PCP
PLANT & CELL PHYSIOLOGY

Searching for a Match: Structure, Function and Application of Sequence-Specific RNA-Binding Proteins

Lauren K. Dedow and Julia Bailey-Serres*

University of California-Riverside, Riverside, CA 92521, USA
*Corresponding author: E-mail, julia.bailey@ucr.edu; Fax, (951)827-4437.
(Received January 21, 2019; Accepted April 11, 2019)

Plants encode over 1800 RNA-binding proteins (RBPs) that modulate a myriad of steps in gene regulation from chromatin organization to translation, yet only a small number of these proteins and their target transcripts have been functionally characterized. Two classes of eukaryotic RBPs, pentatricopeptide repeat (PPR) and pumilio/*fem-3* binding factors (PUF), recognize and bind to specific sequential RNA sequences through protein–RNA interactions. These modular proteins possess helical structural units containing key residues with high affinity for specific nucleotides, whose sequential order determines binding to a specific target RNA sequence. PPR proteins are nucleus-encoded, but largely regulate post-transcriptional gene regulation within plastids and mitochondria, including splicing, translation and RNA editing. Plant PUFs are involved in gene regulatory processes within the cell nucleus and cytoplasm. The modular structures of PPRs and PUFs that determine sequence specificity has facilitated identification of their RNA targets and biological functions. The protein-based RNA-targeting of PPRs and PUFs contrasts to the prokaryotic cluster regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas) that target RNAs in prokaryotes. Together the PPR, PUF and CRISPR-Cas systems provide varied opportunities for RNA-targeted engineering applications.

Keywords: CRISPR-Cas • Pentatricopeptide repeat protein • Pumilio/fem-3 binding factor • RNA binding protein.

Introduction

Eukaryotic RNA-binding proteins (RBPs) participate in a diverse set of processes including: epigenetic regulation, rRNA and protein coding transcript biogenesis, pre-mRNA capping, canonical and alternative splicing, 3′ cleavage and polyadenylation, nuclear export, cytoplasmic localization and trafficking, stabilization, degradation, storage and translation (Glisovic et al. 2008, Bailey-Serres et al. 2009, Lee and Kang 2016, Dykes and Emanueli 2017, Chantarachot and Bailey-Serres 2018). RBPs also function in the biogenesis of regulatory RNAs including microRNAs (miRNA) and short-interfering RNAs involved in transcriptional and post-transcriptional gene silencing (Borges and Martienssen 2015, Gorski et al. 2017). In plastids and mitochondria, nucleus-encoded RBPs guide synthesis, splicing, processing, C-to-U editing, turnover and translation of transcripts (Jacobs and Kück 2011).

The binding of an RBP to an individual transcript is determined by both the physical structure and specific amino acids of the protein, as well as, the primary, secondary and tertiary structure, and RNA modifications, such as N^6-methyladenosine, of the RNA target (Zhao et al. 2017). Several highly conserved protein domains, called RNA-binding domains (RBDs), determine the binding specificity of RNA interaction in eukaryotes. These include the RNA recognition motif (RRM), K-homology (KH) and zinc-finger domains (ZnF) (Glisovic et al. 2008). In addition, recent studies have identified novel RBDs (Hentze et al. 2018, Köster and Meyer 2018). Based on in vitro binding assays, RBPs typically bind short consensus motifs of single-stranded RNA, often with flexibility (Ray et al. 2013). Although RBP–RNA interactions have been determined in vivo and in vitro, few rules exist that describe and predict sequence-specific interactions, except for two families of eukaryotic RBPs, the pentatricopeptide repeat (PPR) and pumilio/*fem-3* binding factor (PUF) proteins. PPRs and PUFs can be found in the genomes of all sequenced eukaryotes, although the majority of PPRs are found in plants and algae. Plants encode for more PUFs than animals, but the number of PUF proteins per plant is far fewer than PPRs: *Arabidopsis thaliana* has 26 PUFs compared to almost 500 PPRs (**Table 1**). Both families of RBPs bind specific sequences within transcripts in a manner defined by the structure and regional amino acid sequences of the protein (Barkan et al. 2012, Yagi et al. 2013, Tanaka Hall 2014).

There is significant understanding of the rules governing the sequence-specific binding of PPRs to RNA, their targets and functional significance in the organelles of plants (reviewed by Barkan and Small 2014). Far less is known about the targets of the nucleocytoplasmic plant PUFs, their physical nature and consequences of PUF–RNA interactions. Here, we review these two classes of sequence-specific RBPs and contrast them to the sub-group of prokaryotic cluster regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated proteins (Cas) proteins that target RNA; comparing the protein structure determinants for RNA recognition, relationship with post-transcriptional regulation, and their potential uses as molecular tools for functional manipulation of RNA dynamics and phenotypes.

Plant Cell Physiol. 60(9): 1927–1938 (2019) doi:10.1093/pcp/pcz072, Advance Access publication on 7 May 2019,
available online at https://academic.oup.com/pcp

Table 1 Structural features and technological applications of PPRs and PUFs

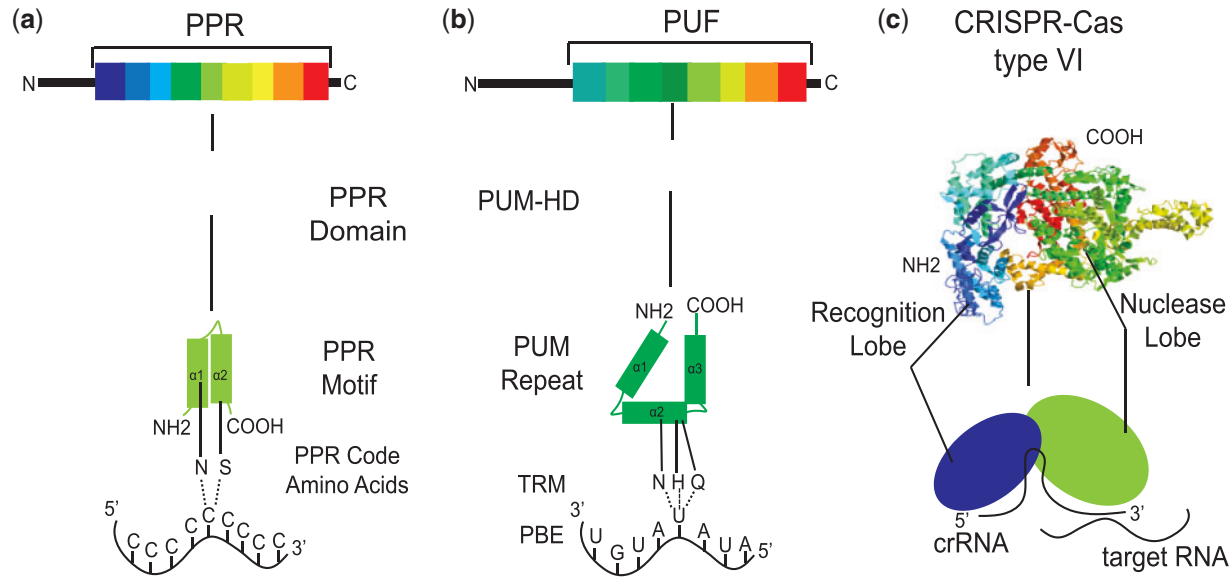
	Pentatricopeptide repeat	Pumilio/FBF
Abbreviations	PPR	PUF
Discovery	2000 (*Arabidopsis thaliana*)	1960s/1997 (*Drosophila melanogaster*)
Conservation	Eukaryotic (some prokaryotic)	eukaryotic
Classes	P & PLS	Group 1-5
RNA-binding domain	PPR domain	PUM-HD
Domain size	478–1472 aa	345-1200 aa
Domain structure	α-Solenoid	Right-handed superhelix
3D shape	Rectangular Spiral Hand-shape V-shape Cylinder	 Crescent C-shape L-shape
Repeat types	P, L and S motifs	PUM repeat
Number per domain	2–30	6–11
Size	31–36 aa	36 aa
Structure	2 α-helix hairpin	3 α-helix triangular array
Key residue motif	PPR code amino acids	Triparite Repeat Motif (TRM)
Number per motif	2	3
Residue location per motif	5, 35 (6, 1′)	12, 13, 16
RNA target motif	PPR domain	PUF binding element (PBE)
Length	2–30 nt	6–11 nt
Repeat:nucleotide ratio	1:1	1:1
Orientation with domain	Parallel	Antiparallel
Conserved sequence		5′-UGUAUAUA-3′
Localization	Organellar Nuclear	Cytoplasmic Nuclear
Mechanism	RNA editing RNA splicing RNA stability Translation regulation tRNA maturation	 RNA stability rRNA processing
Arabidopsis thaliana family members	496	26
Examples	THA8L CLB19 RARE1	APUM5 APUM23 APUM24

