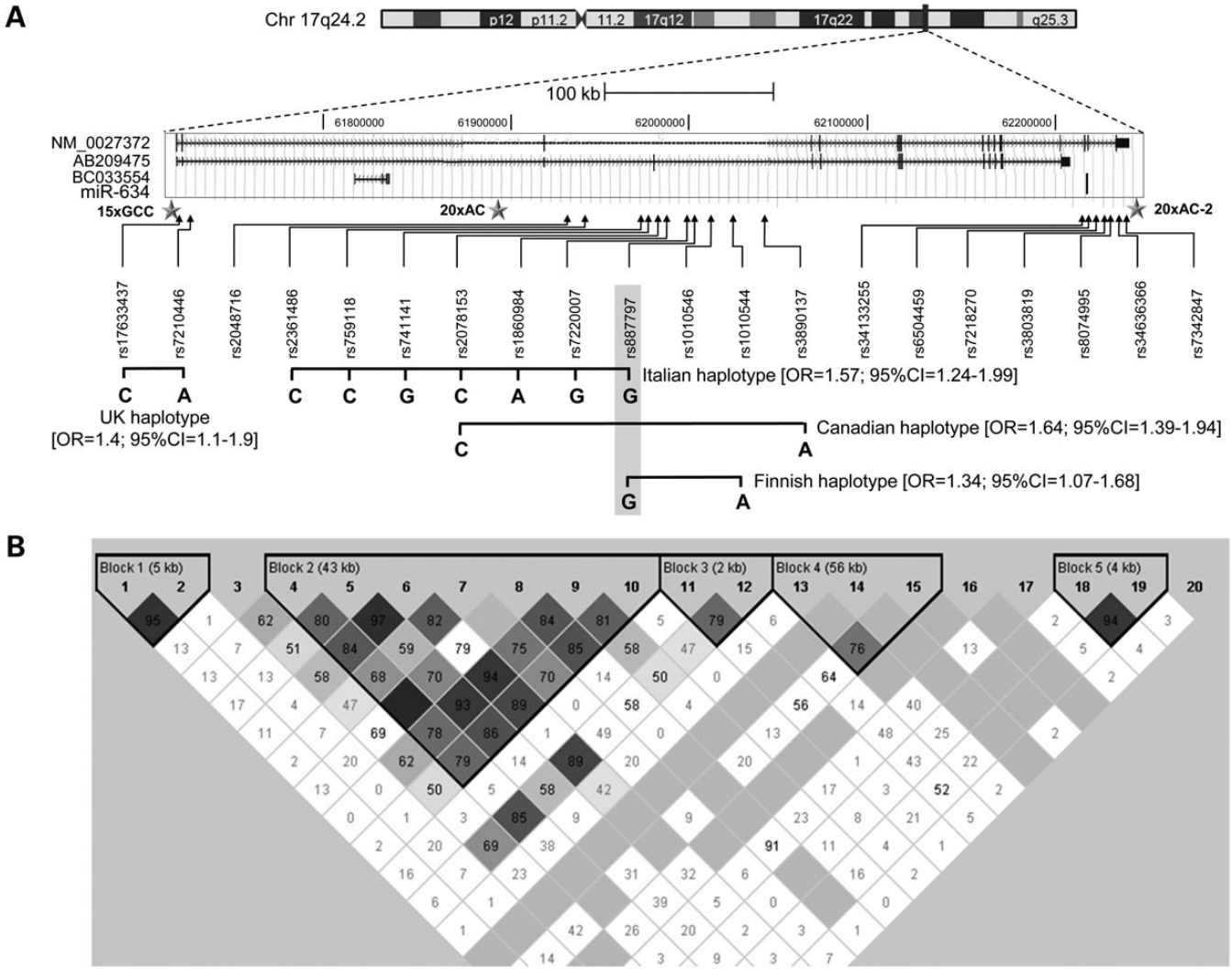

Figure 1. Structure of the *PRKCA locus*. (**A**) Genomic position and exon–intron structure of the *PRKCA* gene (RefSeq NM_0027372), as well as of the annotated AB209475 isoform, of the BC033554 cDNA clone and of *hsa*-miR-634 are shown (drawn to scale). Genotyped microsatellites and SNPs are listed, and their locations relative to PRKCA are shown by stars and arrows, respectively. *P*-values of association relative to microsatellite markers are shown in squared brackets. Risk haplotypes identified in this or previous works are indicated as horizontal lines. OR and CI for the Finnish and Canadian haplotypes were calculated on the basis of published data (12). The overlapping region of risk haplotypes is shaded in gray. (**B**) LD structure of the *PRKCA locus*. Pairwise LD values, estimated for the genotyped SNPs, are represented by boxes (black indicates strong LD, light gray indicates intermediate, white denotes no LD). D' values are shown within the boxes, empty cells are in complete LD ($D' = 100$).

The number of GCC repeats in the promoter modulates *PRKCA* transcription levels

To determine whether the variable length of the MS-associated $(GCC)_n$ repeat in the promoter is able to modulate *PRKCA* transcription, we cloned six promoter fragments, corresponding to the most common alleles, in a luciferase reporter vector. The luciferase assays were performed in HeLa cells, which constitutively express *PRKCA* (see further). In general, we observed a significant variation in *PRKCA* expression levels depending on the allele length ($P = 1.4 \times 10^{-7}$). In particular, the highest level was associated with the allele carrying 11 GCC repeats (i.e. the protective allele), which showed a 1.3- to 5-fold increase in the transcriptional activity respect to other alleles (Fig. 2A). This observation suggests that higher expression levels of *PRKCA* may confer protection against MS.

To confirm this hypothesis, we evaluated *PRKCA* expression levels in peripheral blood mononuclear cells (PBMCs) of 49 RR-MS patients (free from treatments) and 61 controls by semi-quantitative real-time reverse-transcription (RT)–PCR. Indeed, *PRKCA* resulted significantly overexpressed (on average 2-fold) in healthy individuals respect to MS patients ($P < 1 \times 10^{-4}$) (Fig. 2B, left). Mean expression levels were also plotted stratifying individuals according to their 15xGCC genotypes. In view of the low frequency of homozygous genotypes, we considered only subjects heterozygous for the most common allele (i.e. 9GCC repeats), and grouped them as carriers of alleles that should drive lower (six repeats) versus higher expression

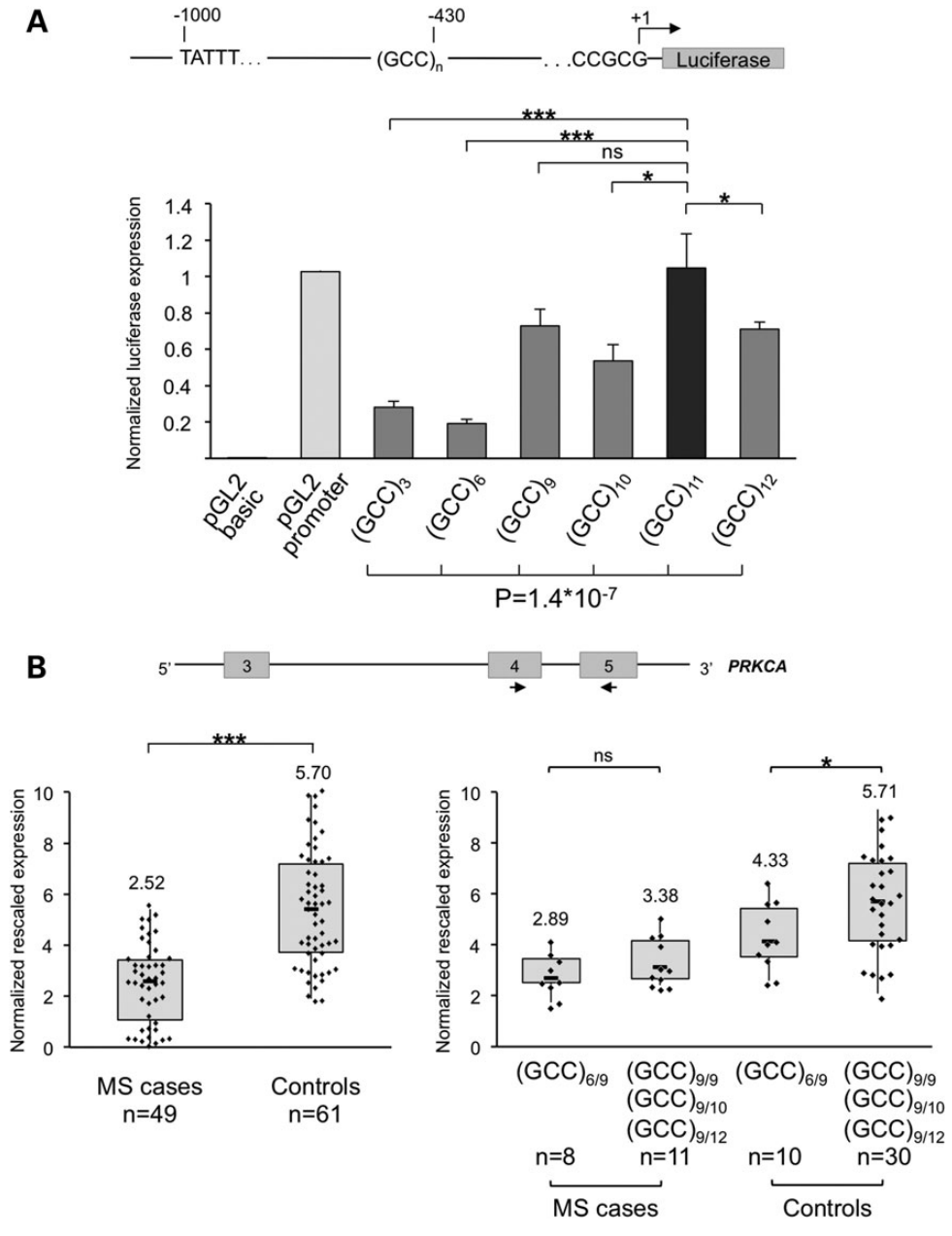

Figure 2. Allele-specific modulation of *PRKCA* expression by the promoter 15xGCC microsatellite. (**A**) *In vitro* experiments. A schematic representation of *PRKCA* cloned promoters is reported. The position of the GCC microsatellite relative to the transcription start site is indicated. In the lower part of the panel, relative luminescence units are reported as normalized values (*Y*-axis). The pGL2-basic and -promoter vectors were used as controls. The transcriptional activity of the pGL2-promoter plasmid was set as 1. Bars represent means ± SEM of at least three independent experiments, each performed in triplicate. Significance levels of *t*-tests [all comparisons performed against the $(GCC)_{11}$-promoter activity] and of the one-way ANOVA are shown above and below the bars, respectively. (**B**) Total *PRKCA* expression levels in PBMCs of MS cases and controls. Expression levels were measured by semi-quantitative real-time RT–PCRs, using primers located in exons 4 and 5 (see the scheme above boxplots). Boxplots show *PRKCA* expression levels according to disease status (left), or grouping subjects according to their 15xGCC genotype (right). Results are presented as normalized rescaled values. Mean values are shown above each boxplot. *t*-Test significance: $^{*}P < 0.05$; $^{***}P < 1 \times 10^{-4}$; ns, not significant.

levels (9, 10 and 12 repeats) on the basis of reporter assay experiments. Both for cases and controls, the six-repeat allele was associated with lower *PRKCA* expression levels, even though the difference was statistically significant only in controls (Fig. 2B, right). This prompted us to come back to genetics, reclassifying 15xGCC alleles as high risk (low *PRKCA* levels; 3 and 6 repeats) or low risk (high *PRKCA* levels; 9, 10, 11 and 12 repeats). In the subsequent association analysis, a significant difference in genotype frequency distribution between MS cases and controls was observed ($P = 0.035$). Moreover, the frequency of individuals carrying two high-risk alleles among patients resulted significantly higher than among controls (8.4 versus 3.8%, respectively; recessive mode of inheritance, $P = 0.036$; OR = 2.20, 95% CI = 1.04–4.66).

