Cecchini Nicolas M (Orcid ID: 0000-0002-2894-7744)

Alvarez María Elena (Orcid ID: 0000-0001-7001-5898)



# Alternative splicing of an exitron determines the subnuclear localization of the Arabidopsis DNA-glycosylase MBD4L under heat stress

Nicolás Miguel Cecchini, José Roberto Torres[#], Ignacio Lescano López[#], Santiago Cobo, Florencia Nota, María Elena Alvarez*

Centro de Investigaciones en Química Biológica de Córdoba, CIQUIBIC, CONICET, Departamento de Química Biológica Ranwel Caputto, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Haya de la Torre y Medina Allende, Ciudad Universitaria, Córdoba, X5000HUA, Argentina.

* To whom correspondence should be addressed.
E-mail: maria.elena.alvarez@unc.edu.ar

[#] These authors equally contributed to this work.

**Running title**

An exitron drives the MBD4L subnuclear destination

**Significance statement**

Exitrons (exonic introns) are a poorly studied class of protein-coding introns within exons expected to increase genome plasticity and protein diversity. Many exitrons have been recently identified in eukaryotic genomes but only few of them have been associated with specific functions. Our work provides the first example of an exitron defining the targeting of a DNA repair enzyme to the nucleoplasm or the nucleolus.

**Keywords**:

DNA glycosylases, MBD4L, exitron, DNA repair, subnuclear distribution, nucleolus.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.15675

**SUMMARY**

DNA glycosylases are critical enzymes that recognize small base lesions in DNA and trigger their repair to preserve genome integrity. The Arabidopsis MBD4-like (MBD4L) DNA glycosylase improves tolerance to genotoxic stress. This enzyme is encoded by a single gene carrying an exitron at its 5´ region. Although alternative splicing (AS) of exitrons (protein-coding cryptic introns within exons) is suspected to increase protein diversity, phenotypes associated to exitron removal or retention are only known for a few genes. Here, we show that alternative splicing of the *MBD4L* exitron determines the generation of two enzyme isoforms with different subnuclear localization. Both isoforms conserve the catalytic domain but are directed either to the nucleoplasm or the nucleolus. Interestingly, heat-stress regulates the AS of the *MBD4L* exitron and increases the relative abundance of the nucleolar variant. This process depends on the splicing factors NTR1 and RS31. Notoriously, we generated transgenic plants expressing a mutated *MBD4L-GFP* gene version that abolished exitron splicing and found that nucleolar protein targeting was impaired in these plants. Our findings suggest that AS of the MBD4L exitron can function as a mechanism to drive the enzyme to the nucleolus during heat stress. Several DNA repair enzymes reach the nucleolus under particular conditions although AS of exitrons has not been so far associated with this process. To our knowledge, this is the first example of an exitron mediating enzyme localization in eukaryotes.

## INTRODUCTION

DNA glycosylases catalyze the first step in the base excision repair (BER) pathway. These enzymes recognize non-bulky modifications, such as inappropriate or mispaired bases, or specific single-base DNA lesions generated by oxidation, alkylation, and deamination, and remove them by hydrolyzing the N-glycosylic bond (Roldán-Arjona *et al.*, 2019). Subsequently, other BER enzymes such as AP-endonucleases, DNA polymerases, and DNA ligases, allow restoration of the healthy base (Jacobs and Schär, 2012). DNA repair enzymes are often adapted to respond to stress since DNA bases are reactive molecules whose damage is exacerbated under adverse conditions. Many mammalian and plant DNA glycosylases are transcriptionally regulated and enhance tolerance to environmental or genotoxic stress when over-expressed (Bramson *et al.*, 1995; Chen *et al.*, 2012; Nota *et al.*, 2015; Frosina, 2000).

MBD4 (methyl-CpG-binding domain protein 4) is a DNA glycosylase belonging to the HhH (helix–hairpin–helix motif) superfamily that excises U and T opposite to G, preferentially at CpG sites (Sjolund *et al.,* 2012). In metazoan, MBD4 contains an N-terminal methyl binding domain (MBD), whereas in plants it lacks this domain and is called MBD4-like (MBD4L). The *Arabidopsis thaliana* (hereafter Arabidopsis) MBD4L excises *in vitro* U, U-halogenated derivatives, and T mispaired to G, with preference for CpG contexts (Ramiro-Merina *et al.*, 2013). Like other genes encoding for BER components, DNA glycosylase genes display alternative splicing (AS) (Boldinova *et al.*, 2019; Owen *et al.*, 2007; Nishioka *et al.*, 1999; Ohtsubo, 2000; Nilsen *et al.*, 1997; Nimeth *et al.*, 2020). In humans, different splicing variants of the DNA glycosylases MutY and OGG1 target either the nucleus or mitochondria (Takao *et al.*, 1999; Furihata, 2015), while those of MBD4 act over distinct DNA lesions as the short isoform lacking the MBD domain recognizes the U:G but not T:G mismatch (Owen *et al*., 2007). In contrast, splicing variants of plant DNA-glycosylases have been poorly characterized (Boldinova *et al.*, 2019; Nimeth *et al.*, 2020). In Arabidopsis, MBD4L is encoded by a single gene (At3g07930) with four predicted alternative transcripts (https://jbrowse.arabidopsis.org/). *MBD4L.3* and *MBD4L.4* are present in leaves and flowers while *MBD4L.1* and *MBD4L.2* have not been detected in these or other tissues so far. *MBD4L.3* and *MBD4L.4* differ in an exitron and they encode for 445 (MBD4L.3) or 329 (MBD4L.4) amino acid proteins that conserve the entire catalytic domain, and at least one of these proteins is nuclear (Nota *et al*., 2015).

Exitrons are introns located within exons with characteristics of both elements (Marquez *et al*., 2015). They contain all splicing canonical signals but no stop codons, and when retained, they do not trigger the nonsense-mediated decay pathway as the transcripts derived from other intron retention (IR) events. Exitrons have great potential to increase protein diversity through AS. They

encode for sequences with a marked enrichment of predicted post-translational modifications sites, domains and motifs, and intrinsically disordered regions (Staiger and Simpson, 2015; Marquez *et al.*, 2015). Curiously, gene ontology (GO) classification of exitron-containing genes indicates a strong association to stress responses, transcription, and development. Despite this, functional studies of protein isoforms derived from AS of exitrons are scarce. The first study describing a biological function associated to these elements in plants demonstrated that an exitron spliced-isoform of the AUXIN RESPONSE FACTOR 8 has tissue-specific expression and is required for anther filament elongation and early stomium opening (Ghelli *et al.*, 2018). More recently, isoforms generated by AS of exitrons of the *FLS2* immune receptor gene were detected in dicot plants, and proposed to function as suppressors of defenses (Cheng *et al.*, 2020).

Interestingly, the ratio between *MBD4L.3* and *MBD4L.4* transcripts changes under water stress conditions and in response to methyl-jasmonate, and is modulated by the levels of the RS31 splicing factor (Marquez *et al.*, 2015). The short variant *MBD4L.4* loses a region that includes a nuclear localization signal (NLS), which is present in *MBD4L.3* (Nota *et al*., 2015).Then, it is possible that exitron AS affects the nuclear enzyme targeting. Here, we examined this possibility by studying the location of the two *MBD4L* splicing isoforms at basal and stress conditions. Unexpectedly, we found that AS of the exitron does not affect the nuclear enzyme localization but instead its subnuclear distribution. The larger exitron-retained version MBD4L.3 is located in the nucleoplasm, while the shorter spliced isoform MBD4L.4 is found in the nucleolus. Moreover, the distribution pattern of MBD4L is dynamic and is modified by heat-stress, which triggers an increase of the nucleolar isoform. To our knowledge, this is the first example showing that the AS of an exitron determines a protein subnuclear localization.

## RESULTS

### *Two MBD4L exitron-spliced variants reach different subnuclear locations*

Arabidopsis activates the BER pathway requiring DNA glycosidases not only in the nucleus but also in mitochondria and chloroplasts (Roldán-Arjona *et al.*, 2019). MBD4L.3 and MBD4L.4 include the DNA glycosylase domain and are suspected to be active enzymes (Ramiro-Merina *et al.*, 2013). As MBD4L.3 contains two predicted NLS, and MBD4L.4 has only one (Nota *et al.*, 2015), it is possible that AS of the exitron determines that MBD4L.3 but not MBD4L.4 is targeted to the nucleus. MBD4L.3 seems to be nuclear since plants transformed with the MBD4L.3 coding region fused to GFP accumulate the MBD4L.3-GFP protein and show nuclear GFP fluorescence. However, these

plants also contain *MBD4L.4-GFP* transcripts and the MBD4L.4 location has not been determined so far (Nota *et al.*, 2015). Using laser scanning confocal microscopy we here determined the subcellular localization of the MBD4L.3-GFP and MBD4L.4-GFP fusion proteins expressed in *Nicotiana benthamiana* and Arabidopsis.

