Genomics 99 (2012) 36–43

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identification and characterization of novel splice variants of the human *EPM2A* gene mutated in Lafora progressive myoclonus epilepsy

Deepti Dubey, Rashmi Parihar, Subramaniam Ganesh *

Department of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, 208016, India

ARTICLE INFO

Article history:
Received 11 January 2011
Accepted 3 October 2011
Available online 19 October 2011

Keywords:
Phosphatase
Alternative splicing
Epilepsy
Protein diversity

ABSTRACT

The *EPM2A* gene, defective in the fatal neurodegenerative disorder Lafora disease (LD), is known to encode two distinct proteins by differential splicing; a phosphatase active cytoplasmic isoform and a phosphatase inactive nuclear isoform. We report here the identification of three novel *EPM2A* splice variants with potential to code for five distinct proteins in alternate reading frames. These novel isoforms, when ectopically expressed in cell lines, show distinct subcellular localization, interact with and serve as substrates of malin ubiquitin ligase—the second protein defective in LD. Two phosphatase active isoforms interact to form a heterodimeric complex that is inactive as a phosphatase *in vitro*, suggesting an antagonistic function for laforin isoforms if expressed endogenously in significant amounts in human tissues. Thus alternative splicing could possibly be one of the mechanisms by which *EPM2A* may regulate the cellular functions of the proteins it codes for.

1. Introduction

Lafora's progressive myoclonus epilepsy or Lafora disease (LD) is a teenage-onset and fatal form of neurodegenerative disorder [1,2]. In addition to symptoms like epileptic seizures, progressive loss of intellectual abilities, loss of muscle coordination and imperfect articulation of speech, LD patients also manifest pathognomonic periodic acid-Schiff-positive (PAS+) staining intracellular inclusion bodies called the Lafora bodies [1,2]. Hence an abnormal glycogen metabolic process is thought to underlie the pathogenesis in LD [1,2]. The genetic basis of LD has been deciphered and characterized. LD can be caused by defects in the *EPM2A* gene encoding a dual-specificity protein phosphatase named laforin, or by defects in the *NHLRC1* gene encoding an E3 ubiquitin ligase named malin [reviewed in [2]]. Laforin and malin interact with each other and colocalize in endoplasmic reticulum [3,4]. Besides being a substrate to malin, laforin also work together with malin as a complex to promote the degradation of malin targets through the ubiquitin proteasome system [3–5].

Alternative splicing of mRNA enables generation of multiple gene products with distinct functions from a single coding sequence. It has been recently proposed that nearly 95% of human genes undergo alternative splicing resulting in the functional complexity in the human proteome [6]. Indeed studies have shown important functional roles for the spliced forms and their products both in normal and in disease conditions [6,7]. In this regard it is interesting to note that the *EPM2A* gene undergoes alternative splicing [8,9]. We and others have reported earlier the existence of two isoforms for laforin; a major isoform localized in the cytoplasm, and a minor isoform that is targeted to the nucleus as well [8,9]. We have further demonstrated that the nuclear isoform is an inactive phosphatase, and it regulates the phosphatase activity of cytoplasmic isoform by interacting with it and forming a heterodimer [10]. Here, we report the identification and characterization of three novel splice variants of the *EPM2A* gene and their functional significance.

2. Materials and methods

2.1. Extraction of RNA and reverse transcriptase-polymerase chain reaction (RT-PCR) analysis

Human brain tissues (obtained from the Human Brain Bank, National Institute of Mental Health and Neuroscience, Bangalore, India), human cell lines (HEK293 and SHSY-5Y), and mouse tissues were processed for RNA extraction using the RNAwiz reagent (Ambion Inc.) as recommended by the manufacturer. For the RT-PCR, first strand cDNAs were synthesized from the total RNA using first strand cDNA synthesis kit (Fermentas Life Sciences). Transcript specific primers (see Supplementary Table S1) were used for the expression studies.

2.2. Northern blot hybridization

The human multiple tissue northern blot membrane (Human 8-Lane MTN® Blot) was commercially obtained from Clontech

* Corresponding author. Fax: +91 512 259 4010.
E-mail address: sganesh@iitk.ac.in (S. Ganesh).

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.10.001

Laboratoreis Inc., and hybridized overnight with (α^{32}P)-dCTP labeled DNA probe (see Supplementary Table S1). The blots were washed finally in 0.1× SSC/0.1% SDS at 55 °C and exposed to X-ray film at −80 °C.

2.3. Expression constructs

Details of the expression construct for the Myc/HIS or GFP-tagged laf331, FLAG-laf331, and Myc/His-tagged-laf331 mutant C266S are reported in our previous study [10]. The coding region of laf152 or POCR (type-3 transcript) was PCR amplified from cDNA and cloned in-frame into pEGFP-N1 (laf152-GFP) or pcDNA3.1/Myc-HIS (POCR-Myc/HIS) or N-FLAG vector using primers listed in Supplementary Table 1. An expression construct for Myc/His-tagged laf224 was made by amplifying laf224 from IMAGE clone IMAGE: 3829460 and then cloning it into the pcDNA vector. For laf177, cDNA from HEK293 cells was used to amplify the coding region and then cloned in pcDNA to create laf177-Myc/His construct. Expression constructs for Myc/His-tagged malin (wild-type and the mutant form) or GFP-tagged malin (wild-type) are reported in our previous study [10]. Similarly, the shRNA-mediated knockdown constructs for malin are reported in our previous study [5], and the efficiency of the knockdown was established by immunoblot using a reporter construct (see Supplemental Figure S3).

2.4. Cell culture, transfections, immunostaining, immunoblotting and antibodies

Cell lines were grown and maintained in Dulbecco's modified Eagle's medium as reported earlier [10]. Transfections were performed using Polyfect transfection reagent (Qiagen) according to the manufacturer's protocols. Wherever mentioned, 10 μM of MG132 treatment was given to cells post transfections for 12 h. Cells grown on gelatin coated coverslips were processed for immunofluorescence staining as reported previously [10]. For immunoblotting, protein samples were run on a 12% SDS-PAGE, and transferred onto a nitrocellulose filter as described [10]. After blocking, the membranes were processed through sequential incubations with primary antibody followed by horseradish peroxidase (HRP) tagged secondary antibody and signals were visualized using a chemiluminiscent substrate (SuperSignal West PICO, Thermo Scientific, USA). Antibodies used in the present study are: rabbit anti-Myc (Cell Signaling Technology Inc), mouse anti-Myc (Sigma-Aldrich), rabbit anti-FLAG (Sigma-Aldrich), mouse anti-GFP (Roche Applied Science), rabbit anti-GST (Sigma-Aldrich), and mouse anti-γ-tubulin (Sigma-Aldrich). The FITC, TRITC or HRP tagged secondary antibodies were obtained from Jackson Immunoresearch, USA.

