Principles of mRNA targeting and regulation via the Arabidopsis m^{6}A-binding proteins ECT2 and ECT3

Laura Arribas-Hernández[1,6,7], Sarah Rennie[2,6], Tino Köster[3], Michael Schon[4], Carlotta Porcelli[1], Martin Lewinski[3], Balaji Enugutti[4], Michael Nodine[4,5], Dorothee Staiger[3,7], Robin Andersson[2,7] and Peter Brodersen[1,7,8]

[1] University of Copenhagen, Copenhagen Plant Science Center, Ole Maaløes Vej 5, DK-2200 Copenhagen N

[2] University of Copenhagen, Department of Biology, Ole Maaløes Vej 5, DK-2200 Copenhagen N

[3] University of Bielefeld, Faculty of Biology, RNA Biology and Molecular Physiology, D-33615 Bielefeld

[4] Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[5] Laboratory of Molecular Biology, Wageningen University, Wageningen, 6708 PB, the Netherlands

[6] These authors contributed equally to this work

[7] Corresponding author

[8] To whom requests for biological material should be addressed

Email LA-H: laura.arribas@bio.ku.dk; DS: dorothee.staiger@uni-bielefeld.de; RA: robin@binf.ku.dk; PB: pbrodersen@bio.ku.dk

Running title: m^{6}A-mRNA targeting by plant YTHDF proteins

Keywords: m^{6}A, ECT2, ECT3, YTHDF, iCLIP, HyperTRIBE, plant, Arabidopsis.

Abstract

Gene regulation dependent on *N6*-methyladenosine (m^6A) in mRNA involves RNA-binding proteins that recognize m^6A through a YTH domain. The Arabidopsis YTH-domain protein ECT2 is thought to influence mRNA 3'-end formation via binding to URU(m^6A)Y sites, an unexpected conclusion given that ECT2 functions require its m^6A binding activity, and that RR(m^6A)CH is the m^6A consensus site in all eukaryotes. Here, we apply the orthogonal techniques individual nucleotide-resolution UV-crosslinking and immunoprecipitation (iCLIP) and HyperTRIBE to define high-quality target sets of the YTH-domain proteins ECT2 and ECT3. The results show that *in vivo*, ECT2 does in fact bind to RR(m^6A)CH. URUAY and other pyrimidine-rich motifs are enriched around, but not at m^6A-sites, reflecting a preference for *N6*-adenosine methylation of RRACH islands in pyrimidine-rich regions. Such regions may also be implicated in ECT2-binding. In particular, a series of properties unique to the URUAY motif suggest that URUAY-type sequences act as sites of competition between unknown RNA-binding proteins and the intrinsically disordered region of ECT2. We also show that the abundance of many ECT2/3 mRNA targets is decreased in meristematic cells devoid of ECT2/3/4-activity. In contrast, loss of ECT2/3/4 activity has no effect on polyadenylation site usage in ECT2/3 targets, consistent with the exclusive cytoplasmic localization of ECT2 observed by super-resolution confocal microscopy. Our study reconciles conflicting results between genetic observations on *N6*-adenosine methylation and ECT2/3/4 function on the one side, and ECT2 target identification on the other, and point to regulation of cytoplasmic mRNA function, including abundance, as a mechanism of plant YTHDF action.

Introduction

N6-methyladenosine (m^6A) is the most abundant modified nucleotide in eukaryotic mRNA bodies. It is required for embryonic development and stem cell differentiation in several animals and plants (Zhong et al. 2008; Batista et al. 2014; Ping et al. 2014; Geula et al. 2015; Zhang et al. 2017) and for the control of the meiotic program in yeast (Shah and Clancy 1992; Clancy et al. 2002; Agarwala et al. 2012). Most *N6*-methylation of mRNA is catalyzed in the nucleus (Salditt-Georgieff et al. 1976; Ke et al. 2017; Huang et al. 2019) by a highly conserved, multimeric methylase (Balacco and Soller 2019) whose catalytic core consists of the heterodimer METTL3/METTL14 (MTA/MTB in plants) (Bokar et al. 1994; Bokar et al. 1997). In addition, a number of highly conserved proteins are required for *N6*-methylation *in vivo* (Balacco and Soller 2019). The strong conservation of these core factors suggests that the biochemical basis of *N6*-adenosine methylation is common in eukaryotes and indeed, m^6A occurs in the consensus site RR(m^6A)CH (R=G/A, H=A/C/U) primarily in 3'-UTRs in all animals, plants and fungi examined to date (Dominissini et al. 2012; Meyer et al. 2012; Schwartz et al. 2013; Luo et al. 2014; Wan et al. 2015; Lence et al. 2016; Shen et al. 2016; Duan et al. 2017; Anderson et al. 2018; Miao et al. 2019; Parker et al. 2020).

m^6A may impact mRNA function by different mechanisms, including the creation of binding sites for reader proteins that specifically recognize m^6A in mRNA (Fu et al. 2014; Meyer and Jaffrey 2014). The best understood class of readers contains a so-called YT521-B homology (YTH) domain (Stoilov et al. 2002) in which a hydrophobic methyl-binding pocket increases the affinity of m^6A-containing RNA by more than 10-fold over unmethylated RNA (Dominissini et al. 2012; Li et al. 2014b; Luo and Tong 2014; Theler et al. 2014; Wang et al. 2014; Xu et al. 2014; Zhu et al. 2014). YTH domain proteins often contain long intrinsically disordered regions (IDRs) and fall into two phylogenetic groups, YTHDC and YTHDF (Patil et al. 2018; Balacco and Soller 2019).

YTHDF proteins are typically cytoplasmic and consist of a long N-terminal IDR followed by the globular YTH domain. They may control mRNA fate via accelerated mRNA decay and translational activation in mammalian cells (Patil et al. 2018; Zaccara et al. 2019), but the mechanisms involved are not clear. Early reports seemed to indicate functional specialization of specific YTHDF proteins for either translational activation or mRNA decay (Wang et al. 2014; Wang et al. 2015), whereas recent studies support functional redundancy among the

three YTHDFs in mammals and zebrafish (Kontur et al. 2020; Lasman et al. 2020; Zaccara and Jaffrey 2020), similar to the functional overlap described earlier for plant YTHDFs (Arribas-Hernández et al. 2018).

In plants, the YTHDF family is greatly expanded, with 11 members in *Arabidopsis*, referred to as EVOLUTIONARILY CONSERVED C-TERMINAL REGION1-11 (ECT1-11), compared to 3 in humans (Li et al. 2014a; Scutenaire et al. 2018). ECT2, ECT3 and ECT4 have important functions in post-embryonic development, because simultaneous knockout of *ECT2* and *ECT3* results in delayed organogenesis and defective morphology of leaves, roots, stems, flowers, and fruits; defects that are exacerbated by additional mutation of *ECT4* in most cases (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020). Importantly, the biological functions of ECT2/3/4 described thus far are shared with those of m^6A writer components and, where tested, have been shown to depend on intact m^6A-binding pockets, strongly suggesting that the basis for the observed phenotypes in *ect2/3/4* mutants is defective regulation of m^6A-modified mRNA targets (Bodi et al. 2012; Shen et al. 2016; Růžička et al. 2017; Arribas-Hernández et al. 2018; Scutenaire et al. 2018; Wei et al. 2018; Arribas-Hernández et al. 2020). It remains unclear, however, which mRNA targets of ECT2/3/4 are responsible for the above-mentioned phenotypes, and what the effects of ECT2/3/4 binding to them may be (Arribas-Hernández and Brodersen 2020). Satisfactory answers to these questions require a robust identification of the mRNA targets of individual ECT proteins, a systematic characterization of the determinants of ECT-mRNA binding, and an assessment of the ECT2/3/4-dependent properties of those mRNAs. Towards that goal, formaldehyde crosslinking and immunoprecipitation (FA-CLIP) was used to identify mRNA targets of ECT2 (Wei et al. 2018). Nonetheless, because formaldehyde, in contrast to UV illumination, generates both protein-protein and protein-RNA crosslinks, it is not an ideal choice for identification of mRNAs bound directly by a protein of interest (see Arribas-Hernández and Brodersen (2020) for a discussion). In particular, this problem concerns the unexpected conclusion that ECT2 binds to the plant-specific consensus motif URU(m^6A)Y, not RR(m^6A)CH (Wei et al. 2018), and indeed, subsequent Nanopore-based single-nucleotide resolution maps of m^6A in Arabidopsis do not support the occurrence of m^6A in URUAY motifs (Parker et al. 2020). Finally, because the URUAY-ECT2 binding sites identified by FA-CLIP are located in the same region as far-upstream elements implicated in poly(A) site (PAS)

selection (~30-150 nucleotides upstream of PAS) and have similar nucleotide composition (Wei et al. 2018; Bernardes and Menossi 2020), and because ECT2 was reported to localize to the nuclear periphery in addition to the cytoplasm, a function of ECT2 in 3'-end formation and PAS selection was proposed (Wei et al. 2018). Thus, the field of gene regulation via m^6A-YTHDF modules in plants is in a state of confusion: On the one hand, single-nucleotide resolution m^6A mapping and phenotypes of mutants defective in m^6A writing or m^6A-binding of ECT2/3/4 suggest that these YTHDF proteins should act, presumably in the cytoplasm to which they largely localize, via recognition of m^6A in the RRACH context. On the other hand, the only attempt at a mechanistic understanding of ECT2 function via mRNA target identification concluded that ECT2 binds to a sequence element different from RRACH, and acts in the nuclear periphery to influence PAS selection in mRNA targets.

To clarify the principles underlying mRNA recognition and regulation by ECT2 and ECT3, we undertook rigorous analysis of its mRNA binding sites using two orthogonal methods, the proximity-labeling method HyperTRIBE (targets of RNA binding proteins identified by editing) (McMahon et al. 2016), and iCLIP (individual nucleotide resolution crosslinking and immunoprecipitation) (König et al. 2010) (see Supplemental Fig. S1). We used the resulting high-quality target sets as a basis to establish five fundamental properties of mRNA targeting by ECT2 and ECT3: (1) m^6A sites and the canonical RRACH motif are enriched immediately downstream of ECT2 crosslinking sites as determined by iCLIP; (2) U-rich motifs including, but not limited to, URUAY are depleted at m^6A sites, but enriched around them, including at ECT2 iCLIP sites; (3) Most targets are shared between ECT2 and ECT3, in agreement with their similar expression patterns and functional redundancy (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020); (4) ECT2/3/4 do not appreciably influence 3'-end formation of target mRNAs, consistent with the absence of ECT2-mCherry from the nucleoplasm; and (5) in ECT2-expressing cell populations, the abundance of the majority of ECT2/ECT3-target mRNAs is reduced upon loss of ECT2/3/4 activity.

Results

ADARcd fusions to ECT2 are functional in vivo

HyperTRIBE uses fusion of RNA binding proteins to the hyperactive E488Q mutant of the catalytic domain of the *Drosophila melanogaster* adenosine deaminase acting on RNA (*Dm*ADAR^{E488Q}cd) (Kuttan and Bass 2012) to achieve proximity labeling *in vivo* (McMahon et al. 2016; Xu et al. 2018). Targets are identified as those mRNAs that contain adenosine-inosine sites significantly more highly edited than background controls, measured as A-G changes upon reverse transcription and sequencing (Supplemental Fig. S1). To develop material suitable for ECT2 HyperTRIBE, we cloned and expressed *AtECT2pro:AtECT2-FLAG-DmADAR^{E488Q}cd-AtECT2ter* (henceforth "*ECT2-FLAG-ADAR*") in the single *ect2-1* and triple *ect2-1/ect3-1/ect4-2* (*te234*) knockout backgrounds (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020). We identified lines exhibiting nearly complete rescue of *te234* mutant seedling phenotypes, indicating that the fusion protein was functional (Fig. 1A). We then used the expression level in complementing lines as a criterion for selection of lines in the *ect2-1* single mutant background, for which no easily scorable phenotype has been described (Supplemental Fig. S2A). To control for background editing, we produced lines expressing free ADAR^{E488Q}cd under the control of the endogenous *ECT2* promoter (*AtECT2pro:FLAG-DmADAR^{E488Q}cd-AtECT2ter;* henceforth *FLAG-ADAR*) at levels similar to or higher than those of the fusion lines (Supplemental Fig. S2A). Expression of *FLAG-ADAR* did not result in phenotypic abnormalities in Col-0 WT plants (Fig. 1A).

The ECT2-ADARcd fusion imparts adenosine-to-inosine editing of target mRNAs in planta

To identify ECT2 HyperTRIBE targets (HT-targets), we sequenced mRNA from dissected root tips and shoot apices of 10-day-old seedlings of *ect2-1*/*ECT2-FLAG-ADAR* and *FLAG-ADAR* transgenic lines. We used five independent lines of each type as biological replicates to prevent line-specific artifacts and increase the confidence in the detection of ECT2-dependent editing sites. To enable an assessment of the correlation between ADAR editing efficiency and abundance, the reads were also mapped to the *FLAG-ADAR* sequence (Supplemental Fig. S2B). Next, we generated nucleotide base counts for all positions with at least one mismatch across the full set of samples of mapped reads (Fig. 1B), resulting in a raw list of

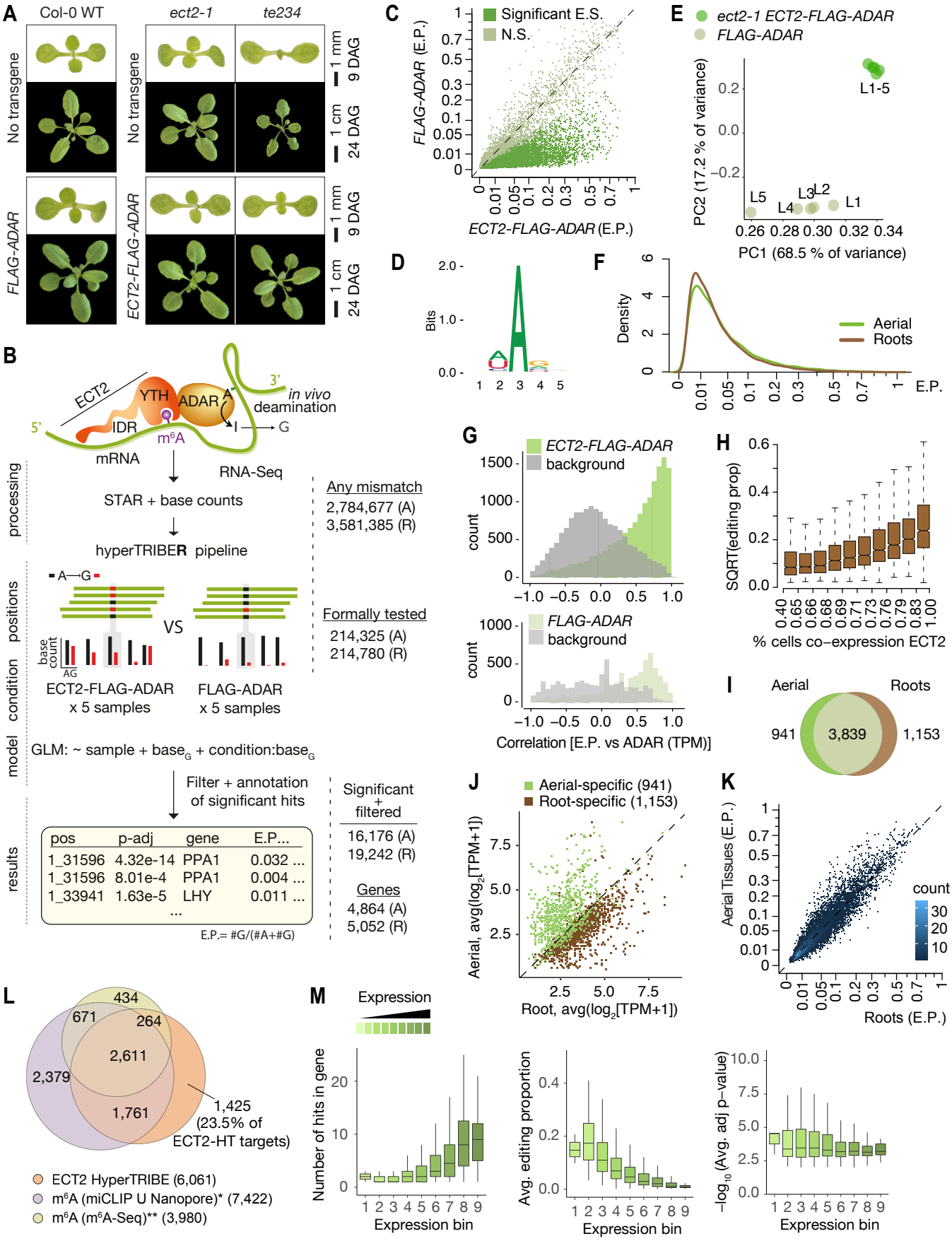

Figure 1. HyperTRIBE identifies m^{6}A-reader targets in plants

(*Continues on the next page*)