Transient expression of MBD4L.3-GFP in *N. benthamiana* leaves showed GFP fluorescence predominantly in the nucleoplasm, either with a homogenous –uniform- (48%) or non-homogenous -dotted- (52%) distribution (Figure 1a). *MBD4L.4-GFP* transcripts were not detected in these tissues (Figure S1) indicating that these patterns corresponded to MBD4L.3-GFP. These patterns could eventually be associated with a particular cell cycle stage, the transgene expression level or another trait. In turn, tissues expressing MBD4L.4-GFP displayed a strong GFP signal in a nucleolus-like region in the vast majority of nuclei (87%), with a few of them (13%) forming nucleolar aggregates. In contrast, the nucleolus of MBD4L.3-GFP and the nucleoplasm of MBD4L.4-GFP showed very weak GFP signals. As expected, transformation with GFP (control) resulted in a homogenous nucleoplasmic fluorescence pattern in all analyzed cells.

To confirm the presence of MBD4L.4-GFP in the nucleolus, we co-expressed this variant with the nucleolus protein marker Nucleolin2-mCherry (Liu *et al.*, 2014). In parallel, we performed the same analysis with MBD4L.3-GFP. As expected, we only observed a robust co-localization between MBD4L.4-GFP and the nucleolar marker (Figure 1b). In fact, the GFP and mCherry signal profile plots indicated that MBD4L.3-GFP is reduced in the nucleolus (Figure 1b, right). Next, we evaluated the subcellular localization of both MBD4L-GFP isoforms in stably transformed Arabidopsis plants. As found in *N. benthamiana*, enrichment of MBD4L.4-GFP in the nucleolus was reproduced in both root and stomata cells (Figure 1c-d). As expected, MBD4L.3-GFP was mostly nucleoplasmic according to previous observations (Nota *et al.*, 2015).

These results indicate that the AS of the *MBD4L* exitron allows the generation of two enzyme isoforms that reside at different subnuclear regions.

***Subnuclear localization of MBD4L splicing isoforms depends on their N-terminal region***

We next searched for the presence of nucleolar localization sequences (NoS) in both MBD4L variants using the Nucleolar Sequence Detector software (NoD; (Scott *et al.*, 2011)). We detected one NoS at the N-terminus of both isoforms (amino acids 1-25), and a second NoS present only in MBD4L.4 (residues 94-113) (Figure 2a). These NoS included positively charged residues, they depended on the sequence context, and were present at the unstructured protein region preceding the DNA-glycosylase domain (Figure S2). Therefore, we fused the N-terminal region of either MBD4L.3 (amino acids 1 to 306) or MBD4L.4 (amino acids 1 to 189) to GFP and analyzed the

subnuclear localization of these proteins. As shown in Figure 2b, when expressed in *N. benthamiana* MBD4L.3$^{1-306}$-GFP was distributed in the nucleoplasm while MBD4L.4$^{1-189}$-GFP was enriched in the nucleolus. Thus, both truncated versions recapitulated the patterns of full-length GFP-fusion proteins.

To further characterize the nucleolar MBD4L.4-GFP protein, we compared its mobility at the nucleolus and the nucleoplasm by fluorescence recovery after photobleaching (FRAP) assays (Figure 2c). For this purpose, *N. benthamiana* epidermal cells expressing this protein were photobleached in a small nuclear area, and GFP signal was subsequently monitored to determine the half-time for fluorescence recovery. Interestingly, the half-time fluorescence recovery was lower for the nucleoplasmic than the nucleolar fraction (37 and 82 sec, respectively) showing that the nucleoplasmic pool constitutes a faster pool. In turn, the nucleolar MBD4L.4-GFP fraction showed similar recovery times than nucleoplasmic MBD4L.3-GFP (82 and 98 sec, respectively). Similarly, other nucleolar proteins fused to GFP have differences in mobility when are located (and retained) either at the nucleolus or the nucleoplasm (Chen and Huang, 2001; Musinova *et al.*, 2015).

Together, these results indicate that the region preceding the DNA glycosylase domain contains signals/motifs that determine the MBD4L nuclear fate, and that removal or retention of the exitron defines the subnuclear enzyme destination.

***Heat stress regulates MBD4L exitron alternative splicing***

Splicing of the *MBD4L* exitron is affected by water deprivation and JA application (Marquez *et al.*, 2015). Looking for other conditions that modify the AS of *MBD4L*, we exposed Arabidopsis wild type Col-0 seedlings to different treatments and analyzed the relative *MBD4L.3* and *MBD4L.4* transcript content. We defined the splicing index (SI) as the ratio of *MBD4L.4* / *MBD4L.3*. Treatments included bacterial infection (*Pseudomonas syringae* pv *tomato -Pst-*), osmotic (mannitol) and heat stress, and exogenous application of hydrogen peroxide ($H_2O_2$), methyl viologen (MV), methyl-jasmonate (JA) or salicylic acid (SA). The SI was similar in the control condition (mock) and in response to MV and $H_2O_2$, but tended to increase with JA, mannitol, SA or *Pst* treatments, although these changes were not consistent between replicates and thus no statistically significant (Figure S3).

Heat stress triggered robust and reproducible AS changes (Figure 3a, Figure S3). At control conditions (22ºC), the SI values were not significantly altered throughout the whole analyzed period (6.5 h) (SI: 0.86 ± 0.42 at 0 hpt -hours post-treatment-; 0.91 ± 0.36 at 6.5 hpt; Figure 3a). In contrast, these values changed at 37ºC where they evidenced the relative increase of *MBD4L.4* over *MBD4L.3* from 1.5 to 6.5 hours hpt (SI: 7.68 ± 4.96 at 6.5 hpt; Figure 3a). We next tested if heat-induced *MBD4L* transcript changes were reversible. For this, we analyzed the ratio of *MBD4L* transcripts in seedlings incubated for 6.5 h at 37°C and recovered for 17.5 h at 22°C. After recovery, the *MBD4L* transcripts

ratio returned to basal levels (0 hpt) (SI: 0.56 ± 0.15 at 0 hpt; 0.48 ± 0.15 at 17.5 hpt after recovery; Figure 3b). These results suggest that AS of the *MBD4L* exitron is tightly controlled by temperature.

If heat stress regulates AS of the *MBD4L* exitron, changes in relative transcripts abundance would depend on the AS machinery. We tested if this process requires the spliceosomal disassembly factor NTR1 that widely affects AS (Jones *et al.*, 2012; Dolata *et al.*, 2015). In addition, we evaluated if the splicing factor RS31, whose overexpression increases *MBD4L* exitron removal at basal conditions (Marquez *et al.*, 2015), preserves this capacity under heat stress. To do this, we analyzed the levels of *MBD4L.3* and *MBD4L.4* transcripts in the *ntr1-1* mutant (Dolata *et al.*, 2015) and RS31 overexpressing (RS31ox; (Petrillo *et al.*, 2014)) plants after 6.5 h incubation at 37°C. As shown in Figure 3c, heat shock did not induce changes on the relative *MBD4L.3* and *MBD4L.4* transcript levels in *ntr1-1* plants, indicating that NTR1 is required to activate MBD4L AS. On the other hand, the RS31ox seedlings showed an increased SI value at control conditions, and this was maintained after heat shock.

Therefore, heat stress alters the ratio between *MBD4L* transcripts promoting exitron removal, and this process involves at least the splicing factors NTR1 and RS31.

***Heat-stress induces changes in MBD4L subnuclear localization***

Based on our previous results, we evaluated the MBD4L localization after heat treatment. We exposed Arabidopsis seedlings expressing MBD4L.3-GFP to 37°C for 6.5 h and analyzed the GFP subnuclear localization in root cells. As observed in Figure 4a, heat stress triggered a markedly redistribution of GFP from the nucleoplasm (22ºC, control) to the nucleolus (37ºC). Indeed, nucleolar GFP fluorescence was detected in 11% of nuclei under control conditions and in 70% of nuclei after heat stress (Figure 4a, right). In contrast, other types of stress and hormone treatments did not substantially change GFP subnuclear localization (Figure S4), consistent with their weak effect on AS changes.

Next, we asked whether the observed GFP patterns correlated with the AS of the exitron of the transgene. To evaluate this, we compared the *MBD4L.3-GFP* and *MBD4L.4-GFP* transcript levels and the SI values in MBD4L.3-GFP seedlings exposed to 37°C for 1.5, 3.5 or 6.5 h. As previously observed for the native transcripts, the ratio of *MBD4L.4-GFP* and *MBD4L.3-GFP* transcripts changed in response to high temperature (Figure 4b). The shorter product started to accumulate from 1.5 hpt reaching the highest level at 6.5 hpt, whereas the *MBD4L.3-GFP* transcript was significantly reduced at this later time. Consistently, the SI showed a rapid increase, although changes in this value were statistically significant only at 6.5 hpt at 37°C (Figure 4c, bottom). Probably, at early stages of heat shock the transgene has a more variable behavior than the endogenous gene (Figure 4b, Figure 3a).