2.5. Pull-down assay

For Ni-affinity pull down assays, cells were transfected with suitable tagged constructs as described [10]. Forty eight hours after transfection, the cells were lysed with RIPA lysis buffer and then incubated with EZview Red His-Select HC Nickel Affinity Gel (Sigma-Aldrich Chemicals Pvt Ltd) at 4 °C for 2 h as reported. The beads were then washed with RIPA buffer to remove non-specific proteins. The bead bound proteins were either used for phosphatase activity assay or loaded on SDS-PAGE after boiling with laemmli buffer as described [10].

2.6. Phosphatase activity assay

His-tagged proteins were purified using nickel resin and the bead bound proteins were used for the phosphatase activity assay as described [10]. Approximately 100 ng of purified protein was used and the assay was performed in 75 μl reaction buffer containing 25 mM para-nitro phenyl phosphate (pNPP; Bangalore Genei). The reaction was incubated at 37 °C for 30 min and then stopped using 200 μl of 0.2 N NaOH. Absorbance was taken at 405 nm. Experiments were done in triplicate and average absorbance readings were plotted in graphs. A fraction of the bead bound protein was tested in immunoblot analysis.

2.7. Glycogen binding assay

For the production of glutathione S-transferase (GST)-tagged fusion protein, desired constructs were expressed in *E. coli* and the protein was purified as described [11]. The purified proteins were then added to the glycogen binding assay in the presence or absence of 10 mg/ml glycogen, incubated on ice for 30 min and then centrifuged at 17,000 *g* for 90 min at 4 °C as described [11]. The supernatant and pellet fractions were separated and immunoblotted with anti-GST antibody.

3. Results

3.1. Identification and characterization of novel splice variants of the EPM2A gene

The *EPM2A* gene is known to undergo alternative splicing, resulting in at least two splice-variants, coding for two distinct proteins [8,9] (referred to here as type-1 and type-2 transcripts respectively; see Fig. 1). To identify novel splice variants of the *EPM2A* gene, we searched the sequence database of NCBI for partial or complete cDNA sequences representing the transcripts of *EPM2A* gene but having unique exon organization. The AceView database (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/), lists 17 distinct splice variants for the *EPM2A* gene out of which the expression status of only two was established (referred to here as type- and type-2 transcripts; Fig. 1). We therefore designed primers to specifically amplify each of the splice variants that are represented by more than one independent cDNA clones and tested their expression in the cDNA derived from the adult human brain. Our reverse transcriptase polymerase chain reaction (RT-PCR) assays established the expression of three novel splice variants for the *EPM2A* gene, and they were named as type-3, type-4 and type-5 respectively (see Fig. 1). The type-4 transcript is represented by two cDNA clones (GenBank accession # BC005286; cDNA clones, BE268120 and BF032161), expressed in the adult brain tissue, and is characterized by the skipping of exon 3. Clone BC005286 harbors the poly-A tail but the 5′-end appears to be incomplete; a sequence alignment with the full length clone of type-1 transcript (GenBank accession # NM_005670) reveals that BC005286 harbors the first exon but lacks the 5′ sequence from +4 position of the coding sequence (Fig. 1). RT-PCR analysis, using primers flanking the exon 3 (F2-R4), resulted in the amplification of two distinct amplicons from the human adult brain cDNA (Figs. 1A and B). The higher molecular weight amplicon represent the type-1 or type-2 transcript while the lower molecular weight amplicon represent the novel transcript type-4 lacking exon 3 in its coding sequence (Fig. 1B). The type-5 transcript is represented by the cDNA with GenBank accession # AK022721 and is supported by four independent Expression Sequence Tag (EST) clones (AU125766, BI561221, DA920514, and AA400516). This transcript appears to have a novel 5′-end sequence thus defining a novel first exon for this transcript, located between the exon 2 and 3 of the type-1 transcript (see Fig. 1). We have used a specific forward primer from the 5′-end unique exon of the type-5 transcript region for the RT-PCR and confirmed the expression of this transcript in the adult human brain (Figs. 1A and B). The third transcript characterized in the present study is type-3, represented by a cDNA clone with GenBank accession # AF454494 (Fig. 1). This clone appears to lack the 5′-, and 3′-ends of the messenger as it lacked poly-A sequence as well. A