The association signal in *PRKCA* intron 3 is driven by an ins/del polymorphism

All the identified risk signals map within *PRKCA* intron 3 (Fig. 1A; Supplementary Material, Fig. S1). This region includes an additional exon (hereafter referred to as exon 3*), which is part of an alternative *PRKCA* transcript, lacking the last 3 exons (GenBank accession number AB209475) (Fig. 1A). Sequencing of the region surrounding exon 3* in 30 MS patients carrying at least 1 risk haplotype revealed the existence of an ins/del polymorphism (i.e. presence of 1 or 2 GGTG tandem repeats) located at position +4 of the donor splice site of the exon 3*. Bioinformatics analyses evidenced that the presence of 2 GGTG repeats could determine the activation of an alternative 5′ splice site, located 5 nucleotides downstream of exon 3* donor splice site (Supplementary Material, Fig. S2). The ins/del polymorphism later appeared in dbSNP 131 database as rs35476409/rs61762387.

The association of the ins/del polymorphism with MS was evaluated in our case–control cohort. Allelic analysis revealed only a slight increase of the 1-repeat allele frequency in healthy individuals (31.5% in cases versus 34.5% in controls; $P = 0.11$). However, when genotype distributions were analyzed, a significant difference between cases and controls was found both in the genotypic analysis ($P = 0.042$) and under the recessive model of inheritance ($P = 0.013$) (Supplementary Material, Table S6). All together, these results suggest that the two-repeat allele could be a risk factor for MS, acting in a dominant fashion (OR = 1.65; 95% CI = 1.11–2.46).

The inclusion of the two-repeat allele within the risk haplotype improved the significance of the association signal (from $P = 7.7 \times 10^{-4}$ to $P = 2.8 \times 10^{-4}$) (Supplementary Material, Table S4): the extended haplotype (CCG2CAGG) was found in 29.2% of cases versus 21.2% of controls. In this analysis, the complementary haplotype TTA1GAGA also emerged as significantly associated, and showed a protective effect (4.5% of cases, 9.6% of controls; $P = 2.2 \times 10^{-4}$) (Supplementary Material, Table S4). Unfortunately, it was not possible to impute the ins/del polymorphism in the GWAS cohorts for further replication using the 1000Genome Project data.

The ins/del polymorphism modulates the splicing of the alternative exon 3*

To test whether the 1-repeat (1R) and 2-repeat (2R) alleles of rs35476409/rs61762387 could affect the splicing of exon 3*, we used a combination of *ex vivo* (hybrid minigenes) and *in vivo* (RT–PCRs on PBMC RNAs) approaches.

For *ex vivo* splicing assays, a *PRKCA* region of 433 bp, containing exon 3* along with intronic flanking sequences, was PCR amplified from the genomic DNA of a healthy individual, heterozygous for the risk haplotype, and both 1R and 2R alleles were cloned into the α-globin-fibronectin minigene plasmid (pBS-KS modified) (19). Considering that the strongest association signal was detected by incorporating the putative splicing polymorphism into an extended haplotype, a longer construct of 4.5 kb was also produced. This minigene comprises the two closest SNPs belonging to the risk haplotype upstream and downstream of the ins/del polymorphism (Fig. 3A).

The short minigene constructs were transiently transfected into nine different cell lines. Analysis of transcripts generated from the two allelic variants showed that exon 3* skipping/inclusion is strictly dependent on the ins/del genotype, with the MS-associated 2R allele always resulting in exon 3* skipping (Fig. 3B). This "all-or-nothing" effect was observed all the analyzed cell lines, except GM06895 and EAhy296, in which a low level of exon 3* skipping was found in the presence of the 1R allele (Fig. 3B).

Transfection experiments of the long minigene constructs confirmed that exon 3* skipping/inclusion depends on the ins/del genotype: in this more physiologic construct, exon 3* was less efficiently included in transcripts produced by the 1R allele, better mimicking the *in vivo* situation (see further). However, the unexpected activation of additional pseudoexons, undetectable *in vivo*, prevented a reliable quantitative evaluation of exon 3* levels (data not shown). For this reason, a different primer couple, anchored to exon 3*, was used in RT–PCR experiments (assay B; Fig. 3A). These assays in HeLa cells revealed that, only in the presence of the additional intronic sequences, low levels of exon 3* inclusion could be detected also in the 2R transcript (Fig. 3C). Moreover, a 5 nt-longer exon 3* (exon 3*+5nt), previously predicted by *in silico* analysis (Supplementary Material, Fig. S2), was detected only in cells expressing the long construct carrying the 2R allele (Fig. 3C). These data were confirmed by fluorescent RT–PCR: capillary electrophoresis analysis of the obtained fragments and peak area measurements revealed an ∼20% of the 3*+5nt isoform in cells transfected with the 2R construct (Fig. 3C).

Summarizing, our experiments in transfected cells demonstrated the presence, besides the wild-type one, of 2 *PRKCA* mRNA Variants, V* (including exon 3*) and V*+5nt (including exon 3*+5nt). The presence of exon 3*/3*+5nt including transcripts as well as the relative balance between V* and V*+5nt isoforms is dependent on the ins/del polymorphism genotype (i.e. 2R is associated with less 3* inclusion and allows the recognition of the V*+5nt donor splice site).

The actual existence *in vivo* of V* and V*+5nt isoforms and their relative amount to the total *PRKCA* transcript were first evaluated by fluorescent competitive RT–PCRs on RNA extracted from PBMCs of a healthy 1R/1R individual (Fig. 4A and b). A very low, but measurable, level of 3*-containing transcript was detected (Fig. 4B). This is substantially different from what observed in minigene experiments, where the 1R genotype was associated with an almost complete exon 3* inclusion. This is not totally unexpected considering the very large size of *PRKCA* intron 3 (∼150 kb), which may include additional splicing regulatory elements, and the possible effect *in vivo* of the nonsense-mediated mRNA-decay (NMD) pathway (see further). Next, we measured V*/V*+5nt expression levels by semi-quantitative real-time RT–PCRs in 110 PBMC RNA samples. Stratifying individuals on the basis of their ins/del genotype, the amount of V*/V*+5nt transcript ranged from 0.3 to 1.2% of the total *PRKCA* mRNA (Fig. 4C). In particular, 1R homozygotes showed a significantly higher expression of isoforms containing exon 3* (V* and V*+5nt), when compared with 1R/2R heterozygous ($P < 1.0 \times 10^{-4}$ and $P = 0.022$, for cases and controls, respectively) and 2R homozygous individuals ($P < 1.0 \times 10^{-4}$ and $P = 0.027$, for cases and controls,