Figure 1. HyperTRIBE identifies m^{6}A-reader targets in plants *(continues from previous page)*

(A) Phenotypes of wild type, *ect2-1* and *te234* mutants with (lower panels) or without (upper panels) expression of *ECT2-FLAG-ADAR* or *FLAG-ADAR* transgenes, at 9 or 24 days after germination (DAG). **(B)** Experimental design for ECT2-HyperTRIBE (ECT2-HT) target identification and hyperTRIBE**R** pipeline. Nucleotide base counts quantified from mapped RNA-seq libraries were passed into the HyperTRIBE**R** pipeline to call significant editing sites, which were further filtered and annotated. The number of sites in either aerial (A) or root (R) tissues considered at each stage of the analysis is indicated. E.P., editing proportion. **(C)** Scatterplot of the editing proportions (defined as E.P. = G/(A+G)) of potential and significant editing sites (E.S.) in aerial tissues of *ect2-1/ECT2-FLAG-ADAR* lines compared to the *FLAG-ADAR* controls. Significant sites are highlighted in vivid green. N.S., not significant. **(D)** Consensus motif identified at significant editing sites in aerial tissues of *ect2-1/ECT2-FLAG-ADAR* lines. **(E)** Principal component analysis of editing proportions at significant editing sites in all the samples with aerial tissue used for ECT2-HT. **(F)** Density of editing proportions for significant editing sites in aerial tissues and roots of *ect2-1/ECT2-FLAG-ADAR* lines. **(G)** Distribution of the correlations between editing proportions and ADAR expression (TPM) for significant editing sites in aerial tissues of either *ect2-1 ECT2-FLAG-ADAR* or *FLAG-ADAR* lines. Background correlations are based on randomly shuffling ADAR expression for each site. **(H)** Boxplots showing the mean editing proportions as a function of the proportion of cells co-expressing *ECT2*, calculated based on single cell RNA-seq in roots (Denyer et al., 2019). **(I)** Venn diagram showing the overlap between ECT2-HT targets in roots and aerial tissues, based on genes commonly expressed in both tissues. **(J)** Scatterplot showing the expression in roots and aerial tissues (mean $\log_2$(TPM+1) over the 5 ECT2-HT control samples) of the genes that are identified as targets only in aerial tissues or only in roots. **(K)** Scatterplot of the editing proportions of significant editing sites in ECT2-HT for aerial vs root tissues. **(L)** Venn diagram showing the overlap between ECT2-HT targets and m^{6}A-containing genes. *Parker et al. (2020); ** Shen et al. (2016). **(M)** Number of hits (significant editing sites), average editing proportions and average significance of editing sites per gene in ECT2-HT targets (aerial tissues), split according to their expression levels (from the 5 ECT2-HT control samples). Analogous plots from root samples are shown in Supplemental Fig. S2.

potential editing positions (see Methods). This revealed that the amount of editing was clearly higher in the lines expressing the ECT2-FLAG-ADAR fusion protein than in the negative control lines (Fig. 1C and Supplemental Fig. S2C). To identify positions from this set with significantly higher editing rates in the ECT2-FLAG-ADAR fusions lines compared to the FLAG-ADAR background controls, we developed a new approach to detect differential editing (Fig. 1B) that will be described in detail in a subsequent report. Briefly, the hyperTRIBE**R** (https://github.com/sarah-ku/hyperTRIBER) method of detecting differential editing exploits the powerful statistical capabilities of a method originally designed to detect differential exon usage (Anders et al. 2012). It efficiently takes replicates and possible differences in expression into account, resulting in high power to detect sites despite low editing proportions. As expected, the tendency towards higher editing proportions in fusion lines compared to controls was even more pronounced after filtering non-significantly edited sites (Fig.1C and Supplemental Fig. S2C). Three additional properties of the resulting editing sites indicate that they are the result of ADARcd activity guided by its fusion to ECT2. First, the vast majority of significant hits corresponded to A-to-G edits (Supplemental Fig. S2D). Second, the consensus motif at the edited sites matched the sequence preference of *Dm*ADAR^{E488Q}cd (5' and 3' nearest-neighbor preference of U>A>C>G and G>C>A~U,

respectively (Xu et al. 2018)) (Fig. 1D and Supplemental Fig. S2E), with highly edited sites more closely matching the optimal sequence context than lowly edited ones (Supplemental Fig. S2F). Third, principal component analysis of editing proportions at significant sites over the different lines clearly separated the ECT2-FLAG-ADAR fusion lines from the control lines, which clustered tightly together (Fig. 1E and Supplemental Fig. S2G). Application of subsequent minor filtering steps, including removal of non-(A-to-G) edits and potential line-specific single-nucleotide variants (see Methods), resulted in a final list of 16,176 edited sites for aerial tissues and 19,242 for roots, corresponding to 4,864 and 5,052 genes (ECT2 HT-targets), respectively (Fig. 1B and Supplemental Dataset 1). In both cases, this corresponds to 27% of all expressed genes. We note that the editing proportions were generally low (Fig. 1F) compared to previous work in Drosophila (Xu et al. 2018), perhaps in part due to the limited number of cells that express ECT2 at root and shoot apices (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020). Indeed, the ECT2-ADAR expression level (TPMs) correlates strongly with editing proportions among ECT2-FLAG-ADAR lines (Fig. 1G and Supplemental Fig. S2H), and editing proportions are higher for target mRNAs co-expressed with ECT2 according to single-cell RNA-seq data of root cells (Denyer et al. 2019) (Fig. 1H), lending further support to the conclusion that the observed editing is ADAR-specific and driven to target mRNAs by ECT2.

HyperTRIBE is highly sensitive and identifies primarily m^{6}A-containing transcripts as ECT2 targets

We first noted that the majority of ECT2 HT-target genes were common between root and aerial tissues (Fig. 1I), as expected given the recurrent function of ECT2 in stimulating cell division in organ primordia (Arribas-Hernández et al. 2020). Most of the targets specific to root or aerial tissue were simply preferentially expressed in either tissue (Fig. 1J). Moreover, the significant editing sites in roots and aerial tissues had a considerable overlap (Supplemental Fig. S2I), and their editing proportions were similar in the two tissues (Fig. 1K). Of most importance, we observed a large and significant overlap between the ECT2 HT-targets and m^{6}A-containing transcripts mapped by different methods in seedlings (Shen et al. 2016; Parker et al. 2020) (~76.5% of ECT2 HT-targets) (Fig. 1L). These results validate our HyperTRIBE experimental setup and data analysis, and confirm that ECT2 binds

predominantly to m^6A-containing transcripts *in vivo*. Interestingly, the average statistical significance over all edited sites per target mRNA did not change with the expression level (Fig. 1M, right panel), indicating that our experimental setup and data analysis identify lowly expressed targets. Nonetheless, highly significant individual sites are enriched among highly expressed targets (Supplemental Fig. S2J), perhaps because low read coverage results in higher replicate variation and thereby decreases inferred significance for editing sites in lowly expressed genes. Contrary to the average editing significance, the number of detected hits increases with the expression level of the targets while, unexpectedly, the editing proportion decreases (Fig. 1M, left panels). This may be caused by dilution, as the *ECT2* promoter is active only in highly dividing cells (Arribas-Hernández et al. 2020) while some abundant target mRNAs may be ubiquitously expressed. Taken together, our analysis shows that ECT2 HyperTRIBE identifies targets across a wide range of expression levels, and suggests that the significance of differential editing versus a negative control, rather than raw editing proportion or number of editing sites is the preferred parameter for definition and ranking of targets.

ECT2-mCherry can be specifically UV-crosslinked to target RNA in vivo

We next moved on to independent target and binding site identification via iCLIP (Fig. 2A). We used transgenic lines expressing functional ECT2-mCherry under the control of the endogenous *ECT2* promoter (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020) to co-purify mRNAs crosslinked to ECT2 for iCLIP. Lines expressing the m^6A-binding-deficient *ECT2*^{W464A}*-mCherry* variant were used as negative controls, because this point mutant behaves like a null allele despite its wild type-like expression pattern and level (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020).

To test the feasibility of iCLIP, we first assessed the specificity of RNA co-purified with ECT2-mCherry after UV-illumination of whole seedlings by 5'-radiolabeling of the immunoprecipitated RNP complexes followed by SDS-PAGE. These tests showed that substantially more RNA co-purifies with wild type ECT2 than with ECT2^{W464A} upon UV-crosslinking, and that no RNA is detected without UV irradiation, or from irradiated plants of non-transgenic backgrounds (Fig. 2B and Supplemental Fig. S3A). RNAse and DNAse treatments also established that the co-purified nucleic acid is RNA (Supplemental Fig. S3B).

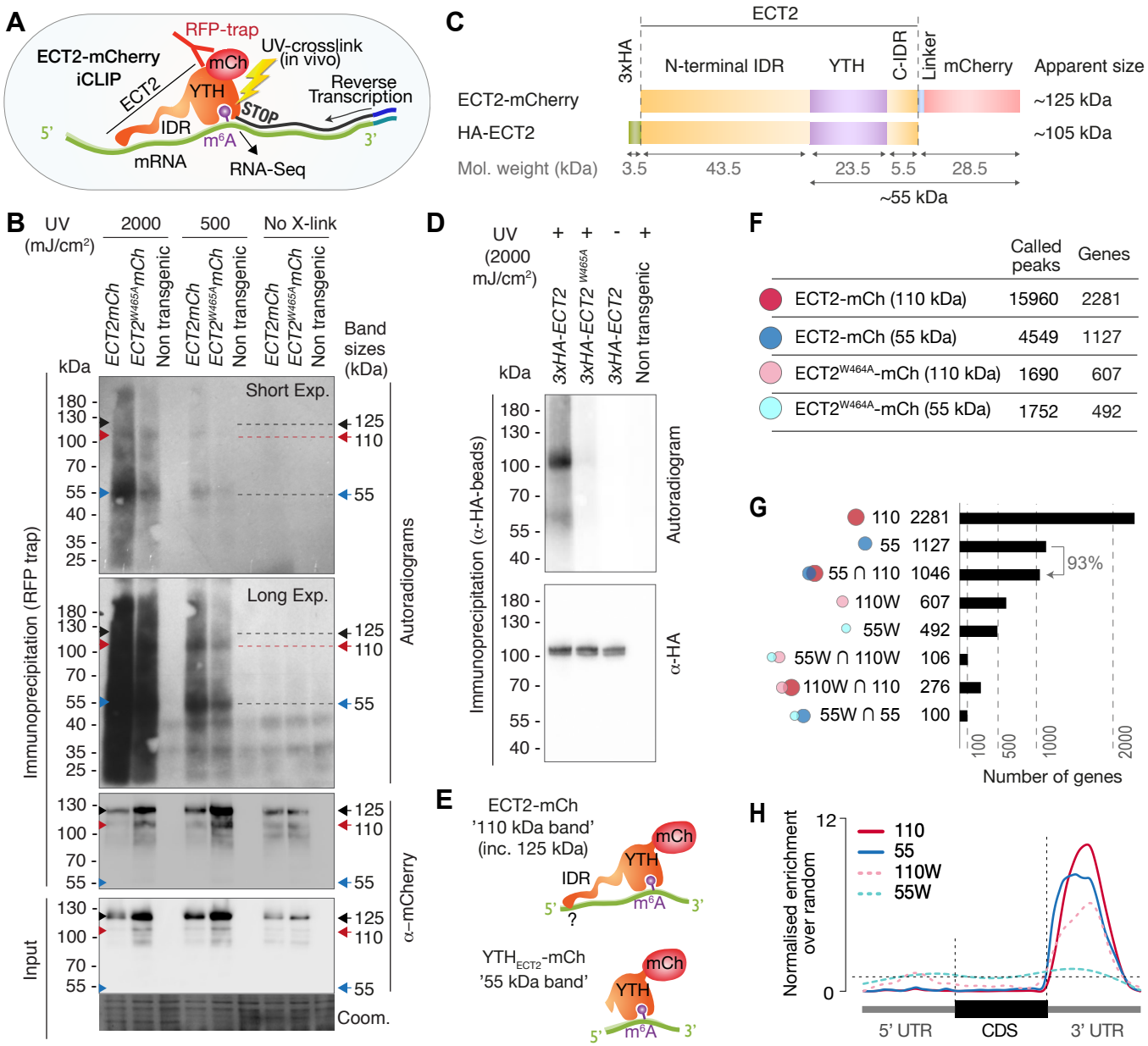

Figure 2. RNA-binding properties of ECT2 revealed by CLIP

(A) iCLIP experimental design. **(B)** Upper panels: autoradiogram (top) and α-mCherry protein blot (below) of RFP-trap immuno-purifications. Samples are cell extracts from 12-day-old seedlings expressing *ECT2-mCherry* or *ECT2^{W464A}-mCherry* in the *ect2-1* mutant background after in vivo UV-crosslinking as indicated, and subjected to DNase digestion, partial RNase digestion, and 5'-^{32}P labeling of RNA. Non-transgenic, Col-0 wild type. Lower panels: α-mCherry protein blot of the same extracts before immunoprecipitation (input) and Coomassie staining of the membrane. Sizes corresponding to full length ECT2-mCherry (~125 kDa) and the most apparent RNA bands are indicated with arrows. A repeat of the experiment with independently grown and crosslinked tissue is shown in Supplemental Fig. S3A. **(C)** Schematic representation of ECT2-mCherry and HA-ECT2 fusion proteins with their apparent size (electrophoretic mobility). The molecular weight of each region is indicated. Notice that IDRs tend to show higher apparent sizes (low electrophoretic mobility) than globular domains. **(D)** Equivalent to B with lines expressing *3xHA-ECT2* variants in the *ect2-1* background, α-HA immuno-purifications and α-HA detection by western blot. **(E)** Cartoon illustrating the 2 most apparent bands of labelled RNA co-purifying with ECT2-mCherry. **(F)** Number of called peaks and genes detected from the 4 iCLIP libraries sequenced for this study (Supplemental Fig. S4). **(G)** Upset plot showing single and pairwise combinations of genes for the 4 sequenced iCLIP libraries. **(H)** Metagene profiles depicting the enrichment along the gene body (5'UTR, CDS or 3'UTR) of the called iCLIP peaks detailed in E.