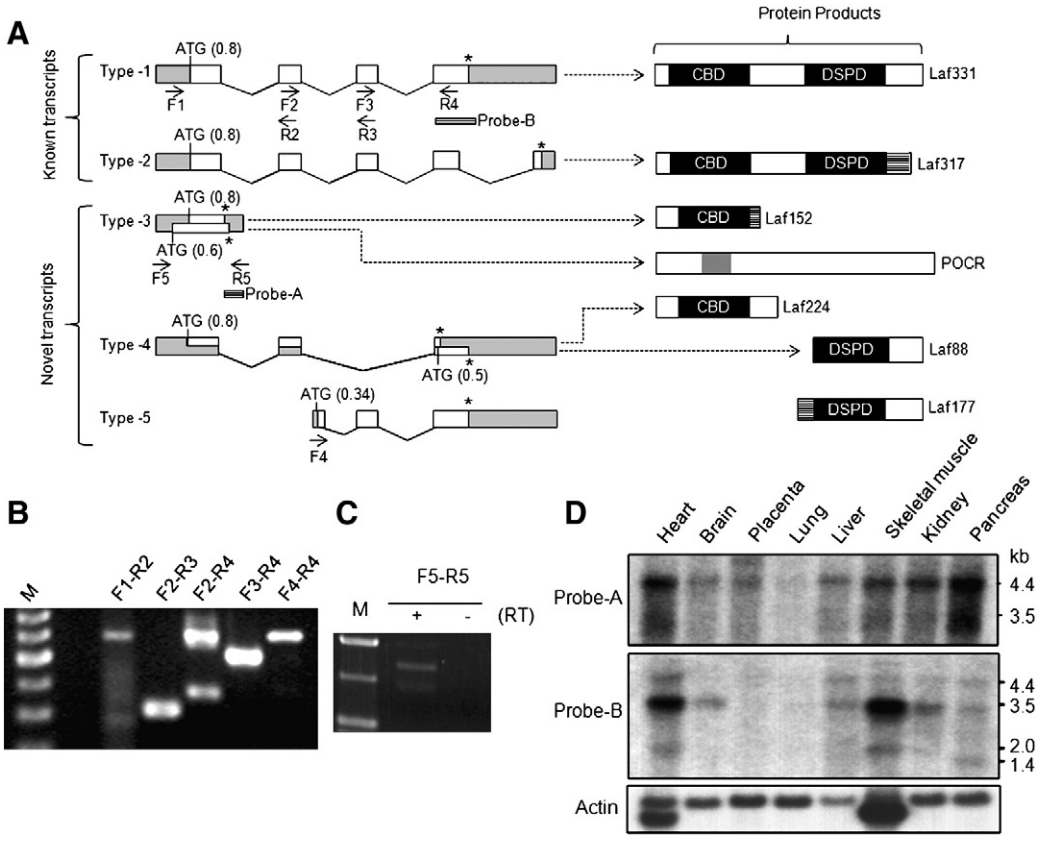

Fig. 1. *Identification of three novel transcript variants of the EPM2A gene and proteins they encode.* (A) Schematic showing the known (type-1 and -2) and novel transcripts (types 3–5) of *EPM2A* on the left hand side. The empty boxes represent the coding sequence of the exons while the grey boxes represent untranslated regions in the transcript. The position of the start codon and the stop codon in the open reading frames are depicted by ATG and asterisk (*), respectively. The relative strength of the predicted start codons (Kozak reliability score) [14] is shown close to the start codons identified as "ATG" in the schematic. The start codon shown for type-4 transcript is predicted based on type-1 transcript since the sequence of cDNA clone lacks the 5′ sequence from +4 position of the coding sequence of type-1 transcript. The arrows below each transcript show the position of the forward (F) and the reverse (R) primers used in the RT-PCRs (see Supplementary Table S1 for primer sequence). The small, lined boxes (identified as *probe A* and *probe B*) below type-3 and type-1 transcripts show the positions of the probes used for the northern blot hybridization. The predicted protein products coded by these transcripts are identified on the right side (dotted arrow). The domain structure of the putative peptides predicted to code by each ORF is shown. Here, conserved domains are identified by shaded boxes, and region that is unique to each isoform is identified by a lined box. CBD—carbohydrate binding domain; DSPD—dual-specificity phosphates domain. POCR sequence does not harbor CBD or DSPD but shows a weak similarity API5 domain (identified by a shaded box). (B and C) Gel pictures showing the result of the RT-PCRs. Human brain tissues were used to isolate RNA and then cDNA was synthesized. Different combinations of primer pairs (positions identified in A) were used to perform RT-PCRs as indicated. In C, the cDNA derived from the reaction done in the presence or absence of the reverse transcriptase (RT) is indicated with "+" and "–" symbols, respectively. (D) The human multiple tissue northern blot was hybridized with probe-A (specific to type-3 transcript) or probe-B (could detect all transcripts except type-3) as indicated. Hybridization with actin probe served as control. The amino acid sequence of individual isoforms is given in Supplemental Figure S1.

sequence alignment with the type-1 transcript and the genomic sequence spanning the *EPM2A* locus reveal that this cDNA clone could possibly represent an unspliced mRNA as it shared exact identity with exon 1 coding region and the adjacent 649 bp novel sequence was found to be originated from the intronic region present contiguous to exon 1 (see Fig. 1A). To confirm that this clone indeed represent transcribed region, northern and RT-PCR analyses were done (Figs. 1C and D). A 214 bp fragment from the 3′-end unique region (probe-A) yielded prominent signal at 4.4 kb position in the human multiple tissues northern blot (Fig. 1D). Based on the sequence of type-1 transcript, primers were designed to amplify putative 5′-flanking region and expected size fragment was detected by RT-PCR confirming the production of such a novel isoform, designated as type-3 transcript (Fig. 1C). Since the size of the transcript detected in the northern blot was found to be larger than the length of the cDNA clone, it could be presumed that this transcript might harbor additional regions not represented in the cDNA. Our repeated attempts to extend the 5′- and 3′-ends of the cDNA clone using the RACE approach did not yield any specific product. Nonetheless, the detection of more than one band for the *EPM2A* gene in the northern blot suggests the presence of multiple transcript variants (Fig. 1D). Alternative splicing is known to be regulated by SR family proteins by binding to the exonic splicing enhancers (ESEs) present in the constitutively and alternatively spliced exons [12]. Indeed a sequence search for ESEs for the *EPM2A* transcript, using the online tool ESEfinder [13] (http://rulai.cshl.edu/tools/ESE), identified 5 potential binding sites for the SR family proteins SF2/ASF, SRp40, SRP55 and SC53.

We have also tested whether such differentially spliced transcripts for the *Epm2a* gene are transcribed in mouse. Primers flanking individual exons were used for the amplification of differentially spliced transcript by RT-PCR but none of these attempts identified addition transcripts other than the known *Epm2a* transcript (Type-1 equivalent) (see Supplemental Fig. S2).

In order to understand the coding potential of these novel transcripts identified for the human *EPM2A* gene, we analyzed the potential open reading frames (ORFs) using the online tool, ATGpr (http://atgpr.dbcls.jp/) [14], and found that five laforin isoforms can be encoded by these three novel transcripts (Fig. 1A and Supplemental Fig. S1). The consensus sequence of the type-3 transcript contains two potential open reading frames (ORFs) that overlap each other (Fig. 1A and Supplemental Fig. S1). One of them encodes a truncated form of laforin composed of 152 amino acids containing the putative carbohydrate binding domain (CBD) (Fig. 1A and Supplemental Fig. S1). This truncated laforin was named as laf152, denoting its

amino acid length. A second ORF, present out-of-frame with the laf152 coding sequence, could produce an entirely distinct protein of 344 amino acids with a molecular mass of 35 kDa (Fig. 1A and Supplemental Fig. S1). POCR does not harbor CBD or DSPD (Dual-Specificity Phosphatase Domain) but showed a weak similarity, in the region aa55–135, with the pfam 05918 (E-value: 4e−05) in a CD-Domain search of NCBI (see Fig. 2A and Supplemental Fig. S1). Pfam 05918 represents a protein family containing apoptosis inhibitory protein 5 (API5) sequences and such proteins, known in several organisms, were shown to function as anti-apoptotic proteins [15]. Similar to type-3, the type-4 transcript also showed two overlapping ORFs. Although, the cDNA clone representing the type-4 transcript starts from 4 nucleotides upstream of the first exon of the type-1 transcript, we assume here that the 5′-end of the type-4 transcript is similar to that of the type-1 transcript. In one of the ORFs it encodes for a 224 amino acid long protein which harbors the CBD while in the other reading frame, it could encode a small 88 amino acid long peptide having a truncated DSPD (Fig. 1A and Supplemental Fig. S1). Unlike the other two novel transcripts, type-5 transcript has only one ORF wherein it codes for a 177 amino acid long protein (laf177) having the DSPD (Fig. 1A and Supplemental Fig. S1).

We then checked the translation ability of these ORFs by expressing the potential coding sequence in mammalian cell lines. For this, we made expression constructs containing the coding region of POCR, laf224, laf88 or laf177 with C-terminal Myc/His tag or N-termianl FLAG tag for laf152, transiently transfected them in COS-7 cells and then analyzed by immunoblotting (Fig. 2A). All the constructs produced expected size proteins. However, laf88 did not obtain any signal despite several attempts with several independent clones. Intriguingly, the subcellular distribution of all other isoforms was found to be distinct (Fig. 2B). Whereas the signal for POCR was found restricted to the nucleus in a majority of the cells, laf152 and laf224 were distributed both in nucleus and cytoplasm. On the other hand, laf177 showed majorly cytoplasmic localization similar to laf331 (Fig. 2B).