Protein Structure and Sequence-Targeted RNA Binding by PPRs and PUFs

PPR proteins are structurally diverse and employ a binary code to target diverse RNA sequences

PPRs were first systematically classified in *A. thaliana* (*At*) (Aubourg et al. 2000, Small and Peeters 2000). These RBPs are characterized by a series of centrally localized, tandem PPR motifs and are subdivided into two major classes, the P-class and PLS-class. *Arabidopsis thaliana* (*Columbia-0*), encodes 283P-class and 213 PLS-class PPRs (Cheng et al. 2016). The P-class PPRs contain, on average, nine to ten P-type motifs, each 35 amino acids long and generally ending in a proline. The PLS-class PPRs have repeats that alternate between the P-type and either long (L, 35–36 amino acid) or short (S, 31 amino acid)-type motifs. PLS-class PPRs can be further categorized based on variable domains found downstream of the PPR motifs: E1, E2, E+ and DYW (aspartic acid-tyrosine-tryptophan), which can be present singularly or in combination (Lurin et al. 2004, Cheng et al. 2016). The PLS-class PPRs generally contain more PPR motifs than P-class: 14–15 PPR motifs (Cheng et al. 2016). A summary of PPRs characteristics is detailed in **Table 1**.

Regardless of the subclass of PPR protein, each PPR motif folds into two parallel α-helices (**Fig. 1A**). These repeat motifs align to allow the key nucleotide-specifying residues, called the PPR code amino acids, to interact with a single nucleotide (nt) (Yin et al. 2013). These code amino acids are located at positions 5 and 35 of the PPR motif (alternatively numbered positions 6 and 1′ based on the designation of the first amino acid of the repeat) and interact with RNA via hydrogen bonds in a 1:1 parallel fashion, leading to a binary binding code (Barkan et al. 2012, Yagi et al. 2013). There is no conserved PPR binding motif, at the RNA level, as the unique variation of motif types, PPR code amino acids and motif order of an individual PPR



Molecular Applications	PPR	PUF	CRISPR-Cas
Location/Movement Tracking	Proposed	Validated	Proposed
Directed Localization	Proposed	Validated	Proposed
Cleavage/Degradation	Proposed	Validated	Validated
Stabilization	Proposed	Validated	Proposed
Editing	Proposed	Validated	Validated
Disease Diagnotics	Proposed	Validated	Validated

Fig. 1 PUF, PPR and type VI Cas proteins bind RNA in a sequence-specific manner. Schematic of PUF, PPR and type VI CRISPR-Cas protein structures, target RNA interactions and molecular tool applications (a) Protein structure of a 10-motif PPR protein (synthetic cPPR, Coquille etal. 2014). (b) Protein structure of an eight-repeat PUF protein (APUM2) (c) Protein structure of *Leptotrichia shahii* type VI Cas protein C2c2. Structures generated using Iterative Threading ASSEmbly Refinement (Yang et al. 2015). Molecular applications include validated techniques (dark green), proposed techniques (light green).

allows for a high degree of variation in bound RNA sequence. PPR–RNA binding is further specified by flexibility influenced by residues 2 and 2′, the binding affinity of L and S-class versus P-class motifs, the context of the PPR motif within the protein, and a RNA structure phenomenon called base flipping, wherein individual RNA bases within a target sequence orient away from the protein and therefore do not contact the nucleotide-specifying residues. Nonetheless, the PPR-RNA code has been determined through predictive modeling, in vitro, and in vivo assays (Barkan et al. 2012, Yagi et al. 2013, Coquille et al. 2014, Gully et al. 2015a, Shen et al. 2015, Shen et al. 2016). In addition to facilitating the investigation of the functional role of PPRs, this knowledge has allowed for the modular design and assembly of synthetic PPR domains to generate RNA sequence-specific molecular tools (Coquille et al. 2014, Yagi et al. 2013, Manna 2015, Gully et al. 2015a).

The variation in PPR motif number and domain length confers considerable adaptability in the three-dimensional (3-D) structure of PPR binding pockets. To date, there are numerous crystal structures of P-class PPRs, but only one PLS-class PPR has been resolved, a synthetically designed $(PLS)_3$PPR (Yan et al. 2017). PPR domain-RNA co-crystal structures range from short rectangular proteins, such as *At*THA8L (Ban et al. 2013) to cylindrical dimer complexes, such as *Zea mays* (*Zm*) PPR10. The most common structure seen in synthetic PPRs and the monomeric *Zm*PPR10 is a right-handed α-solenoid structure (Coquille et al. 2014, Gully et al. 2015a, Gully et al. 2015b, Yan et al. 2017). The PPR motifs spiral around a central axis exposing the code amino acids within the spiral, thereby allowing the target RNA molecule to thread through the cylinder (Coquille et al. 2014, Li et al. 2014, Gully et al. 2015b, Shen et al. 2016). Thus, PPRs can position an RNA molecule within a protein complex, facilitating a specific process, such as RNA editing, processing or translation.

An additional level of structural complexity can be seen in the crystal structure of *Zm*PPR10. Although *Zm*PPR10 was found to bind RNA as a monomeric protein, it requires dimerization to form its final 3-D structure, a cylindrical tube (Barkan et al. 2012, Yin et al. 2013). Dimerization is not necessary for RNA binding, but it may be important for protein tertiary structure and/or catalytic activity.

PUFs have a strict repeat structure and RNA specificity

As observed for PPR proteins, PUFs bind to specific linear RNA sequences defined by protein repeat modules that interact in a nucleotide-specific manner. PUFs are characterized by a single, carboxy-terminal positioned pumilio homology domain (PUM-HD) comprised of multiple PUM repeats that stack to create a concave PUM-HD pocket (**Fig. 1B**). Each PUM repeat is approximately 36 amino acids long and when folded creates a triangular array of three stacked α-helices. The central helix (α2) of each repeat is oriented along the inner face of the binding pocket and contains three nucleotide-specifying amino acids at positions 12, 13 and 16, whose sequence is called the tripartite repeat motif (TRM). The TRM interacts with a single RNA nucleotide via hydrophobic and stacking interactions. The key residues of the TRM, like PPR code amino acids, determine the RNA base to which the PUM repeat binds (Zamore et al. 1997, Koh et al. 2011, Zhang et al. 2016). PUFs also bind nucleotides in a 1:1 repeat/TRM to nucleotide ratio, but this binding is antiparallel, in contrast to PPRs. The 6–11 nucleotide (nt) target sequence of the RNA is called the PUF binding element (PBE). The deciphering of the PUF–RNA binding code through in vivo and in vitro assays has enabled the synthetic production of proteins that bind specific sequences within individual mRNAs, enabling visualization, control of localization and other aspects of post-transcriptional regulation (Tilsner et al. 2009, Porter et al. 2015).

PUM-HD:RNA structural studies have defined three distinct configurations of PUFs: an eight PUM repeat crescent shape, a 10 PUM repeat C-shape and an 11 PUM repeat L-shape (Edwards et al. 2000, Qiu et al. 2014, Zhang et al. 2016). The PUF crescent form was discovered in 2001 when the crystal structures of *Homo sapiens* (*Hs*) PUM1 and *Drosophila melanogaster* (*Dm*) PUM and their RNA targets were resolved (Edwards et al. 2000, Wang et al. 2001). Using published crystal structures, most PUFs can be modeled with this crescent shape and a predicted 8 nt PBE. Most animal PUM-HDs have identical TRMs and bind a conserved eight nucleotide PBE (5′-UGUAUA UA-3′; Zamore et al. 1997). Plant PUM-HDs vary more in TRM sequence and thus may not bind this conserved sequence (Tam et al. 2010). Despite the high similarity in TRMs and structure, there are known exceptions in PBE sequence, most often due to base flipping. Generally, this results in an eight-repeat PUF that binds to a nine nucleotide sequence, with a variable nucleotide at the site of the base flipping, examples include Caenorhabditis elegans (*Ce*)FBF-1, *Ce*FBF-2 and Saccharomyces cerevisiae (*Sc*)PUF4 (Zhang et al. 1997, Valley et al. 2012).