Thus, UV irradiation of intact *Arabidopsis* seedlings followed by immunopurification successfully captures ECT2-RNA complexes that exist *in vivo*. Curiously, although the pattern of ECT2-RNA complexes with bands migrating at ~110 and 55 kDa is highly reproducible, it does not correspond to the majority of the purified ECT2-mCherry protein which runs at ~125 kDa in SDS-PAGE (Fig. 2B,C). A variety of control experiments (Supplemental Fig. S3C-E), most importantly the disappearance of additional bands with use of an N-terminal rather than a C-terminal tag (Fig. 2D), indicates that the band pattern arises as a consequence of proteolytic cleavage of the N-terminal IDR in the lysis buffer, such that fragments purified using the C-terminal mCherry tag include the YTH domain with portions of the IDR of variable lengths (Supplemental Fig. S3F). Still, it is noteworthy that RNA co-migrating with the most abundant ECT2-mCherry fragment (full-length, ~125 kDa) is barely labeled while the strongest signal appears at ~55 kDa (the size of the YTH domain fused to mCherry, Fig. 2C), where protein abundance is below the western blot detection limit (Fig. 2B and Supplemental Fig. S3A). This observation suggests limited accessibility of 5'-ends of ECT2-bound RNA to polynucleotide kinase (Supplemental Fig. S3F). Comparative analysis of RNA migrating at 55 and 110-125 kDa may, therefore, provide insight into the mode of RNA interaction of ECT2 (Fig. 2E). Thus, we prepared separate iCLIP libraries from RNA crosslinked to ECT2-mCherry/ECT2^{W464A}-mCherry that migrates at ~110-280 kDa ('110 kDa band'), and at ~55-75 kDa ('55 kDa band') (Supplemental Fig. S4A).

ECT2-mCherry iCLIP peaks are enriched in the 3'UTR of mRNAs

During iCLIP library preparation (Supplemental Fig. S4A,B; see Methods), we noticed that compared to their W464A or 55-kDa-band counterparts, the samples containing wild type and full-length protein required fewer PCR cycles to obtain similar amounts of library DNA (Supplemental Fig. S4C) and generated more unique reads (Supplemental Fig. S4D). This further supports the idea that most of the RNA co-purifies with full-length wild type ECT2-mCherry. We identified a total of 15,960 iCLIP 'peaks' or crosslink sites (i.e. single nucleotide positions called by PureCLIP from mapped iCLIP reads (Krakau et al. 2017)) in 2,281 genes from the 110-kDa band of wild type ECT2-mCherry (henceforth referred to as ECT2 iCLIP peaks and targets, respectively). In the corresponding 55-kDa band, 4,549 crosslink sites in 1,127 genes were called, with 93% of them being contained in the 110-kDa target set (Fig.

2F,G, Supplemental Fig. S4D-F and Supplemental Dataset 2). Importantly, the majority of crosslink sites for both types mapped to the 3'-UTRs of mRNAs (Fig. 2H, see Fig. 3A and Supplemental Fig. S5 for examples), confirming that ECT2 binds to mRNA in the transcript region where m^6A is deposited (Fig. 3B) (Parker et al. 2020). Accordingly, the 3'-UTR bias was largely lost in RNA isolated from 55 kDa $ECT2^{W464A}$ (Fig. 2H), and called iCLIP targets in ECT2 WT and $ECT2^{W464A}$ overlapped only marginally (Fig. 2G), providing molecular proof of the dependence of m^6A-binding activity for ECT2 function.

iCLIP sites tend to be in the vicinity of HyperTRIBE editing sites

To evaluate the congruence of the results obtained by iCLIP and HyperTRIBE, we investigated the cumulative number of iCLIP sites as a function of distance to the nearest editing site determined by HyperTRIBE. This analysis showed a clear tendency for iCLIP peaks called with $ECT2^{WT}$-mCherry, but not for $ECT2^{W464A}$-mCherry, to be in the vicinity of editing sites (Fig. 3C), strongly supporting the idea that the majority of called iCLIP peaks identify genuine ECT2 binding sites on mRNAs. Similar tendencies of proximity between iCLIP peaks and HyperTRIBE editing sites were previously observed for a *Drosophila* hnRNP protein (Xu et al. 2018). Although manual inspection of individual target genes (Fig. 3A, Supplemental Fig. S5) confirmed these tendencies, it also revealed that ADAR-edited sites are too dispersed around iCLIP peaks to give precise information on the actual ECT2-binding sites. Therefore, we used HyperTRIBE and iCLIP for gene target identification, but relied only on iCLIP peaks for motif analyses.

ECT2 targets identified by iCLIP and HyperTRIBE overlap m^6A–containing transcripts

To examine the quality of our target identification in further detail, we analyzed the overlap between ECT2 targets identified by iCLIP and HyperTRIBE. We also included in this analysis m^6A mapping data obtained with either m^6A-seq (Shen et al. 2016) or the single-nucleotide resolution methods miCLIP and Nanopore sequencing (Parker et al. 2020), as young seedlings were used in all cases. ECT2 targets identified by iCLIP and HyperTRIBE showed clear overlaps, both with each other and with m^6A-containing transcripts, regardless of the method used to identify them (Fig. 3D and Supplemental Fig. S6). This observation supports the conclusion that ECT2 target identification via combined iCLIP and HyperTRIBE

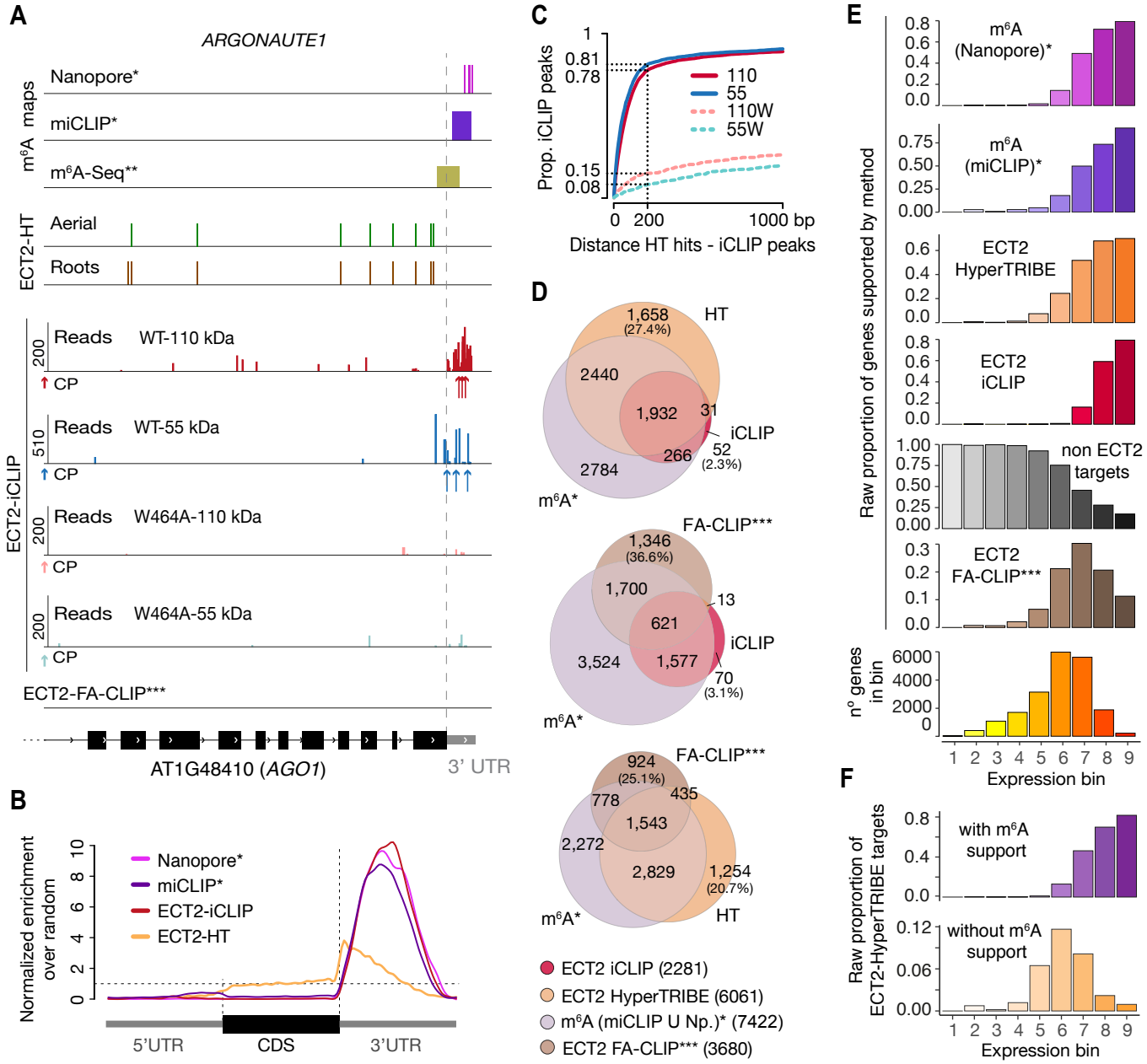

Figure 3. iCLIP identifies bona-fide ECT2 targets

(A) Example of an ECT2 target (AGO1) showing the distribution of m^{6}A sites, ECT2-iCLIP reads and peaks, ECT2-HT edited sites, and FA-CLIP peaks along the transcript. CP, called peaks. More examples are shown in Supplemental Fig. S5. **(B)** Metagene profiles comparing the distributions along the gene body of ECT2-mCherry iCLIP peaks (wild type, 110-kDa band), ECT2-HT editing sites (in roots and aerial tissues), ECT2 FA-CLIP sites, and m^{6}A miCLIP and Nanopore-identified sites. **(C)** Proportion of ECT2 iCLIP peaks within a given distance from the nearest ECT2-HT edited site. Numbers indicated on the y-axis show the proportion of ECT2 iCLIP peaks less than or equal to 200 nt from the nearest ECT2-HT edited site. **(D)** Overlap between genes supported as containing m^{6}A or ECT2 targets by the different techniques indicated. The ECT2-HT target set includes the sum of targets identified in root and aerial tissues. Additional overlaps are shown in Supplemental Fig. S6. **(E)** Proportions of genes in each expression bin either containing m6A or supported as ECT2 targets by the same techniques as in A and C. **(F)** Proportion of ECT2-HT targets with or without support from m^{6}A data (Nanopore*, miCLIP* or m^{6}A-Seq**) in each expression bin.

* Parker et al. (2020); ** Shen et al. (2016); *** Wei et al. (2018).

approaches is of high quality. Furthermore, as m^{6}A maps were generated using loss-of-function mutants of the core *N6*-adenosine methyltransferase components FIP37 and VIR (Shen et al. 2016; Parker et al. 2020) as background controls, it confirms the previously challenged assumption that ECT2 binds to mRNAs with m^{6}A deposited by the canonical methyltransferase. Importantly, although some m^{6}A-targets are expected not to be bound by ECT2 because of the presence of MTA in cells that do not express ECT2 (Arribas-Hernández et al. 2020), only 18% of the high-confident set of m^{6}A-containing genes (with support from miCLIP and Nanopore) did not overlap with either ECT2 iCLIP or HT target sets (Supplemental Fig. S6, arrow). We also observed that HyperTRIBE identifies ~3 times more ECT2 targets than iCLIP, possibly because of the bias towards high abundance inherent to purification-based methods like iCLIP. To test this idea, we compared the distribution of target mRNAs identified by the different techniques across 9 expression bins defined similarly to those in Fig. 1M. As expected, a bias towards highly abundant transcripts was evident for iCLIP-identified targets compared to HyperTRIBE (Fig. 3E). We also observed a similar bias for m^{6}A-containing transcripts detected by miCLIP, another purification-based method, and to a lesser extent in the Nanopore dataset (Fig. 3E), probably explained by its relatively low sequencing depth (Parker et al. 2020). These observations also suggest that the higher sensitivity of HyperTRIBE might explain the lack of m^{6}A support (by Nanopore or miCLIP) for 28% of ECT2 HT-targets (1,689) compared to only 4% (83) of ECT2 iCLIP targets (Fig. 3D and Supplemental Fig. S6, upper row): HT-targets may simply include genes that escape detection by m^{6}A mapping methods due to low expression. Indeed, ECT2-HT targets without any m^{6}A support were distributed in lower-expression bins compared to those with m^{6}A support (Fig. 3F). Intriguingly, ECT2 FA-CLIP targets (Wei et al. 2018) did not show a bias towards highly expressed genes as their distribution over expression bins largely reflected that of the total number of genes (Fig. 3E), and as many as 37% of FA-CLIP targets did not have m^{6}A support (Fig. 3D and Supplemental Fig. S6, upper row). In summary, ECT2 iCLIP and HT target sets are in excellent agreement with each other and with independently generated m^{6}A maps, and HyperTRIBE identifies targets below the detection limit of other techniques.

ECT2 crosslink sites coincide with m^{6}A miCLIP-sites and are immediately upstream of Nanopore m^{6}A-sites

To analyze patterns of sequence composition around ECT2 binding sites, we first examined the positions of m^{6}A sites at single-nucleotide resolution (Parker et al. 2020) relative to ECT2 iCLIP crosslink sites (peaks). This analysis showed that full length ECT2 crosslinks in the immediate vicinity, but preferentially upstream (~11 nt) of Nanopore-determined m^{6}A sites, with a mild depletion at the exact m^{6}A site, and vice versa (Fig. 4A, upper left panels). Furthermore, while m^{6}A-miCLIP sites corresponded to m^{6}A Nanopore sites overall, a subset of miCLIP sites were located upstream of Nanopore sites and coincided well with ECT2-iCLIP peaks (Fig. 4A, left panels). Because UV illumination used in both iCLIP and miCLIP preferentially generates RNA-protein crosslinks involving uridine (Hafner et al. 2021), also detectable in the datasets analyzed here (Fig. 4A, right panels, Fig. 4B), the depletion of ECT2-iCLIP sites at Nanopore-, but not miCLIP-determined m^{6}A sites might be explained in part by the absence of uridine within the RR<u>A</u>C core of the m^{6}A consensus motif, and perhaps in part by reduced photoreactivity of the m^{6}A base stacking with indole side chains of the YTH domain.

DRACH, GGAU and U/Y-rich motifs are the most enriched around m^{6}A/ECT2-sites

The 5' shift observed for iCLIP and miCLIP sites relative to Nanopore sites could be explained by a higher occurrence of uridines upstream of m^{6}A sites, a particularly interesting possibility given the numerous reports of U-rich motifs enriched around m^{6}A sites in plants (Li et al. 2014c; Anderson et al. 2018; Miao et al. 2019; Zhang et al. 2019; Zhou et al. 2019; Luo et al. 2020) and animals (Patil et al. 2016). To investigate the sequence composition around m^{6}A and ECT2 sites, we first performed exhaustive unbiased *de novo* motif searches using Homer (Heinz et al. 2010) with a variety of settings and backgrounds, and extracted candidate motifs, including the m^{6}A consensus motif RRACH, as well as GGAU (Anderson et al. 2018), URUAY (Wei et al. 2018) and several further U-rich motifs (Supplemental Fig. S7). Combined with manually derived candidate motifs and those identified from published studies, we then calculated position weight matrices (PWM) for the final set of 48 motifs (Supplemental Fig. S8) and scanned for their occurrences genome-wide using FIMO (Grant et al. 2011) (Supplemental Fig. S7). This allowed us to determine three key properties (Supplemental