3.2. Novel isoform laf177 is an active phosphatase and it interacts with laf331, laf317 and POCR

Laf331 is an active dual-specificity protein phosphatase [16] while its isoform laf317 is an inactive phosphatase [10]. The type-5 transcript discovered in the present study is predicted to encode a protein (laf177) with the dual-specificity phosphatase domain (DSPD). In order to test whether laf177 will be active as a phosphatase or not, we transiently expressed Myc/His tagged laf177 in COS-7 cells and affinity purified laf177 using Nickel resin and used the protein for an *in vitro* phosphatase assay. Similarly, we expressed the wild-type laf331 or its catalytically inactive mutant (Laf331-C266S), purified and used them in the phosphatase assay as positive and negative controls, respectively. Around 100 ng of the bead-bound protein was incubated with the chromogenic substrate pNPP to detect the phosphatase activity. As shown in Fig. 3A, the novel isoform laf177 showed robust phosphatase activity which was comparable to that of laf331. The laf331 mutant C266S did not show any activity as expected. Thus laf177 appears to be an active phosphatase.

We have shown earlier that the laforin isoforms laf331 and laf317 physically interact with each other and form a heterodimer that is enzymatically inactive [10]. Thus, to investigate the possibility that the laf177 could form heterodimer with laf331 or laf317, we coexpressed Myc/His tagged laf177 with FLAG tagged laf331, laf317 or laf152 and pulled the proteins using Nickel resin, and the interaction was analyzed using an immunoblot. As shown in Fig. 3B, laf177 was able to pull-down both laf331 and laf317 suggesting that this novel isoform of laforin could physically interact with the two known isoforms of laforin. Similar observations were made when POCR was coexpressed (Fig. 3B). However, no signal was detected for laf152 in the pull-down product, suggesting that laf177 does not interact with laf152 (Fig. 3B). Similarly, the Nickel resin did not pull laf331 in the absence of any His-tagged protein (Fig. 3B, last lane), thus establishing the specificity of the assay. We next tested the phosphatase activity of laf177–laf331 or laf177–laf317 heterodimers. For this, we coexpressed Myc/His tagged laf177 with FLAG tagged laf331 or laf317 and pulled the proteins using Nickel resin (Fig. 3C). Thereafter, the resin bound protein complex was subjected to the *in vitro* phosphatase assay. Intriguingly, the phosphatase activity of the laf177–laf331 and laf177–laf317 complexes was found to be lower than that of laf177 (Fig. 3C). This could possibly mean that laf177 as a heterodimer with laf331 or laf317 is inactive as a phosphatase. Since the complexes were pulled using the His-tag attached to laf177, the pull-down product is likely to have laf177 as a monomer (or as homodimer) and the laf177–laf331 or laf177–laf317 complex as a heterodimer. Thus, the decreased activity (~50% to that of laf177 only complex; see Fig. 3C) observed in these two pull-down products could possibly be due to the presence of laf177 monomer or its homodimer. This suggestion is strengthened by the observation that the mixture having laf177–laf317 complex also showed phosphatase activity albeit at lower level (see Fig. 3C). Laf317 is inactive as phosphatase and the activity observed here should have come from the laf177 monomer or its homodimer. This suggestion is further strengthened by the observation that the level of phosphatase active laf331 and phosphatase inactive laf317 was similar in the laf177–laf331 and laf177–laf317 complexes respectively (bars/lanes 5 and 6 in Fig. 3C) yet there was no difference in the phosphatase activity of these two complexes. Since we were unable to make a construct for laf177 with any other tag, we were not able

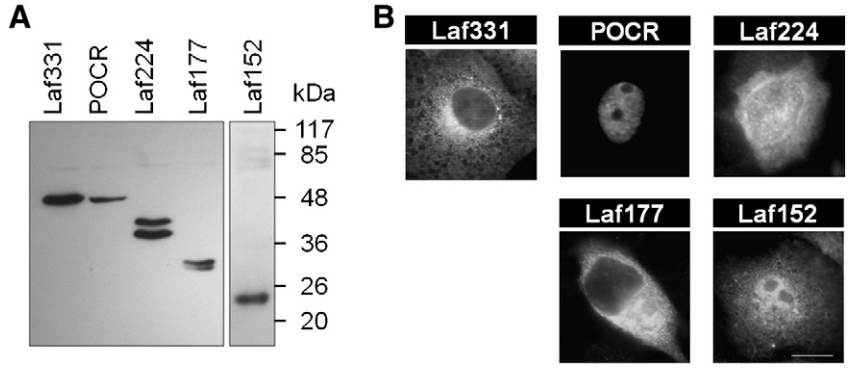

Fig. 2. *Expression analyses of laforin isoforms:* Expression constructs that code for laf331, POCR, laf224, and laf177 with Myc/His tag at the C-terminal or with N-terminal FLAG tag for laf152, were transiently expressed in COS-7 cells and processed for immnunoblot analysis (A) or indirect immunofluorescence analysis (B) using anti-Myc or anti-FLAG antibody. These images in A represent the localization pattern of a majority of transfected cells (~80–90%). Scale, 10 μm.