More recently the crystal structures of PUFs with 10 and 11 PUM repeat were shown to form C- and L-shapes, respectively (Qiu et al. 2014, Zhang et al. 2016, Bao et al. 2017). The founding member of the C-shape PUFs, *Sc* nucleolar protein 9 (NOP9), recognizes an 11 nt PBE within the 18S pre-rRNA, 5′-GGA AUUGACGG-3′, where the underlined nucleotides do not interact with the PUM-HD (Zhang et al. 2016, Wang and Ye 2017). The *Sc*NOP9 PUM-HD only physically contacts seven nt as the central three nucleotides are base flipped and PUM repeats 5–7 do not contact the RNA (Wang and Ye 2017). The *Sc*NOP9 ortholog of *A. thaliana*, APUM23, was also resolved and its PUM-HD binds the same PBE, with nucleotide 4 lacking direct interaction with the protein, while nucleotides 5–6, which base flip in *Sc*NOP9, are recognized by repeat 6. The similarity in PBE sequence between *Sc*NOP9 and APUM23 despite differences in their domain sequence and structure speaks to their highly conserved function in pre-rRNA processing.

The L-shape PUF founder, *Hs*PUM3/Puf-A, has 11 PUM repeats and does not bind RNA in a sequence-specific manner. The TRMs of L-shaped PUFs are highly variable and show little similarity to TRMs found in crescent or C-shape PUFs. There is no known PBE for L-shaped PUFs, although they are capable of binding both RNA and DNA (Qiu et al. 2014). *Hs*PUM3/Puf-A orthologs include APUM24, which also lacks sequence specificity in RNA binding (Shanmugam et al. 2017). A summary of PUF characteristics conserved throughout eukaryotes is summarized in **Table 1**.

CRISPR-Cas13 uses RNA-directed targeting to identify DNA and RNA targets

The advent of CRISPR-Cas as a molecular tool has led to an explosion of DNA editing techniques, but a less well-known function of a subset of CRISPR-Cas systems is RNA cleavage. Multiple types (type I–VI) of prokaryotic CRISPR-Cas systems have been characterized and placed into two classes (Class 1 and 2). The Class 1 systems include type I, which relies on a complex of Cas proteins, and type III, which is similar to type I, but utilizes distinct Cas proteins. Class 2 systems include type II, exemplified by the well-known Cas9, type V, a single protein system distinct from Cas9, and type VI (Abudayyeh et al. 2016). Of these, only type III and VI Cas proteins specifically target RNA (Hale et al. 2009, Abudayyeh et al. 2016). The type VI CRISPR-Cas system is characterized by the single-component effector, Cas13/C2c2, that contains two higher eukaryotes and prokaryotes nucleotide binding (HEPN) domains that provide RNA cleavage functionality (East-Seletsky et al. 2016). There are four known sub-categories of Cas13 proteins, Cas13a-d, which vary in size, domain structure and CRISPR guide RNA (crRNA) length (Shmakov et al. 2017). Overall, Cas13, as well as other Class 2 proteins, exhibit a distinct bi-lobed 3-D structure, wherein one lobe (Recognition lobe) recognizes the crRNA through a N-terminal domain and helical-1 domain, while the other (Nuclease lobe) contains a helical-2 domain and two HEPN domains and is responsible for target cleavage (**Fig. 1C;** Liu et al. 2017b, Zhang et al. 2018). Recognition of type VI crRNA by Cas13 involves both RNA base contacts and secondary structure recognition. Cas13 crRNA is transcribed as pre-crRNA and processed by the RNA nuclease activity of the Cas13 protein, as compared to Cas9 which relies on the hybridization of the pre-crRNA with a trans-activating crRNA (tracrRNA) that is cleaved by the Host FACTOR NUCLEASE RNase III (Hochstrasser and Doudna 2015).

The type VI CRISPR-Cas systems use a modular guide-RNA based targeting system to find and bind target RNAs, as compared to the direct protein–RNA interactions of PPRs and PUFs. This is accomplished through the use of a mature type VI

crRNA which consist of a small (4–14 nt) hairpin loop flanked on both the 5′ and 3′ by a 20–30 nt spacer sequence, with variation in the loop and spacer sequences based on the co-evolved Cas13 protein (Abudayyeh et al. 2017, Cox et al. 2017, Shmakov et al. 2017). Through a series of non-covalent nucleotide interactions with the Recognition and Nuclease lobes, the crRNA is oriented to expose the spacer region of the crRNA for target binding (Liu et al. 2017a, Liu et al. 2017b, Knott et al. 2017). These nucleotide interactions appear to be specific to the individual crRNA and Cas13, as there is no consensus between the currently resolved Cas13-crRNA structures. While the spacer allows for some mismatching with the target, mismatches in the center of the spacer can have effects on both binding and activation of nuclease activity (Abudayyeh et al. 2016, Tambe et al. 2018). Cleavage of target RNAs is accomplished by the set of HEPN domains in the Nuclease lobe of the protein. The HEPN domains of different Cas13 proteins have cleavage sight preference for a variety of dinucleotide regions (Gootenberg et al. 2018). Interestingly, Cas13 exhibits a cleavage phenomenon called collateral cleavage, wherein once the protein's nuclease activity is activated through binding and cleavage of one target RNA it continues to cleave both unbound target and off-target RNAs (Abudayyeh et al. 2016, East-Seletsky et al. 2016). Collateral cleavage can be enhanced and reduced by mismatch binding in the spacer (Tambe et al. 2018). To date, collateral cleavage has not been observed in human or plant cells when Cas13 has been used as a molecular tool, but can be seen in endogenous systems (Abudayyeh et al. 2017, Cox et al. 2017, Aman et al. 2018). The structural components of Cas13 binding have been nicely reviewed by O'Connell (2018).

Functional activities of modular sequence-specific PPRs

Despite similarities in their structures and methods of target specification, PPRs and PUFs show distinction in subcellular localization and biological function. PPR proteins are almost exclusively localized to the mitochondria and plastids and function in processes such as targeted C-to-U RNA editing (reviewed in Sun et al. 2016), group I and II intron splicing (de Longevialle et al. 2008), transcript processing, stabilization and turnover (Manavski et al. 2018), and tRNA and rRNA maturation (Wu et al. 2016, Klemm et al. 2017). Based on studies in yeast and metazoa, PUFs are nucleocytoplasmic and exhibit localization segregation based on shape: crescent-shape PUFs are cytoplasmic and function in mRNA turnover and translation (reviewed in Quenault et al. 2011), whereas C and L-shape PUFs localize to the nucleolus and contribute to rRNA processing (Qiu et al. 2014, Zhang et al. 2016, Shanmugam et al. 2017). And while only one crystal structure has been resolved for APUMs, their localization and function is consistent with their predicted structures. For both PPRs and PUMs, the distinction in functions beyond RNA binding is largely explained by interactions with other proteins and is defined by the functional domains of these proteins.