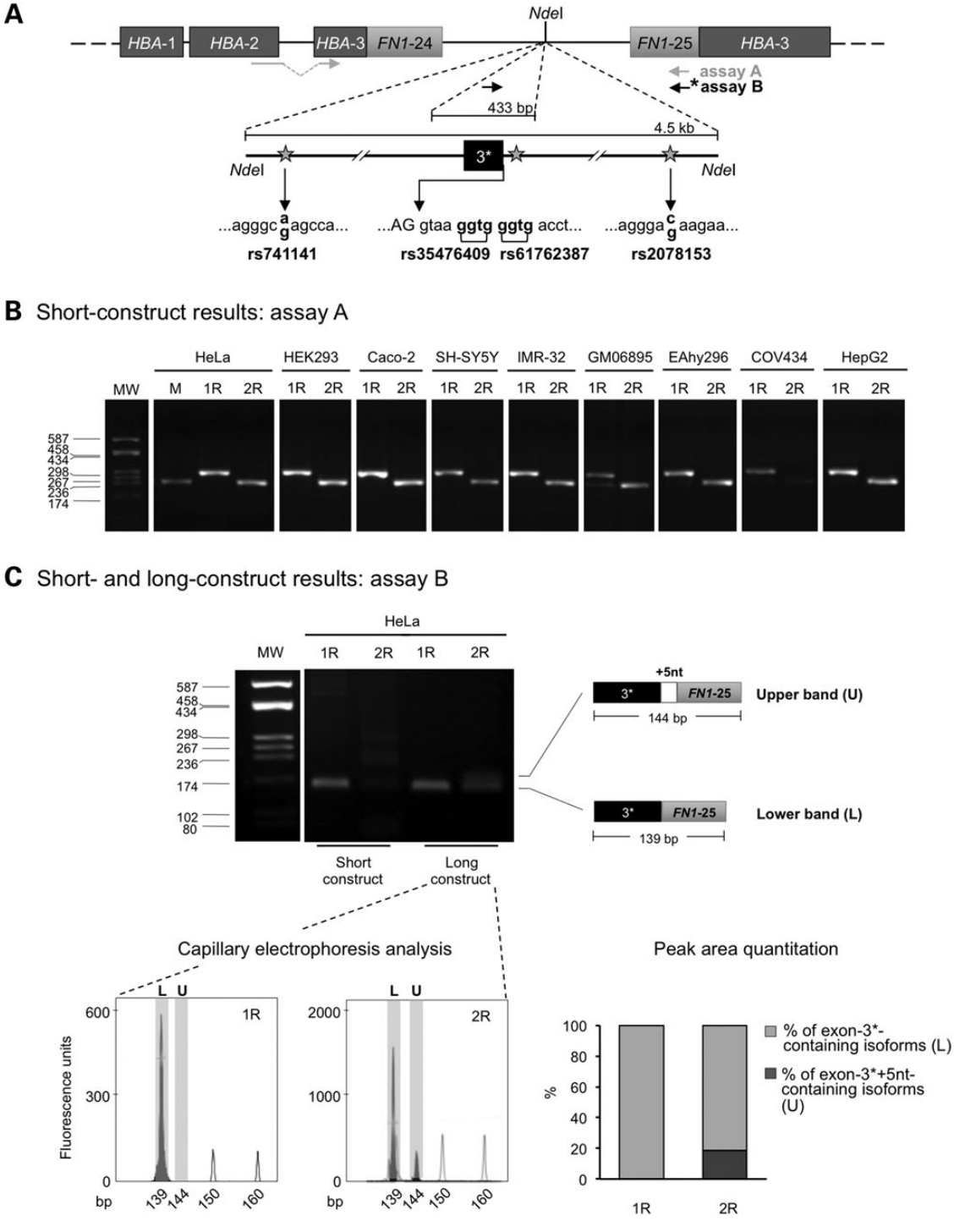

Figure 3. *Ex vivo* functional characterization of the ins/del polymorphism. (**A**) Schematic representation of the hybrid pBS-KS minigenes: α-globin (*HBA*) and fibronectin (*FN1*) exons are represented by boxes, introns by lines (not to scale). *PRKCA* exon 3* (black box) and its flanking intronic regions are also reported. The ins/del polymorphism and the two closest SNPs belonging to the risk haplotype are indicated by a star and their dbSNP ID. The positions of the primer couples used for RT–PCR experiments is shown by gray (assay A) or black arrows (assay B; asterisk indicates the FAM-labeled primer). (**B**) RT–PCR products (assay A) obtained from RNA of nine cell lines transfected with the short (433 bp-long insert) minigene constructs, separated on a 2% agarose gel. MW: molecular-weight marker (pUC8-*Hae*III). M (mock): cells transfected with the empty vector. 1R and 2R: cells transfected with the vector containing 1 or 2 GGTG repeats. The longer band (307 bp long) corresponds to exon 3* inclusion, the shorter band (246 bp long) is obtained when exon 3* is skipped from the mature transcript. (**C**) Upper-left panel: RT–PCR products (fluorescent assay B) obtained from RNA of HeLa cells transfected with either the long or the short constructs, separated on a 2% agarose gel. Upper-right panel: schematic representation of the obtained RT–PCR products. The 139 and 144 bp-long bands correspond to the inclusion of exon 3* (lower band, L) and of a 5 bp-longer exon 3* (upper band, U), respectively. Lower-left panels: GeneMapper windows displaying fluorescence peaks corresponding to the two molecular species (U and L) are shown. Empty peaks represent the size standard (ROX-500 HD), filled peaks correspond to the RT–PCR-labeled products. The *X*-axis indicates GeneMapper data points and the *Y*-axis represents fluorescence units (FUs). Lower-right panel: histograms indicate the quantitative analysis on fluorescence peak areas. For each sample, the sum of all the peak areas was set as 100%.

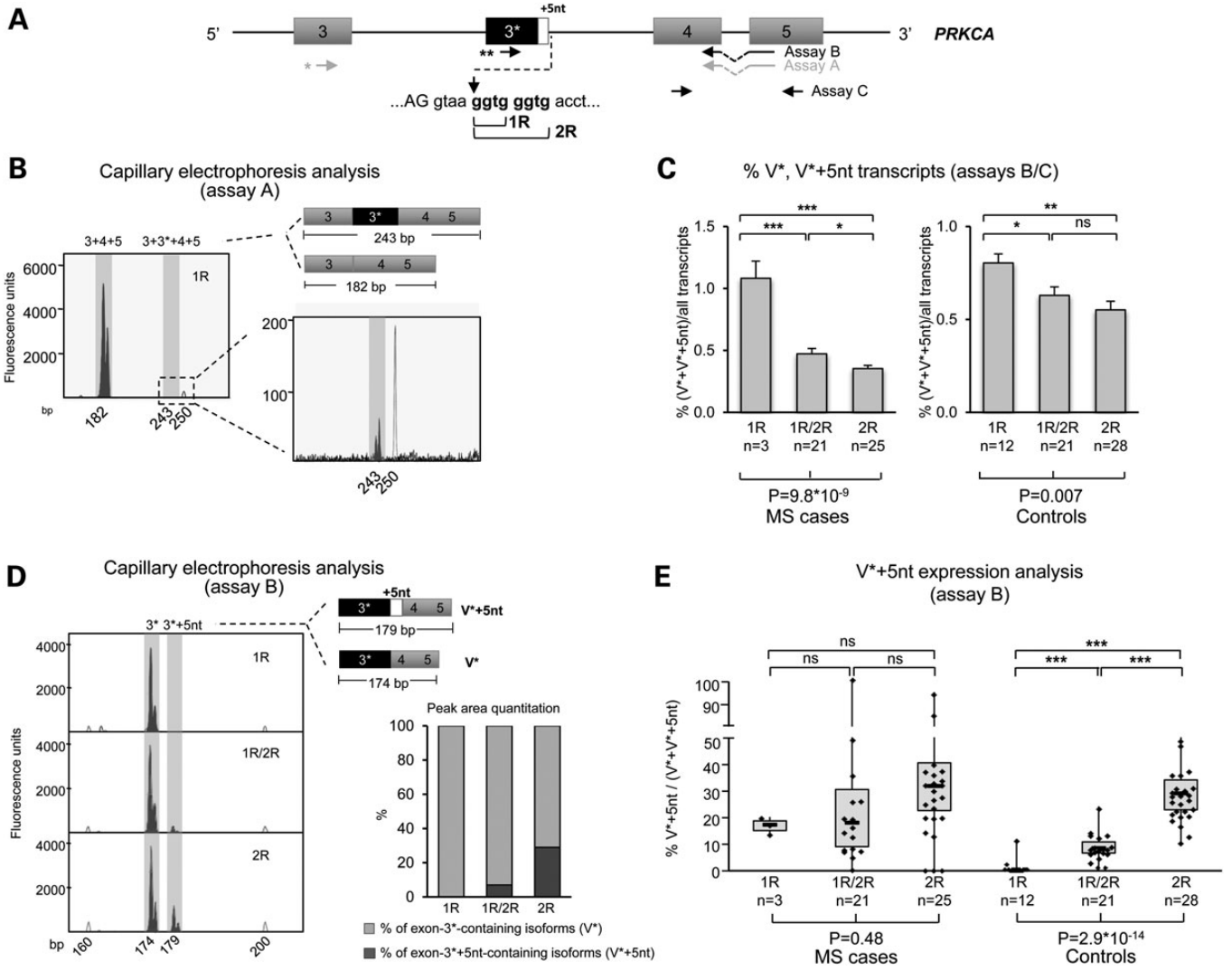

Figure 4. *In vivo* analysis of the ins/del polymorphism. (**A**) Schematic representation of the *PRKCA* genomic region spanning from exon 3 to exon 5 (not to scale). The positions of the primer couples used for RT–PCR experiments is shown by arrows (assays A, B or C); assays A and B were used both in real-time experiments and in competitive PCRs (in this case the forward primer was labeled with the HEX or FAM fluorophore). The donor splice-site region of exons 3*/3*+5nt is also reported, together with the ins/del polymorphism determining the 1R and 2R genotypes. (**B**) Left panel: RT–PCR products (fluorescent competitive assay A), obtained from the RNA extracted from PBMCs of a control individual (homozygous 1R for the ins/del polymorphism). Right panel: close-up view of the GeneMapper window encompassing the peak of the exon 3*-containing isoform. A schematic representation of the two obtained RT–PCR products is also depicted. (**C**) Specific V* and V*+5nt (assay B) as well as "total" *PRKCA* (assay C) expression levels were measured by semi-quantitative real-time RT–PCRs in PBMCs of 61 healthy individuals and 49 MS cases, and data were stratified according to the ins/del genotype. Histograms represent percentages of V* plus V*+5nt over total *PRKCA* transcripts, calculated by the ΔΔCt method. (**D**) Left panels: RT–PCR products (fluorescent competitive assay B), obtained from RNA extracted from PBMCs of three controls (homozygous 1R or 2R and heterozygous 1R/2R for the ins/del polymorphism). A schematic representation of the two obtained RT–PCR products is also depicted. In the right panel, histograms indicate the quantitative analysis on fluorescence peak areas. (**E**) Expression levels of the specific *PRKCA* V*+5nt isoforms in PBMCs of MS cases and controls. Boxplots show the percentage of the V*+5nt transcript among all V*-containing mRNAs, according to disease status and by stratifying subjects on the basis of their ins/del genotype. Expression levels were measured by fluorescent competitive RT–PCRs. Significance levels of *t*-tests and of the one-way ANOVA analyses are shown above and below the boxplots, respectively. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 1 \times 10^{-4}$; ns, not significant.

respectively). The effect of the genotype seemed more pronounced in cases (Fig. 4C).

To specifically detect and quantitate the V*+5nt isoform, we first used the previously described 3*-specific fluorescent competitive RT–PCR assay on RNA extracted from three healthy individuals with different ins/del genotypes. The *in vivo* results mirrored those obtained *in vitro*, showing a clear dependence of the alternative 5′ splice-site usage on the 2R allele (Fig. 4D). Hence, we extended the measurement of the V* and V*+5nt ratio to the whole RNA sample cohort. In general, the level of the V*+5nt variant resulted significantly higher in MS cases versus controls ($P = 0.002$), with the mean V*+5nt percentage of inclusion in cases being about twice that measured in controls (24.5 ± 3.2% versus 13.8 ± 1.6%). More importantly, a striking difference was evidenced when we compared cases and healthy subjects grouped on the basis of their ins/del

genotype: while a clear dependence of V*+5nt level upon the number of GGTG repeats was seen in controls (difference among groups $P = 2.9 \times 10^{-14}$), the distribution observed in MS cases resulted completely flattened ($P = 0.48$) (Fig. 4D). In particular, MS patients showed percentages of the V*+5nt isoform of 14.1, 22.1 and 27.5% for the 1R, 1R/2R and 2R genotypes, respectively (Fig. 4E, left); conversely, in controls the mean percentage of the V*+5nt isoform resulted close to zero (1.1%) for 1R homozygous individuals and of 7.7 and 25.1% for heterozygotes and 2R homozygotes, respectively (Fig. 4E, right).

Exon 3* inclusion is modulated by the heterogeneous nuclear ribonucleoprotein H

The clear dependence of exon 3*/3*+5nt inclusion upon the ins/del genotype, prompted us to further dissect the underlying molecular mechanism. In particular, the presence of 2 GGTG repeats is predicted to create a well-known intronic splicing silencer (ISS) site, i.e. a stretch of 3 G (Fig. 5A), which is potentially recognized by the splicing-inhibitory factor heterogeneous nuclear ribonucleoprotein (hnRNP) H (20). This ISS is absent in the protective 1R allele, thus possibly explaining the observed increase in exon 3* inclusion.