In addition, the *MBD4L-GFP* transcripts recovered their basal ratio (0 h) after heat stress release (17.5 h) (Figure 4c, bottom). To know how transcriptional changes affect the protein content we analyzed MBD4L-GFP variants by western blot using α-GFP antibodies. Here again, we observed that heat stress altered the ratio between isoforms, producing a relative increase of MBD4L.4-GFP up to 3.5 hpt (Figure 4d). At 6.5 hpt, both isoforms showed similar levels, for reasons still unknown which could include stabilization of MBD4L.3-GFP by some post-translational modification. Therefore, heat stress induces MBD4L-GFP subnuclear re-localization by affecting the relative abundance of both *MBD4L.3-GFP* and *MBD4L.4-GFP* transcripts and protein isoforms.

***Exitron splicing drives MBD4L nucleolar localization under heat stress***

Our data suggest that heat stress induces the splicing of the *MBD4L* exitron increasing the content of *MBD4L.4-GFP* transcripts and the nucleolar protein isoform. To confirm this, we designed a non-spliceable *MBD4L* gene version. We used a *MBD4L.3* synthetic gene with synonymous point mutations not only in donor and acceptor splicing sites, but also in most of the possible alternative splicing sites inside the exitron and surrounding regions (Figure 5a) (see methods). This approach reduced the chances of using alternative/cryptic (weaker) donor/acceptor sites generated after mutation (Gelli *et al.*, 2019; Jian *et al.*, 2014). This construct was fused to GFP (MBD4L.3$^{exit}$-GFP) and transiently expressed in *N. benthamiana* leaves. GFP fluorescence was detected in the nucleoplasm, either with homogeneous or non-homogeneous distribution (Figure 5b), with a pattern indistinguishable from that of wild type MBD4L.3-GFP (Figure 1a). As the wild type *MBD4L.3-GFP* transgene did not undergo splicing in *N benthamiana* (Figure S1), we could not evaluate *MBD4L.3$^{exit}$-GFP* splicing in this model. Instead, we generated transgenic Arabidopsis plants expressing the *MBD4L.3$^{exit}$-GFP* construct and analyzed the *MBD4L.3-GFP* and *MBD4L.4-GFP* transcript levels. Notoriously, although *MBD4L.3$^{exit}$-GFP* transcripts accumulated after heat stress (37ºC, 6.5 h), removal of exitron was not observed at this condition (Figure 5c). Furthermore, the GFP signal was not accumulated at the nucleolus in samples exposed to 37°C (Figure 5d).

These and previous results indicate that the *MBD4L.3-GFP* exitron splicing and the subsequent generation of MBD4L.4 is the cause for MBD4L-GFP localization changes from nucleoplasm to nucleolus in response to heat stress.

**DISCUSSION**

In plants, nearly 70% of multi-exonic genes display AS, and IR event is the predominant AS type (Chaudhary *et al.*, 2019; Syed *et al.*, 2012; Reddy, 2007). AS is believed to play vital roles in plant growth, development, and stress tolerance, and some studies have begun to address the effect of IR under particular conditions (Chaudhary *et al.*, 2019). However, how AS of exitrons increases protein diversity is widely unexplored (Marquez *et al.*, 2015). Here, we provide the first characterization of an exitron whose splicing determines the subnuclear localization of a plant enzyme. We show that the Arabidopsis DNA glycosylase MBD4L has two isoforms that differ in sequences encoded by an exitron, which defines the nucleoplasmic or nucleolar enzyme fate. We confirmed the predominant nucleoplasmic localization of MBD4L.3 by expressing the *MBD4L.3$^{exit}$-GFP* construct in Arabidopsis and *N. benthamiana* plants (Figure 1, Figure 5, Figure S1). Moreover, we showed that MBD4L.4 is mostly nucleolar since plants expressing MBD4L.4-GFP alone showed GFP fluorescence at this compartment (Figure 1), and *MBD4L.4-GFP* transcripts and nucleolar GFP signal increased in *MBD4L.3-GFP,* but not *MBD4L.3$^{exit}$-GFP* plants, after heat stress (Figure 4, Figure 5). The presence of these localization patterns in both Arabidopsis and *N. benthamiana* suggests that mechanisms determining the enzyme fate can be conserved between *Solanaceae* and *Brassicaceae*. As both isoforms may be active (Ramiro-Merina *et al.*, 2013), AS of the exitron may operate as a mechanism to regulate the abundance of each fraction in Arabidopsis. Moreover, as discussed later, this mechanism may be used to re-localize MBD4L at the nucleolus after heat stress.

The MBD4 protein subfamily contains a poorly conserved N-terminal region dispensable for product binding and catalytic activity (Ramiro-Merina *et al.*, 2013). No specific functions have been assigned to this region for the Arabidopsis enzyme. Here we show that the N-terminus mediates nuclear targeting of both isoforms but also distinguishes them by defining their subnuclear localization (Figure 1). The first NLS shared between isoforms is apparently sufficient to direct MBD4L.3 and MBD4L.4 to the nucleus. The reasons why MBD4L.4 but not MBD4L.3 is targeted to the nucleolus are unknown. Moreover, the requirements for other proteins to locate in this compartment are not yet clear. One possibility is that the presence of an extra predicted NoS in MBD4L.4 generated by exitron splicing could help determine this fate (Scott *et al.*, 2011). Alternatively, differences in the N-terminal protein charge could mediate protein targeting since enrichment in basic amino acids is associated with nucleolar protein localization and RNA binding (Martin *et al*., 2015). Consistently, the N-terminal charge value is higher for MBD4L.4 (16.96 for aa 1-189 at pH=7) than for MBD4L.3 (3.39 for aa 1-306 at pH=7). Another possibility is that splicing of the exitron changes the interaction of MBD4L.3 or MBD4L.4 with proteins that define their destination, as described for other enzymes that reach the nucleoplasm or nucleolus (Vascotto *et al*., 2009). Supporting this, the MBD4L N-terminal region is intrinsically disordered and contains repeats of the phosphorylation motif SPxh that differ in length

and number between isoforms (Manke *et al.*, 2003; Ramiro-Merina *et al.*, 2013). Future studies including mutagenesis of NLSs, NoSs, positive charged amino acids and/or the phosphorylation sites will shed light into these alternatives.

The AS of the MBD4L exitron is altered by high temperature. The relative abundance of *MBD4L.3* and *MBD4L.4* transcripts changes after heat treatment and is recovered after stress release. The reversibility of this process suggests that relocation of the enzyme in the nucleolus could be advantageous in responding to stress. RS31 is a splicing factor that affects the AS of the *MBD4L* exitron. Interestingly, truncated variants of RS31 are generated by AS in response to temperature changes (Palusa *et al.*, 2007). Then, heat stress-mediated *MBD4L* splicing could be sensitive to changes in RS31 variants. On the other hand, we found that NTR1 is necessary for generation of the nucleolar MBD4L isoform after heat shock. This result is somehow surprising since *ntr1-1* plants have broad defects on AS of genes with strong splicing sites (Jones *et al.*, 2012; Dolata *et al.*, 2015), and exitrons are predicted to contain weak splice site signals (Marquez *et al.*, 2015). As NTR1 seems to promote *MBD4L* splicing only under heat stress, it is possible that it can work over weak AS sites aided by other factors under this condition. Alternatively, splicing sites of exitrons may become stronger under heat stress. Another unexpected result was that independently of the promoter that drives the expression of *MBD4L* (native or 35S; Figure 3, Figure 4), heat stress-induced AS of the exitron was similar. This may suggest that this splicing is not occurring co-transcriptionally (Naftelberg *et al.*, 2015). Accordingly, most IR events are post-transcriptionally spliced introns (Jia *et al.*, 2020). However, differences between exitron and classical intron processing deserve to be investigated.