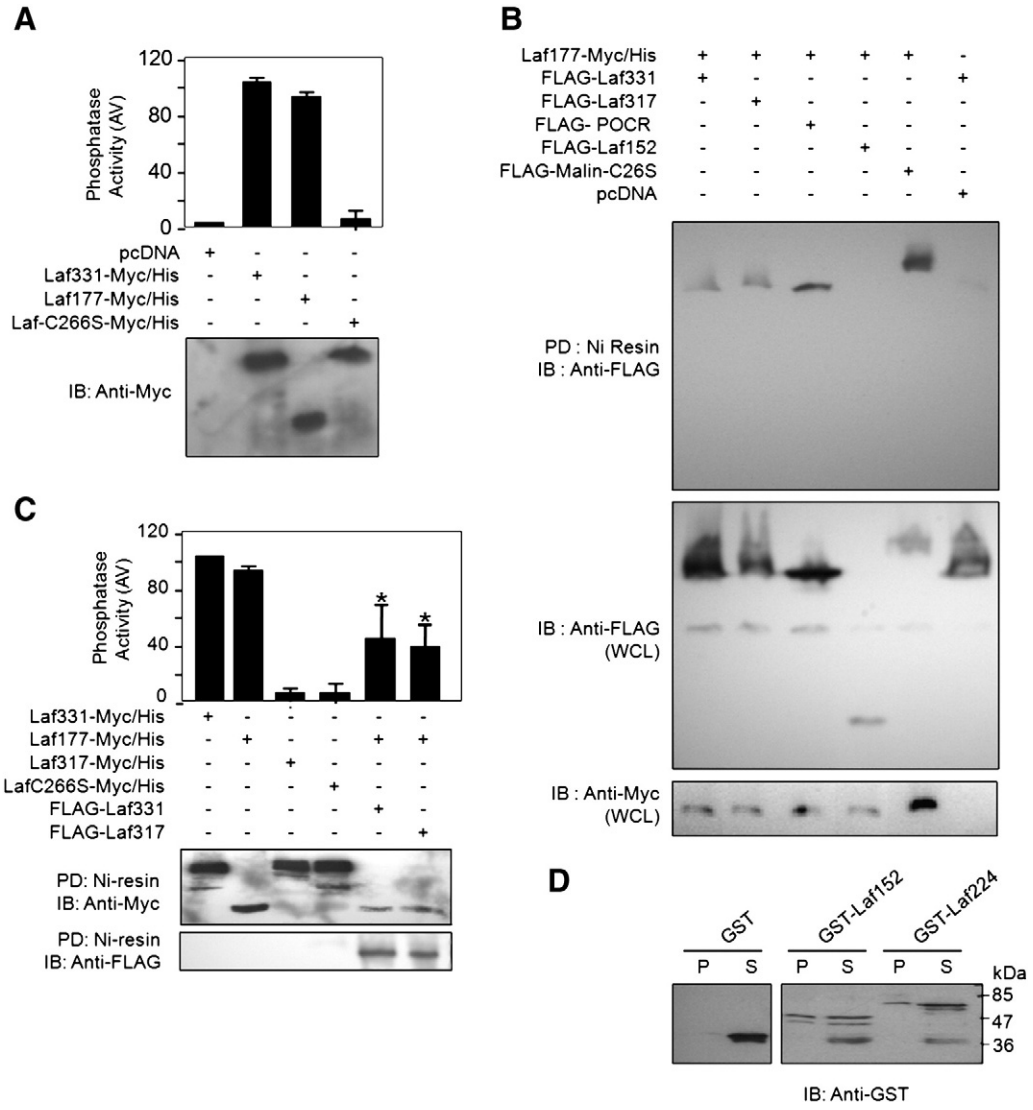

Fig. 3. *Phosphatase activity and glycogen binding property of laforin isoforms*: (A) Expression constructs that code for Myc/His-tagged laf331, laf177, and laf331 mutant C266S or an empty vector (pcDNA) were transiently transfected in COS-7 cells, His-tagged proteins were affinity purified using Nickel resin and subjected to phosphatase assay using the chromogenic substrate pNPP. Values shown are means of three independent reactions (y-axis: arbitrary vales [AV]). A fraction of the eluate was used for immunoblotting with anti-Myc antibody to show the elution/purity (below). (B) To check the interaction between laf177 and lf33, laf317, Laf152, POCR or malin, COS-7 cells were co-transfected with Myc/His-tagged laf177 and FLAG-tagged isoforms as indicated, and processed for the pull-down assay using the Ni-resin. The pull-down products (PD) and whole cell lysates (WCL) were immunoblotted (IB) with anti-FLAG and anti-Myc antibodies, as indicated. Cells co-transfected with empty vector (pcDNA) served as control. (C) To check the phosphatase activity of heterodimers of laforin isoforms, COS-7 cells were transiently transfected with the expression construct as indicated and the lysate was processed for the Ni-affinity resin pull-down assays. The resin bound proteins were used for the phosphatase assay using the pNPP as substrate. Values shown are means of three independent reactions. The purity and expression levels of the pulled-down proteins (the bait as well as the interacting partner) in each experiment were examined by immunoblotting, as shown below the graph ($p > 0.005$ [*]; *t*-test). (D) To check the glycogen binding property of laf224 and laf152, the coding regions of these two peptides were cloned into a bacterial expression vector to produce the protein as a fusion to the GST in *E. coli.* The protein was affinity purified and then incubated with glycogen, centrifuged and the proteins in the glycogen pellet (*P*) or the supernatant (*S*) fraction were visualized by western blot using anti-GST antibody. The GST protein without any fusion was used as a negative control for the assay.

to pull laf177 using the His-tagged laf317 and check for the phosphatase activity. Nonetheless, the observation that the decreased phosphatase activity was seen in the laf177–laf331 complex strongly suggests that laf177 as heterodimer is inactive as a phosphatase.

3.3. Laf224 and laf152 bind to glycogen in vitro

Although both laf331 and laf317 harbor CBD, only laf331 binds to glycogen *in vitro*[10]. Since the two of the novel isoforms of laforin — laf152 and laf224, harbor CBD, we next wanted to check whether or not these proteins bind to glycogen. For this we made an expression construct that code for glutathione S-transferase (GST) fused with laf152 or laf224 at the carboxyl terminal. Thereafter, we expressed GST (control) or the GST fusion protein in *E. coli*, purified using glutathione beads, and affinity of the purified proteins to the glycogen was tested in an *in vitro* glycogen binding assay. Both laf152 and laf224 precipitated with glycogen and were detected in the pellet fraction in western analysis using anti-GST antibody (Fig. 3D). As expected, the GST protein did not bind to the glycogen (Fig. 3D).

3.4. Malin interacts with laforin isoforms and promotes their degradation

Malin and laf331 are known to interact with each other and malin is known to promote its degradation [5]. Similar to laf331, isoform laf317 has also been shown to interact with malin and undergo degradation [10]. To check whether laforin isoforms laf152, laf177, and

laf224 would interact with malin or not, we coexpressed these proteins with distinct tags, in COS-7 cells, and processed for the pull-down assays using Nickel resin. As shown in Fig. 3B (lane 5) and Fig. 4A, Myc/His-tagged laf224 and laf177 were able to pull FLAG-tagged malin, suggesting that malin physically interacts with these isoforms. Since Myc/His-tagged construct was not available for laf152, we coexpressed FLAG-tagged laf152 with Myc/His tagged malin. Malin was able to pull Flag-tagged laf152. Taken together with previous reports [5,10], it could be suggested that malin interacts with all 5 isoforms of laforin that bear DSPD and/or CBD domains.

Having known that the isoforms of laforin interact with malin, we next checked whether malin can promote the degradation of the novel isoforms of laforin. To test this possibility, we coexpressed laf331 and the three novel isoforms (laf224, laf177 and laf152) along with a shRNA (RNAi) knockdown construct for malin or GFP (control) in COS-7 cells and measured the cellular level of laforin isoforms by immunoblotting. The shRNA knockdown constructs are validated in our previous studies [5] and controls blots are included here to show their efficiency (see Supplemental Fig. S3). As shown in Fig. 4B, knockdown of malin resulted in an increase in the signal intensity for each of the four isoforms. To further confirm the notion that malin indeed promote the degradation of laforin isoforms, we coexpressed laf224, laf177 or laf152 either with the wild-type malin or its inactive mutant C26S and in one set of cells coexpressing isoforms and wild type malin were treated with the proteasomal blocker MG132 (Fig. 4C). As a control, isoforms were expressed alone in the