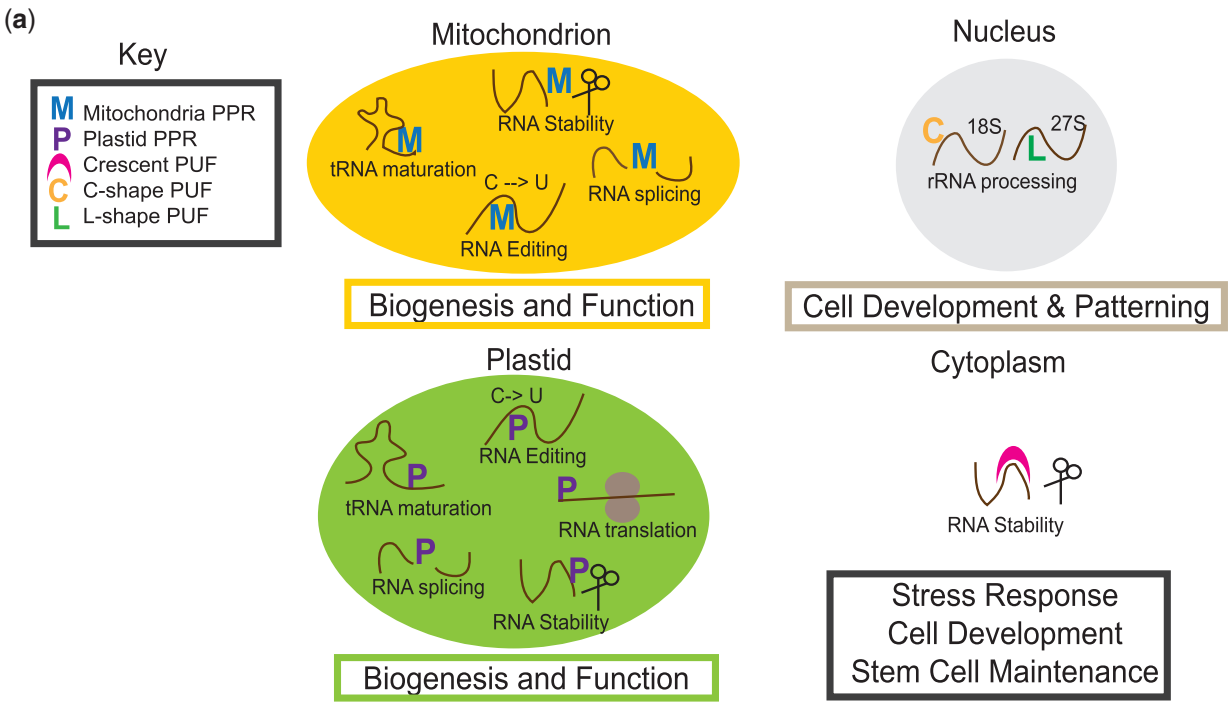
Despite similarity in protein structure, P and PLS-class PPRs have distinct biological roles (Fig. 2). P-class PPRs generally only contain a PPR domain; notable exceptions include *A. thaliana* PROTEIN-ONLY RNASE Ps (Gobert et al. 2010), SUPPRESSOR OF VARIEGATION 7 (Schelcher et al. 2016) and the small MutS-related domain containing PPRs (PPR-SMRs) (Liu et al. 2013). P-class PPRs function in organellar RNA stabilization, splicing, intergenic RNA cleavage and translation regulation (Manavski et al. 2018). For example, *Zm*PPR10, essential for chloroplast development, binds the 5′ region of its target transcripts to protect against 5′ to 3′ exonucleases and enhances translation by exposing a ribosome-binding site (Prikryl et al. 2011).

PLS-class PPR molecular mechanisms have been studied more extensively than that of P-class due to their role in C-to-U RNA editing in both mitochondria and plastids (Hammani et al. 2009, Okuda et al. 2009) (Fig. 2). In contrast to the P-class, PLS-class PPRs often contain at least one C-terminal variable domain. The DYW has been proposed to provide the deaminase activity of C-to-U editing, thus allowing these RBPs to locate and position editing sites, recruit the multisubunit editosome complex and directly edit target RNAs (Boussardon et al. 2014, Okuda et al. 2014). The other major variable domain of PLS-class PPRs, the E (1/2/+) domains, are less well characterized. E1/2 domains share sequence similarity to PPR motifs and have been suggested to serve as additional RNA-binding motifs (Cheng et al. 2016). The E+ domain, which shares sequence similarity to a truncated DYW domain, of *At*CHLOROPLAST BIOGENESIS 19 (CLB19) facilitates RNA editing activity in vitro via interaction with MULTIORGANELLAR RNA EDITING FACTOR 2/RNA-EDITING FACTOR INTERACTING PROTEIN 2 (MORF2/RIP2) (Takenaka et al. 2012). This interaction between PPRs and MORFs was also reported for REQUIRED FOR ACCD RNA EDITING 1 (RARE1) and RIP1 (Bentolila et al. 2012) and the synthetic $(PLS)_3$ PPRs and MORF9 (Yan et al. 2017).

PPRs also work in concert. The E+ PPR proteins, *At*CLB19 and *At*SLOW GROWTH 2 (SLO2), which are critical components of the E+ editosome in chloroplasts and mitochondria, respectfully, rely on interactions with DYW-containing PPR, DYW2 and P-class PPR, NUWA, to enact their editing activity (Andrés-Colás et al. 2017, Guillaumot et al. 2017). Genetic loss-of-function alleles of *NUWA* and *DYW2* resulted in hundreds of differentially and commonly non-edited sites (Guillaumot et al. 2017). As only seven NUWA and DYW2 edit sites have also been verified as SLO2 or CLB19 edit sites, E+ PPR proteins, including SLO2 and CLB19, may rely on NUWA and DYW2 for activity (Chateigner-Boutin et al. 2008, Zhu et al. 2012, Andrés-Colás et al. 2017). The current model proposes that E+ PPR editosomes work as a tripartite PPR complex in which a E+ PPR protein (i.e. CBL19 or SLO2) provides the RNA-binding specificity just 5′ of an editing site, a DYW PPR protein (i.e. DYW2) provides the editing activity via its DYW domain and a P-class PPR (i.e. NUWA) provides scaffolding that enhances the interaction between the E+ and DYW PPRs. Other proteins including MORFs and Heat Shock Proteins may also interact with these complexes to modulate the function of PLS-class PPRs (Andrés-Colás et al. 2017).

Functional activities of modular sequence-specific PUFs in plants and other organisms

All PUFs characterized to date solely bind RNA and thus any post-transcriptional processes they modulate are likely to



(b)

	PPR		PUF	
	Mitochondria M	Plastid P	Nucleus	Cytoplasm
rRNA maturation		PPR2, *Os* PGL12 SVR7, PPR-SMRs	*A* PUM23 C *A* PUM24 L	
mRNA stability		*Zm* PPR10 *Zm* PPR103		*A* PUM1-12
mRNA translation	MTL1	*Zm* PPR10		
mRNA splicing	MTL1, OTP43 ZmBek35	THA8L SOT5		
RNA editing	SLO2, DYW2 NUWA	SLO2, DYW1, DYW2 NUWA, RARE1, CLB19		
tRNA maturation	Rnase P PRORP1	PRORP1		

Fig. 2 PPRs and PUFs family members vary in structure, function and roles in development. (a) Schematic of PPR and PUF localization within cells and their molecular functions. C, C-shaped PUFs; L, L-shaped PUFs; crescent, crescent-shaped PUFs; M, mitochondrial localized PPRs; P, plastid localized PPR. (b) Examples of *Arabidopsis* and other plant PUF and PPR molecular functions. *Zm*, *Zea mays*; *Os*, *Oryza sativa*.

involve interaction with other proteins or complexes. PUFs are known to interact with a variety of different proteins in the context of the post-transcriptional regulation of their RNA targets. These proteins also function in multiple post-transcriptional processes involving rRNA and mRNAs in plants and other eukaryotes (Fig. 2).

Roles of crescent-conformation PUFs in post-transcriptional processes

Crescent conformation PUFs function in a variety of processes, including regulation of mRNA stability, through recruitment of the C-C chemokine receptor type 4-Negative Regulator of Transcription 1 (CCR4-NOT) deadenylation complex, and the regulation of translation of target mRNAs. PUFs specifically target transcripts and recruit degradation and translation repressive machinery through interaction with other RBPs. In the classic *Drosophila* example, target transcripts are bound by *Dm*PUM via two canonical PBEs. This complex then recruits Nanos (Nos; a ZnF RBP) (Spassov and Jurecic 2003, Arvola et al. 2017). Nos binds an additional three nucleotides upstream of the PBE and acts as a molecular clamp, strengthening the *Dm*PUM–RNA interaction (Weidmann et al. 2016). A third RBP, Brain Tumor [Brat; a Tripartite Motif-NCL-1, HT2A and LIN-41 (NHL) protein] can be recruited to bind an additional 6 nt upstream of the *Dm*PUM PBE, and contributes to the repressive activity of the complex (Loedige et al. 2015). Multiple methods of translation repression for the PUM-Nos-Brat complex have been proposed, including disruption of Poly(A) Binding Protein poly(A) tail binding (Weidmann et al. 2014), displacement of the mRNA cap-binding protein eukaryotic

Initiation Factor (eIF)4E (Cho et al. 2006), recruitment of the translation initiation inhibitor eIF4E-binding protein (Blewett and Goldstrohm 2012) and recruitment of the CCR4-NOT deadenylation complex (Weidmann et al. 2014, Arvola et al. 2017). Thus, the composition and context of a multi-component protein–RNA complex is likely to regulate the stability of target mRNAs.