The biological significance of MBD4L targeting the nucleolus or nucleoplasm is unknown. The nucleolus has been proposed as a central hub in DNA damage, but the role of DNA repair enzymes at this region remains puzzling (Iarovaia *et al.*, 2019). BER activity has been mostly associated with the nucleoplasm, whereas rRNA synthesis and metabolism have been related to the nucleolus. Interestingly, many DNA repair proteins are present in the nucleolus. Moreover, some of them exchange between nucleolus and nucleoplasm and have moonlighting functions in RNA metabolism (Iarovaia *et al.*, 2019). SMUG1 (single-strand selective monofunctional uracil DNA glycosylase) is a well-characterized BER enzyme that reaches the nucleolus where it can excise modified bases from RNA contributing to rRNA quality control (Jobert *et al.*, 2013). Similarly, the BER enzyme APE1 (apurinic/apyrimidinic endonuclease 1) is targeted to the nucleolus through its interaction with nucleoplasmin 1 (NPM1), where it can cleave abasic RNA helping to remove oxidized RNA molecules (Vascotto *et al.*, 2009). Thus, it is possible that MBD4L.4 is involved in some RNA metabolic

processes at the nucleolus. FRAP studies showed that MBD4L.4-GFP has lower mobility in the nucleolus than in the nucleoplasm (Figure 2). Probably this protein is retained at that nucleolus by association with nucleolar proteins or RNAs, as described for APE1 and others (Vascotto *et al.*, 2009; Lirussi *et al.*, 2012). The nucleolus could also operate as a site of storing and preserving enzymes until they are needed in the nucleoplasm (Lirussi *et al.*, 2012). In mammals, some BER enzymes re-localize from the nucleolus to the nucleoplasm after DNA damage, apparently increasing the nucleoplasmic repair capacity (Poletto *et al.*, 2014). In addition, the nucleolus can transiently hold epigenetic regulators from the Polycomb group during heat shock, allowing to restore the epigenomic landscape after stress recovery (Azkanaz *et al.*, 2019). Therefore, the causes and consequences of targeting MBD4L.4, but not MBD4L.3, to the nucleolus after heat stress remain to be elucidated.

In summary, we describe how an exitron contributes to increase protein diversity under an adverse environmental condition, showing that the exitron of the MBD4L gene encoding a BER DNA glycosylase is subject to AS under heat stress, this controls the subnuclear enzyme targeting promoting its nucleolar accumulation, where it might have a new yet unknown function.

## EXPERIMENTAL PROCEDURES

### Plants

The Arabidopsis (*Arabidopsis thaliana*) Columbia (Col-0), *atntr1-1* mutant (SALK_073187), and the transgenic plants overexpressing *RS31* (*RS31ox*; *mRNA1*-cDNA version), the GFP-MBD4L.3 fusion or eGFP were previously described (Petrillo *et al.*, 2014; Dolata *et al.*, 2015; Nota *et al.*, 2015). Plants expressing the GFP-tagged MBD4L.4 or the non-spliceable MBD4L.3 version (MBD4L.3$^{exit}$) under 35S promoter were obtained by floral-dip transformation method with *Agrobacterium tumefaciens* GV3101 strain and selected with Basta (10 µg ml$^{-1}$) or kanamycin (10 µg ml$^{-1}$). In assays with transgenic plants homozygous MBD4L.3-GFP (F3 generation), MBD4L.4-GFP (F4) or MBD4L.3$^{exit}$-GFP (F2) lines were used. For studies with seedlings, sterile seeds were stratified for 3 days at 4°C, germinated, and grown in ½ MS (1% sucrose) agar plates in a growth chamber with 10 h light (100-120 µmol sec$^{-1}$ m$^{-2}$) and 14 h dark cycles at 20-22°C (Cecchini *et al.*, 2019; Rizzi *et al.*, 2017). Plates were sealed with paper tape micropore™. Adult plants (4-5-week-old) were grown under the same conditions as the seedlings. *Nicotiana benthamiana* plants were grown at 24°C under 16 h day (120 µmol sec$^{-1}$ m$^{-2}$ light) and 8 h night for 4 weeks and used for *A. tumefaciens*-mediated transient transformation.

**Vectors and constructs**

The primers and vectors used in this study are described in Supplementary Table 1. The full *MBD4L.4* coding region and the N-terminal *MBD4L.3* and *MBD4L.4* coding regions (*MBD4L.3*$^{1\text{-}306}$ and *MBD4L.4*$^{1\text{-}189}$) were amplified from leaf-derived cDNA using specific PCR primers (Table S1), to be introduced into TOPO pENTR® vector and then transferred to the destination vector pK7FWGF2 using GATEWAY® technology (Karimi *et al.*, 2002). The non-spliceable *MBD4L.3* variant (MBD4L.3$^{exit}$) was obtained by synthesis of the N-terminal coding region (1-471 bp, including the exitron sequence) carrying 45 synonymous point mutations (Figure 5a) cloned into a pUC57-Mini plasmid (Genscript, Hong Kong). This plasmid was used to amplify an overlapping fragment to be linked by PCR to the *MBD4L.3* C-terminal coding region (472-1335 bp; cloned from *A. thaliana* cDNA). The final construct was amplified by PCR, cloned into TOPO pENTR® and transferred by GATEWAY® system into the destination vector pGWB5 (Nakagawa *et al.*, 2007). The pK7FWGF2-MBD4L.4 and pGWB5-MBD4L.3$^{exit}$ constructs express the transgenes as C-terminal GFP-fusions controlled by the 35S promoter. All constructs were confirmed by sequencing. The construct used as nucleolus marker (Nuc2-mCherry) was kindly provided by Drs. Fang and Spector (Liu *et al.*, 2014).

**Transformation and GFP imaging**

Leaves from 4-week-old *N. benthamiana* plants were infiltrated with *A. tumefaciens* (GV3101 strain) clones carrying the different constructs. For co-expression analysis cultures containing different clones were mixed (1:1 ratio) and immediately infiltrated. For analysis of Arabidopsis MBD4L.3-GFP, MBD4L.4-GFP or MBD4L.3$^{exit}$-GFP transgenic plants, 12-day-old seedling roots and cotyledons were used. *N. benthamiana* and Arabidopsis tissues were mounted for microscopic analysis as previously described (Cecchini *et al.*, 2015; Cecchini *et al.*, 2019). Arabidopsis roots were stained with DAPI solution (1 µg/ml) for 10-15 minutes before mounting. Olympus FV1000 or FV1200 (Olympus, Latin America) laser-scanning inverted spectral confocal microscopes were used to detect GFP (excitation: 488 nm; emission: 505 to 530 nm), mCherry (excitation: 561 nm; emission: 570 to 620 nm), DAPI (excitation: 405 nm; emission: 420-475 nm) and plastid autofluorescence (excitation: 633, emission: 650-750 nm). Images were taken using a LD C-Apochromat 40x/1.1 W Korr or PLAPON 60x/1.42AN (Olympus, Japan) objectives and a sequential acquisition mode. An Axioplan (Zeiss, Germany) direct microscope with the DAPI, Alexa 488/546, CFP, YFP filter set was used for epifluorescence microscopy. Images were processed using Fiji (https://fiji.sc/), FluoView (Olympus)

and Adobe Photoshop software. All microscopy studies were developed at the Center of Microscopy and Nanoscopy Córdoba (CEMINCO; http://ciquibic.fcq.unc.edu.ar/infraestructura/microscopia/).

**Fluorescence recovery after photobleaching (FRAP)**

*N. benthamiana* leaves expressing MBD4L.3-GFP or MBD4L.4-GFP were used for this study. An Olympus FV1000 laser-scanning confocal microscope was used to photobleach a defined area of the nuclei (2 µm diameter circular ROI) and quantify fluorescence recovery in this area over time. As a control, fluorescence was measured in a background (non-photobleached) region inside the same nucleus. Before and after photobleaching images were acquired every 2-3 sec during a 3-min period with a PLAPON 60x/1.42AN (Olympus, Japan) objective, with the confocal pinhole at 211 µm. Recovery half-time kinetic parameter ($t_{0.5}$) was estimated by calculating the time to recover half of the photobleached fluorescence. The $t_{0.5}$ values were calculated from four independent experiments using background fluorescence values to relativize the data. Fluorescence signal in selected regions of interest was quantified using the Fiji software (https://fiji.sc/).

**Stress and chemical treatments**

Assays testing the effect of stress or chemicals on *MBD4L* transcript levels and/or protein subcellular localization used 12-day old seedlings submerged in liquid ½ MS media in a 24-multiwell plate supplemented with: ½ MS medium (mock), 300 mM mannitol (M4125, Sigma-Aldrich, USA), 5 mM $H_2O_2$ (02-003-191, J.T. Baker™, USA), 100 µM methyl viologen dichloride hydrate (856177, Sigma-Aldrich, St. Louis, MO, USA), 100 µM methyl-jasmonate (392707, Sigma-Aldrich, St. Louis, MO, USA), 500 µM salicylic acid (09045, Research Organics, USA), or *Pseudomonas syringae* pv *tomato* DC3000 suspension (*Pst* DC3000, $OD_{600}$=0.8; (Cambiagno *et al.*, 2018)). Samples were collected after different times post-treatment. To examine heat stress, seedlings were incubated at 37°C or 22°C (control) in plates sealed with cling film to prevent dehydration and darkness. After 6.5 h at 37°C, seedlings were returned to 22°C and normal grow conditions for 17.5 hours to analyze stress recovery. Adult Arabidopsis plants grown in soil and exposed to 37°C in darkness were used to analyze heat-induced changes in the content of MBD4L isoforms by Western blot.