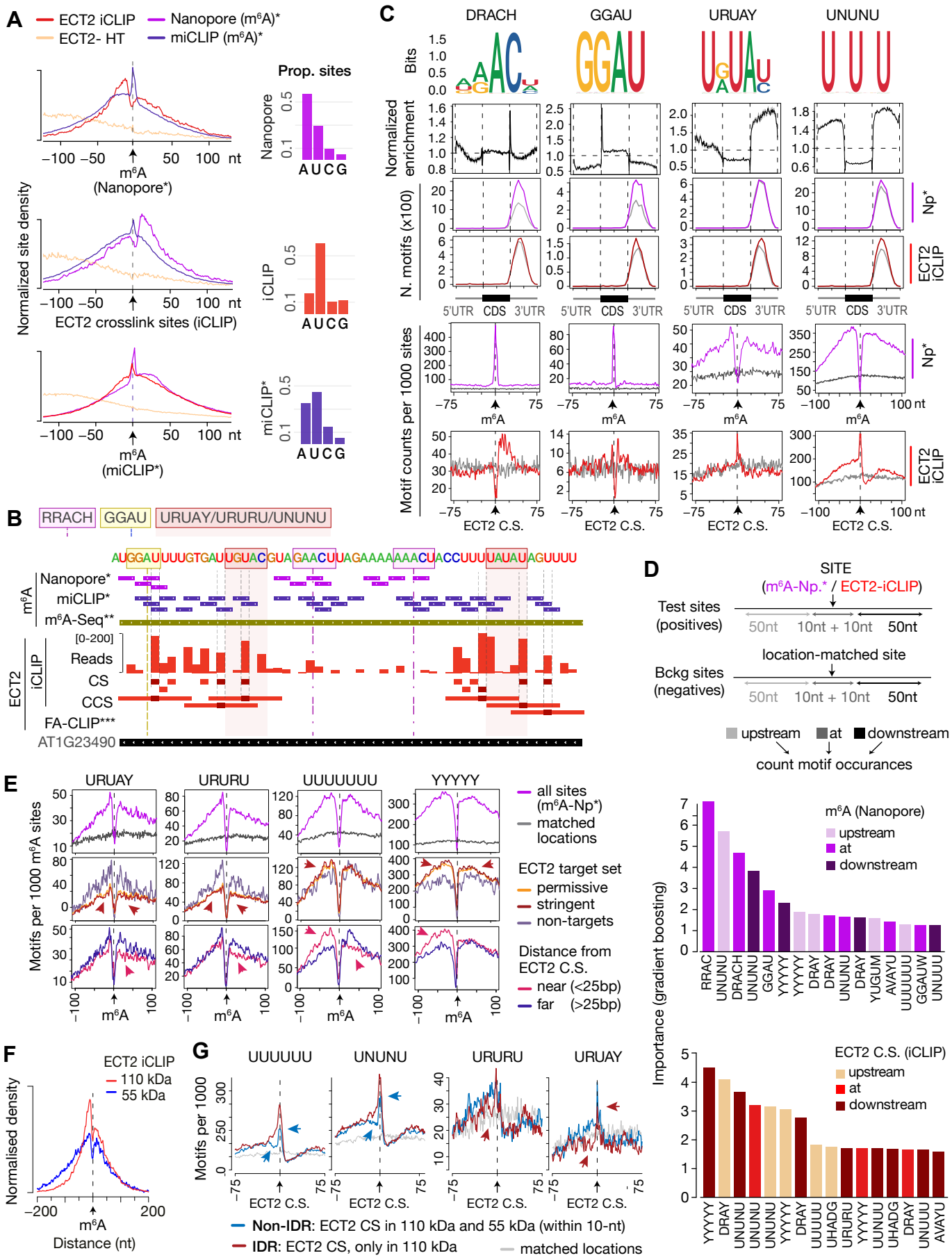

Figure 4. ECT2 preferentially crosslinks to U-rich sequence motifs enriched in the flanking regions of DRACH and GGAU-containing m^{6}A sites.

(continues on the next page)

(continues from previous page)

Figure 4. ECT2 preferentially crosslinks to U-rich sequence motifs enriched in the flanking regions of DRACH and GGAU-containing m^{6}A sites.

(A) Left panels: normalized density of sites (m^{6}A Nanopore* and miCLIP*, and ECT2-iCLIP and -HT) at and up to +/-100 nt of either m^{6}A Nanopore*, m^{6}A miCLIP* or ECT2 iCLIP sites. Density curves are normalized so that sets can be visually compared to one another within the same plot. Right panels: proportion of m^{6}A and ECT2 iCLIP sites at each nucleotide by the different methods. **(B)** View from IGV browser illustrating presence of RRACH, GGAU and URUAY/U-rich motifs in the vicinity of m^{6}A (Nanopore*, miCLIP* or m6A-seq**), ECT2-iCLIP (raw reads, called peaks (crosslink sites) and collapsed crosslink sites (CSS)) or ECT2 FA-CLIP*** positions in the 3'UTR of AT1G23490 (*ARF1*). **(C)** Key motifs analyzed in this study. Numbered from top to bottom: (1) motif logos for derived position weight matrices (PWMs); (2) normalized enrichment of motif locations across gene body (5'UTR,CDS, or 3'UTR); (3-4) total number of the relevant motif found at Nanopore*-m^{6}A (3) or ECT2-iCLIP (4) sites according to gene body location. Grey lines indicate numbers found in a gene-body location-matched background set of sites of equivalent number; (5-6) distribution of the relevant motif relative to m^{6}A-Nanopore (5) or ECT2 –iCLIP (6) sites (number of motifs per 1000 sites). Grey lines represent the distribution for the same gene-body location-matched set as derived in the panels above. **(D)** Top: diagram representing strategy for machine learning model trained to distinguish either m^{6}A nanopore peaks or ECT2-iCLIP peaks from their respective gene-body location matched background sets. Bottom: Bar plots showing top 16 motif feature importance scores from the two models, ordered from left to right by importance. **(E)** Top panels: Distance-based enrichment of motifs at and around m^{6}A-Nanopore* sites, plotted as motif counts per 1000 m^{6}A sites (purple lines). Grey lines indicate the enrichment in a location-matched background set as in C. Middle and bottom panels: sites are split according to whether they sit on ECT2 targets (middle), or to distance from the nearest ECT2 crosslink site (for ECT2-iCLIP targets only) (bottom). **(F)** Motifs per 1000 ECT2 iCLIP crosslink sites split according to whether they are found in libraries from both 110 kDa and 55 kDa bands ('Non-IDR'), or exclusively (distance > 10 nt) in the 110 kDa band ('IDR'). Grey lines indicate the enrichment in a location-matched background set as in C. * Parker et al. (2020); ** Shen et al. (2016); *** Wei et al. (2018)

Dataset 3). First, the global enrichment of the motifs at locations across the gene body. Second, the total number of motifs found at m^{6}A sites and ECT2 iCLIP crosslink sites compared to a set of sites in non-target mRNAs matching the location within gene bodies of m^{6}A/ECT2 iCLIP sites (expected background). Third, the distribution of the motifs relative to m^{6}A and ECT2 iCLIP sites. The results of this systematic analysis were used to select those with a more prominent enrichment at or around m^{6}A and ECT2 sites. This approach defined two major categories of motifs of outstanding interest, RRACH-like and GGAU on one side, and a variety of U/Y-rich motifs on the other. Fig. 4C shows a minimal selection of such motifs, while a more comprehensive compilation is displayed in Supplemental Figs. S9-S10. Not surprisingly, RRACH-like motifs were the most highly enriched at m^{6}A sites, with the degenerate variant DRACH being the most frequently observed (Fig. 4C and Supplemental Fig. S9). Motifs containing GGAU behaved similarly to DRACH, with a sharp enrichment exactly at m^{6}A sites and mild enrichment downstream ECT2 peaks (Fig. 4A), supporting a previous suggestion of GGAU as an alternative methylation site (Anderson et al. 2018). The possible roles of the U/Y-rich motifs in m^{6}A deposition and ECT2 binding are analyzed in the following sections.

Addition of neighboring U/Ys results in enriched RRACH- and GGAU-derived motifs

We first addressed the role of U-rich motifs. The distribution of these motifs relative to m^6A showed a progressive 5'-shift by the addition of Us upstream of DRACH/GGAU (e.g. UGAAC/UGGAU), and a 3'-shift by the addition of Us or C/Us (Ys) downstream (e.g. ACUCU) (Supplemental Fig. S10). Importantly, such motifs showed a clear enrichment over location-matched background sites along the gene body at ECT2-iCLIP sites (Supplemental Fig. S10), indicating that ECT2 preferentially crosslinks to uridines located in the immediate vicinity of DRACH (/GGAU). U/R rich motifs without traces of DRACH (e.g. YUGUM, URUAY, URURU) showed a characteristic enrichment upstream, but depletion at the m^6A site itself (Fig. 4C and Supplemental Fig. S10). The distance between the motif enrichment and the m^6A site roughly coincided with the shift observed in ECT2 crosslink sites relative to m^6A (Fig. 4A) and, accordingly, these motifs were enriched exactly at ECT2 crosslink sites (Fig. 4C and Supplemental Fig. S10), suggesting that they may constitute additional m^6A-independent sites of interaction with ECT2.

Finally, UUUUU/UNUNU and YYYYY motifs were highly enriched ~25 nt upstream and ~5-10 nt downstream of m^6A (Fig. 4C and Supplemental Fig. S10). The 3' enrichment of YYYYY was closer to m^6A than that of UUUUU, indicating a preference for pyrimidines immediately downstream the m^6A site, as suggested by the enrichment of DRACUCU motifs described above. The distribution of these U/Y-rich motifs around ECT2 peaks also confirmed these ideas, because UUUUU was highly enriched upstream, but depleted (relative to background) downstream of ECT2-crosslink sites, while the 3' YYYYY enrichment followed immediately adjacent to the dip at the Y-depleted m^6A consensus site (Supplemental Fig. S10). Taken together, these results suggest that *N6*-adenosine methylation preferentially occurs in DRACH/GGAU sequences surrounded by stretches of pyrimidines, with a preference for YYYYY (e.g. CUCU) immediately downstream, URURU (inc. URUAY) immediately upstream, and UUUUU/UNUNU slightly further away in both directions (Supplemental Fig. S11).

DRACH/GGAU motifs are the most important determinants for m^6A deposition at the site, while flanking U(/Y)-rich motifs are essential for m^6A presence and ECT2 binding

Since our analysis thus far uncovered several motifs of potential importance for m^6A

deposition and ECT2 binding, we employed machine learning to distinguish either m^{6}A sites or ECT2 iCLIP sites from random location-matched background sites using motif-based features. Importantly, the underlying classification model includes all motif features within the same model, allowing an evaluation of the importance of the motif features relative to each other. We used as features the number of matches to each of the 48 motifs (Supplemental Figs. S7 and S8) according to their overlap directly with the methylated or ECT2-bound site (”at”) or within either the upstream or downstream flanking regions (”up” or “down”, respectively, Fig. 4D). The model involving all motifs could successfully distinguish the methylated sites from the background sites as indicated by an area under the receiver operating characteristic curve (true positive rate versus false positive rate, AUC) of 0.93, and even a reduced model incorporating only the top 10 features from the full model classified sites largely correctly (AUC = 0.86; Supplemental Fig. S12). The top 16 features ordered by importance from the full model confirmed that RRAC/DRACH or GGAU at the site was indicative of the presence of m^{6}A (Fig. 4D). Interestingly, U/Y-rich sequences (UNUNU and YYYYY in particular) flanking the site were also strongly indicative. Some motifs showed a skew in their feature importance score, with UNUNU and YUGUM showing an upstream and YYYYY a downstream preference, thus corroborating our previous observations.

An analogous modeling approach using ECT2 iCLIP sites compared to random matched background sites also showed flanking U-rich sequences. In agreement with the strong enrichment of ACUCU-containing motifs immediately downstream of m^{6}A and at ECT2 sites (AUC=0.97, and AUC=0.94 using only the top 10 features, Supplemental Fig. S12), YYYYY (Y=U/C) showed a particularly strong downstream preference and ranked above UUUUU-like motifs, confirming that combinations of U and C downstream of RRACH likely play a role in ECT2 binding (Fig. 4D).

The U(-R) paradox: URURU-like sequences around m^{6}A sites repel ECT2 binding, while U-rich sequences upstream enhance its crosslinking

To investigate the idea of URURU-like motifs as additional sites of ECT2 binding upstream of the m^{6}A-YTH interaction site, we split Nanopore-m^{6}A sites according to two criteria: 1) whether the transcript binds to ECT2 (i.e., it is a target or not), and 2) for ECT2 targets, whether there is an ECT2 crosslink site within 25 nt of the m^{6}A site (near) or not (far).

Although there was no obvious difference between these categories for most of the motifs (Supplemental Dataset 3), U-rich sequences displayed distinctive features (Fig. 4E and Supplemental Fig. S13) that can be summarized as follows. If a transcript has m^{6}A and ECT2 sites in close proximity, it is: i) more likely to have UNUNU/UUUUU/YYYYY sequences upstream of the m^{6}A site than targets with distantly located ECT2 binding sites (or than non-ECT2 targets); ii) less likely to have UUUUU/URURU sequences downstream of the m^{6}A site, possibly because ECT2 prefers CUCU-like sequences downstream as supported by downstream enrichment of YYYYY in ECT2 targets and the results above; iii) less likely to have URURU/URUAY-like motifs upstream of the m^{6}A site. However, for the specific subset of ECT2-bound m^{6}A-sites with URURU/URUAY upstream, these sequences tend to crosslink to ECT2, as seen by the enrichment spike at ECT2 crosslink sites (Fig. 4C and Supplemental Fig. S13, bottom panels). Although these two results seem contradictory at first glance, they may be reconciled by a model in which a URURU/URUAY-binding protein would compete with ECT2 for binding adjacent to m^{6}A. If that protein is absent, ECT2 may bind to the site, potentially via its IDR, to stabilize the low-affinity YTH-m^{6}A interaction and crosslink efficiently due to the U-content. Conversely, if occupied by the alternative interacting protein, the site might repel ECT2 (see discussion and Supplemental Fig. S11).

The N-terminal IDR of ECT2 is involved in preferential crosslinking at U-rich sequences and URURU-repulsion immediately upstream m^{6}A sites

We reasoned that insights into contacts between ECT2 and RNA may be gained by analysis of the iCLIP libraries prepared with the YTH-mCherry truncation ('55 kDa band') (Fig. 2 and Supplemental Fig. S4). Initial inspection of the distribution of ECT2 peaks relative to Nanopore-m^{6}A sites showed that the 5'-3' asymmetry observed with full-length ECT2 ('110 kDa') (Fig. 4A) was greatly reduced with the truncated protein (Fig. 4F), as was the bias towards uridines (Supplemental Fig. S14A,B). These observations suggest that the IDR indeed is implicated in binding to U-rich regions upstream of m^{6}A. We next split the full-length ECT2 iCLIP peaks according to whether they are present in libraries from both full length and truncated (devoid of the N-terminal IDR) forms ('Non-IDR'), or exclusively in the full length ('IDR') (distance > 10 nt), and plotted the enrichment of the studied motifs relative to the crosslink site (Fig. 4G, Supplemental Fig. S14C and Supplemental Dataset 3).

UUUUUU/UNUNU-like motifs were depleted immediately upstream of non-IDR crosslink sites relative to IDR crosslink sites, supporting preferential crosslinking of the IDR to Us in this region. Remarkably, the exact opposite was true for URURU/URUAY motifs that showed modest depletion 5' to IDR crosslink sites relative to their non-IDR counterparts (Fig. 4G). These observations are consistent with a model of an RNA binding protein competing with the ECT2 IDR for interaction with upstream URURU/URUAY motifs (Supplemental Fig. S13).