Crescent PUFs have also been found to collaborate with miRNA and miRNA machinery to control target transcript abundance. In humans, PBEs are significantly enriched in proximity to miRNA-binding sites (Galgano et al. 2008) and PUF binding has been reported to enhance miRNA binding by facilitating secondary structure changes that expose miRNA-binding sites (Kedde et al. 2010, Miles et al. 2012). PUFs can also inhibit the GTPase activity of eEF1A, which is required for factor release from the ribosome after tRNA binding, thus repressing translation elongation and causing ribosome stalling (Friend et al. 2012). Given their wide range of molecular mechanisms, yeast and metazoan PUFs are recognized as essential for a variety of functions, including stem cell maintenance, organ patterning and viral RNA sensing.

Only a few plant crescent PUFs have been studied at the phenotypic or functional level. APUM5 over-expression convers resistance to Cauliflower Mosaic Virus (CaMV) and osmotic stress hypersensitivity through targeting of the conserved PBE found in the CaMV viral RNA and key osmotic response genes, including *Response to Desiccation 22* and *RAB GTPase Homolog B18* (Huh et al. 2013, Huh and Paek, 2014). Other crescent-shaped APUMs have been implicated in a variety of processes including stem cell maintenance (APUM1-6; Francischini and Quaggio 2009), salt stress tolerance (APUM7, 8 and 12; K. C. Huang et al. 2018) and seed dormancy (APUM9, 10 and 11; Xiang et al. 2014) (**Fig. 2**). A role for plant PUFs in miRNA-regulated processes has not yet been discovered, despite the evident role of a number of APUMs in regulation of mRNA stability.

PUF involvement in ribosome biogenesis is conserved across eukaryotes

C- and L-shape PUFs appear to function differently from crescent conformation PUFs. To begin with, C- and L-shape PUFs target the pre-rRNAs of the small and large ribosomal subunits, respectfully, whereas crescent PUFs target mRNA. The C-shape PUF *Sc*NOP9, binds to both the internal transcribed spacer 1 of the 20S pre-rRNA and the central pseudoknot of the 18S rRNA via an 11 nt PBE (5′-GGAAUUGACGG-3′) (Zhang et al. 2016, Wang and Ye 2017). When bound to the 20S pre-rRNA, *Sc*NOP9 blocks the binding of the endonuclease *Sc*NOP1, thus contributing to the spatial regulation of *Sc*NOP1 activity within the nucleolus (Zhang et al. 2016). When bound to the 18S rRNA, *Sc*NOP9 is also positioned to block the binding of the U3 snoRNA, a key component of the 90S pre-ribosome (Wang and Ye 2017). It is unknown if *Sc*NOP9 acts as a U3 block or recruits U3 to the 18S rRNA, but *Sc*NOP9 is critical for 40S subunit maturation.

In *A. thaliana*, the C-shape PUF APUM23, plays a role in rRNA maturation; *apum23* mutants accumulate unprocessed 18S and 5.8S rRNAs (Abbasi et al. 2010, Zhang et al. 2016, Bao et al. 2017). The importance of APUM23's role in rRNA processing is evident by the pleiotropic effects of *apum23* mutants, including alterations in germination, leaf patterning (Huang et al. 2014), ABA response and salt stress tolerance (W. Huang et al. 2018) (**Fig. 2**).

L-shape PUFs have only recently been characterized. The founding member, *Sc*PUF6, binds double- and single-stranded DNA and RNA with no apparent sequence specificity, although key basic residue-containing patches are critical for in vitro nucleic acid binding activity (Qiu et al. 2014). Genetic disruption of *Sc*PUF6 results in over-accumulation of 7S and 27S rRNA precursors (Qiu et al. 2014).

In *A. thaliana*, loss-of-function alleles of the L-shape PUF *APUM24* lead to over-accumulation of the 27Sb precursor and 5.8S rRNAs (Maekawa et al. 2017, Shanmugam et al. 2017). APUM24 binds in vitro and in vivo to the unprocessed ITS2 of 27S and 5.8S rRNAs (Maekawa et al. 2017). It also interacts with a number of ribosomal proteins, as well as the proteins BIOGENESIS OF RIBOSOMES IN *XENOPUS* 1 (BRX1)-1 and *At*BRX1-2, strengthening the evidence for APUM24's role in pre-rRNA processing during ribosome biogenesis. The role of APUM24 in rRNA processing is critical for cell development and patterning such that the majority of mutants are embryo lethal (**Fig. 2**). Although L-shape PUFs may lack sequence specificity, they play a critical role in the processing of rRNAs. It remains to be determined if their interaction with RNA is guided by RNA secondary structure or other protein interactions that have yet to be determined.

Leveraging PPRs, PUFs and type VI CRISPR-Cas as molecular tools in synthetic biology

The characterization of the structure and binding code of sequence-specific RBPs has created opportunities for their use as molecular tools in synthetic biology. The utility of PUFs and PPRs relies on the nucleotide binding code defined by the TRM/PPR code amino acid modules and their modularity. Naturally occurring and in vitro determined TRMs and PPR code amino acids have been determined for all four RNA nucleotides allowing for the generation of synthetic PUFs and PPRs that target specific RNA sequences (Filipovska et al. 2011, Yagi et al. 2013, Coquille et al. 2014, Gully et al. 2015a, Shen et al. 2015, Zhang and Muench 2015, Shen et al. 2016, Bao et al. 2017). By coupling RNA-binding specificity with protein–protein interaction domains or protein domains with defined activities, synthetic PPRs and PUF may be used to track, capture, regulate or modify specific RNA in living cells, including those of plants.

There are pros and cons to the use of PPRs and PUFs as tools to target specific RNAs. The larger size and wide variation in the PPR code found in nature makes them ideal for discrimination between highly similar RNA molecules, such as endogenous and transgenic transcripts. PPRs can be engineered to target longer elements than PUFs due to less restriction in the number of PPR motifs that can be combined within a single protein. Naturally occurring PPRs have been shown to bind *cis*-elements up to 18 nt long in vivo and some have been predicted to bind even longer sequences (Miranda et al. 2017). They are also less

constrained by 3-D shape compared to PUFs. PPRs also could provide a more convenient way to modify RNA targets in vivo as the variable domain of PPRs can be changed to allow for direct editing of RNAs. However, due to the requirement of an editosome to target multiple edit sites within the plastid or mitochondrial transcriptome, design of synthetic PPRs must consider the protein domains important for recruitment and localization if direct editing is desired (Sun et al. 2016). In contrast to PPRs, greater target specificity may be obtained through the use of PUFs to recognize specific RNAs, as the reliance of TRM-PBE code on a 3:1 amino acid to nucleotide interaction provides potentially high specificity. However, crescent-shaped PUFs are significantly limited in the length of the *cis*-element they can recognize. The extension of eight-repeat PUFs to 10–11 repeat PUFs could alter protein conformation from crescent to L- or C-shape, potentially affecting binding. To overcome this challenge, two synthetic PUFs that bind adjacent sequences can be used to reduce off-target binding (Ozawa et al. 2007, Tilsner et al. 2009).

A diversity of synthetic PPR and PUF applications have been proposed over the years, including visualization of single RNA molecules, binding of N6-methyladenosine-modified nucleotides, and manipulation of RNA metabolism, including processing, editing, translation and turnover (Wang et al. 2013, Abil et al. 2014, Shen et al. 2016, Yoshimura 2018). Due to the recent discovery of the PPR binding code, no technologies have been validated using synthetic PPRs in plants or other organisms, although patent applications have been submitted for these technologies (**Fig. 1A**) (Nakamura et al. 2017). In contrast, synthetic PUFs have been created and successfully implemented (**Fig. 1B**). This has been aided by a set of modified Golden Gate based cloning scaffolds developed for easy assembly of synthetic PUF proteins (Abil et al. 2014, Adamala et al. 2016).