We hence examined whether hnRNP H plays a role in exon 3* splicing by performing overexpression experiments in cell lines endogenously expressing *PRKCA* and differing for their ins/del genotype, i.e. HEK293 (1R homozygous) and IMR-32 (2R homozygous). Nicely confirming our hypothesis, hnRNP H overexpression led to a significant downregulation of both 3*/3*+5nt containing transcripts only in IMR-32 cells ($P = 0.0014$; Fig. 5B, left). Concomitantly, a significant shift in the balance between V* and V*+5nt towards the 3*+5nt isoform was evidenced by fluorescent competitive RT–PCRs ($P = 0.0069$; Fig. 5B, right). The same results were obtained in immortalized 1R or 2R homozygous fibroblasts (Fig. 5C), thus excluding cell-specific effects, and confirming that hnRNP H response strictly relies on the presence of 2 GGTG repeats. As a further confirmation, we showed that HeLa cells (1R/2R) silenced for hnRNP H significantly increased exon-3* inclusion (Fig. 5D).

PRKCA shows a complex AS architecture and AS/polyadenylation coordination

To better unravel the functional implications of the ins/del polymorphism in MS pathogenesis, we decided to comprehensively study the *PRKCA* AS pattern and expression profile.

Long-range RT–PCR experiments, performed on RNA extracted from PBMCs of a healthy individual, detected the presence of at least six different transcripts, which differ for the presence/absence of exon 3* and its variant 3*+5nt, and/or for the use of alternative polyadenylation sites (Fig. 6A and b). These transcripts will be hereafter referred to as V_L (*l*ong isoform), V_S (*s*hort isoform), V* (if including exon 3*) and V*+5nt (if including 3*+5nt). The only isoforms already annotated in GenBank correspond to the V_L (wild type, NM_002732) and V_S* transcripts (AB209475) (Figs 1a and 6a).

The tissue distribution of these *PRKCA* variants was analyzed in a panel of 22 human tissues, showing a nearly ubiquitous expression of all transcripts (Supplementary Material, Figs S3 and S4). In all tissues, longer isoforms resulted more expressed than shorter ones (from 10- to 250-fold). In MS-relevant tissues, i.e. brain and PBMCs, shorter isoforms collectively accounted for 0.7 and 5.9% of all transcripts, respectively (Supplementary Material, Fig. S3). Concerning *PRKCA* 3*/3*+5nt including isoforms, the highest levels were measured in brain and PBMCs; in particular, among lymphocyte subpopulations, Th17 cells showed a strikingly high level of these transcripts (Supplementary Material, Fig. S4).

Additionally, we noticed that the expression profile of 3*/3*+5nt including isoforms better correlates with that of *PRKCA* shorter transcripts (two-way ANOVA, $P = 0.06$) than that observed for longer ones (two-way ANOVA, $P = 4.2 \times 10^{-4}$), leading us to hypothesize the existence of an AS/polyadenylation coordination (21). In this view, short isoforms should preferentially include exon 3*/3*+5nt, which conversely should be more frequently skipped from longer transcripts. To compare the relative percentage of exon 3*/3*+5nt inclusion into shorter versus longer *PRKCA* isoforms, we performed semi-quantitative real-time RT–PCRs in PBMCs of healthy individuals using two specific RT primers, anchored to the relevant 3′UTR. As hypothesized, we observed a profound difference in the percentages of inclusion of exon 3*/3*+5nt (0.9 and 43.7% for longer and shorter isoforms, respectively; $P < 1 \times 10^{-4}$) (Fig. 6C), demonstrating a coordination between 3*/3*+5nt AS and the choice of the polyadenylation site during the maturation process of *PRKCA* transcripts.

PRKCA is regulated by the NMD pathway

The existence of AS/polyadenylation coordination, preferentially pushing alternative exon 3*/3*+5nt into shorter transcripts, is predicted to have important mechanistic consequences on the repertoire of translated products. All the five *PRKCA* alternative transcripts are predicted to code for aberrant proteins. Specifically, V_S would give rise to a PKCα in which the last 137 amino acids of the catalytic domain are replaced by a 15-residue-long polypeptide. V_S*+5nt and V_L*+5nt are predicted to code for potentially non-functional proteins containing an aberrant 22 amino acid sequence interrupting the PKCα phorbol-ester/DAG-binding domain. In the light of the AS/polyadenylation coordination, however, V_L*+5nt should be present in low amount among longer isoforms. Finally, V_S* and V_L* (this last one again present in traces among longer isoforms) would both generate a severely truncated protein lacking the entire catalytic domain, due to the inclusion of the 61 nt-long out-of-frame exon 3*. Alternatively, the introduction of a premature termination codon (PTC), might send V_S* and V_L* to NMD degradation, a translation-dependent mRNA-decay process targeting transcripts coding for potentially deleterious truncated proteins (22).

The possible degradation of V* isoforms through NMD was investigated by inhibiting the pathway with the use of puromycin and cycloheximide in HepG2 and HEK293 cells (23). Total RNA was extracted from treated/untreated cells and the relative amount of V* transcripts was assessed by semi-quantitative real-time RT–PCR. This analysis demonstrated a significant increase in the expression levels of exon 3*-containing mRNAs in treated

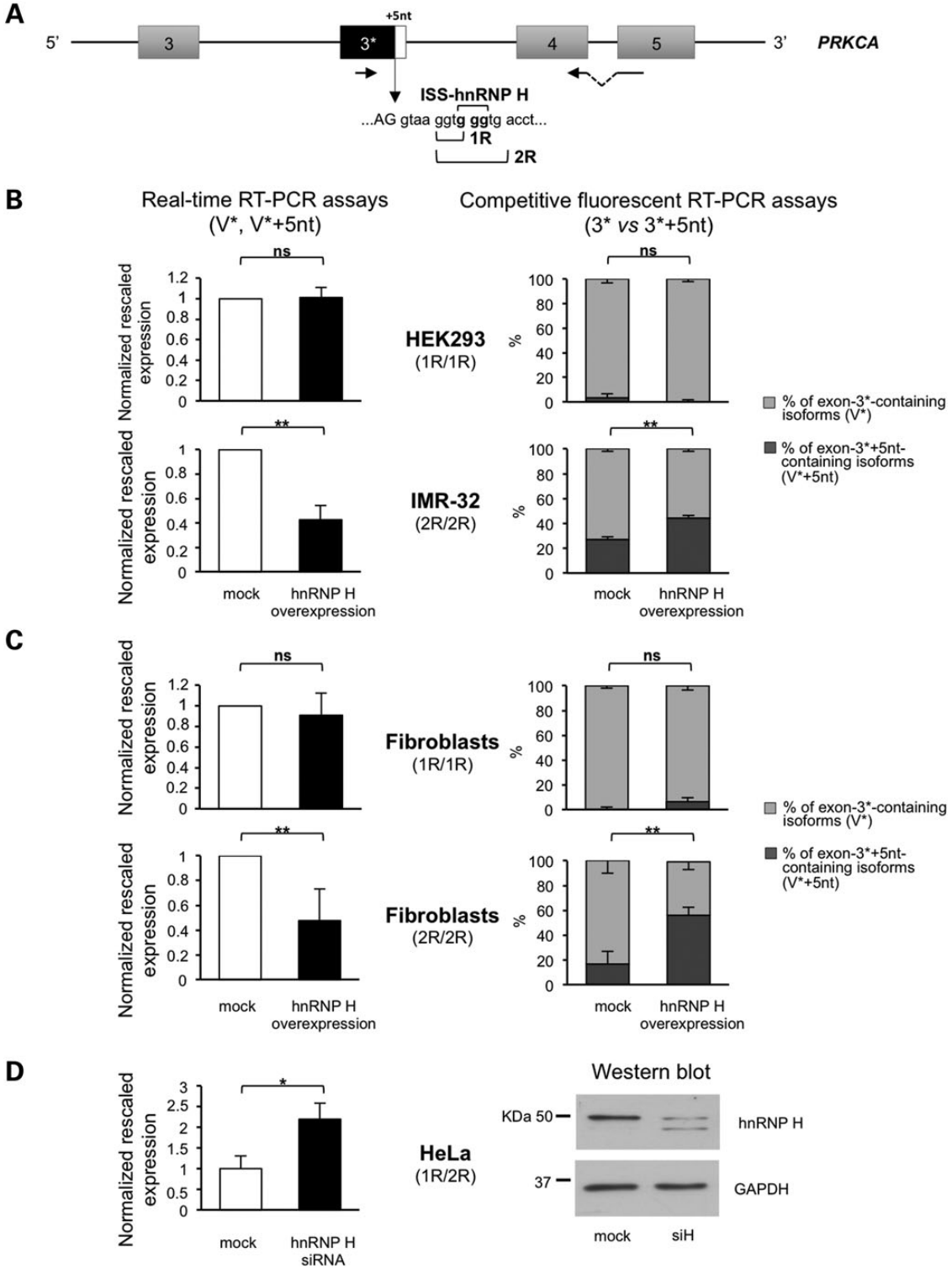

Figure 5. Effect of hnRNP H overexpression and silencing on V* and V*+5nt levels. (**A**) Schematic representation of the *PRKCA* genomic region spanning exons 3–5 (not to scale). RT–PCR primers are indicated by arrows. The ins/del polymorphism, determining the 1R/2R genotypes and the absence/presence of the ISS site, is also reported. (**B** and **C**) A plasmid expressing hnRNP H was transfected into the appropriate cell line (whose genotype for the ins/del polymorphism is reported in brackets). Left panels show expression levels of *PRKCA* isoforms V*/V*+5nt in cells transfected with an empty vector (mock) or overexpressing hnRNP H. Expression levels were measured by semi-quantitative real-time RT–PCRs and results are presented as normalized rescaled values, setting as 1 the value of the mock sample. Bars represent means ± SEM of at least three independent experiments, each performed in triplicate. Right panels show exon 3* and 3*+5nt quantitation on the same transfected cells by using the fluorescent competitive RT–PCRs. SEM is reported for both the 3*+5nt (upon the dark gray portion of the histogram) and the 3* (upside down) quantitation experiments. (**D**) Knockdown of hnRNP H was obtained by performing siRNA-mediated silencing in HeLa cells (heterozygous for the ins/del polymorphism). Left panel shows expression levels of *PRKCA* isoforms V*/V*+5nt (obtained as described above), whereas the right panel displays the western blot analysis demonstrating the actual silencing of hnRNP H in RNAi experiments. Significance levels of *t*-test: $^{**}P < 0.01$; $^{*}P < 0.05$; ns, not significant.