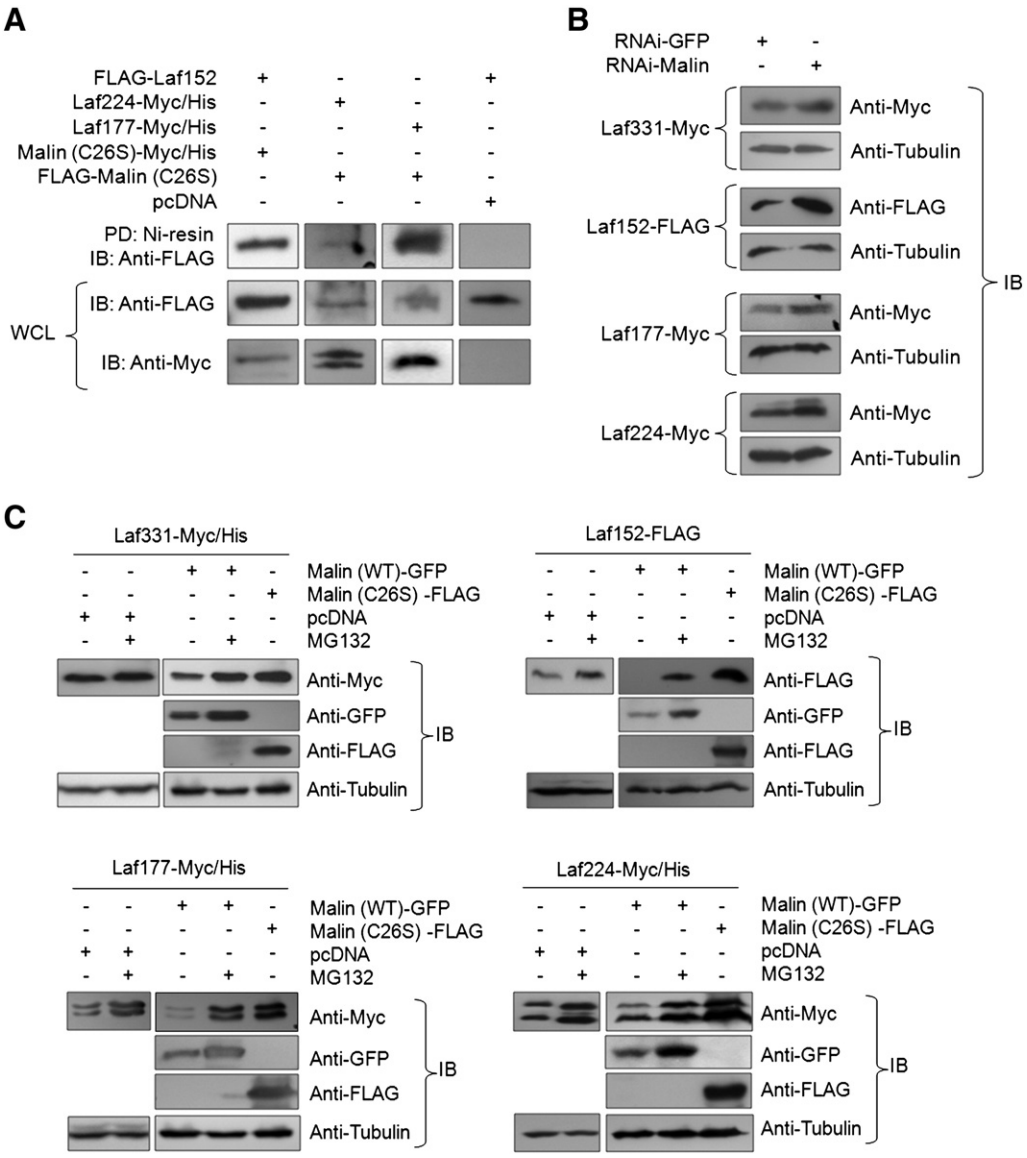

Fig. 4. *Malin interacts with and promotes the degradation of laforin isoforms*: (A) To check the interaction of malin with laforin isoforms laf152, laf224, and laf177, COS-7 cells were cotransfected with expression constructs as indicated. The lysates were then processed for pull down assays using Nickel affinity resin. The pulled-down products (PD) and whole cell lysates (WCL) were immunoblotted (IB) with anti-FLAG and anti-Myc antibodies as indicated. Lysate of cells transfected with an empty vector was used as a negative control (last lane). (B) To check whether malin promotes the degradation of laforin isoform, expression construct coding for laf331, laf152, laf177 or laf224 was coexpressed with an RNAi construct for GFP (control) or malin mutant and the cellular level of laforin isoforms was evaluated by immunoblotting, as indicated. Probing the blot with anti-tubulin antibody served as loading control. (C) As an alternate, laforin isoforms were either expressed alone or with the wild type GFP-malin (WT) or with the activity mutant malin (C26S) and were treated (+) or not treated (−) with the proteasomal blocker MG132 as indicated. The cellular level of laforin isoform was established by immunoblotting (IB). The expression of malin or its mutant was also established by immunoblotting. Tubulin served as loading control. Densitometric measurement confirmed significant decrease in the signal intensity for laforin isoform when coexpressed with wild-type malin as compared to its level when expressed with mutant malin C26S or upon MG132 treatment (data not shown).

presence or absence of MG132 as well (Fig. 4C). Laf331 was also expressed with similar combination for a comparison (Fig. 4C). In each of these sets, the signal intensity of laforin isoform is higher when expressed alone as compared to the one that was expressed with wild-type malin. However when treated with MG132 or when expressed with the mutant malin, its level was significantly higher. These data suggest that, as was known for laf331 and laf317 [10], malin promotes the degradation of the novel isoforms of laforin, laf152, laf224 and laf177, through the proteasomal system.

4. Discussion

A majority of the human genes predicted to have multiple transcriptional units thereby coding for more than one gene product. Supporting this notion, the present study demonstrates that *EPM2A* is a gene of such complexity. Alternative splicing in *EPM2A* generates at least five different transcript variants, named here as type-1 to type-5 transcripts. These include the three novel transcripts discovered and characterized in the present study (type-3, -4, and -5). Among these, the cDNA representing the type-4 and type-5 validated in the present study seem to have the 3′-end intact as both had the poly-A tail. Repeated attempt to extend the 5′-end of the cDNA, using the 5′-RACE approach did not yield any product. It may be recalled here that the first two reports on the discovery of the *EPM2A* gene had carried the partial cDNA sequences [17,18], and only the third report identified the 5′-end of the type-1 transcript [16]. Thus, only one cDNA clone (LDH1 — GenBank accession # NM_005670.3) is known to harbor the 5′-end of the reported coding sequence of the *EPM2A* gene [16]. The 5′-end of the *EPM2A* gene is GC rich (80%) [19]. It is therefore likely, that the sequence complexity at the 5′-region of the *EPM2A* gene might limit copying of the 5′-end of the mRNA into a cDNA and hence, almost all cDNA clones reported in the GenBank lack exon 1 region of type-1 transcript. Nonetheless, the discovery of the cDNA clones representing type-3 to -5 transcripts reveals the existence of novel exons for the *EPM2A* gene. The expression of all these transcripts was confirmed by RT-PCR using exon-specific primers and for the single exon transcript (type-3) by northern hybridization using the unique exon as the probe. It may be noted here that type-5 transcript appears to have a novel exon 1. While the presence of an exon upstream of this region could not be ruled out, the possibility that exons 1 and 2 of the type-1 transcript may join with the exon-1 of type-5 transcripts is less likely. This is because primers amplifying exons 2 and 3, or exons 1-to-3 of the type-1 transcript did not yield any product that was larger than the expected size representing the type-1 transcript. Therefore, it is likely that a few of the *EPM2A* transcripts might harbor novel first exon, suggesting the presence of more than one promoter elements for the *EPM2A* gene. It is possible that the alternate promoters may provide tissue specific expression of transcripts and this has shown for several mammalian genes [20]. While we have not tested the promoter elements, it is tempting to speculate that both alternative promoter usage and alternative splicing may contribute to the remarkable mRNA diversity of the *EPM2A* gene.