PUF-based RNA targeting has been used successfully to localize and track mitochondrial RNA movement within human mitochondria (Ozawa et al. 2007), to localize Tomato Mosiac Virus RNA to viral replication complexes within *Nicotiana benthamiana* cells (Tilsner et al. 2009), to localize telomeric repeat-containing RNAs to the telomere neighboring region, and to localize β-actin to the cytoskeleton leading edge in live mammalian cells (Yamada et al. 2011, Yamada et al. 2016). In one variation, two synthetic PUFs were designed to bind two adjacent RNA sequences and fused to the two halves of a fluorescent protein. When both PUFs bound the target RNA, bimolecular fluorescence complementation allowed for the visualization of the target RNA both spatially and temporally. There is some evidence that the binding of synthetic PUFs to RNA targets does not affect the movement velocity, making synthetic PUFs a nice tool for tracking RNA movement (Yamada et al. 2011). In addition, Tilsner et al. (2009) engineered PBEs into the Potato Mosaic Virus RNA as a way to track infections and proposed that if PBEs can be integrated into viral RNA, that synthetic PUFs can be used as a fluorescent based diagnostic tool in plants.

Expanding on the concept of localization, synthetically designed PUFs can be used to control the localization of target RNAs. This system, called PUF-assisted localization of RNA, uses synthetically designed PUFs fused to FK506-binding protein (FKBP) and GFP, to provide target RNA binding and visualization. A second fusion protein consisting of a FKBP rapamycin binding domain (FRB) and a targeting peptide binds the FKBP and targets the complex to the desired cellular locations. This system has been verified in mammalian cells by specifically targeting firefly luciferase RNA to the cell periphery, centrosomes of microtubules and axonal growth cones (Abil et al. 2017).

The coupling of a PUM-HD to a catalytic domain is also a common theme when using PUFs as molecular tools. Synthetic PUFs have been fused to EST1A, a non-specific endonuclease domain, to create Artificial Site-specific RNA Endonucleases (ASREs) (Choudhury et al. 2012), to various splicing factors to create repressive and activating Engineered Splicing Factors (ESFs) (Wang et al. 2009), and to the deadenylase components of the CCR4-NOT complex, to either activate or repress the translation of target RNAs via poly(A) addition or removal (Cooke et al. 2011). PUF fusions present a myriad of possibilities to regulate RNA editing, stability and translational efficiency (Wang et al. 2013).

Endogenously in prokaryotes, type VI CRISPR seems to function as a viral RNA defense through crRNA-targeted single-stranded RNA cleavage (Abudayyeh et al. 2016). By reprogramming the crRNA, type VI CRISPR-Cas systems using *Leptotrichia shahii* (*Ls*) Cas13a and *Prevotella* sp. *P5-125* Cas13b have been used as a substitute for RNA interference (RNAi) (Abudayyeh et al. 2016, Abudayyeh et al. 2017), for mRNA visualization using a non-catalytic mutant (Abudayyeh et al. 2017), and for A-to-I RNA editing when fused to an ADENOSINE DEAMINASE ACTING ON RNA (ADAR) domain (Cox et al. 2017). Type II *Staphylococcus aureus* and *Neisseria meningitides* Cas9 have also been used to bind RNA via crRNA to target single-stranded RNA for cleavage and provide protection from RNA virus infections. These RNA-binding Cas9 proteins can also be used to regulate mRNA levels in vivo, similar to type VI *Ls*Cas13a (Rousseau et al. 2018, Strutt et al. 2018). In vivo, type VI Cas13 with a specially designed crRNAs have been used to detect bacterial and viral infections and identify individual SNPs (Gootenberg et al. 2018), as well as to negate mammalian disease mutations by correcting mutated SNPs. In plants, Cas13 has been used to control Turnip Mosaic Virus infection in *Nicotiana benthamiana* via RNA cleavage (Cox et al. 2017, Aman et al. 2018), providing a new opportunity for viral suppression.

Type II and VI CRISPR are particularly interesting as a RBP tools, as they can perform mRNA regulation functions similar to endogenous RBPs with potentially fewer off-targets. Direct targeting of viral RNAs by CRISPR has enabled similar or improved knock-down of target RNA abundance, with greater specificity, as compared to techniques like RNAi; CRISPR-Cas13a knock-down can be accomplished with no off-targets, whereas RNAi knocked-down expression of the same targets can result in hundreds of off-targets depending on the RNAi used (Abudayyeh et al. 2017). Cas proteins can also be effectively fused to other protein domains to process RNA targets, as is the case with Cas13b-ADAR (Cox et al. 2017). CRISPR-Cas systems are more easily designed and implemented than RBP tools as only the crRNA needs to be redesigned versus the full

RBD. RNA targeting CRISPR-Cas can conceivably be designed to perform functions similar to synthetic PUFs and PPRs, including tracking RNA movement and localization, affecting mRNA stability and processing, and controlling viral RNA infections.

For all the advantages of CRISPR-Cas RNA targeting, design of crRNA presents similar, but distinct challenges compared to synthetic PUF and PPR design. To start, the RNA target sequence is also limited, targets longer than 30 nt are not efficiently processed (Yang et al. 2013) and Cas13 efficiency is sensitive to RNA secondary structure and mismatches (Abudayyeh et al. 2016). While easier to deploy, as only a new crRNA must be designed, not all Cas proteins function efficiently in all systems (Bortesi and Fischer 2015), thus necessitating optimization of Cas-organism pairs. Also, while it has not been seen, so far, in animals or plants experiments, collateral cleavage of non-targets could create a large, uncontrollable problem when targeting RNAs for degradation (Abudayyeh et al. 2016, East-Seletsky et al. 2016, Abudayyeh et al. 2017, Cox et al. 2017, Aman et al. 2018, Tambe et al. 2018). Finally, the ultimate genome editing advantage of CRISPR-Cas systems, that such mutations are not considered "genetically modified organisms," does not hold true for RNA targeting CRISPR. In genome editing, Cas proteins need only be present for the initial editing of gametes and then can be removed, either by transient expression or back-crossing. For RNA targeting, the Cas protein must always be present to have a long-lasting effect, making CRISPR-based RNA targeting, from a transgenic approach, no more advantageous than other RBP-based techniques.

Conclusion

PPR and PUF proteins are members of two sequence-specific sub-classes of RBPs that use modular domains to target specific RNA sequences to facilitate a myriad of post-transcriptional activities in eukaryotes. The deciphering of PPRs functions in plastids and mitochondria has led to resolution of their RNA-binding code and recognition of synthetic opportunities. The functions of the expanded and diversified PUFs found in vascular plants is relatively limited, yet their elucidation will undoubtedly yield important insights into post-transcriptional regulation. The potential use of PPRs, PUFs and CRISPR/Cas13 based RNA targeting strategies provides new opportunities to fine tune steps of gene regulation within the nucleus, cytoplasm, plastids and mitochondria of plants.

Disclosures

The authors have no conflicts of interest to declare.

Funding

Funding for this research was provided by the National Science Foundation to J.B-S. (MCB-1021969 and MCB-1716913). Funding for L.K.D. was provided by the National Science Foundation Graduate Research Fellowship Program (DGE-1326120).

References

Abbasi, N., Kim, H.B., Park, N.-I., Kim, H.-S., Kim, Y.-K., Park, Y.-I., et al. (2010) APUM23, a nucleolar PUF domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. *Plant J.* 64: 960–976.

Abil, Z., Denard, C.A. and Zhao, H. (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *J Biol Eng.* 8: 7.

Abil, Z., Gumy, L.F., Zhao, H. and Hoogenraad, C.C. (2017) Inducible control of MRNA transport using reprogrammable RNA-binding proteins. *ACS Synth. Biol.* 6: 950–956.

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., et al. (2017) RNA targeting with CRISPR-Cas13. *Nature* 550: 280.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., et al. (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353: aaf5573.

Adamala, K.P., Martin-Alarcon, D.A. and Boyden, E.S. (2016) Programmable RNA-binding protein composed of repeats of a single modular unit. *Proc. Natl. Acad. Sci. USA* 113: E2579–E2588.

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M.Z., et al. (2018) RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* 19: 1.

Andrés-Colás, N., Zhu, Q., Takenaka, M., De Rybel, B., Weijers, D. and Van Der Straeten, D. (2017) Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. *Proc. Natl. Acad. Sci. USA* 114: 8883–8888.

Arvola, R.M., Weidmann, C.A., Tanaka Hall, T.M. and Goldstrohm, A.C. (2017) Combinatorial control of messenger RNAs by pumilio, nanos and brain tumor proteins. *RNA Biol.* 14: 1445–1456.

Aubourg, S., Boudet, N., Kreis, M. and Lecharny, A. (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.* 42: 603–613.