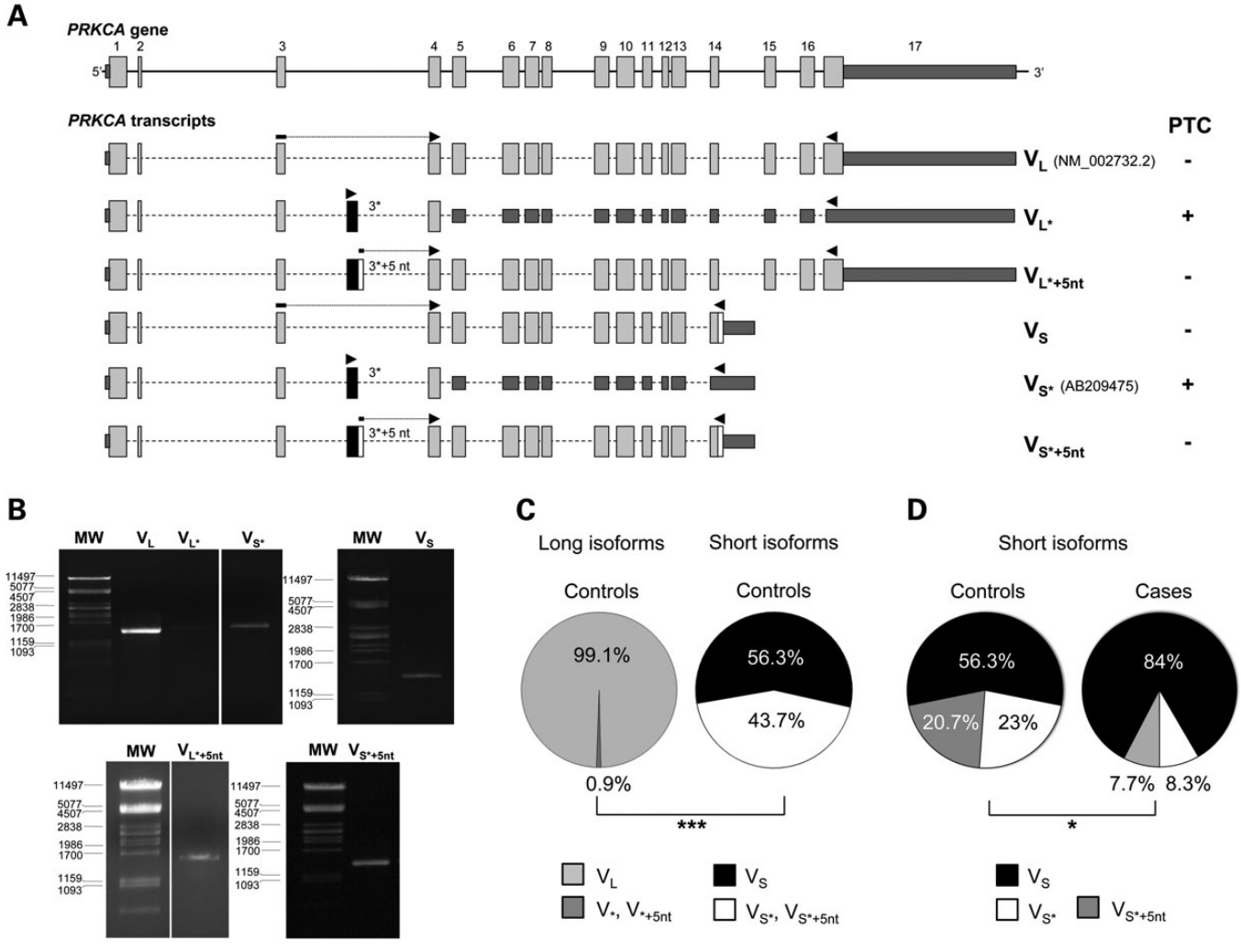

Figure 6. *PRKCA* splicing pattern and AS/polyadenylation coordination. (**A**) Schematic representation of the *PRKCA* gene and its transcripts. Exons and introns (dashed in the transcript schemes) are represented by boxes and lines, and are approximately drawn to scale; smaller boxes represent UTRs. Accession numbers are reported for isoforms annotated in genome databases. The presence in the transcript of a PTC is also indicated. Arrowheads above exons indicate the primers used in RT–PCR experiments. (**B**) Agarose gels of long-range RT–PCRs showing the existence of six *PRKCA* isoforms. RT–PCRs were performed on cDNA obtained from PBMCs of a healthy individual, using primers specific for the different isoforms (Supplementary Material, Table S8). MW, molecular-weight marker (λ-*Pst*I). (**C**) The amount of 3*-containing transcripts in long and short *PRKCA* isoforms was measured by semi-quantitative real-time RT–PCRs in PBMCs of 14 healthy individuals (all homozygous for the 2R risk allele), converted into percentages and presented as pie charts. A specific primer (anchored to the relevant 3′UTR) was used in the RT step. (**D**) Short isoforms were further analyzed by competitive fluorescence RT–PCRs in cases and controls (14 healthy individuals and 11 cases, all 2R/2R) to specifically measure the inclusion of exon 3* and 3*+5nt. Significance levels of *t*-tests are shown below the pie charts and refer to V*/V*+5nt comparisons ($^{*}P < 0.05$; $^{***}P < 1 \times 10^{-4}$).

cells, confirming that they are downregulated by NMD (Supplementary Material, Fig. S5A and B). As expected, the same experiment, performed on the in-frame V*+5nt isoforms, did not reveal any significant upregulation upon NMD inhibition in HepG2 cells, whereas these isoforms were not detected in HEK293 cells, due to their 1R/1R genotype (Supplementary Material, Fig. S5A and B). Data were confirmed by competitive RT–PCR.

Finally, in the light of the existence of AS/polyadenylation coordination, one could expect that NMD inhibition should preferentially upregulate shorter isoforms. Indeed, both in HepG2 and HEK293 cells, we observed a significant increase only of shorter transcripts, due to the preferential inclusion of exon 3* (Supplementary Material, Fig. S5C).

V_S PKCα is expressed and primarily localized to the plasma membrane

Given the *PRKCA* isoform diversity and complex posttranscriptional control, we sought to point out how the identified risk haplotype ultimately affects *PRKCA* function in MS. Normally, the 2R allele, by inhibiting the 3* inclusion (Fig. 4C), is responsible for a decrease in V_S* and V_L* levels, both actually destined to degradation; the 2R genotype also determines a relative increase of the stable V_S*+5nt and V_L*+5nt transcripts (Fig. 4D). This strict control, operated through the 2R sequence, is however partially compromised in MS patients, where an overrepresentation of V*+5nt isoforms was found even in carriers of the 1R allele (Fig. 4E). As a consequence, we expect,

in MS, a relative increase of the stable V_S isoform among shorter *PRKCA* transcripts. We hence measured the relative percentage of the specific V_S transcript respect to the overall amount of shorter *PRKCA* isoforms, revealing a significant increase in MS patients (on average, V_S represents 84% of all short isoforms in cases, and 56% in controls; $P = 0.027$; dosage performed on 2R-homozygous individuals) (Fig. 6D). As expected, given the 2R/2R genotype of the analyzed individuals, no significant difference in the use of the two possible alternative donor splice sites of exon 3* was seen between cases and controls (Fig. 6D).

To determine if the V_S transcript actually gives rise to the corresponding protein, we performed a western blot analysis on lysates of IMR-32 cells. This neuroblastoma line was chosen for its high expression levels of the shorter *PRKCA* isoforms (Supplementary Material, Fig. S6A). To produce a suitable positive control for western blot, the V_S PKCα isoform was overexpressed in COS-1 cells and the recombinant protein immunoprecipitated. After 2 h of exposure, we observed the presence of two distinct bands in IMR-32 cells: the higher one corresponding to the wild-type PKCα isoform, and the lower one compatible with the V_S variant, which hence seems to be physiologically expressed (Supplementary Material, Fig. S6B).

To better characterize the V_S PKCα protein, its subcellular localization was compared with that of the wild-type PKCα in transiently transfected Madin Darby canine kidney (MDCK) cells, both in basal conditions and after stimulation with TPA (4-β-12-*O*-tetradecanoyl-phorbol 13-acetate). The full-length protein was found diffusely distributed in the cytosol in unstimulated cells, whereas a 60 min treatment with TPA induced its massive translocation to the plasma membrane (Fig. 7). Conversely, the V_S protein showed a prevalent localization to the plasma membrane yet in basal conditions, with a 46% of cells characterized by V_S accumulating in cytoplasmic clusters. In cells overexpressing V_S PKCα, the TPA treatment had only a modest effect (Fig. 7). Moreover, at difference with the wild-type protein, V_S PKCα fluorescent signals overlapped with those of Golgi markers; no co-localization with early endosomes was evident (Supplementary Material, Fig. S7).

DISCUSSION

In this study, we identify different *PRKCA* alleles that are significantly associated with MS in the Italian population. Together with previous findings, our results confirm the *PRKCA* gene as a good candidate for MS susceptibility (11,12). Nonetheless, single-marker analysis of SNPs mapping in the *PRKCA* gene never reached significance in previous GWASs (24). This may be due to the fact that the actual functional variants were not genotyped/imputed, because of either the absence of any good proxy in the pool of genotyped polymorphisms, or the need of an interaction of multiple SNPs on a haplotype to generate a significant association. Both explanations fit to the present study. In fact, a signal of association corresponds to the 15xGCC promoter polymorphism, which is multiallelic and presents both protective and predisposing alleles, making it difficult to detect any association using a single SNP. In addition, our second association signal arose from a haplotype in which a functional ins/del polymorphism is necessary but not sufficient to explain the association signal; moreover, this variant shows a high allele frequency, so that it could easily lie within different haplotypes in different populations. Of note, haplotype analysis is not routinely performed in GWASs, principally because of multiple testing problems, even though it could help in dissecting genetic predisposition to complex diseases. This is particularly true when the *locus* under study is large, split into different LD blocks and characterized by the co-existence of multiple alleles with opposite effects (all features characterizing *PRKCA*).

Concerning our association data, we evidenced two protective signals by genotyping microsatellite markers: one located in the promoter of the gene, the other mapping in intron 2, close to an antisense transcript (BC033554; Fig. 1). After having verified that this antisense mRNA is expressed only in uterine cervix and prostate adenocarcinoma tissues (data not shown), the intron 2 region was no further investigated. Conversely, the possible impact of the 15xGCC microsatellite on the regulation of *PRKCA* expression was in-depth explored because of its position within the promoter, and because of its proximity to the 2SNP risk haplotype strongly associated with MS in the UK population (11). Our data showed that 15xGCC regulates *PRKCA* expression and suggested that the protective 11-GCC-repeat allele may decrease MS risk by enhancing transcription levels from 1.3 up to 5-fold respect to other alleles. This hypothesis well fits with the higher *PRKCA* levels measured in PBMCs of controls compared with MS cases ($P < 1 \times 10^{-4}$).