**mRNA and Western blot analysis**

Transcript levels were analyzed by reverse transcription (RT) followed by semi-quantitative PCR (RT-sqPCR) or quantitative PCR (RT-qPCR) performing RNA extraction (Rizzi *et al.*, 2017) and cDNA synthesis as previously described (Cambiagno *et al.*, 2018). *UBQ5* (Ubiquitin 5; *At3g62250*) or *EF1α* (elongation factor 1 alpha, *At5g60390*) were used as internal standards. Oligonucleotide

sequences used as primers are showed in Table S1. The splicing index (SI) was defined as the ratio between the longest and shorter splicing products. The sqPCR set up was as follows: 95°C for 3 min and the respective number of cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s to 1.5 min. Gel-Pro Analyzer™ software was used for densitometric quantification of bands. For qPCR we used 1:5 cDNA dilutions in 15 µL reactions of Luna Universal Probe qPCR Master Mix (New England Biolabs) plus EvaGreen® (Biotium) on a CFX96 Touch™ real time PCR System (Bio-Rad) and the following set up: 95°C for 1 min and 45 cycles at 95°C for 15 s and 60°C for 30 s, and 1 cycle of dissociation from 65°C to 95°C with 1°C temperature increase for 5s. The raw data obtained with CFX Manager 3.1 Software (Bio-Rad) was baseline corrected and the window of linearity was determined using LinRegPCR 2021.1 (Ruijter *et al.*, 2009). The starting concentrations (No) of cDNA of each gene were used to obtain the SI values (*MBD4L.4* No / *MBD4L.3* No), or determine the *MBD4L.3* or *MBD4L.4* expression levels (*MBD4L.3* No / *EF1α* No or *MBD4L.4* No / *EF1α* No, respectively). Expression values were log2-transformed before statistical analysis. Plant nuclei enriched fractions were obtained as previously described (Weigel and Glazebrook, 2002) with modifications. Briefly, leaves (500 mg) were homogenized in 2 volumes of extraction buffer (20mM Tris HCl, pH 7.4; KCl 20 mM; 2mM EDTA, 1mM $MgCl_2$; 250 mM sucrose; 25% glycerol; 1 mM DTT and 0.5x complete protease inhibitor cocktail (Sigma-Aldrich P2714)) and double Miracloth® -filtered. Triton X-100 was added to the filtrate to a final concentration of 0.5% and pelleted (2,000 g), washed with the above buffer plus 0.1% Triton X-100 and centrifuged again. Protein concentration was determined by Bradford assay. Identical sets of samples were separated by 12 % (anti-GFP) or 15 % (anti-histone3) SDS-PAGE and transferred to nitrocellulose. Western blots were developed either with anti-GFP (#10744900, Roche, 1:3000) or anti-histone H3 (AS10 710, Agrisera, 1:3000) antibodies. Infrared fluorescent secondary antibodies (926-32211 and 925-32210, LI-COR, 1:25000) were used to detect the bands (800 ηm channel) in an Odyssey Infrared scanner (LI-COR Bioscience).

**Analysis of amino acid sequences**

The Nucleolar localization sequence detector software was used to detect predicted nucleolar localization sequences ((Scott *et al.*, 2011); http://www.compbio.dundee.ac.uk/www-nod/index.jsp). ClustalW and SnapGene® 5.0 (from Insightful Science; snapgene.com) software were used to perform protein sequence alignments (https://www.snapgene.com/).

**Statistical analysis**

Analyses were done with the software SigmaPlot v11.0 (Systat Software, Inc.) and Infostat v2018 (www.infostat.com.ar). The normality and homogeneity of variance were tested by Shapiro-Wilk's and Levene's, respectively. One-tailed Student's *t*-test and analysis of variance (ANOVA) followed by the Duncan post hoc test were used as indicated in Figure legends.



## DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

## ACKNOWLEDGMENTS

This research was supported by Agencia Nacional de Promoción Científica y Tecnológica (PICTs 2014-3255, 2016-2986 and 2018-4588 to MEA; 2017-0589 to NMC) and by Secretaría de Ciencia y Tecnología, Universidad Nacional de Córdoba to MEA. We thank Dr. Ezequiel Petrillo for helpful discussions and for providing us with *atntr1-1* and RS31ox Arabidopsis seeds. We also thank Drs. Yuda Fang and David L. Spector for the kind gift of the nucleolar marker construct Nuc2-mCherry, and to Dr. Carlos Mas and Dr. Cecilia Sampedro (Centro de Micro y Nanoscopia de Córdoba -CEMINCO-) for their assistance with confocal microscopy and FRAP analysis. NMC, CILL and MEA are members of the research career of CONICET. JT is, and FN was, a CONICET fellow.

## AUTHOR CONTRIBUTIONS

NMC and MEA conceived the study and designed the experiments. FN contributed to the initial stage of research. NMC, JT, ILL, SC and FN performed the experiments. NMC, MEA, JT and ILL analyzed and interpreted data. NMC and MEA wrote the paper.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interests.

## SHORT LEGENDS FOR SUPPORTING INFORMATION

**Figure S1**. RT-sqPCR evaluating *MBD4L* splicing in *N. benthamiana*.

**Figure S2**. Prediction of nucleolar localization sequences (NoS) in MBD4L splicing variants.

**Figure S3**. AS of *MBD4L* exitron in response to stress and hormone treatments.

**Figure S4**. Subnuclear localization of MBD4L.3-GFP in response to stress and hormone treatments.

**Table S1**. Vectors, constructs, and primers used in this study.



## REFERENCES

**Azkanaz, M., López, A.R., Boer, B. De, et al.** (2019) Protein quality control in the nucleolus safeguards recovery of epigenetic regulators after heat shock. *Elife*, **8**, 1–27.

**Boldinova, E.O., Khairullin, R.F., Makarova, A. V. and Zharkov, D.O.** (2019) Isoforms of base excision repair enzymes produced by alternative splicing. *Int. J. Mol. Sci.*, **20**, 3279.

**Bramson, J., O'Connor, T. and Panasci, L.** (1995) Effect of alkyl-N-purine DNA glycosylase overexpression on cellular resistance to bifunctional alkylating agents. *Biochem. Pharmacol.*, **50**, 39–44.

**Cambiagno, D.A., Nota, F., Zavallo, D., Rius, S., Casati, P., Asurmendi, S. and Alvarez, M.E.** (2018) Immune receptor genes and pericentromeric transposons as targets of common epigenetic regulatory elements. *Plant J.*, **96**, 1178–1190.

**Cecchini, N.M., Steffes, K., Schläppi, M.R., Gifford, A.N. and Greenberg, J.T.** (2015) Arabidopsis AZI1 family proteins mediate signal mobilization for systemic defence priming. *Nat. Commun.*, **6**, 7658.

**Cecchini, N.M., Roychoudhry, S., Speed, D.J., et al.** (2019) Underground azelaic acid-conferred resistance to Pseudomonas syringae in Arabidopsis. *Mol. Plant. Microbe. Interact.*, **32**, 86–94.

**Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A.S.N., Byrne, L.J., Wilson, C.M. and Syed, N.H.** (2019) Alternative splicing and protein diversity: Plants versus animals. *Front. Plant Sci.*, **10**, 1–14.

**Chen, D. and Huang, S.** (2001) Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J. Cell Biol.*, **153**, 169–176.

**Chen, H., Chu, P., Zhou, Y., et al.** (2012) Overexpression of AtOGG1, a DNA glycosylase/AP lyase, enhances seed longevity and abiotic stress tolerance in Arabidopsis. *J. Exp. Bot.*, **63**, 4107–21.

**Cheng, Q., Xiao, H. and Xiong, Q.** (2020) Conserved exitrons of FLAGELLIN-SENSING 2 (FLS2) across dicot plants and their functions. *Plant Sci.*, **296**, 110507.

**Dolata, J., Guo, Y., Kołowerzo, A., Smoliński, D., Brzyżek, G., Jarmołowski, A. and Świeżewski,**

**S.** (2015) NTR 1 is required for transcription elongation checkpoints at alternative exons in Arabidopsis . *EMBO J.*, **34**, 544–558.

**Frosina, G.** (2000) Overexpression of enzymes that repair endogenous damage to DNA. *Eur. J. Biochem.*, **267**, 2135–2149.

**Furihata, C.** (2015) An active alternative splicing isoform of human mitochondrial 8-oxoguanine DNA glycosylase (OGG1). *Genes Environ. Off. J. Japanese Environ. Mutagen Soc.*, **37**, 21.

**Gelli, E., Colombo, M., Pinto, A.M., et al.** (2019) Usefulness and limitations of comprehensive characterization of mRNA splicing profiles in the definition of the clinical relevance of BRCA1/2 variants of uncertain significance. *Cancers (Basel).*, **11**, 295.

**Ghelli, R., Brunetti, P., Napoli, N., et al.** (2018) A newly identified flower-specific splice variant of AUXIN RESPONSE FACTOR8 regulates stamen elongation and endothecium lignification in arabidopsis. *Plant Cell*, **30**, 620–637.