ECT2 and ECT3 bind to overlapping sets of targets

Having identified the sequence elements implicated in m^6A deposition and ECT2 mRNA interaction, we moved on to address possible molecular effects of ECT2 mRNA binding. Since observation of the major *in vivo* effect of ECT2 in organogenesis requires simultaneous knockout of *ECT2* and *ECT3*, an effect that is exacerbated by additional mutation of *ECT4* (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020), we reasoned that a meaningful analysis of effects of ECT proteins on target mRNAs necessitates the use of *ect2/ect3*/(*ect4*) mutants, and hence also an assessment of the degree to which ECT2 and ECT3 target mRNAs overlap. We chose HyperTRIBE to identify targets of the less highly expressed ECT3 (Arribas-Hernández et al. 2018), because of its efficient capture of ECT2 targets with little expression bias, and proceeded in exactly the same way as described for ECT2 with selection of suitable transgenic lines, mRNA-seq and data analysis (Supplemental Figs. S15-S17 and Supplemental Dataset 4). Despite the lower expression of *ECT3* compared to *ECT2* and, consequently, generally lower editing proportions in *ECT3-FLAG-ADAR* lines compared to *ECT2-FLAG-ADAR* lines (Supplemental Figs. S15E and S17B), the implementation of our rigorous statistical approach to call significant editing sites successfully identified 2,451 targets in aerial tissues, and 3,498 in roots (ECT3 HT-targets). Remarkably, in both aerial and root tissues, roughly 95% of ECT3 HT-targets overlapped with the larger group of ECT2 HT-targets (Fig. 5A), and for many targets, e.g. *ATP-Q* (AT3G52300), the pattern of editing sites resulting from fusion of ADAR to ECT2 and ECT3 was very similar (Fig. 5B). We also noticed a few examples with apparently preferential targeting by ECT2 (e.g. *TUA4*, AT1G04820) or by ECT3 (e.g. *UBQ6*, AT2G47110) (Fig. 5B), perhaps hinting to molecular explanations for the recently described non-redundant roles of ECT2 and ECT3 in determining root growth directionality (Arribas-Hernández et al. 2020). Overall, however, the

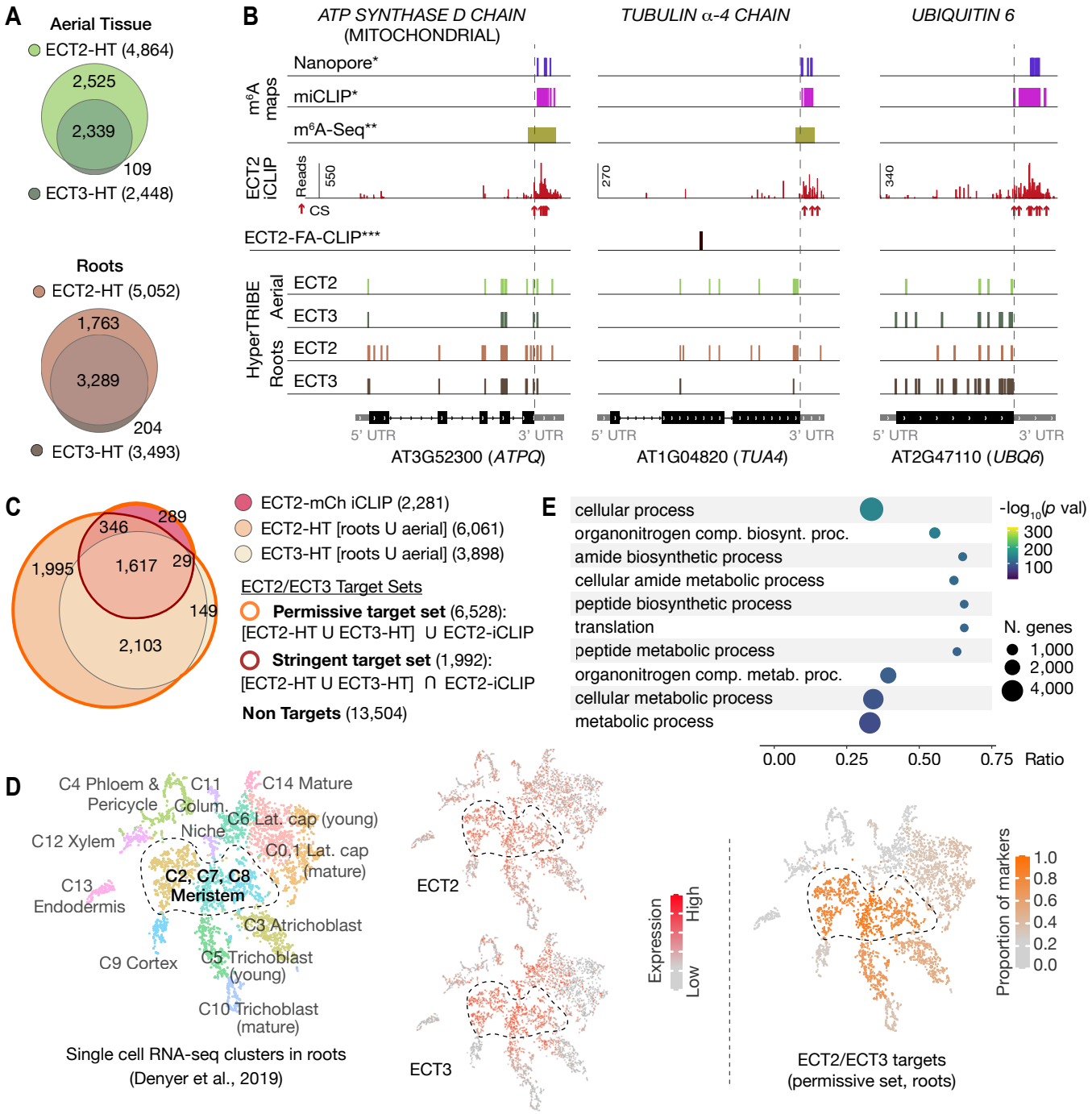

Figure 5. ECT2 and ECT3 target sets overlap to a large extent

(A) Venn diagrams showing overlap of ECT2-HT and ECT3-HT targets, for the set of commonly expressed genes, separately for aerial tissues and roots. **(B)** Examples of common ECT2 and ECT3 targets (*ATPQ, TUA4* and *UBQ6*) showing the distribution of ECT2/3-HT editing sites sites in either roots or shoots along the transcript. m^{6}A sites, ECT2-iCLIP reads and crosslink sites (CS), and FA-CLIP peaks are also indicated. * Parker et al. (2020); ** Shen et al. (2016); *** Wei et al. (2018). **(C)** Venn diagrams showing overlap between ECT2-iCLIP target genes with ECT2-HT and ECT3-HT target gene sets. Regions outlined in bold orange and red indicate the defined permissive and stringent ECT2/3 target sets in whole seedlings, respectively (Supplemental Dataset 4, see also Supplemental Fig. S17E,F for aerial and root-specific target sets). Non-targets are all genes with detectable transcript levels in the ECT2 or ECT3 HT RNA-Seq datasets that are not in the permissive target set. **(D)** Left: t-SNE plot for scRNA-seq data in roots from Denyer et al. (2018), with cells colored according to their cell-type cluster definitions (see Supplemental Fig. S18 for details). Center: ECT2 and ECT3 single cell expression levels overlayed on to the t-SNE plot (Ma et al., 2020). Right: t-SNE plot with cell-type clusters shaded according to the proportion of marker genes from Denyer et al. (2019), which are targets of ECT2 or ECT3 in roots. Dashed enclosed region indicates clusters that contain meristematic cells. **(E)** The 10 most significantly enriched GO terms among ECT2/3 targets (permissive set).

overwhelming overlap between ECT2 and ECT3 HT-targets in both tissues analyzed is consistent with their largely redundant functions in leaf and root formation. We also observed that ECT2 and ECT3 target their own and each other's mRNAs (Supplemental Fig. S17D).

We next used all of our target sets (ECT2 iCLIP, ECT2-HT, ECT3-HT) to define three gene sets of particular interest for functional analysis of ECT2/3/(4): **The stringent target set (**1,992 genes) defined as all ECT2 or ECT3 HT-targets supported by ECT2 iCLIP, **the permissive target set** (6,528 genes) defined as genes with either ECT2 HT, ECT3 HT or ECT2 iCLIP support, and **the non-target set** (13,504 genes) defined as all expressed genes not contained in the permissive target set (Fig. 5C, Supplemental Fig. S17E-F and Supplemental Dataset 5). As an initial check of consistency of the target sets with the biological context in which ECT2 and ECT3 function, we used single-cell transcriptome analysis of Arabidopsis roots (Denyer et al. 2019; Ma et al. 2020) to analyze the overlap of ECT2/3 expression with the enrichment of markers for different cell types in the permissive target set. This analysis showed reassuring congruence between predominant expression of ECT2/3 in meristem clusters and marker enrichment for these same clusters among targets (Fig. 5D and Supplemental Fig. S18). We also analyzed the permissive target sets for groups of functionally related genes, and found that ECT2/3 targets are enriched in housekeeping genes, many related to basic metabolism and protein synthesis (Fig. 5E). Taken together, our results show that the genetic redundancy between ECT2 and ECT3 extends to an overlap in mRNAs directly bound by the two proteins, and provide well-defined common ECT2/ECT3 target sets for further functional analysis.

Recovery of ECT2-expressing cell populations with and without ECT2/(ECT3/ECT4) activity

ECT2, *ECT3* and *ECT4* expression is largely restricted to rapidly dividing cells of organ primordia (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020), and since many ECT2/3 targets are broadly expressed (Fig. 5E), cell populations expressing ECT2-4 need to be isolated prior to transcriptome analyses to avoid confounding effects from cells not expressing the m^6A readers. We therefore used the fact that *ect2-1*/*ECT2-mCherry* exhibits wild type root growth while *te234*/*ECT2*^{W464A}*-mCherry* exhibits defects in root growth nearly identical to *te234* triple knockouts (Arribas-Hernández et al. 2020), and used fluorescence-

associated cell sorting to select mCherry-expressing cell populations from root protoplasts of three independent transgenic lines for each the two genetic backgrounds (Fig. 6A). Because wild type and mutant fluorescent proteins have the same expression level, pattern, and intracellular localization in these lines (Fig. 6B,C), this procedure yielded comparable ECT2-expressing cell populations (Fig. 6D and Supplemental Fig. S19) with (*ECT2-mCherry/ect2-1/ECT3/ECT4*, henceforth "wild type") or without (*ECT2*^{W464A}*-mCherry/ect2/ect3/ect4,* henceforth "mutant") ECT2/3/4 function. We therefore isolated mRNA and constructed Smart-Seq2 libraries for comparison of location of poly(A) clusters (PACs) and abundance of ECT2/3 targets and non-targets in *ECT2*-expressing cells from plants of the two different genetic backgrounds. Compared to standard mRNA-seq, Smart-seq2 recovers more reads with untemplated As in addition to gene-specific sequence and can, therefore, be used for PAS mapping. We note that the selection of ECT2-expressing cells from the root meristem division zones of wild type and mutant lines also circumvents the trouble of preparing comparable samples from intact tissues of plants at different developmental stages.

ECT2(/3/4) do not play a direct role in alternative polyadenylation of targets

We first addressed the conjecture on a nuclear role of ECT2 in PAS selection (Wei et al. 2018). In plants, PASs are not sharply defined but rather spread along localized regions and can be grouped into PAS clusters (PACs) for analysis (Wu et al. 2011; Sherstnev et al. 2012). Using a modification of the nanoPARE analysis pipeline (Schon et al. 2018) to map PASs from reads with ≥ 9 untemplated As, we identified a total of 14,667 PACs belonging to 12,662 genes after filtering possible false positives (see Methods, Supplemental Fig. S20A,B and Supplemental Dataset 6). We found no tendency for ECT2/3 target mRNAs to have more PACs than non-targeted genes (Supplemental Fig. S20C), suggesting that differential PAC location in ECT2/3 targets between mutant and wild type is not prevalent. Nevertheless, we specifically tested whether PASs could be affected by the loss of ECT2/3/4 function in two different ways: either a shift of the dominant PAC to an alternative PAC altogether, or a shift in the most common PAS within clusters. Sorting the 206 genes for which the dominant PAC differed between wild type and mutant samples (18.5% of the 1,114 genes with more than one PAC) into the ECT2/3 target groups (Supplemental Fig. S17F) showed that both the permissive and stringent targets were significantly less likely than non-targets to have a

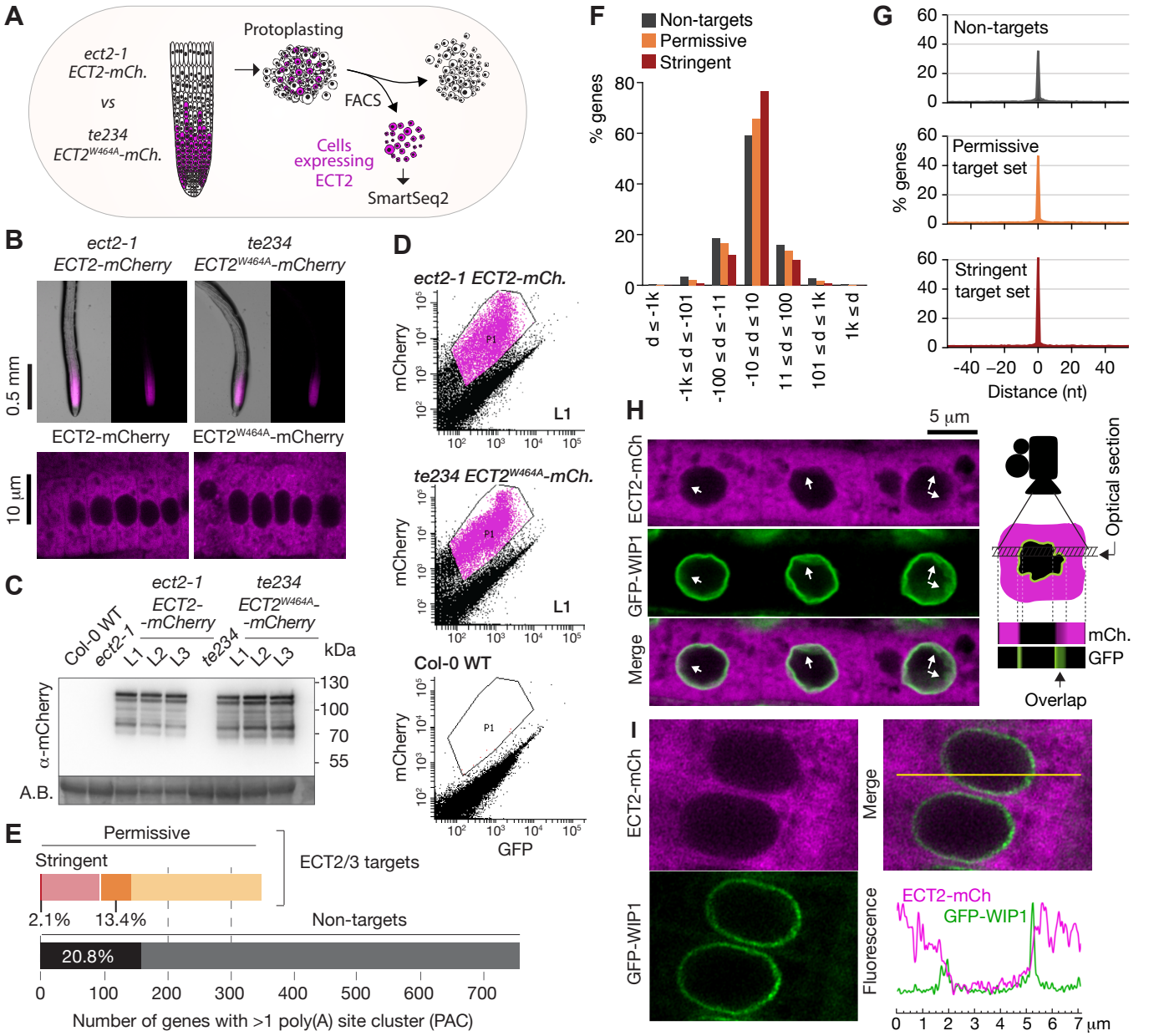

Figure 6. Poly(A) sites in ECT2/3 targets do not change upon loss of ECT2/3/4 function