Emerging evidence suggests that microsatellites may play a key role in susceptibility to complex disorders, thus partially filling the gap left by GWASs (25). Short tandem repeat polymorphisms are widely distributed throughout the genome, with recent bioinformatics analyses suggesting that conserved mammalian microsatellites are overrepresented in the promoter region (within 5 kb from the transcription start site) of genes involved in CNS development and physiology (26). Intriguingly, the third most common motif in promoter regions is CCG/CGG, which instead is very rare in the rest of the genome (27). These motifs have the potential to be sites of epigenetic modifications (CpG methylation) (28) or to form various secondary structures, which are known to be involved in the regulation of gene expression by modulating polymerase activity (29). In both cases, a change in the number of the repeats would alter the modulation of gene expression, either by increasing/decreasing the number of available methylation sites or by modifying the overall secondary structure of the region. Concerning our 15xGCC microsatellite, UCSC Genome browser tracks show that it is not conserved among mammals and apparently it is not site of methylation (ENCODE data). On the other hand, since a certain correlation between the length of the 15xGCC microsatellite and the strength of the *PRKCA* promoter was evidenced in our reporter assays (Fig. 2A), an alternative possibility could be that the GCC repeats represent binding sites directly recognized by transcription factors: the greater the number of repeats, the higher the number of proteins bound. Indeed, the 15xGCC microsatellite is predicted to contain multiple binding sites for members of the activating protein 2 family of transcription factors (TRED—promoter database, http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home), known to be implicated in vertebrate development, apoptosis and cell-cycle control (30).

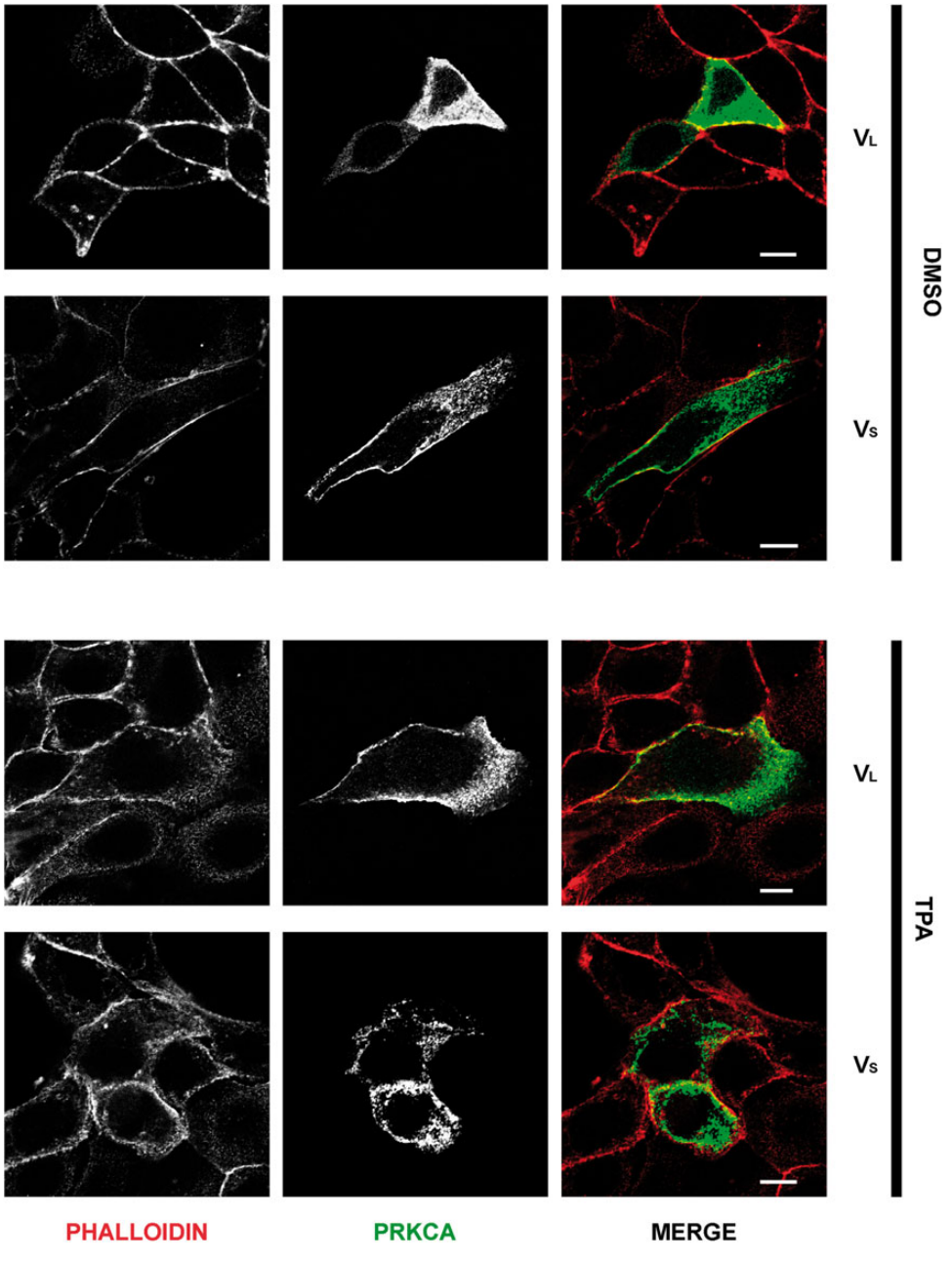

Figure 7. Characterization of the wild-type and V_S PKCα proteins by immunofluorescence. MDCK cells were transiently transfected with plasmids expressing either the wild-type (V_L) or the V_S PKCα, and then incubated with/without TPA for 1 h at 37°C. Cells were fixed and immunostained with the anti-PKCα antibody (green); actin filaments were visualized by Rhodamine phalloidin staining. Single confocal sections are shown. Scale bar: 10 μm.

As for the second association signal, the identification and molecular characterization of the specific genetic variant underlying the 43 kb-long CCGCAGG risk haplotype allowed us to describe a possible pathogenic mechanism, which stems on two interlaced processes: altered splicing and NMD. In particular, the protective 1R allele is associated with an increased inclusion of the out-of-frame exon 3* in *PRKCA* mature transcripts, especially in the shorter ones due to AS/polyadenylation coordination. This unproductive and apparently deleterious splicing event is indeed coupled with NMD, which hence prevents the production of the severely truncated version of PKCα, originating from the V_S isoform. This phenomenon, known as RUST (regulated unproductive splicing and translation), has already been described as a mechanism controlling gene expression for a number of genes (31). The 2R allele, conversely, is associated with two AS events: the skipping of exon 3* and the inclusion of exon 3*+5nt. Both these splicings determine the production of a low amount of stable aberrant transcripts that could be translated in potentially deleterious PKCα isoforms: V_S, V_S*+5nt and, to a lesser extent, V_L*+5nt. This AS profile, strictly depending on the ins/del genotype, is however lost in patients (Fig. 4E). Particularly, intriguing is the results

of competitive RT–PCRs, directly measuring the levels of inclusion of exon 3* versus exon 3*+5nt: the increased inclusion of the extended version of exon 3*, paralleled by a decreased level of exon 3* (Fig. 4E), reflects in the relative augmented levels of all in-frame potentially deleterious *PRKCA* isoforms (especially V_S; Fig. 6D) in MS patients. This altered balance of *PRKCA* isoforms in cases highlights a further layer of complexity, pointing to a possible dysregulation of the splicing process and/or of NMD in MS.

While no evidence involving dysregulation of NMD in MS are available so far, the role of altered splicing in MS pathogenesis is clearly emerging (10,32). The first study monitoring differential AS patterns in MS revealed no substantial changes in full-length transcript levels between 20 RR patients (remitting phase) and 20 controls, but a highly significant differential expression of a large number of alternatively spliced mRNAs (33). Interestingly, clustering of AS exons well-discriminated MS cases from controls. The results of our competitive RT–PCRs well fit with this observation: using the mean expression level of exon 3*+5nt inclusion in heterozygous controls as cut-off value (i.e. 8%), individuals having >8% of inclusion had an OR = 3.1 (95% CI = 1.22–7.85) for MS, suggesting that this parameter could be used as a predictive biomarker, independently from 1R/2R genotype (Fig. 4E).

Besides the abovementioned data, Evsyukova *et al.* (10) well highlighted the numerous evidence supporting a pathogenic role of dysregulated AS in MS, ranging from its involvement in modulating immunogenicity and neurodegeneration, to genetic susceptibility. In this frame, the 1R/2R polymorphism in *PRKCA* adds to the few variants associated with MS that have been characterized at the molecular level, all acting on splicing (10).

Clearly, we are aware of the fact that the causality of *PRKCA* variants in MS pathogenesis has not been formally proven, and the exact role of the increased production of the stable and aberrant V_S variant in MS still needs to be unraveled. However, one can envision some plausible explanations, especially considering our preliminary expression and immunolocalization studies on the V_S PKCα. The preferential localization of V_S PKCα to the plasma membrane, even in the absence of external *stimuli* (such as TPA), may suggest that the V_S protein acts as a sort of competitor of the full-length PKCα for the docking to plasma membrane. For example, it has been proposed that PIP2, which binds to the C2 domain of PKCα and is required for its full enzymatic activity, as well as proteins known to be PKCα effectors, are enriched in lipid rafts (34,35). These enriched rafts could therefore represent organizing centers for signaling complexes in which PKCα lies in close proximity to its substrates. Hence, modulation of the V_S spatiotemporal distribution could play an important role in the wild-type PKCα regulation. This is a particularly intriguing hypothesis in the light of the recent seminal work of Meisel *et al.* (18), which revealed a unique role of PKCα in upregulating IL-17A during Th17-cell-mediated immune responses. The same authors also demonstrated that PKCα is essential for the priming and effector phases of experimental autoimmune encephalomyelitis (EAE, the MS mouse model), suggesting that PKCα complete deficiency in *Prkca* −/− mice protects against EAE. This last observation is apparently in contrast with our results, showing that higher expression levels of *PRKCA* could confer protection against MS. Indeed, the situation in humans is complicated by the presence of multiple *PRKCA* mRNA isoforms, and of at least one aberrant protein (V_S), whose overexpression is potentially deleterious. In this view, *Prkca* −/− mice may represent an oversimplified model, in which the total absence of the gene impairs the expression of all possible isoforms, thus resulting in a protective genotype. Moreover, the human *PRKCA* gene contains potential regulatory elements (the 3* exon, the intronic miR-634 gene) that are not evolutionarily conserved, suggesting that the expression regulation of this gene may have undergone substantial divergence from the mouse homolog.

In conclusion, our study provides evidence that *PRKCA* is a strong MS susceptibility gene, revealing a possible mechanistic basis for disease association, and adds a further layer of complexity in the interpretation of the therapeutic outcome of PKCα inhibitors in immune diseases.

MATERIALS AND METHODS

Subjects

This study was approved by local Ethical Committees and conducted according to the Helsinki Declaration. All subjects signed an informed consent.