Bailey-Serres, J., Sorenson, R. and Juntawong, P. (2009) Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends Plant Sci.* 14: 443–453.

Ban, T., Ke, J., Chen, R., Gu, X., Tan, M.H.E., Zhou, X.E., et al. (2013) Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.* 288: 31540–31548.

Bao, H., Wang, N., Wang, C., Jiang, Y., Liu, J., Xu, L., et al. (2017) Structural basis for the specific recognition of 18S rRNA by APUM23. *Nucleic Acids Res.* 45: 12005–12014.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., et al. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* 8: e1002910.

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65: 415–442.

Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., et al. (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. USA* 109: E1453–E1461.

Blewett, N.H. and Goldstrohm, A.C. (2012) An eIF4E-binding protein promotes mRNA decapping and is required for PUF repression. *Mol. Cell Biol.* 32: 4181–4194.

Borges, F. and Martienssen, R.A. (2015) The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16: 727–741.

Bortesi, L. and Fischer, R. (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33: 41–52.

Boussardon, C., Avon, A., Kindgren, P., Bond, C.S., Challenor, M., Lurin, C., et al. (2014) The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*. *New Phytol.* 203: 1090–1095.

Chantarachot, T. and Bailey-Serres, J. (2018) Polysomes, stress granules, and processing bodies: a dynamic triumvirate controlling cytoplasmic mRNA fate and function. *Plant Physiol.* 176: 254–269.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A., Andrés, C., de la Luz Gutiérrez-Nava, M., Cantero, A., et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant J* 56: 590–602.

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M.F., Bai, F., et al. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* 85: 532–547.

Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P. and Sonenberg, N. (2006) Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. *Curr. Biol.* 16: 2035–2041.

Choudhury, R., Tsai, Y.S., Dominguez, D., Wang, Y. and Wang, Z. (2012) Engineering RNA endonucleases with customized sequence specificities. *Nat. Commun.* 3: 1147.

Cooke, A., Prigge, A., Opperman, L. and Wickens, M. (2011) Targeted translational regulation using the PUF protein family scaffold. *Proc. Natl. Acad. Sci. USA* 108: 15870–15875.

Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J.P., Razif, M.F.M., et al. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* 5: 5729.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., et al. (2017) RNA editing with CRISPR-Cas13. *Science* 358: 1019–1027.

de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., et al. (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana*. *Plant J.* 56: 157–168.

Dykes, I.M. and Emanueli, C. (2017) Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genomics Proteomics Bioinformatics* 15: 177–186.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., et al. (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538: 270–273.

Edwards, T.A., Trincao, J., Escalante, C.R., Wharton, R.P. and Aggarwal, A.K. (2000) Crystallization and characterization of pumilo: a novel RNA binding protein. *J. Struct. Biol.* 132: 251–254.

Filipovska, A., Razif, M.F.M., Nygård, K.K.A. and Rackham, O. (2011) A universal code for RNA recognition by PUF proteins. *Nat. Chem. Biol.* 7: 425–427.

Francischini, C.W. and Quaggio, R.B. (2009) Molecular characterization of Arabidopsis thaliana PUF proteins binding specificity and target candidates. *FEBS J.* 276: 5456–5470.

Friend, K., Campbell, Z.T., Cooke, A., Kroll-Conner, P., Wickens, M.P. and Kimble, J. (2012) A conserved PUF-Ago-eEF1A complex attenuates translation elongation. *Nat. Struct. Mol. Biol.* 19: 176–183.

Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M. and Gerber, A.P. (2008) Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS One* 3: e3164.

Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582: 1977–1986.

Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R.K., et al. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.* 17: 740–744.

Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., Joung, J., Collins, J.J. and Zhang, F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360: 439–444.

Gorski, S.A., Vogel, J. and Doudna, J.A. (2017) RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* 18: 215–228.

Guillaumot, D., Lopez-Obando, M., Baudry, K., Avon, A., Rigaill, G., Falcon de Longevialle, A., et al. (2017) Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc. Natl. Acad. Sci. USA* 114: 8877–8882.

Gully, B.S., Cowieson, N., Stanley, W.A., Shearston, K., Small, I.D., Barkan, A., et al. (2015a) The solution structure of the pentatricopeptide repeat protein PPR10 upon binding *atpH* RNA. *Nucleic Acids Res.* 43: 1918–1926.

Gully, B.S., Shah, K.R., Lee, M., Shearston, K., Smith, N.M., Sadowska, A., et al. (2015b) The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. D Biol. Crystallogr.* 71: 196–208.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139: 945–956.

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.-L., Shikanai, T. and Small, I. (2009) A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell.* 21: 3686–3699.

Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19: 327–341.

Hochstrasser, M.L. and Doudna, J.A. (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* 40: 58–66.

Huang, K.-C., Lin, W.-C. and Cheng, W.-H. (2018) Salt hypersensitive mutant 9, a nucleolar APUM23 protein, is essential for salt sensitivity in association with the ABA signaling pathway in Arabidopsis. *BMC Plant Biol.* 18: 40.

Huang, T., Kerstetter, R.A. and Irish, V.F. (2014) APUM23, a PUF family protein, functions in leaf development and organ polarity in Arabidopsis. *J. Exp. Bot.* 65: 1181–1191.

Huang, W., Zhu, Y., Wu, W., Li, X., Zhang, D., Yin, P., et al. (2018) The pentatricopeptide repeat protein SOT5/EMB2279 is required for plastid *rpl2* and *trnK* intron splicing. *Plant Physiol.* 177: 684–697.

Huh, S.U., Kim, M.J. and Paek, K.-H. (2013) Arabidopsis pumilio protein APUM5 suppresses cucumber mosaic virus infection via direct binding of viral RNAs. *Proc. Natl. Acad. Sci. USA* 110: 779–784.

Huh, S.U. and Paek, K.-H. (2014) APUM5, encoding a pumilio RNA binding protein, negatively regulates abiotic stress responsive gene expression. *BMC Plant Biol.* 14: 75.

Jacobs, J. and Kück, U. (2011) Function of chloroplast RNA-binding proteins. *Cell. Mol. Life Sci.* 68: 735–748.

Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A.F., Elkon, R. and Agami, R. (2010) A pumilio-induced RNA structure switch in *p27*-3′ UTR controls miR-221 and miR-222 accessibility. *Nat. Cell Biol.* 12: 1014–1020.

Klemm, B.P., Karasik, A., Kaitany, K.J., Shanmuganathan, A., Henley, M.J., Thelen, A.Z., et al. (2017) Molecular recognition of pre-tRNA by Arabidopsis protein-only ribonuclease P. *RNA* 23: 1860–1873.

Knott, G.J., East-Seletsky, A., Cofsky, J.C., Holton, J.M., Charles, E., O'Connell, M.R., et al. (2017) Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.* 24: 825–833.

Koh, Y.Y., Wang, Y., Qiu, C., Opperman, L., Gross, L., Tanaka Hall, T.M., et al. (2011) Stacking interactions in PUF-RNA complexes. *RNA* 17: 718–727.

Köster, T. and Meyer, K. (2018) Plant ribonomics: proteins in search of RNA partners. *Trends Plant Sci.* 23: 352–365.

Lee, K. and Kang, H. (2016) Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. *Mol. Cells* 39: 179–185.

Li, Q., Yan, C., Xu, H., Wang, Z., Long, J., Li, W., et al. (2014) Examination of the dimerization states of the single-stranded RNA-recognition protein PPR10. *J. Biol. Chem.* 289: 31503–31512.

Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., et al. (2017a) The molecular architecture for RNA-guided RNA cleavage by Cas13a. *Cell* 170: 714–726.e10.

Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., et al. (2017b) Two distant catalytic sites are responsible for C2c2 RNase activities. *Cell* 168: 121–134.e12.

Liu, S., Melonek, J., Boykin, L.M., Small, I. and Howell, K.A. (2013) PPR-SMRs: ancient proteins with enigmatic functions. *RNA Biol.* 10: 1501–1510.

Loedige, I., Jakob, L., Treiber, T., Ray, D., Stotz, M., Treiber, N., et al. (2015) The crystal structure of the NHL domain in complex with RNA reveals the molecular basis of Drosophila brain-tumor-mediated gene regulation. *Cell Rep*. 13: 1206–1220.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089–2103.