(A) Experimental design. **(B)** Expression pattern of ECT2-mCherry in root tips of *ect2-1 ECT2-mCherry* and *te234 ECT2^{W464A}-mCherry* genotypes by fluorescence microscopy. **(C)** Protein blot showing expression levels of ECT2-mCherry in the 3+3 lines of *ect2-1 ECT2-mCherry* and *te234 ECT2^{W464A}-mCherry* used as biological replicates for FACS selection of ECT2-expressing cells. Amido black (A.B.) is used as loading control. **(D)** Fluorescence profile (mCherry vs. GFP fluorescence) of root cells (protoplasts) from the transgenic lines in C. The complete set of lines/samples is shown in Supplemental Fig. S19. Non-transgenic Col-0 WT is shown as control for background autofluorescence. Cells with a fluorescence profile within the outlined areas were selected for RNA extraction, Smart-seq2 library construction and sequencing. **(E)** Genes with more than one polyA site cluster (PAC) in the different target/non-target sets. Dark shades are genes in which the dominant PAC in *te234 ECT2^{W464A}-mCherry* samples differs from the one in *ect2-1 ECT2-mCherry*. **(F,G)** Distribution of distances (d [nt]) of the most common poly(A) site between *te234 ECT2^{W464A}-mCherry* and *ect2-1 ECT2-mCherry* samples for all genes where the most common poly(A) site could be determined in both genotypes (6,648 non-targets, 4,072 permissive targets, and 1,486 stringent targets). Negative values are upstream (5') and positive values are downstream (3') relative to the gene orientation. (F) Distances are binned by ±10, ±100, ±1000, and >1,000bp. (G) Distances are plotted by nucleotide in a ±40bp window. **(H)** Standard confocal microscopy of root cells co-expressing *ECT2-mCherry* and *GFP-WIP1*. White arrows indicate areas in which apparent spills of ECT2-mCherry signal into the nucleus overlap with blurry GFP signal from the nuclear envelope, a sign of not-perpendicularity between the envelope and the optical plane as exemplified on the right-panel cartoon. **(I)** Airyscan super-resolution confocal microscopy of root cells as in H. The image is cropped from a larger picture shown in Supplemental Fig. S21B. mCherry and GFP fluorescence signals along the yellow line show absence of ECT2-mCherry inside the limits of the GFP-labelled nuclear envelope.

different dominant PAC upon loss of ECT2/3/4 function (p=0.013 and p=1.21e-5 for strictly permissive and stringent targets respectively; Fisher's exact test) (Fig.6E and Supplemental Fig. S20C-D). This significant depletion may be an effect of the higher expression of targets compared to non-targets (Fig. 4G), as accuracy of PAS detection increases with transcript abundance (see Supplemental Fig. S20E for details). The result indicates that the alternative polyadenylation observed upon loss of ECT2/3/4 function is not prevalent among ECT2/3 targets. Finally, we examined changes to the local distribution of PASs within clusters. We defined the most common PAS as the single position in all overlapping PACs with the most reads, and determined the distance between such dominant PASs in wild type and mutant samples. Comparison of the distances revealed that the most common PAS does not change by more than 10 bp in the majority of genes, and is not more likely to be different in ECT2/3 targets than in non-targets (Fig. 6F). In fact, the most common PAS is more likely to be unchanged in targets than in non-targets (Fig. 6G) (p=0.028 and p=2.2e-16 for strictly permissive and stringent targets respectively; Fisher's exact test). Taken together, these analyses show that neither the usage of alternative PACs nor the dominant PASs within clusters have any tendency to change in ECT2/3-targets upon loss of ECT2/3/4 function.

ECT2-mCherry does not localize to the nucleoplasmic side of the nuclear envelope

To further investigate whether ECT2 may have any nuclear functions, we revisited the evidence for localization of ECT2 in the nucleoplasm, which is based on confocal fluorescence microscopy of ECT2-GFP or YFP-ECT2 in DAPI-stained root cells of stable Arabidopsis lines (Scutenaire et al. 2018; Wei et al. 2018). Because (i) the localization of ECT2-mCherry in living root cells of our lines has a general sharp boundary with what we interpreted to be the nucleus (Fig. 6B) (Arribas-Hernández et al. 2018) and does not overlap with nucleoplasmic MTA-TFP (Arribas-Hernández et al. 2020), (ii) paraformaldehyde fixation routinely used to permeate DAPI inside plant tissues (used by Scutenaire et al. and not specified by Wei et al.) can introduce artifacts in the localization of fluorescent proteins (Li et al. 2015), and (iii) the RNA-binding properties of DAPI could yield signal from the RNA-rich rough endoplasmic reticulum surrounding the nucleus (Tanious et al. 1992), we decided to examine the localization of ECT2-mCherry relative to the nuclear envelope in living cells. We therefore crossed our functional *ECT2-mCherry* lines (Arribas-Hernández et al. 2020) with

plants expressing the outer nuclear envelope and nuclear pore complex-associated protein WIP1 fused to GFP (Xu et al. 2007). Confocal fluorescence microscopy of intact roots showed that the sharp boundaries of the ECT2-mCherry expression domain were delimited by the GFP-WIP1 signal from the nuclear envelope (Fig. 6H and Supplemental Fig. S21A). Importantly, the occasional points at which ECT2-mCherry seemed to fuzzily spill into the nucleus (white arrows in Fig. 6H and Supplemental Fig. S21A) overlapped with blurry GFP-WIP1 signal, probably due to lack of perpendicularity between the nuclear envelope and the optical section in these areas. In such cases, the cytoplasm, nucleus and nuclear envelope may be contained in the same region of the optical section and thus appear to be overlapping (Fig. 6H, right panel). To verify this interpretation, we inspected our plants with the super-resolution confocal Airyscan detector (Huff 2015) and, as expected, we did not observe ECT2-mCherry signal inside the GFP-WIP1-delimited nuclei in any instances (Fig. 6I and Supplemental Fig. S21B-D). Based on these analyses, we conclude that ECT2 resides in the cytoplasm and its presence in the nucleus, if any, may be too transient to be detected by fluorescence microscopy. These results agree with the lack of evidence for a function of ECT2/3/4 in choice of poly(A) sites, and strongly suggest that the molecular basis for the importance of ECT2/3/4 should be sought in cytoplasmic properties of their mRNA targets.

ECT2/3 targets tend to show reduced abundance upon loss of ECT2/3/4

We next assessed the effect of loss of ECT2/3/4 function on target mRNA abundance, using the Smart-seq2 data described above. Principal component analysis showed that the three repeats of wild type (*ect2-1/ECT2-mCherry*) were well separated from the three repeats of mutant (*te234/ECT2^{W464A}-mCherry*) along the first principal component (Supplemental Fig. S22), indicating that differential gene expression analysis with mutant to wild type comparison was meaningful. We focused on stringent, permissive and non-ECT2/3 targets (Supplemental Fig. S17F and Supplemental Dataset 5), and visualized their differential expression between mutant and wild type by scatter, volcano and box plots (Fig. 7A-C and Supplemental Dataset 7). These approaches showed that stringent targets have a clear tendency towards down-regulation upon loss of ECT2/3/4 function. This trend is maintained, but is less pronounced in permissive targets, and is not detectable in non-targets (Fig. 7A-C). Indeed, of the significantly differentially expressed stringent ECT2/3 targets, nearly all were down-regulated

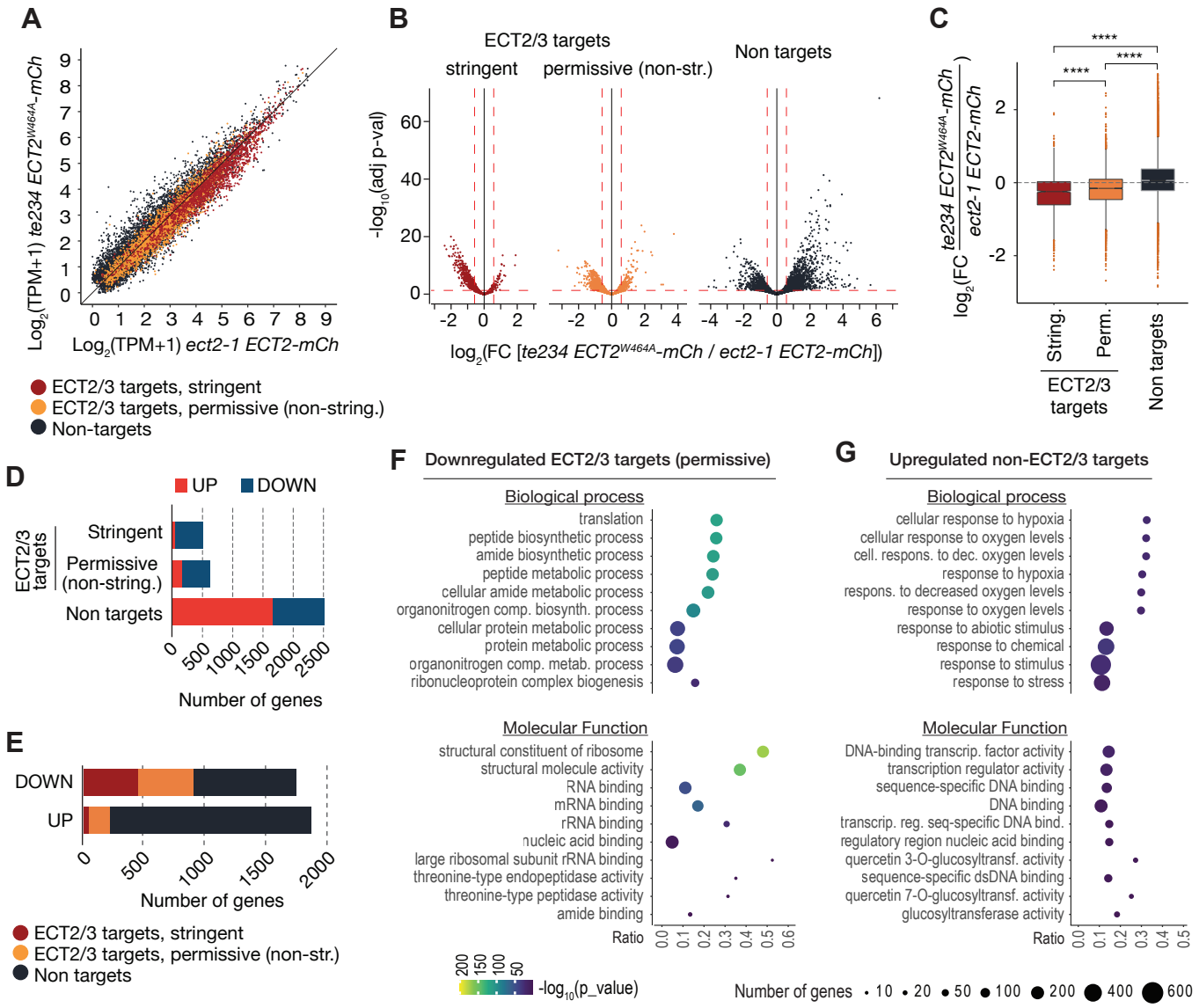

Figure 7. ECT2/3 targets are generally less abundant in cells without ECT2/3/4 function

(A) Scatterplot of TPM expression values in Smart-seq2 libraries of root protoplasts expressing ECT2-mCherry in *te234/ECT2^{W464A}-mCherry* vs. *ect2-1/ECT2-mCh* samples. **(B)** Volcano plots reveal genes differentially expressed between the genotypes described in A. **(C)** Boxplots of $\log_2$ fold change expression values between *te234/ECT2^{W464A}-mCherry* and *ect2-1/ECT2-mCh samples*. **(D,E)** Bar plots showing the amount of significantly up- and downregulated genes in ECT2/3 targets and non-targets. **(F,G)** List with the 10 most significantly enriched GO terms among significantly upregulated ECT2/3 targets (permissive set) (F), or downregulated non-targets (G) upon loss of ECT2/3/4 function.

in the mutant, while the majority of differentially expressed non-targets were up-regulated compared to wild type (Fig. 7D). Furthermore, ECT2/3 targets accounted for more than half of all significantly downregulated genes, but only about 15% of upregulated genes (Fig. 7E). In contrast, highly upregulated genes tended to be non-targets (Fig. 7A,B). To test if these differentially regulated gene sets represented subsets of functionally related genes within target and non-target groups, we analyzed their potential enrichment of GO terms. This analysis revealed that down-regulated ECT2/3 targets were particularly enriched in genes related to ribosome biogenesis and translation, while upregulated non-targets were enriched in “stress responses” with molecular function “transcription factor” (Fig. 7F). Because cell wall digestion required for protoplast isolation is a cellular stress, we tested the trivial possibility that loss of ECT2/3/4 function renders cells more susceptible to stress, and that such potential hyper-susceptibility underlies the observed differences of gene expression in ECT2-expressing root protoplasts. To this end, we isolated RNA from intact root apices of 4-day old plants of Col-0 wild type and *te234* mutants, and performed mRNA-seq analysis. These results recapitulated the trends of downregulation of stringent ECT2/3 targets and upregulation of stress-responsive non-targets, albeit with less pronounced differences than observed in the selected ECT2-expressing cell populations (Supplemental Fig. S21). These data confirm that the observed patterns of differential gene expression are genuine and biologically meaningful, and that the selection of ECT2-expressing cells ensures the most accurate description of differential gene expression resulting from loss of ECT2/3/4 function. We note that while the differential gene expression analysis suggests that ECT2/3/4 formally act to enhance abundance of their mRNA targets, it does not allow conclusions to be drawn on how such activation is brought about: a direct stabilizing effect of ECT2/3/4 binding to their targets is consistent with the observed results, but indirect effects via transcriptional repression cannot be excluded, especially given the presence of stress-related transcription factors in the set of up-regulated non-targets.

Discussion

Methodology for mapping protein-RNA interactions in plants

Our work establishes experimental and computational approaches to implement HyperTRIBE for unbiased and sensitive mapping of direct targets of RNA binding proteins in plants. Two

points are particularly relevant in this regard. First, the examples studied here show that stable transgenic expression of ADARcd does not lead to detrimental phenotypes, perhaps because of the generally low editing proportions obtained *in vivo*. Second, the rigorous statistical approach developed to call editing sites makes HyperTRIBE powerful, despite the low editing proportions observed. We also note that ECT2 and ECT3 are well suited to verify that HyperTRIBE mostly recovers directly bound target RNAs, because of the possibility to cross-reference the data with independently obtained m^6A maps (Parker et al. 2020). The combination of iCLIP and HyperTRIBE for unbiased mapping of targets proved particularly attractive for at least two reasons. First, the convergence on overlapping target sets by orthogonal methods strengthens the confidence that the identified targets are biologically meaningful. Second, HyperTRIBE, especially with the novel computational approach for calling of editing sites developed here, offers higher sensitivity than iCLIP, while iCLIP is unmatched in providing information on binding sites within target RNAs. It is possible that better positional information on binding sites may be obtained from HyperTRIBE data using maximal editing proportions rather than statistical significance as the parameter to call editing sites. Indeed, recent work on the use of HyperTRIBE to identify targets of the RNA-binding protein MUSASHI-2 (MSI-2) in leukemic stem cells recovered the known MSI-2 binding site as enriched around editing sites in targets (Nguyen et al. 2020). Nonetheless, our data shows that highly edited sites match the ADAR substrate consensus site better than lowly edited sites, suggesting that site proximity to ADAR is not the only determinant of editing proportions. Finally, our work also clearly indicates that FA-CLIP, now used in at least two studies involving YTH domain proteins (Wei et al. 2018; Song et al. 2021), is not a recommendable technique, as it recovers many false positives and fails to include many genuine targets. Thus, with the possible exception of cases in which evidence for indirect association is specifically in demand, such as the recent study in human cells of mixed tailing of viral RNA by the cellular terminal nucleotidyl transferase TENT4 (Kim et al. 2020), FA-CLIP should not be used for identification of RNAs associating with a particular RNA-binding protein of interest.

Cytoplasmic reading of RR(m^{6}A)CH in 3'-UTRs of target mRNAs by ECT2

It is a major conclusion of the present work that ECT2 is exclusively cytoplasmic and binds to m^{6}A predominantly in the RR(m^{6}A)CH sequence context *in vivo*, consistent with cytoplasmic reading of m^{6}A written by the conserved nuclear MTA/MTB methyltransferase. These observations refute the claim by Wei et al. (2018) that ECT2 binds to the supposedly plant-specific m^{6}A-containing sequence motif URU(m^{6}A)Y, and they thereby reconcile knowledge on m^{6}A-YTHDF axes in plants specifically and in eukaryotes more broadly. The phenotypic similarity of plants defective in MTA/MTB writer and ECT2/3/4 reader function is now coherent with the locations MTA/MTB-written m^{6}A- and ECT2/(3) binding sites transcriptome-wide, and it is now clear that plants do not constitute an exception to the general biochemical framework for eukaryotic m^{6}A-YTHDF function in which cytoplasmic YTHDF proteins read the m^{6}A signal written by the MTA/MTB methylase in the nucleus.