A total of 409 unrelated patients affected by clinically definite MS according to the revised McDonald's criteria (36) and classified according to Lublin's criteria (37) as having RR, SP or PP courses, entered the study. Their main clinical details are summarized in Supplementary Material, Table S7.

As controls, 723 healthy volunteers were recruited: 409 were matched for age, gender and geographic origin with the MS patients; the remaining 314 had the same female/male distribution of the patients, and an age >52 years (mean 57.6 ± 5.4). Controls declared no signs of neurological disease or familial history for MS.

DNA extraction was performed from peripheral blood using an automated DNA extractor (Maxwell 16 System; Promega, Madison, USA).

Microsatellite and SNP genotyping

Genotyping was performed in all cases by PCR-based techniques. For details, see Supplementary Material, Materials and Methods.

Statistical analysis

All analyses were performed using different tools freely available on line. For details, see Supplementary Material.

Promoter cloning and luciferase reporter assays

PRKCA promoter sequence variants were PCR amplified from genomic DNA of individuals with different repeats. PCR primers were designed to incorporate each a restriction site (*Kpn*I or *Xho*I), used for the directional cloning into the pGL2-basic vector. The resulting constructs, containing 3, 6, 9, 10, 11 and 12 repeats, were named pGL2-$(GCC)_3$ to pGL2-$(GCC)_{12}$, and were verified by sequencing.

Luciferase reporter assays were performed by transiently transfecting 4 μg of each recombinant vector into HeLa cells. For the normalization step, 100 ng of the control vector

pRL-TK was also cotransfected. In each experiment, an equal number of cells (2.5×10^5) were transfected with the FuGENE 6 reagent (Roche, Indianapolis, USA) in six-well plates, as described by the manufacturer.

The activities of firefly/renilla luciferase were measured 48 h after transfection by using the Dual-Luciferase Reporter Assay System (Promega) and the Wallac 1420 VICTOR^3 V reader (Perkin Elmer, Waltham, USA). The values of firefly luciferase were normalized against the corresponding values of renilla luciferase (all values expressed as relative luminescence units).

Blood-derived RNA samples

PBMCs were collected from 49 RR patients (remitting phase) and 61 healthy subjects. Among patients (mean age 42.6 ± 7.2 years; range 27–53), 36 were females. They had not received any immunomodulatory therapy within the month prior to blood withdrawal. Controls were selected within the same age range of the patients and with the same female–male ratio. PBMCs were isolated immediately after phlebotomy from heparinized blood, by centrifugation on a Lympholyte Cell separation media (Cederlane Laboratories Limited, Hornby, Canada) gradient.

Th1, Th17, B lymphocytes and natural killer cells were isolated from pooled buffy coats obtained from healthy donors by using a FACSAria Cell Sorter (BD Biosciences, San Jose, USA) and various combinations of surface markers, as described (38).

In all cases, total RNA was isolated using the Eurozol kit (Euroclone, Wetherby, UK). RNA concentration/quality were assessed using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA).

RT–PCR, semi-quantitative real-time RT–PCR and competitive fluorescent RT–PCR

Random nonamers or *PRKCA*-specific primers and the Superscript-III Reverse Transcriptase (Invitrogen, Carlsbad, USA) were used to perform first-strand cDNA synthesis starting from 1 μg of total RNA, according to the manufacturer's instructions. Of a total of 20 μl of the RT reaction, 1 μl (undiluted or 1 : 10 diluted) was used as template for amplifications. RT–PCRs were performed under standard conditions using the FastStart *Taq* DNA Polymerase on a Mastercycler EPgradient (Eppendorf, Hamburg, Germany).

To quantify *PRKCA* isoforms, semi-quantitative real-time RT–PCRs were carried out using the FastStart SYBR Green Master Mix and a LightCycler 480 (Roche), following a touchdown thermal protocol. Expression levels were normalized using two housekeeping genes (hydroxymethylbilane synthase, *HMBS*; β-actin, *ACTB*). Data were analyzed and rescaled using the GeNorm software (39).

To quantify the relative amount of *PRKCA* 3*-containing versus 3*+5nt-containing isoforms, we took advantage of their length differences, and performed competitive RT–PCRs by using a 6-FAM-labeled primer. Amplified fragments were separated by capillary electrophoresis on an ABI-3130XL Genetic Analyzer and quantitated by the GeneMapper v4.0 software (Applied Biosystems, Foster City, USA). The sum of all fluorescence peak areas in a single run was set equal to 100%, and the relative quantity of each transcript expressed as a fraction of the total. The identical amplification efficiency of the two amplicons was verified by real-time PCR.

Assays were all performed at least in triplicate. Primer couples for these assays are listed in Supplementary Material, Table S8.

Splicing assays

To generate hybrid minigene constructs, *PRKCA* regions were PCR amplified from the DNA of a healthy individual heterozygous for the risk haplotype. Oligonucleotides used for the amplification carried an *Nde*I restriction site in their 5′ ends, in order to clone the products into the α-globin-fibronectin minigene plasmid. Recombinant plasmids were used for *ex vivo* splicing assays, by transfecting 4 μg of each minigene in the appropriate cell line. Total RNA was isolated from cells 48 h after transfection, using the Eurozol kit. RT–PCRs were carried out by means of primers mapping in the flanking fibronectin exonic regions of the plasmid; competitive RT–PCR assays were accomplished by anchoring primers to *PRKCA* exon 3* and to the flanking fibronectin exon 25.

HnRNP-H overexpression was performed in HEK293, IMR-32 and in two human fibroblast cell lines. Transfection experiments and RNA extractions were performed as described above.

For hnRNP H knockdown, 150 000 HeLa cells were seeded on 3.5 cm-well plates. After 24 h, 5 μl Oligofectamine (Life Technologies, Foster City, USA) was mixed with 15 μl Opti-MEM I reduced serum medium (Life Technologies), incubated at room temperature for 7 min and added to 2.5 μl (25 pmol) of siRNA duplex, which had been mixed with 175 μl of Opti-MEM I. The mixture was incubated at room temperature for 20 min, and then added to the cells. After 48 h, RNA and protein extractions were performed. The 20 nt target sequence in hnRNP H was 5′-GGAAATAGCTGAAAAGGCT-3′. The specificity of the siRNA treatment was previously assayed using a predesigned siRNA, targeting luciferase (target sequence: 5′-CGTACGCGGAATACTTCGA-3′) (Euroclone) (40). Silencing efficiency was assessed by western blotting performed according to standard protocols. Antibodies against hnRNP H were a kind gift of Dr. Buratti (41).

The effect of hnRNP H overexpression, as well as siRNA treatment against hnRNP H, was assessed by real-time RT–PCR or fluorescent competitive RT–PCR aimed at quantitating the 3*/3*+5nt endogenous transcripts.

PRKCA splicing pattern, expression profile and sensitivity to NMD

PRKCA splicing pattern was analyzed using the RNA isolated from PBMCs of a healthy individual. A set of RT–PCR assays was designed to catch the vast majority of possible AS events (Supplementary Material, Table S8). Long-range RT–PCRs were performed using the Go*Taq* kit (Promega) according to the manufacturer's instruction. All RT–PCR products were directly sequenced to confirm their identity.

Expression profiles of *PRKCA* isoforms were obtained by real-time RT–PCRs on RNAs from: PBMCs, sorted PBMCs, 21 human tissues (Ambion, Austin, USA), 25 human brain districts (Clontech laboratories, Palo Alto, USA) and 9 cell lines.

Analysis of *PRKCA* susceptibility to NMD was undertaken in HepG2 and HEK293 cells. Cells were plated at a density of 5.5×10^6 per 10 cm dish and, after 72 h, treated for 8 h with puromycin or cycloheximide (100 μg/mL). Untreated samples were incubated with the corresponding drug solvent. After the treatment, cells were washed twice with phosphate-buffered saline and total RNA was extracted. RT–PCRs were performed as described, and variations in the expression levels of the AS transcripts upon treatments quantified using as reference an NMD-resistant transcript (i.e. the mRNA for Connexin 32, whose coding sequence is all contained in a single exon) (22).

Protein expression studies

The plasmid expressing the PKCα-V_S isoform was produced by mutagenizing in two steps the pEFneo-PKCα plasmid with the QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies): the HA-tag was removed, and a sequence of 48 nucleotides, corresponding to the first ones of intron 14 followed by two stop codons, was introduced after codon 535 of the wild-type cDNA. The mutant plasmid (pEFneo-PKCα-V_S) was verified by sequencing.

COS-1 cells were transfected with the pEFneo-PKCα-V_S plasmid. After 24 h, 0.4 mg of total lysate was immunoprecipitated either with antibodies directed against the N-terminus of PKCα (Ab86715; Abcam, Cambridge, UK), or with non-specific immunoglobulins (IgG; negative control). For western blot, immunoprecipitated proteins from COS-1 cells, as well as 10 μg of total proteins extracted from untransfected IMR-32 cells, were loaded onto a 10% non-reducing polyacrylamide gel. A rabbit anti-human antibody (Ab32122, Abcam; diluted 1 : 1000 in skim milk, and directed against the PKCα N-terminus) was used as primary antibody. For the detection step, a secondary anti-rabbit peroxidase-conjugated antibody (diluted 1 : 5000 in skim milk) was used in addition to Super-Signal West Dura Extended Duration Substrate kit (Pierce, Rockford, USA).

Immunolocalization studies

PKCα V_L and V_S proteins were expressed into MDCK cells, chosen for the absence of cross-reactivity between the antibody directed against human PKCα and the endogenous dog ortholog. The cellular localization of PKCα was studied by immunofluorescence staining of isoforms and of different subcellular compartments either in basal conditions or after stimulation with TPA (1 μM in DMSO for 1 h, 37°C).

A total of 3×10^5 MDCK cells were plated on glass coverslips, grown to 70% confluence, and transfected with either the pEFneo-PKCα or the pEFneo-PKCα-V_S plasmid. After 24 h, cells were fixed, washed, preincubated and incubated with primary and secondary antibodies as described (42). The images were acquired by means of an LSM 510 Meta confocal microscope (Carl Zeiss, Oberkochen, Germany).