Maekawa, S., Ishida, T. and Yanagisawa, S. (2017) Reduced expression of APUM24, encoding a novel rRNA processing factor, induces sugar-dependent nucleolar stress and altered sugar responses in *Arabidopsis thaliana*. *Plant Cell*. 30: 209–227.

Manavski, N., Schmid, L.-M. and Meurer, J. (2018) RNA-stabilization factors in chloroplasts of vascular plants. *Essays Biochem*. 62: 51–64.

Manna, S. (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* 113: 93–99.

Miles, W.O., Tschöp, K., Herr, A., Ji, J.-Y. and Dyson, N.J. (2012) Pumilio facilitates miRNA regulation of the *E2F3* oncogene. *Genes Dev*. 26: 356–368.

Miranda, R.G., Rojas, M., Montgomery, M.P., Gribbin, K.P. and Barkan, A. (2017) RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA* 23: 586–599.

Nakamura, T., Yagi, Y. and Kobayashi, K. (2017) Method For Designing RNA Binding Protein Utilizing PPR Motif, and Use Thereof. 20170124252.

O'Connell, M.R. (2018) Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR-Cas systems. *J. Mol. Biol*. 431: 66–87.

Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., et al. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell*. 21: 146–156.

Okuda, K., Shoki, H., Arai, M., Shikanai, T., Small, I. and Nakamura, T. (2014) Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J*. 80: 870–882.

Ozawa, T., Natori, Y., Sato, M. and Umezawa, Y. (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat. Methods* 4: 413–419.

Porter, D.F., Koh, Y.Y., VanVeller, B., Raines, R.T. and Wickens, M. (2015) Target selection by natural and redesigned PUF proteins. *Proc. Natl. Acad. Sci. USA* 112: 15868–15873.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA* 108: 415–420.

Qiu, C., McCann, K.L., Wine, R.N., Baserga, S.J. and Hall, T.M.T. (2014) A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proc. Natl. Acad. Sci. USA* 111: 18554–18559.

Quenault, T., Lithgow, T. and Traven, A. (2011) PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol*. 21: 104–112.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499: 172–177.

Rousseau, B.A., Hou, Z., Gramelspacher, M.J. and Zhang, Y. (2018) Programmable RNA cleavage and recognition by a natural CRISPR-Cas9 system from *Neisseria meningitidis*. *Mol. Cell* 69: 906–914.e4.

Schelcher, C., Sauter, C. and Giegé, P. (2016) Mechanistic and structural studies of protein-only RNase P compared to ribonucleoproteins reveal the two faces of the same enzymatic activity. *Biomolecules* 6: 30.

Shanmugam, T., Abbasi, N., Kim, H.-S., Kim, H.B., Park, N.-I., Park, G.T., et al. (2017) An Arabidopsis divergent pumilio protein, APUM24, is essential for embryogenesis and required for faithful pre-rRNA processing. *Plant J*. 92: 1092–1105.

Shen, C., Wang, X., Liu, Y., Li, Q., Yang, Z., Yan, N., et al. (2015) Specific RNA recognition by designer pentatricopeptide repeat protein. *Mol. Plant* 8: 667–670.

Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., et al. (2016) Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun*. 7: 11285.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol*. 15: 169–182.

Small, I.D. and Peeters, N. (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci*. 25: 46–47.

Spassov, D.S. and Jurecic, R. (2003) The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* 55: 359–366.

Strutt, S.C., Torrez, R.M., Kaya, E., Negrete, O.A. and Doudna, J.A. (2018) RNA-dependent RNA targeting by CRISPR-Cas9. *Elife* 7: e32724.

Sun, T., Bentolila, S. and Hanson, M.R. (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci*. 21: 962–973.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B. and Brennicke, A. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. USA* 109: 5104–5109.

Tam, P.P.C., Barrette-Ng, I.H., Simon, D.M., Tam, M.W.C., Ang, A.L. and Muench, D.G. (2010) The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC Plant Biol*. 10: 44.

Tambe, A., East-Seletsky, A., Knott, G.J., Doudna, J.A. and O'Connell, M.R. (2018) RNA binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a. *Cell Rep*. 24: 1025–1036.

Tanaka Hall, T.M. (2014) Expanding the RNA-recognition code of PUF proteins. *Nat. Struct. Mol. Biol*. 21: 653–655.

Tilsner, J., Linnik, O., Christensen, N.M., Bell, K., Roberts, I.M., Lacomme, C., et al. (2009) Live-cell imaging of viral RNA genomes using a pumilio-based reporter. *Plant J*. 57: 758–770.

Valley, C.T., Porter, D.F., Qiu, C., Campbell, Z.T., Hall, T.M.T. and Wickens, M. (2012) Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc. Natl. Acad. Sci. USA* 109: 6054–6059.

Wang, B. and Ye, K. (2017) Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Res*. 45: 3559–3567.

Wang, X., Zamore, P.D. and Hall, T.M. (2001) Crystal structure of a Pumilio homology domain. *Mol. Cell* 7: 855–865.

Wang, Y., Cheong, C.-G., Hall, T.M.T. and Wang, Z. (2009) Engineering splicing factors with designed specificities. *Nat. Methods* 6: 825–830.

Wang, Y., Wang, Z. and Tanaka Hall, T.M. (2013) Engineered proteins with Pumilio/*fem-3* mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J*. 280: 3755–3767.

Weidmann, C.A., Raynard, N.A., Blewett, N.H., Van Etten, J. and Goldstrohm, A.C., et al. (2014) The RNA binding domain of Pumilio antagonizes poly-adenosine binding protein and accelerates deadenylation. *RNA* 20: 1298–1319.

Weidmann, C.A., Qiu, C., Arvola, R.M., Lou, T.-F., Killingsworth, J., Campbell, Z.T., et al. (2016) Drosophila nanos acts as a molecular clamp that modulates the RNA-binding and repression activities of pumilio. *Elife* 5: e17096.

Wu, W., Liu, S., Ruwe, H., Zhang, D., Melonek, J., Zhu, Y., et al. (2016) SOT1, a pentatricopeptide repeat protein with a small MutS-related domain, is required for correct processing of plastid 23S-4.5S rRNA precursors in *Arabidopsis thaliana*. *Plant J*. 85: 607–621.

Xiang, Y., Nakabayashi, K., Ding, J., He, F., Bentsink, L. and Soppe, W.J.J. (2014) Reduced Dormancy5 encodes a protein phosphatase 2C that is required for seed dormancy in Arabidopsis. *Plant Cell*. 26: 4362–4375.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8: e57286.

Yamada, T., Yoshimura, H., Inaguma, A. and Ozawa, T. (2011) Visualization of nonengineered single mRNAs in living cells using genetically encoded fluorescent probes. *Anal. Chem.* 83: 5708–5714.

Yamada, T., Yoshimura, H., Shimada, R., Hattori, M., Eguchi, M., Fujiwara, T.K., et al. (2016) Spatiotemporal analysis with a genetically encoded fluorescent RNA probe reveals TERRA function around telomeres. *Sci. Rep.* 6: 38910.

Yan, J., Zhang, Q., Guan, Z., Wang, Q., Li, L., Ruan, F., et al. (2017) MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat. Plants* 3: 17037.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., et al. (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* 41: 9049–9061.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12: 7–8.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504: 168–171.

Yoshimura, H. (2018) Live cell imaging of endogenous RNAs using pumilio homology domain mutants: principles and applications. *Biochemistry* 57: 200–208.

Zamore, P.D., Williamson, J.R. and Lehmann, R. (1997) The pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* 3: 1421–1433.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., et al. (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 390: 477–484.

Zhang, C., Konermann, S., Brideau, N.J., Lotfy, P., Wu, X., Novick, S.J., et al. (2018) Structural basis for the RNA-guided ribonuclease activity of CRISPR-Cas13d. *Cell* 175: 212–223.e17.

Zhang, C. and Muench, D.G. (2015) A nucleolar PUF RNA-binding protein with specificity for a unique RNA sequence. *J. Biol. Chem.* 290: 30108–30118.

Zhang, J., McCann, K.L., Qiu, C., Gonzalez, L.E., Baserga, S.J. and Hall, T.M.T. (2016) Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nat. Commun.* 7: 13085.

Zhao, B.S., Roundtree, I.A. and He, C. (2017) Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* 18: 31–42.

Zhu, Q., Dugardeyn, J., Zhang, C., Takenaka, M., Kühn, K., Craddock, C., et al. (2012) SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J.* 71: 836–849.