The role of URUAY and other pyrimidine-rich motifs in m^{6}A writing and reading

Despite the conclusions that URUAY does not contain m^{6}A in Arabidopsis, and that ECT2 binds to DR(m^{6}A)CH, our detailed analysis of sequence motifs enriched around m^{6}A and ECT2 iCLIP crosslink sites shows that U/Y-rich motifs, including URUAY, are likely to be implicated in m^{6}A writing and/or reading. In particular, m^{6}A occurs in DRACH/GGAU islands embedded in U/Y-rich regions. Such U/Y-rich regions were identified by machine learning as important features around m^{6}A-sites, suggesting their implication in MTA/MTB-catalyzed adenosine methylation (Supplemental Fig. S11). This, in turn, may also explain the pronounced 3'-UTR bias of m^{6}A occurrence, as extensive poly-pyrimidine tracts are rare in coding regions. We note that the observed pattern is similar to that observed in mammalian cells, in which guidance of the methyltransferase depends, at least in part, upon the associated poly(U)-interacting proteins RBM15A/B (Patil et al. 2016). Whether a similar mechanism operates in plants, potentially via the distant RBM15A/B homologue FPA (Arribas-Hernández and Brodersen 2020), remains to be investigated. In contrast to the multitude of pyrimidine-rich sequences surrounding m^{6}A-sites, URUAY appears also to have ties more specifically to ECT2 binding thanks to three properties. **(1) When present 5' to m^{6}A sites, it crosslinks to ECT2,** suggesting that some part of the protein can be in contact with URUAY. **(2) The 5'/3' asymmetry of ECT2 crosslinks to URUAY is observed only**

with RNA crosslinked to full-length ECT2 ("IDR-only"), not with RNA crosslinked to the YTH-mCherry fragment. This suggests that the N-terminal IDR, not the YTH domain, makes contacts to URUAY; an idea consistent with the only reduced, but not abolished, 3'-UTR bias of crosslink sites found with the m^{6}A-binding-deficient ECT2 mutant (see Fig. 2H). **(3) URUAY is more enriched close to m^{6}A-sites for which there is no evidence of ECT2-binding**, suggesting that it weakens ECT2 binding. Although these observations may be explained by multiple scenarios, we find a simple, yet at present speculative, model attractive: URUAY may be a site of competition between the IDR of ECT2 and another, as yet unknown, RNA binding protein. We note that the notion of RNA-interaction by IDRs has precedent (Corley et al. 2020), and that in the case of ECT2, it helps explain the poor 5'-labeling efficiency of RNA in complex with full-length ECT2. We also note that the idea of a URUAY-binding protein influencing ECT2-binding and/or regulation is consistent with the recovery of formaldehyde crosslinks between ECT2 and URUAY (Wei et al. 2018), in this case indirectly. Finally, it is intriguing that URUAY resembles part of a Pumilio binding site (Hafner et al. 2010; Huh et al. 2013), raising the tantalizing possibility of functional interaction between YTHDF and Pumilio proteins. In any event, the functional dissection of the URUAY element in m^{6}A-reading now constitutes a subject of major importance, emphasized by the broad conservation of its enrichment around m^{6}A sites across multiple plant species (rice (Li et al. 2014c; Zhang et al. 2019), maize (Luo et al. 2019; Miao et al. 2019), tomato (Zhou et al. 2019), Arabidopsis (Miao et al. 2019)).

Molecular functions of ECT2/ECT3

Although our main focus here was to use robust target identification of ECT2 and ECT3 to establish fundamental properties of targeting by these major plant YTHDF proteins, important aspects of their molecular functions were also revealed. First, the pronounced genetic redundancy in organogenesis (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020) – the first such example of YTHDF redundancy in any eukaryote - is reflected in a strong overlap in target mRNAs. This observation suggests that many target mRNAs can bind to either ECT2 or ECT3 with similar consequences; i.e. they can exhibit redundant function *sensu stricto*, not just the ability to replace function in the absence of the other protein. Second, ECT2 is not nuclear, and ECT2/3/4 do not appreciably affect PAS location in their

direct mRNA targets. Instead, ECT2/3/4 clearly affect the abundance of direct targets. Whether this effect is exclusively a consequence of direct ECT2/3/4 function at the post-transcriptional level, for example protection from endonucleolysis as previously suggested (Anderson et al. 2018), or whether more indirect effects also play a role, potentially related to transcriptional repression of ECT2/3/4 targets via stress responses activated upon their loss of function, is an important question for future studies.

Methods

All data analyses were carried out using TAIR 10 as the reference genome and Araport11 as the reference transcriptome. Unless otherwise stated, data analyses were performed in R (https://www.R-project.org/) and plots generated using either base R or ggplot2. (https://ggplot2.tidyverse.org).

Plant material

All lines used in this study are in the *Arabidopsis thaliana* Col-0 ecotype. The mutant alleles or their combinations: *ect2-1* (SALK_002225) (Arribas-Hernández et al. 2018; Scutenaire et al. 2018; Wei et al. 2018), *ect3-1* (SALKseq_63401), *ect4-2* (GK_241H02), and *ect2-1/ect3-1/ect4-2* (*te234*) (Arribas-Hernández et al. 2018) have been previously described. The transgenic lines *GFP:WIP1* (Xu et al. 2007) and those expressing *ECT2pro:ECT2-mCherry-ECT2ter*, *ECT2pro:ECT2^{W464A}-mCherry-ECT2ter*, *ECT2pro:3xHA-ECT2-ECT2ter*, or *ECT2pro:3xHA-ECT2^{W464A}-ECT2ter* have also been described or generated by floral dip in additional mutant backgrounds using the same plasmids and methodology (Arribas-Hernández et al. 2018; Arribas-Hernández et al. 2020). Plants co-expressing *ECT2-mCherry* and GFP-*WIP1* used for fluorescence microscopy were the F1 progeny of a genetic cross between GFP-*WIP1* and *ECT2-mCherry*-expressing plants.

Growth conditions

Seeds were surface-sterilized, germinated and grown on vertically disposed plates with Murashige and Skoog (MS)-agar medium (4.4 g/L MS, 10 g/L sucrose, 10 g/L agar) pH 5.7 at 20°C, receiving ~70 µmol m^{-2} s^{-1} of light in a 16 hr light/8 hr dark cycle as default. To assess

phenotypes of adult plants, ~8-day-old seedlings were transferred to soil and maintained in Percival incubators also under long day conditions. Additional details and variations of growth conditions for specific experiments can be found in Supplemental Methods.

Generation of transgenic lines for HyperTRIBE

We employed USER cloning (Bitinaite and Nichols 2009) to generate *ECT2pro:ECT2-FLAG-ADAR-ECT2ter, ECT2pro:FLAG-ADAR-ECT2ter, ECT3pro:ECT3-FLAG-ADAR-ECT3ter* and *ECT3pro:FLAG-ADAR-ECT3ter* constructs in pCAMBIA3300U (pCAMBIA3300 with a double PacI USER cassette inserted between the *Pst*I-*Xma*I sites at the multiple cloning site (Nour-Eldin et al. 2006)). Details on the cloning procedure can be found in Supplemental Methods. *Arabidopsis* stable transgenic lines were generated by floral dip transformation (Clough and Bent 1998) of Col-0 WT, *ect2-1*, *ect3-1* or *te234*, and selection of primary transformants (T1) was done on MS-agar plates supplemented with glufosinate ammonium (Fluka) (10 mg/L). We selected 5 independent lines of each type based on segregation studies (to isolate single T-DNA insertions), phenotypic complementation (in the *te234* background) and transgene expression levels assessed by FLAG western blot (see Supplemental Methods).

HyperTRIBE library preparation

We extracted total RNA (see Supplemental Methods) from manually dissected root tips and apices (removing cotyledons) of 5 independent lines (10-day-old T2 seedlings) of each of the lines used for ECT2- and ECT3-HT, to use as biological replicates. Illumina mRNA-Seq libraries were then prepared by Novogene (see Supplemental Methods) after mRNA enrichment with oligo(dT) beads (18-mers).

HyperTRIBE editing site calling

Significant differentially edited sites between *ECT2-FLAG-ADAR* (fusion) and *FLAG-ADAR* (control) samples were called according to our hyperTRIBE**R** pipeline (https://github.com/sarah-ku/hyperTRIBER), testing all nt positions with some evidence of differential editing across multiple samples. Significant (adjusted-*p*-value < 0.01 and $\log_2$FC > 1) A-to-G hits were further filtered and annotated according to Araport11 by integrating quantification information generated using Salmon (Patro et al. 2017), based on the Araport11

transcriptome with addition of *FLAG-ADAR* sequence. Similar steps were carried out for ECT3-HT. See Supplemental Methods for additional details.

Analysis of HyperTRIBE sites

For all significant editing sites (sig. E.S.), editing proportions (E.P.) where calculated as G/(A+G) where A, G are the number of reads covering the E.S. with A or G at the site, respectively. Sample-specific E.P.s for all sig. E.S. were used for principal component analyses and correlations between *FLAG-ADAR* TPM and E.P., and replicate-averaged E.P. was used for density plots, condition- or cell-type-based comparisons and comparisons over expression bins. For expression bins, the $\log_2$(TPM+1) values for all expressed genes in either aerial tissues, roots or combined were split into 9 bins of increasing expression. The Support of ECT2 target or m^6A gene sets was calculated by the proportion of genes in a given bin out of the total number of genes in that bin. See Supplemental Methods for additional details and methods.

CLIP experiments and iCLIP library preparation

In vivo UV-crosslinking of 12-day-old seedlings and construction of iCLIP libraries were optimized for ECT2-mCherry from the method developed by Prof. Staiger's group for Arabidopsis GRP7-GFP (Meyer et al. 2017; Köster and Staiger 2020). Details can be found in Supplemental Methods.

iCLIP data analysis and peak calling

Sequenced reads were mapped to TAIR10 after being processed by trimming, demultiplexing and discarding short reads. Peak calling of uniquely mapped reads was done using PureCLIP 1.0.4 (Krakau et al., 2017) after removal of PCR duplicates. Gene annotation of peaks was carried out using the hyperTRIBE**R** pipeline. Details can be found in Supplemental Methods.

Analysis of publicly available data

Single cell expression data and marker genes associated with 15 clusters annotated to cell types in roots was downloaded from Denyer et al. (2019). Single nucleotide resolution locations of m^6A sites (defined according to nanopore or miCLIP) were downloaded from

Parker et al., 2020. Intervals defining m^6A locations based on m^6A-seq were downloaded from Shen et al. 2016, and intervals defining locations of ECT2-boundistes as determined by FA-CLIP was downloaded from Wei et al., 2018. For consistency with HyperTRIBE and ECT2-iCLIP, all sets of m^6A or ECT2-bound sites were gene annotated using the hyperTRIBE**R** pipeline.

Motif analysis

A list of 48 motifs was compiled from multiple sources and for each motif a position weight matrix (PWM) was generated based on local sequence frequencies around ECT2-iCLIP peaks and used as input to FIMO (Grant et al. 2011) to detect genome-wide occurrences. In order to account for location-specific sequence contexts (typically 3'UTR), each site from iCLIP or m^6A (Parker et al. 2020) sets was assigned a random 'matched background' site, in a non-target gene, at the same relative location along the annotated genomic feature (5'UTR, CDS or 3'UTR) of the site. Distributions of motifs per 1000 sites over distance, centering on ECT2-iCLIP or m^6A sites and the respective matched backgrounds were generated using a custom R-script (https://github.com/sarah-ku/targets_arabidopsis). Sets were further split according to IDR or target status (see Supplemental Methods for further details).

For machine learning, features were generated from motifs according to their relative locations in windows from m^6A or ECT-iCLIP sites. Importance scores were generated using gradient boosting *gbm* (https://github.com/gbm-developers/gbm), with performance statistics based on the AUC calculated from held-out data. See Supplemental Methods for additional details.

Preparation and sorting of protoplasts

We prepared protoplasts from roots of 5-day-old T4 seedlings grown on vertical square plates (MS-1% agar) following the procedure by Benfey's lab (Birnbaum et al. 2005; Bargmann and Birnbaum 2010), adjusting the amount of material for the expression of the fluorescent marker in our lines. Details can be found in Supplemental Methods.

Smart-seq2

Smart-seq2 libraries were generated according to Picelli *et al.* (2013) using the Illumina DNA Nextera Flex kit from total RNA extracted with the RNeasy Plus Micro kit (QIAGEN) from FACS-sorted root protoplasts (Birnbaum et al. 2005). The libraries were sequenced in PE75 mode on an Illumina NextSeq550 sequencer. Nextera transposase adapters were trimmed from all reads using Cutadapt.

Polyadenylation Site Analysis

Polyadenylation site clusters (PACs) were identified from Smart-seq2 reads with at least 9 3'-terminal A nucleotides or 5'-terminal T nucleotides, after trimming, mapping and filtering steps, using a modification of the nanoPARE analysis pipeline (https://github.com/Gregor-Mendel-Institute/nanoPARE) (Schon et al. 2018). See Supplemental Methods for details.

Differential gene expression analysis

Differential gene expression analysis was performed from processed and quantified Smart-seq2 data using DESeq2 (Love et al. 2014), for all genes with at least 1 TPM in all six samples and a total sum of at least 5 TPM. Significantly differentially expressed genes (FDR < 0.05) were upregulated in the *ECT2/3/4*-deficient mutants if the fold change between *te234 ECT2W464A-mCh* and *ect2-1 ECT2-mCh* samples was higher than 1.5, or downregulated if lower than 1/1.5.

GO Term enrichment analysis

The functional enrichment analysis was carried out using the R package gprofiler2 version 0.2.0 (Raudvere et al. 2019).

Fluorescence Microscopy

Entire root tips were imaged with a Leica MZ16 F stereomicroscope mounted with a Sony α6000 camera. Standard confocal fluorescence microscopy images of cells in root meristems were acquired with a Zeiss LSM700 confocal microscope as described in Arribas-Hernández et al. (2018) using ~7-day-old seedlings grown on MS-agar plates and freshly mounted in water. For super-resolution fluorescence microscopy, we used a Zeiss LSM900 equipped with

the Airyscan detector (Huff 2015). Fluorescence intensity plots were obtained with the tool "Plot Profile" of the image-processing package ImageJ (Schindelin et al. 2012).

Data Access

Accession numbers

The raw and processed data for HyperTRIBE (ECT2-HT and ECT3-HT), ECT2-iCLIP, Smart-seq2 from root protoplasts and RNA-seq from root tips have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the accession number PRJEB44359.

Code availability

The code for running the hyperTRIBE**R** pipeline and for bioinformatics analyses is available at https://github.com/sarah-ku/targets_arabidopsis.

Acknowledgements

We thank Lena Bjørn Johansson and Phillip Andersen for technical assistance in the construction of transgenic lines, Theo Bølsterli, René Hvidberg Petersen and their teams for plant care, and Anna Fossum and Rajesh Somasundaram for their assistance with FACS. Kim Rewitz is thanked for providing the *Drosophila* larvae and flies used for cDNA extraction to clone *Dm*ADARcd and, together with Kenneth Halberg, for assistance with Airyscan microscopy. We acknowledge Maria Louisa Vigh for cloning of FLAG-*Dm*ADARcd, Katja Meyer and Kristina Neudorf for support during iCLIP library construction in Bielefeld, and Simon Bressendorff, Mathias Tankmar, Alexander J.H. Andersen and Freja Asmussen for experimental support. We acknowledge Tom Denyer and Marja Timmermans for their input and support in the analyses of ECT2/3 expression and target enrichment in their scRNAseq data. We are grateful to Norman R. Groves and Iris Meier for the kind donation of GFP-WIP1 seeds. This work was supported by a Consolidator Grant from the European Research Council (PATHORISC, ERC-2016-COG 726417) and a Research Grant from the Independent Research Fund Denmark (9040-00409B) to P.B.; an EMBO Short Term Fellowship (STF 7614) to L.A.-H.; a Research Grant from DFG (STA653/14-1) to D.S.; a Starting Grant from the European Research Council (638173) and a Sapere Aude Starting Grant from the Independent Research Fund Denmark (6108-00038B) to R.A.; and a European Research

Council under the European Union's Horizon 2020 Research and Innovation Program grant 63788 to M.N.