Among the transcripts characterized in the present study, the type-3 transcript reveals an unusual genomic organization in having two ORFs that overlap with each other. Although such features are unusual for the eukaryotic genome, it has been predicted that approximately 7% of alternatively spliced genes in humans contain dual (multiple) coding regions [21]. Indeed, the predicted start codons of each of the four ORFs of these transcripts have reliable Kozak score to initiate the translation, and the transfection studies using expression constructs specific to three of the ORFs did confirm that they could possibly code for proteins. However, whether or not these transcripts would use both ORFs and code for both proteins at the same time has not been tested in the present study. Although the occurrence of such overlapping coding sequences are known among viruses and are rare in mammals, few such genes have been reported to have functional overlapping reading frames; these are $p16^{INK4A}$/$p19^{ARF}$ (CDKN2A) gene [22], growth hormone(GH1)/GHDTA gene [23], the LGALS3/GALIG gene [24], the human molybdopterin synthase (MOCS2) gene [25], XLαs(Gnas)/ALEX locus [26], XBP1 ORF1/ORF2 [27] and 4E-BP3(EIF4EBP3)/MASK-$BP3^{ARF}$[28]. It has been demonstrated for these genes, using ectopic expression constructs, that both the ORFs are functional and that they produce proteins having varied properties. The detection of protein products for both ORFs for type-3 transcript in transfection studies implies an alternative initiation of translation at different AUG codons *in vivo*, which could be attributed to a leaky scanning process or an internal entry of ribosomes [29]. We were unable to confirm the endogenous expression of laforin isoforms since antibodies that could detect the individual isoforms are not available. Hence we have used expression constructs for each coding region and tested its coding potential in transient expression conditions. It may be noted here that similar approach was used in a majority of studies that documented the coding potential of transcripts with alternative reading frames [22–28].

One of the interesting aspects of ectopically expressed laforin isoforms is their subcellular localization: POCR shows nuclear localization, laf331 and laf177 are cytoplasmic, and both laf224 and laf152 are localized in nucleus as well as in the cytoplasm. Indeed, proteins with isoforms are known to localize at different subcellular compartments and play different functional roles. For example, Nek2Kinase (NEK2) has three isoforms localized differentially, enabling it to have both nuclear and cytoplasmic functions [30]. Differentially spliced transcripts that encode isoforms lacking the catalytic domain have been identified in other members of phosphatase family [31,32] and it has been suggested that such truncated polypeptides may either present substrate molecules to their active counterpart or instead may protect them from dephosphorylation [31]. The findings on laf331 and laf317 in our previous study [10] fit well with the aforementioned model. The present study demonstrates that isoform laf177, though active on its own as a phosphatase, functions similar to the inactive laf317 in regulating the activity of laf331 by forming a heterodimer *in vitro*. The present study also establishes that malin is a critical regulator for the laforin isoforms since malin reduces the half-life of three novel isoforms of laforin when ectopically expressed. As suggested earlier for laf317 [10], each isoform of laforin could regulate the cellular levels of the other isoform by presenting itself as a substrate to malin and thus saving the other isoform from degradation. Clearly further work on the synergistic and antagonistic functions of the two active phosphatases (laf331 and laf177) and the physiological significance of malin in regulating the cellular levels should shed more light on the complex interplay between the proteins coded by the two LD genes. Nonetheless, it should be noted here that all these studies were performed on proteins that were ectopically overexpressed in cell lines and that the interpretations are based on the assumption that these isoforms are expressed endogenously in significant amounts in human tissues. It is equally likely that these transcripts may function as non-coding regulatory RNAs, a possibility not tested in the present study.

Since its discovery, several groups have identified LD associated mutations (over 50 distinct mutations) in the *EPM2A* gene from diverse populations [2]. Interestingly, the site of the mutations rather than the type of mutations was found to show some positive correlations with distinct clinical subgroups of LD in a few reports [33, 34]. Given the observations that individual exons are alternatively skipped in *EPM2A*, and given the possibility that not all mutations would affect all transcripts or their protein products, it is tempting to speculate that the functional status of individual isoforms might underlie some of the symptoms in LD. It would therefore be of importance to carryout prospective genotype–phenotype correlation study to understand the possible role of alternative splicing in the clinical symptoms of LD.

Acknowledgments

This work was supported by a research grant from the Department of Biotechnology (DBT), Government of India, to SG. DD and RP received research fellowships from the Council of Scientific and Industrial Research, Government of India. The authors thank Dr S.K. Shankar of the National Institute of Mental Health and Neuroscience (NIMHANS), Bangalore, for providing the human brain samples. The Human Brain Bank at NIMHANS is supported by a grant from the DBT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2011.10.001.