**Iarovaia, O. V., Minina, E.P., Sheval, E. V., Onichtchouk, D., Dokudovskaya, S., Razin, S. V. and Vassetzky, Y.S.** (2019) Nucleolus: a central hub for nuclear functions. *Trends Cell Biol.*, **29**, 647–659.

**Jacobs, A.L. and Schär, P.** (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1–20.

**Jia, J., Long, Y., Zhang, H., et al.** (2020) Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat. Plants*, **6**, 780–788.

**Jian, X., Boerwinkle, E. and Liu, X.** (2014) In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet. Med.*, **16**, 497–503.

**Jobert, L., Skjeldam, H.K., Dalhus, B., Galashevskaya, A., Vågbø, C.B., Bjørås, M. and Nilsen, H.** (2013) The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. *Mol. Cell*, **49**, 339–345.

**Jones, M.A., Williams, B.A., McNicol, J., Simpson, C.G., Brown, J.W.S. and Harmer, S.L.** (2012) Mutation of Arabidopsis SPLICEOSOMAL TIMEKEEPER LOCUS1 causes circadian clock defects. *Plant Cell*, **24**, 4066–4082.

**Karimi, M., Inzé, D. and Depicker, A.** (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.*, **7**, 193–5.

**Lirussi, L., Antoniali, G., Vascotto, C., et al.** (2012) Nucleolar accumulation of APE1 depends on charged lysine residues that undergo acetylation upon genotoxic stress and modulate its BER activity in cells. *Mol. Biol. Cell*, **23**, 4079–4096.

**Liu, Y., Liu, Q., Yan, Q., Shi, L. and Fang, Y.** (2014) Nucleolus-tethering system (NoTS) reveals that assembly of photobodies follows a self-organization model. *Mol. Biol. Cell*, **25**, 1366–73.

**Manke, I.A., Lowery, D.M., Nguyen, A. and Yaffe, M.B.** (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*, **302**, 636–639.

**Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A. and Kalyna, M.** (2015) Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res.*, **25**, 995–1007.

**Musinova, Y.R., Kananykhina, E.Y., Potashnikova, D.M., Lisitsyna, O.M. and Sheval, E. V.** (2015) A charge-dependent mechanism is responsible for the dynamic accumulation of proteins inside nucleoli. *Biochim. Biophys. Acta - Mol. Cell Res.*, **1853**, 101–110.

**Naftelberg, S., Schor, I.E., Ast, G. and Kornblihtt, A.R.** (2015) Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.*, **84**, 165–198.

**Nakagawa, T., Kurose, T., Hino, T., et al.** (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.*, **104**, 34–41.

**Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F. and Krokan, H.E.** (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res.*, **25**, 750–755.

**Nimeth, B.A., Riegler, S. and Kalyna, M.** (2020) Alternative splicing and DNA damage response in plants. *Front. Plant Sci.*, **11**, 91.

**Nishioka, K., Ohtsubo, T., Oda, H., Fujiwara, T., Kang, D., Sugimachi, K. and Nakabeppu, Y.** (1999) Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternatively spliced OGG1 mRNAs. *Mol. Biol. Cell*, **10**, 1637–1652.

**Nota, F., Cambiagno, D.A., Ribone, P. and Alvarez, M.E.** (2015) Expression and function of AtMBD4L, the single gene encoding the nuclear DNA glycosylase MBD4L in Arabidopsis. *Plant Sci.*, **235**, 122–129.

**Ohtsubo, T.** (2000) Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria. *Nucleic Acids Res.*, **28**, 1355–1364.

**Owen, R.M., Baker, R.D., Bader, S., Dunlop, M.G. and Nicholl, I.D.** (2007) The identification of a novel alternatively spliced form of the MBD4 DNA glycosylase. *Oncol. Rep.*, **17**, 111–116.

**Palusa, S.G., Ali, G.S. and Reddy, A.S.N.** (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.*, **49**, 1091–1107.

**Petrillo, E., Godoy Herz, M.A., Fuchs, A., et al.** (2014) A chloroplast retrograde signal regulates nuclear alternative splicing. *Science*, **344**, 427–430.

**Poletto, M., Lirussi, L., Wilson, D.M. and Tell, G.** (2014) Nucleophosmin modulates stability, activity, and nucleolar accumulation of base excision repair proteins. *Mol. Biol. Cell*, **25**, 1641–1652.

**Ramiro-Merina, Á., Ariza, R.R. and Roldán-Arjona, T.** (2013) Molecular characterization of a putative plant homolog of MBD4 DNA glycosylase. *DNA Repair (Amst).*, **12**, 890–898.

**Reddy, A.S.N.** (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.*, **58**, 267–94.

**Rizzi, Y.S., Cecchini, N.M., Fabro, G. and Alvarez, M.E.** (2017) Differential control and function of Arabidopsis ProDH1 and ProDH2 genes on infection with biotrophic and necrotrophic pathogens. *Mol. Plant Pathol.*, **18**, 1164–1174.

**Roldán-Arjona, T., Ariza, R.R. and Córdoba-Cañero, D.** (2019) DNA base excision repair in plants: an unfolding story with familiar and novel characters. *Front. Plant Sci.*, **10**, 1–18.

**Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., hoff, M.J.B. van den and Moorman, A.F.M.** (2009) Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.*, **37**.

**Scott, M.S., Troshin, P. V. and Barton, G.J.** (2011) NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics*, **12**, 317.

**Staiger, D. and Simpson, G.G.** (2015) Enter exitrons. *Genome Biol.*, **16**, 136.

**Syed, N.H., Kalyna, M., Marquez, Y., Barta, A. and Brown, J.W.S.** (2012) Alternative splicing in plants--coming of age. *Trends Plant Sci.*, **17**, 616–23.

**Takao, M., Zhang, Q.M., Yonei, S. and Yasui, A.** (1999) Differential subcellular localization of human MutY homolog (hMYH) and the functional activity of adenine:8-oxoguanine DNA glycosylase. *Nucleic Acids Res.*, **27**, 3638–44.

**Vascotto, C., Fantini, D., Romanello, M., et al.** (2009) APE1/Ref-1 Interacts with NPM1 within Nucleoli and Plays a Role in the rRNA Quality Control Process. *Mol. Cell. Biol.*, **29**, 1834–1854.

**Weigel, D. and Glazebrook, J.** (2002) Arabidopsis: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

## FIGURE LEGENDS

**Figure 1.** Subnuclear localization of MBD4L splicing isoforms in *N. benthamiana* and Arabidopsis. *35S:MBD4L.3-GFP* and *35S:MBD4L.4-GFP* constructs were used for transient transformation of *N.*

*benthamiana* **(a,b)** and stable transformation of Arabidopsis **(c,d)** via *Agrobacterium*. **(a)** Laser scanning confocal microscopy (confocal) micrographs showing localization of control GFP (eGFP), MBD4L.3-GFP and MBD4L.4-GFP. The percentage of nuclei showing different localization patterns for MBD4L.3-GFP (n=88) and MBD4L.4-GFP (n=69) is indicated above each representative micrograph. **(b)** Confocal micrographs showing localization of co-expressed MBD4L.3-GFP or MBD4L.4-GFP and the nucleolus marker Nuc2-mCherry. Right: fluorescence profiles for green and red channels along the yellow lines shown in the merge micrographs. a.u.: arbitrary units. **(c)** Confocal micrographs showing localization of MBD4L.3-GFP and MBD4L.4-GFP in DAPI stained nuclei from root elongation zones and **(d)** in occlusive stomata cells of 12-day-old WT-Col-0 seedlings. White arrowheads: nucleolus-like structure; yellow arrowheads: nucleolus. In **(a)**, **(b)** and **(d)** micrographs show GFP (green), mCherry (red) and plastid autofluorescence (blue). In **(c)** micrographs show GFP (green) and DAPI (blue). Scale bar = 5 μm in **(a)**, **(b)** and **(d)** and 10 μm in **(c).**

**Figure 2.** Motifs affecting the subnuclear localization of MBD4L.3 and MBD4L.4 and differences in their mobility. **(a)** Scheme of the MBD4L.3 and MBD4L.4 variants showing the exitron, predicted nucleolar and nuclear localization sequences (NoS, NLS), and DNA glycosylase domain (DNA Glyc.). Amino acid positions delimiting different regions/domains are shown in the upper part. **(b)** Confocal micrographs showing localization of MBD4L.3-GFP, MBD4L.4-GFP and deleted fusions proteins MBD4L.3$^{1\text{-}306}$-GFP and MBD4L.4$^{1\text{-}189}$-GFP lacking the DNA glycosylase domain expressed in *N. benthamiana*. White arrowheads: nucleolus-like structure. Micrographs show GFP (green) and plastid autofluorescence (blue). **(c)** Representative confocal micrographs illustrating the photobleaching fluorescence recovery of MBD4L.3-GFP and MBD4L.3-GFP expressed in *N. benthamiana*. White dotted circle: photobleached area; yellow dotted circle: nucleolus. Right: quantification of the fluorescence recovery half-life time ($t_{0.5}$) for MBD4L.3-GFP in nucleoplasm (L3) and for MBD4L.4-GFP in nucleoplasm (L4_Np) or nucleolus (L4_Nc). The average +/- standard error from four biological replicates is shown. Asterisk indicates statistically significant differences between L3 and L4_Np or L4_Nc as determined by using *t*-test (*P<0.05, n = 4). Scale bar = 5 μm in **(b,c)**.