Author Contributions

P.B. and L.A.-H. designed and coordinated the study. L.A.-H. built the biological material for HyperTRIBE and S.R. developed the hyperTRIBE**R** pipeline, called edited sites to define HyperTRIBE target sets and assessed their veracity. L.A.-H. and T.K. performed iCLIP experiments and produced iCLIP libraries, M.L. analyzed iCLIP data, C.P. and S.R. executed *de novo* motif discovery. and S.R. studied the overlap between m^6A and ECT2 target sets and performed motif enrichment analyses. L.A.-H. produced protoplast RNA samples, B.E. generated Smart-seq2 libraries, M.S. analyzed poly(A) sites, and C.P. performed differential expression and GO-term enrichment analyses. M.N. supervised work related to PAS mapping, R.A. supervised work related to hyperTRIBE**R** development and sequence motif analysis, and D.S. supervised work related to iCLIP data acquisition and analysis. P.B., L.A.-H. and S.R. wrote the manuscript with input from all authors.

References

Agarwala SD, Blitzblau HG, Hochwagen A, Fink GR. 2012. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS genetics* **8**: e1002732.

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-Seq data. *Nature Precedings* doi:10.1038/npre.2012.6837.2.

Anderson SJ, Kramer MC, Gosai SJ, Yu X, Vandivier LE, Nelson ADL, Anderson ZD, Beilstein MA, Fray RG, Lyons E et al. 2018. N(6)-Methyladenosine Inhibits Local Ribonucleolytic Cleavage to Stabilize mRNAs in Arabidopsis. *Cell Rep* **25**: 1146-1157.

Arribas-Hernández L, Bressendorff S, Hansen MH, Poulsen C, Erdmann S, Brodersen P. 2018. An m6A-YTH Module Controls Developmental Timing and Morphogenesis in Arabidopsis. *Plant Cell* **30**: 952-967.

Arribas-Hernández L, Brodersen P. 2020. Occurrence and functions of m6A and other covalent modifications in plant mRNA. *Plant Physiol* **182**: 79-96.

Arribas-Hernández L, Simonini S, Hansen MH, Paredes EB, Bressendorff S, Dong Y, Østergaard L, Brodersen P. 2020. Recurrent requirement for the m6A-ECT2/ECT3/ECT4 axis in the control of cell proliferation during plant organogenesis. *Development* **147**: dev189134.

Balacco DL, Soller M. 2019. The m6A Writer: Rise of a Machine for Growing Tasks. *Biochemistry* **58**: 363-378.

Bargmann BOR, Birnbaum KD. 2010. Fluorescence Activated Cell Sorting of Plant Protoplasts. *JoVE* doi:doi:10.3791/1673: e1673.

Batista Pedro J, Molinie B, Wang J, Qu K, Zhang J, Li L, Bouley Donna M, Lujan E, Haddad B, Daneshvar K et al. 2014. m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**: 707-719.

Bernardes WS, Menossi M. 2020. Plant 3' Regulatory Regions From mRNA-Encoding Genes and Their Uses to Modulate Expression. *Front Plant Sci* **11**: 1252.

Birnbaum K, Jung JW, Wang JY, Lambert GM, Hirst JA, Galbraith DW, Benfey PN. 2005. Cell type–specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat Meth* **2**: 615-619.

Bitinaite J, Nichols NM. 2009. DNA cloning and engineering by uracil excision. *Curr Protoc Mol Biol* **Chapter 3**: Unit 3 21.

Bodi Z, Zhong S, Mehra S, Song J, Graham N, Li H, May S, Fray RG. 2012. Adenosine methylation in Arabidopsis mRNA is associated with the 3' end and reduced levels cause developmental defects. *Front Plant Sci* **3**: 48.

Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F. 1994. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *Journal of Biological Chemistry* **269**: 17697-17704.

Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. 1997. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **3**: 1233-1247.

Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA. 2002. Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N(6)-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res* **30**: 4509-4518.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**: 735-743.

Corley M, Burns MC, Yeo GW. 2020. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell* **78**: 9-29.

Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP. 2019. Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev Cell* **48**: 840-852.e845.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M et al. 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**: 201-206.

Duan H-C, Wei L-H, Zhang C, Wang Y, Chen L, Lu Z, Chen PR, He C, Jia G. 2017. ALKBH10B Is an RNA N6-Methyladenosine Demethylase Affecting Arabidopsis Floral Transition. *Plant Cell* **29**: 2995.

Fu Y, Dominissini D, Rechavi G, He C. 2014. Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat Rev Genet* **15**: 293-306.

Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, Hershkovitz V, Peer E, Mor N, Manor YS et al. 2015. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**: 1002-1006.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.

Hafner M, Katsantoni M, Köster T, Marks J, Mukherjee J, Staiger D, Ule J, Zavolan M. 2021. CLIP and complementary methods. *Nature Reviews Methods Primers* **1**: 20.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M et al. 2010. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**: 129-141.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576-589.

Huang H, Weng H, Zhou K, Wu T, Zhao BS, Sun M, Chen Z, Deng X, Xiao G, Auer F et al. 2019. Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature* **567**: 414-419.

Huff J. 2015. The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. *Nat Meth* **12**: i-ii.

Huh SU, Kim MJ, 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.

Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vågbø CB, Geula S, Hanna JH, Black DL, Darnell JE, Darnell RB. 2017. m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev* **31**: 990-1006.

Kim D, Lee Y-s, Jung S-J, Yeo J, Seo JJ, Lee Y-Y, Lim J, Chang H, Song J, Yang J et al. 2020. Viral hijacking of the TENT4–ZCCHC14 complex protects viral RNAs via mixed tailing. *Nat Struct Mol Biol* **27**: 581-588.

König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909-915.

Kontur C, Jeong M, Cifuentes D, Giraldez AJ. 2020. Ythdf m6A Readers Function Redundantly during Zebrafish Development. *Cell Rep* **33**: 108598.

Köster T, Staiger D. 2020. Plant Individual Nucleotide Resolution Cross-Linking and Immunoprecipitation to Characterize RNA-Protein Complexes. In *RNA Tagging: Methods and*

Protocols, doi:10.1007/978-1-0716-0712-1_15 (ed. M Heinlein), pp. 255-267. Springer US, New York, NY.

Krakau S, Richard H, Marsico A. 2017. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome biology* **18**: 240-240.

Kuttan A, Bass BL. 2012. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci USA* **109**: E3295.

Lasman L, Krupalnik V, Viukov S, Mor N, Aguilera-Castrejon A, Schneir D, Bayerl J, Mizrahi O, Peles S, Tawil S et al. 2020. Context-dependent functional compensation between Ythdf m6A reader proteins. *Genes Dev* **34**: 1373-1391.

Lence T, Akhtar J, Bayer M, Schmid K, Spindler L, Ho CH, Kreim N, Andrade-Navarro MA, Poeck B, Helm M et al. 2016. m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540**: 242.

Li D, Zhang H, Hong Y, Huang L, Li X, Zhang Y, Ouyang Z, Song F. 2014a. Genome-wide identification, biochemical characterization, and expression analyses of the YTH domain-containing RNA-binding protein family in Arabidopsis and rice. *Plant Mol Biol Report* **32**: 1169-1186.

Li F, Zhao D, Wu J, Shi Y. 2014b. Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Res* **24**: 1490-1492.

Li M-W, Zhou L, Lam H-M. 2015. Paraformaldehyde Fixation May Lead to Misinterpretation of the Subcellular Localization of Plant High Mobility Group Box Proteins. *PLOS ONE* **10**: e0135033.

Li Y, Wang X, Li C, Hu S, Yu J, Song S. 2014c. Transcriptome-wide N6-methyladenosine profiling of rice callus and leaf reveals the presence of tissue-specific competitors involved in selective mRNA modification. *RNA Biology* **11**: 1180-1188.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.

Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J et al. 2014. Unique features of the m6A methylome in Arabidopsis thaliana. *Nat Commun* **5**: 5630.

Luo J, Wang Y, Wang M, Zhang L, Peng H, Zhou Y, Jia G, He Y. 2019. Natural variation in RNA m6A methylation and its relationship with translational status. *Plant Physiol* doi:10.1104/pp.19.00987: pp.00987.02019.

Luo J-H, Wang Y, Wang M, Zhang L-Y, Peng H-R, Zhou Y-Y, Jia G-F, He Y. 2020. Natural Variation in RNA m6A Methylation and Its Relationship with Translational Status. *Plant Physiol* **182**: 332.

Luo S, Tong L. 2014. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc Natl Acad Sci USA* **111**: 13834-13839.

Ma X, Denyer T, Timmermans MCP. 2020. PscB: A Browser to Explore Plant Single Cell RNA-Sequencing Data Sets1[OPEN]. *Plant Physiol* **183**: 464-467.

McMahon Aoife C, Rahman R, Jin H, Shen James L, Fieldsend A, Luo W, Rosbash M. 2016. TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* **165**: 742-753.

Meyer K, Köster T, Nolte C, Weinholdt C, Lewinski M, Grosse I, Staiger D. 2017. Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. *Genome biology* **18**: 204.

Meyer KD, Jaffrey SR. 2014. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nature reviews Molecular cell biology* **15**: 313-326.

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**: 1635-1646.

Miao Z, Zhang T, Qi Y, Song J, Han Z, Ma C. 2019. Evolution of the RNA N6-methyladenosine methylome mediated by genomic duplication. *Plant Physiol* doi:10.1104/pp.19.00323: pp.00323.02019.

Nguyen DTT, Lu Y, Chu KL, Yang X, Park S-M, Choo Z-N, Chin CR, Prieto C, Schurer A, Barin E et al. 2020. HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nat Commun* **11**: 2026.

Nour-Eldin HH, Hansen BG, Norholm MH, Jensen JK, Halkier BA. 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res* **34**: e122.

Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, Hall AJW, Barton GJ, Simpson GG. 2020. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. *eLife* **9**: e49658.

Patil DP, Chen C-K, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. 2016. m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**: 369.

Patil DP, Pickering BF, Jaffrey SR. 2018. Reading m6A in the Transcriptome: m6A-Binding Proteins. *Trends Cell Biol* **28**: 113-127.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Meth* **14**: 417-419.

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Meth* **10**: 1096-1098.

Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS et al. 2014. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res* **24**: 177-189.

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**: W191-W198.

Růžička K, Zhang M, Campilho A, Bodi Z, Kashif M, Saleh M, Eeckhout D, El-Showk S, Li H, Zhong S et al. 2017. Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol* **215**: 157-172.

Salditt-Georgieff M, Jelinek W, Darnell JE, Furuichi Y, Morgan M, Shatkin A. 1976. Methyl labeling of HeLa cell hnRNA: a comparison with mRNA. *Cell* **7**: 227-237.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Meth* **9**: 676-682.

Schon MA, Kellner MJ, Plotnikova A, Hofmann F, Nodine MD. 2018. NanoPARE: parallel analysis of RNA 5' ends from low-input RNA. *Genome Res* **28**: 1931-1942.

Schwartz S, Agarwala SD, Mumbach MR, Jovanovic M, Mertins P, Shishkin A, Tabach Y, Mikkelsen TS, Satija R, Ruvkun G et al. 2013. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**: 1409-1421.

Scutenaire J, Deragon J-M, Jean V, Benhamed M, Raynaud C, Favory J-J, Merret R, Bousquet-Antonelli C. 2018. The YTH Domain Protein ECT2 Is an m6A Reader Required for Normal Trichome Branching in Arabidopsis. *Plant Cell* **30**: 986-1005.

Shah JC, Clancy MJ. 1992. IME4, a gene that mediates MAT and nutritional control of meiosis in Saccharomyces cerevisiae. *Mol Cell Biol* **12**: 1078-1086.

Shen L, Liang Z, Gu X, Chen Y, Teo Zhi Wei N, Hou X, Cai Weiling M, Dedon Peter C, Liu L, Yu H. 2016. N6-methyladenosine RNA modification regulates shoot stem cell fate in Arabidopsis. *Dev Cell* **38**: 186-200.

Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton GJ, Simpson GG. 2012. Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation. *Nat Struct Mol Biol* **19**: 845-852.

Song P, Yang J, Wang C, Lu Q, Shi L, Tayier S, Jia G. 2021. Arabidopsis N6methyladenosine reader CPSF30-L recognizes FUE signals to control polyadenylation site choice in liquid-like nuclear bodies. *Molecular Plant* doi:10.1016/j.molp.2021.01.014.

Stoilov P, Rafalska I, Stamm S. 2002. YTH: a new domain in nuclear proteins. *Trends Biochem Sci* **27**: 495-497.

Tanious FA, Veal JM, Buczak H, Ratmeyer LS, Wilson WD. 1992. DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **31**: 3103-3112.

Theler D, Dominguez C, Blatter M, Boudet J, Allain FH. 2014. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res* **42**: 13911-13919.

Wan Y, Tang K, Zhang D, Xie S, Zhu X, Wang Z, Lang Z. 2015. Transcriptome-wide high-throughput deep m6A-seq reveals unique differential m6A methylation patterns between three organs in Arabidopsis thaliana. *Genome biology* **16**: 272.

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G et al. 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**: 117-120.

Wang X, Zhao Boxuan S, Roundtree Ian A, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C. 2015. N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**: 1388-1399.

Wei L-H, Song P, Wang Y, Lu Z, Tang Q, Yu Q, Xiao Y, Zhang X, Duan H-C, Jia G. 2018. The m6A Reader ECT2 Controls Trichome Morphology by Affecting mRNA Stability in Arabidopsis. *Plant Cell* **30**: 968-985.

Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG. 2011. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proc Natl Acad Sci USA* **108**: 12533.

Xu C, Wang X, Liu K, Roundtree IA, Tempel W, Li Y, Lu Z, He C, Min J. 2014. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat Chem Biol* **10**: 927-929.

Xu W, Rahman R, Rosbash M. 2018. Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA* **24**: 173-182.

Xu XM, Rose A, Muthuswamy S, Jeong SY, Venkatakrishnan S, Zhao Q, Meier I. 2007. NUCLEAR PORE ANCHOR, the Arabidopsis homolog of Tpr/Mlp1/Mlp2/megator, is involved in mRNA export and SUMO homeostasis and affects diverse aspects of plant development. *Plant Cell* **19**: 1537-1548.

Zaccara S, Jaffrey SR. 2020. A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. *Cell* **181**: 1582-1595.e1518.

Zaccara S, Ries RJ, Jaffrey SR. 2019. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **20**: 608-624.

Zhang C, Chen Y, Sun B, Wang L, Yang Y, Ma D, Lv J, Heng J, Ding Y, Xue Y et al. 2017. m6A modulates haematopoietic stem and progenitor cell specification. *Nature* **549**: 273-276.

Zhang F, Zhang Y-C, Liao J-Y, Yu Y, Zhou Y-F, Feng Y-Z, Yang Y-W, Lei M-Q, Bai M, Wu H et al. 2019. The subunit of RNA N6-methyladenosine methyltransferase OsFIP regulates early degeneration of microspores in rice. *PLoS genetics* **15**: e1008120.

Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG. 2008. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**: 1278-1288.

Zhou L, Tian S, Qin G. 2019. RNA methylomes reveal the m6A-mediated regulation of DNA demethylase gene SlDML2 in tomato fruit ripening. *Genome biology* **20**: 156.

Zhu T, Roundtree IA, Wang P, Wang X, Wang L, Sun C, Tian Y, Li J, He C, Xu Y. 2014. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res* **24**: 1493-1496.