References

[1] S. Ganesh, R. Puri, S. Singh, S. Mittal, D. Dubey, Recent advances in the molecular basis of Lafora's progressive myoclonus epilepsy, J. Hum. Genet. 51 (2006) 1–8.
[2] S. Singh, S. Ganesh, Lafora progressive myoclonus epilepsy: a meta-analysis of reported mutations in the first decade following the discovery of the EPM2A and NHLRC1 genes, Hum. Mutat. 30 (2009) 715–723.
[3] E.M. Chan, E.J. Young, L. Ianzano, I. Munteanu, X. Zhao, C.C. Christopoulos, et al., Mutations in NHLRC1 cause progressive myoclonus epilepsy, Nat. Genet. 35 (2003) 125–127.
[4] M.S. Gentry, C.A. Worby, J.E. Dixon, Insights into Lafora disease: malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin, Proc. Natl Acad. Sci. USA 102 (2005) 8501–8506.
[5] P. Garyali, P. Siwach, P.K. Singh, R. Puri, S. Mittal, S. Sengupta, R. Parihar, S. Ganesh, The malin–laforin complex suppresses the cellular toxicity of misfolded proteins by promoting their degradation through the ubiquitin-proteasome system, Hum. Mol. Genet. 18 (2009) 688–700.
[6] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, et al., Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476.
[7] B. Hartmann, J. Valcárcel, Decrypting the genome's alternative messages, Curr. Opin. Cell Biol. 21 (2009) 377–386.
[8] S. Ganesh, T. Suzuki, K. Yamakawa, Alternative splicing modulates subcellular localization of laforin, Biochem. Biophys. Res. Commun. 291 (2002) 1134–1137.
[9] L. Ianzano, E.J. Young, X.C. Zhao, E.M. Chan, M.T. Rodriguez, M.V. Torrado, et al., Loss of function of the cytoplasmic isoform of the protein laforin (EPM2A) causes Lafora progressive myoclonus epilepsy, Hum. Mutat. 23 (2004) 170–176.
[10] D. Dubey, S. Ganesh, Modulation of functional properties of laforin phosphatase by alternative splicing reveals a novel mechanism for the EPM2A gene in Lafora progressive myoclonus epilepsy, Hum. Mol. Genet. 17 (2008) 3010–3020.
[11] S. Ganesh, N. Tsurutani, T. Suzuki, Y. Hoshii, T. Ishihara, A.V. Delgado-Escueta, K. Yamakawa, The carbohydrate-binding domain of Lafora disease protein targets Lafora polyglucosan bodies, Biochem. Biophys. Res. Commun. 313 (2004) 1101–1109.
[12] B.R. Graveley, Sorting out the complexity of SR protein functions, RNA 6 (2000) 1197–1211.
[13] L. Cartegni, J. Wang, Z. Zhu, M.Q. Zhang, A.R. Krainer, ESEfinder: a web resource to identify exonic splicing enhancers, Nucleic Acids Res. 31 (2003) 3568–3571.
[14] T. Nishikawa, T. Ota, T. Isogai, Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences, Bioinformatics 16 (2000) 960–967.
[15] P. Rigou, V. Piddubnyak, A. Faye, J.C. Rain, L. Michel, F. Calvo, J.L. Poyet, The anti-apoptotic protein AAC-11 interacts with and regulates Acinus-mediated DNA fragmentation, EMBO J. 28 (2009) 1576–1588.
[16] S. Ganesh, K.L. Agarwala, K. Ueda, T. Akagi, K. Shoda, T. Usui, et al., Laforin, defective in the progressive myoclonus epilepsy of Lafora type, is a dual-specificity phosphatase associated with polyribosomes, Hum. Mol. Genet. 9 (2000) 2251–2261.
[17] B.A. Minassian, J.R. Lee, J.A. Herbrick, J. Huizenga, S. Soder, A.J. Mungall, et al., Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy, Nat. Genet. 20 (1998) 171–174.
[18] J.M. Serratosa, P. Gómez-Garre, M.E. Gallardo, B. Anta, D.B. de Bernabé, D. Lindhout, et al., A novel protein tyrosine phosphatase gene is mutated in progressive myoclonus epilepsy of the Lafora type (EPM2), Hum. Mol. Genet. 8 (1999) 345–352.
[19] S. Ganesh, K. Shoda, K. Amano, A. Uchiyama, S. Kumada, N. Moriyama, et al., Mutation screening for Japanese Lafora's disease patients: identification of novel sequence variants in the coding and upstream regulatory regions of EPM2A gene, Mol. Cell. Probes 15 (2001) 281–289.
[20] T.A. Ayoubi, W.J. Van De Ven, Regulation of gene expression by alternative promoters, FASEB J. 10 (1996) 453–460.
[21] H. Liang, L.F. Landweber, A genome-wide study of dual coding regions in human alternatively spliced genes, Genome Res. 16 (2006) 190–196.
[22] D.E. Quelle, F. Zindy, R.A. Ashmun, C.J. Sherr, Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest, Cell 83 (1995) 993–1000.
[23] N. Labarrière, P.L. Selvais, F.P. Lemaigre, A. Michel, D.M. Maiter, G.G. Rousseau, A novel transcriptional activator originating from an upstream promoter in human growth hormone gene, J. Biol. Chem. 270 (1995) 19205–19208.
[24] M. Guittaut, S. Charpentier, T. Normand, M. Dubois, J. Raimond, A. Legrand, Identification of an internal gene to the human Galectin-3 gene with two different overlapping reading frames that do not encode Galectin-3, J. Biol. Chem. 276 (2001) 2652–2657.
[25] B. Stallmeyer, G. Drugeon, J. Reiss, A.L. Haenni, R.R. Mendel, Human molybdopterin synthase gene: identification of a bicistronic transcript with overlapping reading frames, Am. J. Hum. Genet. 64 (1999) 698–705.
[26] M. Klemke, R.H. Kehlenbach, W.B. Huttner, Two overlapping reading frames in a single exon encode interacting proteins—a novel way of gene usage, EMBO J. 20 (2001) 3849–3860.
[27] H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor, Cell 107 (2001) 881–891.
[28] F. Poulin, A. Brueschke, N. Sonenberg, Gene fusion and overlapping reading frames in the mammalian genes for 4E-BP3 and MASK, J. Biol. Chem. 278 (2003) 52290–52297.
[29] N.K. Gray, M. Wickens, Control of translation initiation in animals, Annu. Rev. Cell Dev. Biol. 14 (1998) 399–458.
[30] W. Wu, J.E. Baxter, S.L. Wattam, D.G. Hayward, M. Fardilha, A. Knebel, et al., Alternative splicing controls nuclear translocation of the cell cycle-regulated Nek2 kinase, J. Biol. Chem. 282 (2007) 26431–26440.
[31] A. Bult, F. Zhao, R. Dirkx Jr., A. Raghunathan, M. Solimena, P.J. Lombroso, STEP: a family of brain-enriched PTPs. Alternative splicing produces transmembrane, cytosolic and truncated isoforms, Eur. J. Cell Biol. 72 (1997) 337–344.
[32] K. Aoyama, T. Matsuda, N. Aoki, Characterization of newly identified four isoforms for a putative cytosolic protein tyrosine phosphatase PTP36, Biochem. Biophys. Res. Commun. 266 (1999) 523–531.
[33] S. Ganesh, A.V. Delgado-Escueta, T. Suzuki, S. Francheschetti, C. Riggio, G. Avanzini, et al., Genotype-phenotype correlations for EPM2A mutations in Lafora's progressive myoclonus epilepsy: exon 1 mutations associate with an early-onset cognitive deficit subphenotype, Hum. Mol. Genet. 11 (2002) 1263–1271.
[34] G. Annesi, V. Sofia, A. Gambardella, I.C. Candiano, P. Spadafora, F. Annesi, N. Cutuli, E.V. De Marco, D. Civitelli, S. Carrideo, P. Tarantino, R. Barone, M. Zappia, A. Quattrone, A novel exon 1 mutation in a patient with atypical Lafora progressive myoclonus epilepsy seen as childhood-onset cognitive deficit, Epilepsia 45 (2004) 294–295.