**Figure 3.** AS of the *MBD4L* exitron is sensitive to high temperature and requires known splicing components. **(a,b)** Splicing Index obtained after quantifying *MBD4L.3* and *MBD4L.4* transcripts by RT–qPCR in 12-day-old WT Col-0 Arabidopsis seedlings maintained at 22°C (control) or 37°C for the indicated times **(a)**, or treated for 6.5 h at 37°C and recovered at 22°C for 17.5 h **(b)**. Values represent average +/- standard error of three independent experiments. Individual data points are presented as scatter-dots. **(c)** RT–sqPCR showing *MBD4L.3 (L.3)* and *MBD4L.4 (L.4)* transcripts in WT-Col-0, *ntr1-*

*1* and RS31 overexpressing (*RS31ox*) seedlings maintained at 22ºC or 37°C for 6.5 h. *UBQ5* was used as reference gene. Bottom panel: SI values shown as average +/- standard deviation of four (Col-0) or three (*ntr1-1*, *RS31ox*) independent experiments. Data from each experiment are shown in the same color as scatter-dots. Asterisks indicate significant differences between treatments in **(a)**, with respect to 0 hours in **(b)** or to Col-0 22 ºC in **(c)** (ANOVA, Duncan's Test, *p<0.1; **p<0.05; ***p<0.001).

**Figure 4.** Heat stress alters the AS of the MBD4L-GFP transgene and the subnuclear localization of the fusion proteins. **(a)** Epifluorescence microscopy micrographs showing localization of eGFP (control) and MBD4L.3-GFP in nuclei from the root elongation zone of 12-day-old transgenic Arabidopsis seedlings after 6.5 h incubation at 22°C (control) or 37°C. White arrowheads: nucleolus-like structure. Micrographs show GFP (green). Right: percentage of nuclei showing nucleoplasmic or nucleolar GFP fluorescence enrichment at each condition. The average +/- standard error from two independent experiments (at least six plants each) is shown. Asterisk indicates statistically significant differences according to *t*-test (*P<0.01, n=25; **P<0.001, n=30). **(b)** RT–sqPCR showing *MBD4L.3-GFP* (*L.3-GFP*) and *MBD4L.4-GFP* (*L.4-GFP*) transcript levels in 12-day-old transgenic Arabidopsis seedlings after different incubation times at 37°C. **(c)** *MBD4L-GFP* transcript levels in 12-day-old transgenic seedlings incubated at 37°C for 6.5 h and recovered for 17.5 h at 22°C. In **(b)** and **(c)** *UBQ5* was used as internal standard. The bottom panels shown in **(b,c)** are the SI values obtained by band densitometry analysis, expressed as average +/- standard error from four independent experiments, with data from each experiment shown as scatter-dots of the same color. Asterisks show significant differences with respect to 0 hours **(b,c)** (ANOVA, Duncan's Test, **p<0.05). **(d)** Western blot revealed with anti-GFP antibodies showing MBD4L-GFP isoforms in nuclear extracts from 4-weeks old transgenic Arabidopsis plants expressing *MBD4L.3-GFP* exposed to 37°C. Identical samples were run in parallel and developed with anti-H3 antibodies for loading control. Similar results were observed in two independent experiments. Scale bar = 10 µm in **(a)**.

**Figure 5.** Expression and subnuclear localization of a non-spliceable *MBD4L.3* version in response to high temperature. **(a)** ClustalW alignment between the N-terminal region of *MBD4L* and a version carrying synonymous mutations in most of the possible exitron donor and acceptor splicing sites (*MBD4L.3*$^{exit}$). Black bar: *MBD4L* exitron. Gray boxes: synonymous mutations. Red boxes: exitron donor and acceptor splicing sites. The amino acid sequence is shown under the alignment. **(b)** Laser scanning confocal microscopy (confocal) micrographs showing localization of MBD4L.3$^{exit}$-GFP in *N. benthamiana* epidermal cells transformed with *A. tumefaciens*. Micrographs show GFP (green) and

plastid autofluorescence (blue). **(c)** RT–sqPCR showing *MBD4L.3-GFP* and *MBD4L.4-GFP* transcripts in 12-day-old transgenic Arabidopsis WT-Col-0 seedlings expressing *MBD4L.3-GFP* (*L.3-GFP*) or *MBD4L.3$^{exit}$-GFP* (*L.3$^{exit}$-GFP*) constructs after 6.5 h incubation at 22ºC or 37°C. *UBQ5* transcript was used as internal standard. Similar results were observed in three independent experiments. **(d)** Epifluorescence microscopy micrographs showing localization of MBD4L.3$^{exit}$-GFP in nuclei from the root elongation zone of 12-day-old Arabidopsis seedlings after 6.5 h incubation at 22 °C or 37°C. White arrowheads: nucleolus-like structure. Micrographs show GFP (green). Bar = 5 µm. Right: percentage of nuclei showing nucleoplasmic or nucleolar GFP fluorescence enrichment in the same conditions. The average +/- standard error from three independent experiments (at least six seedlings each one) is shown. Scale bar = 5 µm in **(b,d)**.

**Figure S1.** End-point RT–sqPCR showing *MBD4L* splicing variants (*MBD4L.3-GFP* and *MBD4L.4-GFP)* in *N. benthamiana* transformed via *A. tumefaciens* with MBD4L.3-GFP or MBD4L.4-GFP constructs. Two independent agroinfiltration assays are shown (rep1 and rep2).

**Figure S2.** Prediction of nucleolar localization sequences (NoS; red letters) in the N-terminal region of MBD4L.3 **(a)** and MBD4L.4 **(b)** according to the Nucleolar localization sequence detector software (NoD; http://www.compbio.dundee.ac.uk/www-nod/index.jsp). The bar above the sequences indicates the MBD4L glycosylase domain.

**Figure S3.** Splicing Index obtained after quantifying *MBD4L.3* and *MBD4L.4* transcripts by RT-qPCR in 12-day-old WT Col-0 Arabidopsis seedlings at 6.5 hours post treatment with mock, hydrogen peroxide ($H_2O_2$), methyl viologen (MV), mannitol, salicylic acid (SA), methyl jasmonate (JA) and *Pseudomonas syringae* pv. *tomato* (*Pst* DC3000; OD=0.8). Values represent average +/- standard deviation of four biological replicates. Individual data points are shown as scatter-dots. No significant differences between treatments and mock samples were observed (ANOVA, Duncan's Test).

**Figure S4.** Subnuclear localization of MBD4L.3-GFP at different conditions. Epifluorescence microscopy micrographs show the localization of MBD4L.3-GFP in nuclei from the root elongation zone of 12-day-old Arabidopsis seedlings at 6.5 hours post treatment with mock, mannitol, hydrogen peroxide, methyl viologen, methyl jasmonate, salicylic acid, and *Pseudomonas syringae* pv. *tomato*

(*Pst* DC3000; OD=0.8). White arrowhead: nucleolus like structure. Micrographs show GFP (green). Scale bar = 5 μm.