Primary antibodies were mouse and rabbit anti-PKCα (Ab86715, Ab32122; Abcam), mouse anti-EEA1 (Sigma, St Louis, USA), rabbit anti-Giantin (provided by Dr. Renz, Institute of Immunology and Molecular Genetics, Karlsruhe, Germany). Actin cytoskeleton was stained with Rhodamine Phalloidin (Cytoskeleton, Denver, USA). Secondary antibodies were DyLight 549-conjugated anti-rabbit or anti-mouse antibodies (Jackson ImmunoResearch, West Grove, USA) and Alexa Fluor 488 anti-rabbit antibody (Life Technologies).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Dr. Ronzoni (Azienda Ospedaliera "Guido Salvini"—Garbagnate Milanese) and Dr. Sola (Nuovo Ospedale Civile S. Agostino Estense AUSL—Modena) for their precious help in collecting PBMCs from MS patients. The "Associazione Volontari Del Sangue" (Dr. Verducci and Dr. Galastri), Dr. Uccellini, the "Pubblica Assistenza Croce Bianca di Piacenza" (Dr. Grana), the "Coro di Turate" and Dr. Mantegazza are acknowledged for their invaluable collaboration in collecting control samples. Fibroblast cell lines were obtained from the "Cell Line and DNA Biobank from Patients Affected by Genetic Diseases" (Istituto G. Gaslini) and "Parkinson Institute Biobank" (Milan, http://www.parkinson.it/dnabank.html) of the Telethon Network of Genetic Biobanks (http://www.biobanknetwork.org, project No. GTB07001). We also thank the Monzino Foundation (Milan, Italy) for its generous gift of the Zeiss LSM 510 Meta confocal microscope. This paper has greatly benefited from helpful discussions with Dr. Francolini (Department of Medical Biotechnologies and Translational Medicine, University of Milan) and Dr. Buratti (International Centre for Genetic Engineering and Biotechnology, Trieste, Italy). Roberta Marotta, Giulia Rizzo, Elena Giardino, Gaia Anelli, Egle Radice and Simone Digregorio are acknowledged for their assistance, technical support and enthusiasm.

The authors wish to acknowledge the International Multiple Sclerosis Genetics Consortium (IMSGC; https://www.imsgc.org/) for allowing access to genotype and imputation data.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by FISM—Fondazione Italiana Sclerosi Multipla (grants 2008/R/1 and "Progetto Speciale Immunochip" 2011/R/14); by Fondazione CARIPLO—Cassa di Risparmio delle provincie Lombarde (grant n° 2010-0728); and by Fondazione CRT—Cassa di Risparmio di Torino (Turin, Italy).

REFERENCES

1. Thompson, A.J. (2008) Multiple sclerosis – a global disorder and still poorly managed. *Lancet Neurol.*, **7**, 1078–1079.
2. Greenstein, J.I. (2007) Current concepts of the cellular and molecular pathophysiology of multiple sclerosis. *Dev. Neurobiol.*, **67**, 1248–1265.
3. Hauser, S.L. and Oksenberg, J.R. (2006) The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron*, **52**, 61–76.
4. Pugliatti, M., Harbo, H.F., Holmøy, T., Kampman, M.T., Myhr, K.M., Riise, T. and Wolfson, C. (2008) Environmental risk factors in multiple sclerosis. *Acta Neurol. Scand. Suppl.*, **188**, 34–40.
5. Oksenberg, J.R. and Barcellos, L.F. (2005) Multiple sclerosis genetics: leaving no stone unturned. *Genes Immun.*, **6**, 375–387.

6. Sawcer, S., Ban, M., Wason, J. and Dudbridge, F. (2010) What role for genetics in the prediction of multiple sclerosis?. *Ann. Neurol.*, **67**, 3–10.
7. Patsopoulos, N.A., Barcellos, L.F., Hintzen, R.Q., Schaefer, C., van Duijn, C.M., Noble, J.A., Raj, T., IMSGC, ANZgene, Gourraud, P.A. *et al.* (2013) Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.*, **9**, e1003926.
8. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, **476**, 214–219.
9. Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kemppinen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., Goris, A. *et al.* (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.*, **45**, 1353–1360.
10. Evsyukova, I., Somarelli, J.A., Gregory, S.G. and Garcia-Blanco, M.A. (2010) Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol.*, **7**, 462–473.
11. Barton, A., Woolmore, J.A., Ward, D., Eyre, S., Hinks, A., Ollier, W.E., Strange, R.C., Fryer, A.A., John, S., Hawkins, C.P. *et al.* (2004) Association of protein kinase C alpha (PRKCA) gene with multiple sclerosis in a UK population. *Brain*, **127**, 1717–1722.
12. Saarela, J., Kallio, S.P., Chen, D., Montpetit, A., Jokiaho, A., Choi, E., Asselta, R., Bronnikov, D., Lincoln, M.R., Sadovnick, A.D. *et al.* (2006) PRKCA and multiple sclerosis: association in two independent populations. *PLoS Genet.*, **2**, e42.
13. Dempsey, E.C., Newton, A.C., Mochly-Rosen, D., Fields, A.P., Reyland, M.E., Insel, P.A. and Messing, R.O. (2000) Protein kinase C isozymes and the regulation of diverse cell responses. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **279**, L429–L438.
14. Landgraf, K.E., Malmberg, N.J. and Falke, J.J. (2008) Effect of PIP2 binding on the membrane docking geometry of PKC alpha C2 domain: an EPR site-directed spin-labeling and relaxation study. *Biochemistry*, **47**, 8301–8316.
15. Pfeifhofer, C., Gruber, T., Letschka, T., Thuille, N., Lutz-Nicoladoni, C., Hermann-Kleiter, N., Braun, U., Leitges, M. and Baier, G. (2006) Defective IgG2a/2b class switching in PKC alpha −/− mice. *J. Immunol.*, **176**, 6004–6011.
16. Iwamoto, T., Hagiwara, M., Hidaka, H., Isomura, T., Kioussis, D. and Nakashima, I. (1992) Accelerated proliferation and interleukin-2 production of thymocytes by stimulation of soluble anti-CD3 monoclonal antibody in transgenic mice carrying a rabbit protein kinase C alpha. *J. Biol. Chem.*, **267**, 18644–18648.
17. Han, Y., Meng, T., Murray, N.R., Fields, A.P. and Brasier, A.R. (1999) Interleukin-1-induced nuclear factor-kappaB-IkappaBalpha autoregulatory feedback loop in hepatocytes. A role for protein kinase calpha in post-transcriptional regulation of ikappabalpha resynthesis. *J. Biol. Chem.*, **274**, 939–947.
18. Meisel, M., Hermann-Kleiter, N., Hinterleitner, R., Gruber, T., Wachowicz, K., Pfeifhofer-Obermair, C., Fresser, F., Leitges, M., Soldani, C., Viola, A. *et al.* (2013) The kinase PKCα selectively upregulates interleukin-17A during Th17 cell immune responses. *Immunity*, **38**, 41–52.
19. Rimoldi, V., Straniero, L., Asselta, R., Mauri, L., Manfredini, E., Penco, S., Gesu, G.P., Del Longo, A., Piozzi, E., Soldà, G. *et al.* (2014) Functional characterization of two novel splicing mutations in the OCA2 gene associated with oculocutaneous albinism type II. *Gene*, **537**, 79–84.
20. Caputi, M. and Zahler, A. (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.*, **276**, 43850–43859.
21. Luo, W., Ji, Z., Pan, Z., You, B., Hoque, M., Li, W., Gunderson, S.I. and Tian, B. (2013) The conserved intronic cleavage and polyadenylation site of CstF-77 gene imparts control of 3′ end processing activity through feedback autoregulation and by U1 snRNP. *PLoS Genet.*, **9**, e1003613.
22. Popp, M.W. and Maquat, L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.
23. Asselta, R., Rimoldi, V., Guella, I., Soldà, G., De Cristofaro, R., Peyvandi, F. and Duga, S. (2010) Molecular characterization of in-frame and out-of-frame alternative splicings in coagulation factor XI pre-mRNA. *Blood*, **115**, 2065–2072.
24. Patsopoulos, N.A., Esposito, F., Reischl, J., Lehr, S., Bauer, D., Heubach, J., Sandbrink, R., Pohl, C., Edan, G., Kappos, L. *et al.* (2011) Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.*, **70**, 897–912.
25. Hannan, A.J. (2010) Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet.*, **26**, 59–65.
26. Sawaya, S.M., Lennon, D., Buschiazzo, E., Gemmell, N. and Minin, V.N. (2012) Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth-death model. *Genome Biol. Evol.*, **4**, 636–647.
27. Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A. and Gemmell, N. (2013) Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS ONE*, **8**, e54710.
28. Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
29. Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y. and Maizels, N. (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.*, **39**, 4975–4983.
30. Hilger-Eversheim, K., Moser, M., Schorle, H. and Buettner, R. (2000) Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene*, **260**, 1–12.
31. Lewis, B., Green, R. and Brenner, S. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. USA*, **100**, 189–192.
32. Mills, J.D. and Janitz, M. (2012) Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiol. Aging*, **33**, 1012. e1011–e1024.
33. Tian, Y., Apperson, M.L., Ander, B.P., Liu, D., Stomova, B.S., Jickling, G.C., Enriquez, R., Agius, M.A. and Sharp, F.R. (2011) Differences in exon expression and alternatively spliced genes in blood of multiple sclerosis compared to healthy control subjects. *J. Neuroimmunol.*, **230**, 124–129.
34. Laux, T., Fukami, K., Thelen, M., Golub, T., Frey, D. and Caroni, P. (2000) GAP43, MARCKS, and CAP23 modulate PI(4,5)P(2) at plasmalemmal rafts, and regulate cell cortex actin dynamics through a common mechanism. *J. Cell Biol.*, **149**, 1455–1472.
35. Uchino, M., Sakai, N., Kashiwagi, K., Shirai, Y., Shinohara, Y., Hirose, K., Iino, M., Yamamura, T. and Saito, N. (2004) Isoform-specific phosphorylation of metabotropic glutamate receptor 5 by protein kinase C (PKC) blocks Ca2+ oscillation and oscillatory translocation of Ca2+-dependent PKC. *J. Biol. Chem.*, **279**, 2254–2261.
36. Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W. *et al.* (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann. Neurol.*, **58**, 840–846.
37. Lublin, F.D. (2005) Clinical features and diagnosis of multiple sclerosis. *Neurol. Clin.*, **23**, 1–15, v.
38. Rossi, R.L., Rossetti, G., Wenandy, L., Curti, S., Ripamonti, A., Bonnal, R.J., Birolo, R.S., Moro, M., Crosti, M.C., Gruarin, P. *et al.* (2011) Distinct microRNA signatures in human lymphocyte subsets and enforcement of the naive state in CD4+ T cells by the microRNA miR-125b. *Nat. Immunol.*, **12**, 796–803.
39. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, RESEARCH0034.
40. Rimoldi, V., Soldà, G., Asselta, R., Spena, S., Stuani, C., Buratti, E. and Duga, S. (2013) Dual role of G-runs and hnRNP F in the regulation of a mutation-activated pseudoexon in the fibrinogen gamma-chain transcript. *PLoS ONE*, **8**, e59333.
41. Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y. and Baralle, F. (2004) hnRNP H binding at the 5′ splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.*, **32**, 4224–4236.
42. De Silvestris, M., D'Arrigo, A. and Borgese, N. (1995) The targeting information of the mitochondrial outer membrane isoform of cytochrome b5 is contained within the carboxyl-terminal region. *FEBS Lett.*, **370**, 69–74.