**Figure 1.**

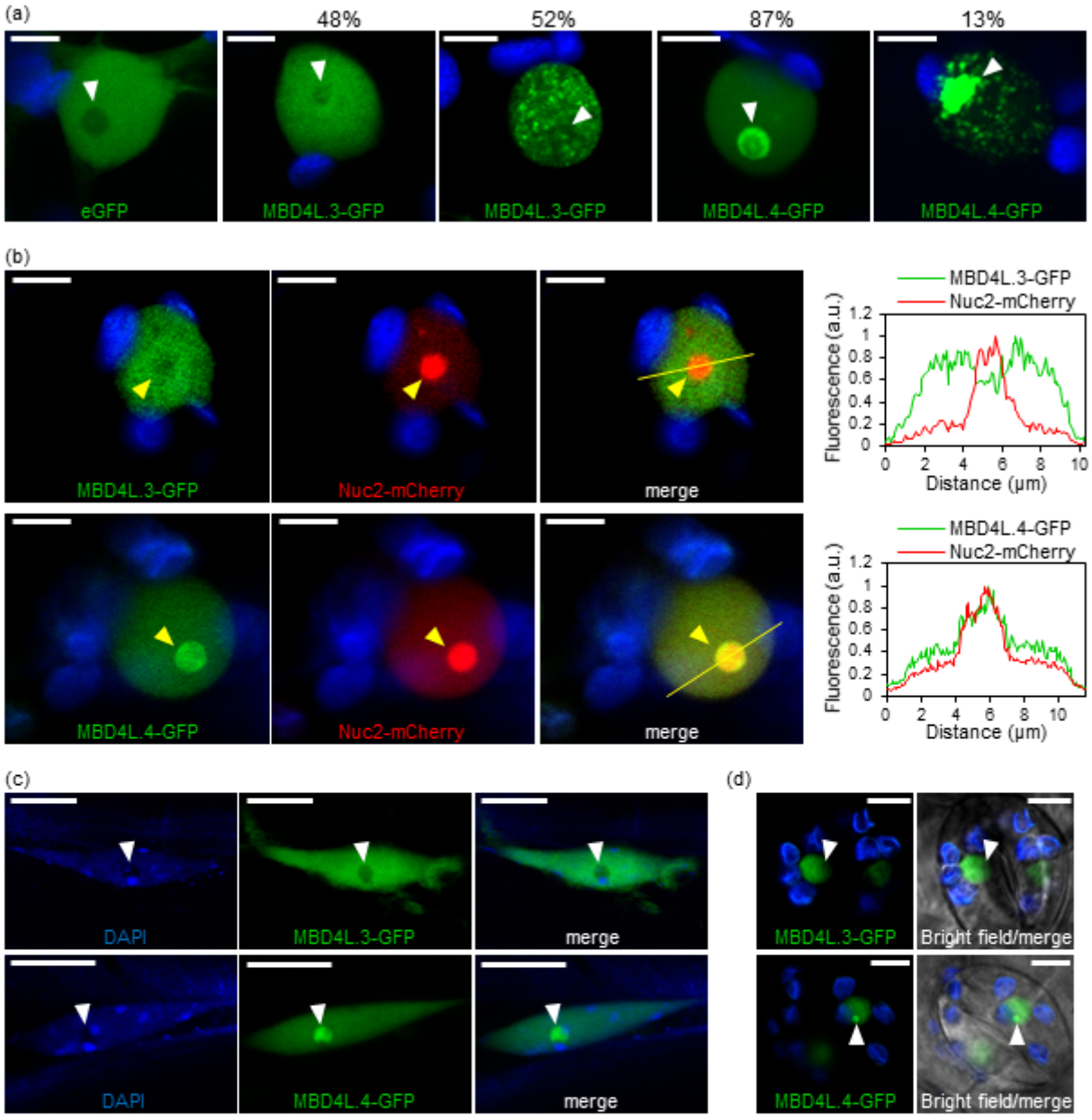

Figure 2.

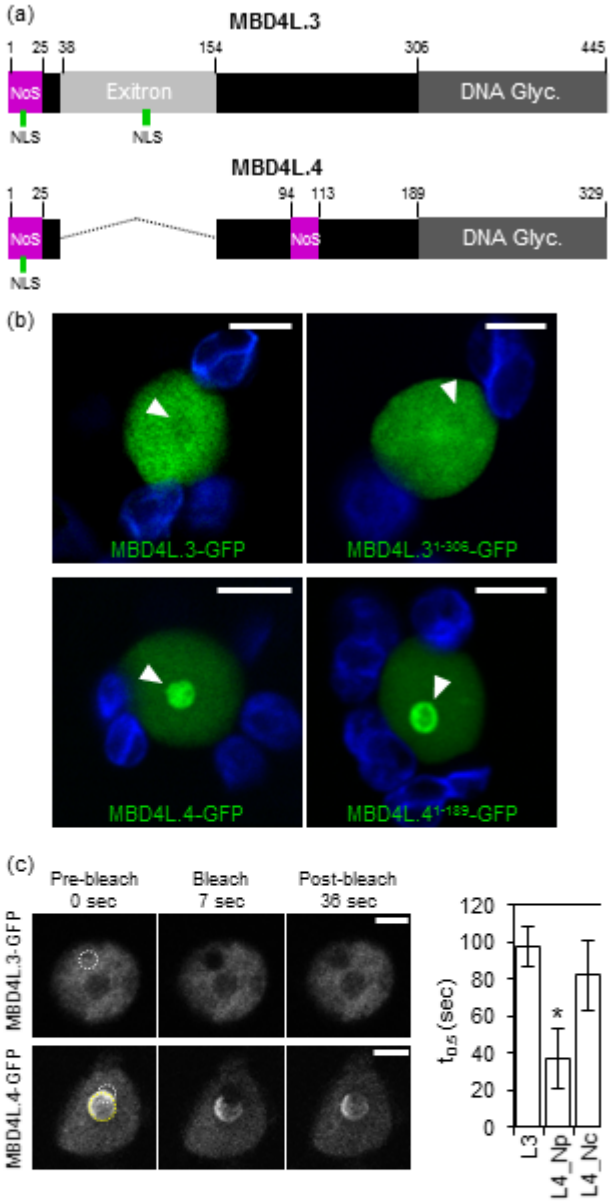

**Figure 3.**

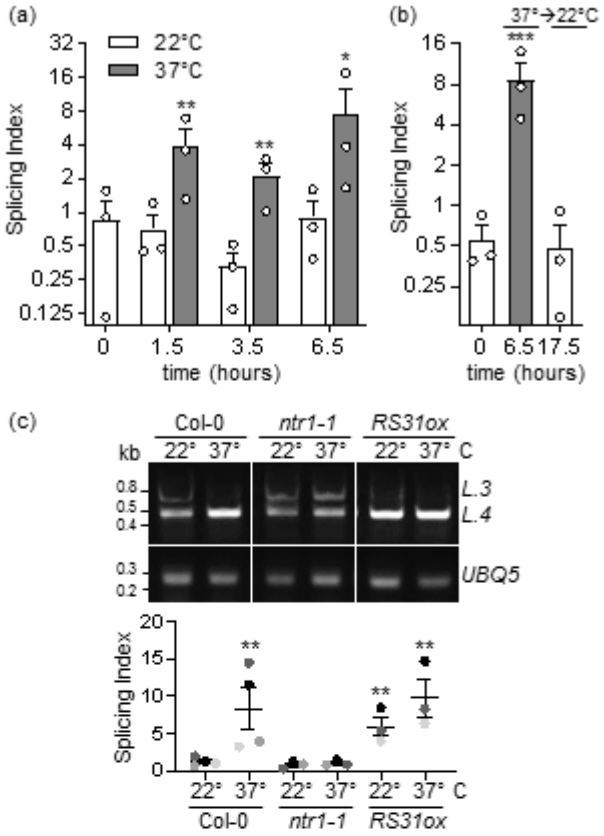

**Figure 4.**

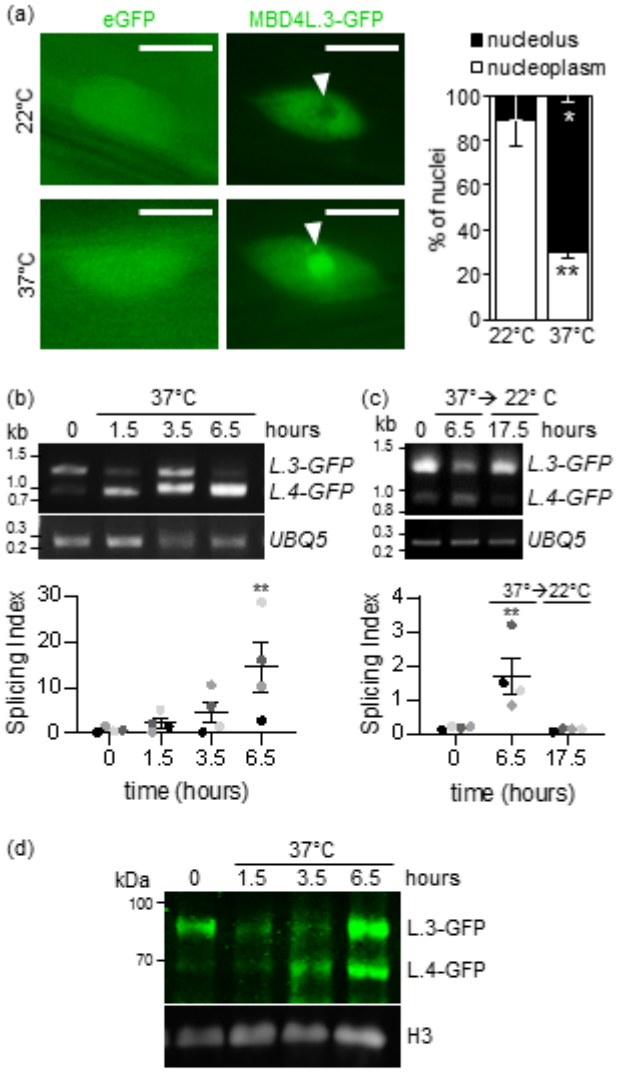

Figure 5.