Widespread premature transcription termination of *Arabidopsis thaliana* NLR genes by the spen protein FPA

Matthew T. Parker[1*], Katarzyna Knop[1*], Vasiliki Zacharaki[1*], Anna V. Sherwood[1], Daniel Tome[2], Xuhong Yu[3], Pascal Martin[3], Jim Beynon[2], Scott Michaels[3], Geoffrey J. Barton[1] and Gordon G. Simpson[1,4^]

[1]School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK.
[2]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.
[3]Department of Biology, Indiana University, Bloomington, IN 47405, USA.
[4]The James Hutton Institute, Invergowrie DD2 5DA, UK.

*Contributed equally

^Correspondence g.g.simpson@dundee.ac.uk

Keywords: NLR, Nanopore, direct RNA sequencing, spen, FPA, m^6A, alternative polyadenylation, chimeric RNA, disease resistance, transcription termination

Abstract

Genes involved in disease resistance are some of the fastest evolving and most diverse components of genomes. Large numbers of nucleotide-binding, leucine-rich repeat receptor (NLR) genes are found in plant genomes and are required for disease resistance. However, NLRs can trigger autoimmunity, disrupt beneficial microbiota or reduce fitness. It is therefore crucial to understand how NLRs are controlled. Here we show that the RNA-binding protein FPA mediates widespread premature cleavage and polyadenylation of NLR transcripts, thereby controlling their functional expression and impacting immunity. Using long-read Nanopore direct RNA sequencing, we resolved the complexity of NLR transcript processing and gene annotation. Our results uncover a co-transcriptional layer of NLR control with implications for understanding the regulatory and evolutionary dynamics of NLRs in the immune responses of plants.

Introduction

In plants and animals, NLR (nucleotide-binding, leucine-rich repeat receptor) proteins function to detect the presence and activity of pathogens (Barragan and Weigel, 2020; Jones et al., 2016; Tamborski and Krasileva, 2020). Plant genomes can encode large numbers of NLR genes, which often occur in physical clusters (Jiao and Schneeberger, 2020; Wei et al., 2016). Powerful selective pressure drives the rapid birth and death of NLR genes, resulting in intraspecific diversity in NLR alleles and gene number. Consequently, the near-complete repertoire of Arabidopsis NLR genes was only recently revealed using long-read DNA sequencing of diverse Arabidopsis accessions (Van de Weyer et al., 2019).

In plants, NLR proteins generally comprise an N-terminal Toll/interleukin receptor (TIR), coiled-coil (CC) or RPW8 domain that facilitates signalling; a central nucleotide-binding NB-ARC domain that acts as a molecular switch; and C-terminal leucine-rich repeats (LRRs) that interact with target proteins. NLRs can recognise pathogen effectors either directly by binding to them through LRR domains or indirectly by detecting modifications to host proteins caused by effector action. In some cases, domains of host proteins targeted by pathogen effectors have been incorporated into NLRs as integrated domains (or decoys) (Le Roux et al., 2015). NLRs that interact directly with effectors are under high levels of diversifying selection to modify their recognition specificities, resulting in significant allelic polymorphism (Prigozhin and Krasileva, 2021). Genomic variation also yields diversity in NLR protein organisation, through domain swapping or truncating mutations, and NLR isoforms that lack NB-ARC or LRR domains can function in plant immune responses (Nishimura et al., 2015; Swiderski et al., 2009; Zhang and Gassmann, 2007). The consequence of this diversity is that there is no one-size-fits-all explanation of how NLR proteins function (Barragan and Weigel, 2020).

The benefit of NLRs to the host is disease resistance, but the costs of increased NLR diversity or activity can include detrimental autoimmunity (Rodriguez et al., 2016), reduced association with beneficial microbes (Yang et al., 2010) and a general reduction in fitness

(Tian et al., 2003). In some cases, autoimmunity caused by epistatic interactions involving NLRs can cause hybrid necrosis (Li et al., 2020). Therefore, a key question is how NLRs are regulated to enable limited expression for pathogen surveillance but enhanced expression during defence responses. This problem is compounded by selective pressure on NLRs because regulatory processes must keep pace with the emergence of new NLR genes and gain or loss of function in others. Consequently, the regulation of NLRs is one of the most important and difficult challenges faced by plants.

NLR control measures occur at different stages of gene expression (Lai and Eulgem, 2018). For example, microRNAs limit the expression of many NLRs by targeting conserved regions encoded in NLR mRNAs and triggering cascades of phased siRNAs that broadly suppress NLR activity (Cai et al., 2019; Canto-Pastor et al., 2019; Shivaprasad et al., 2012; Zhai et al., 2011). Alternative splicing, which promotes the simultaneous expression of more than one NLR isoform, is required for the functions of both the *N* gene which provides resistance to tobacco mosaic virus (Dinesh-Kumar and Baker, 2000), and *RECOGNITION OF PSEUDOMONAS SYRINGAE 4* (*RPS4*), which confers resistance to *Pseudomonas syringae* DC3000 in Arabidopsis (Zhang and Gassmann, 2007). Alternative polyadenylation at intragenic heterochromatin controls the expression of Arabidopsis *RECOGNITION OF PERONOSPORA PARASITICA 7* (*RPP7*), with functional consequences for immunity against the oomycete pathogen *Hyaloperonospora arabidopsidis* (Tsuchiya and Eulgem, 2013). Finally, RNA surveillance pathways control NLRs. For example, null mutants defective in nonsense mediated RNA decay (NMD) are lethal in Arabidopsis because they trigger NLR *RPS6*-dependent autoimmunity (Gloggnitzer et al., 2014). Conversely, mutations in the RNA exosome, which degrades RNAs in a 3′ to 5′ direction, suppress *RPS6*-dependent autoimmune phenotypes (Takagi et al., 2020). Fine tuning of different levels of NLR control may be integrated to produce quantitative patterns of disease resistance (Corwin and Kliebenstein, 2017), but our understanding of how this occurs globally is fragmentary and incomplete (Adachi et al., 2019).

The RNA-binding protein FPA was first identified as a factor required for the control of Arabidopsis flowering time (Koornneef et al., 1991). Loss-of-function *fpa* mutants flower late due to elevated levels of the floral repressor, FLC (Schomburg et al., 2001). However, this cannot be the only function of FPA because it is much more widely conserved than *FLC*. FPA is a member of the spen family of proteins, which are defined by three N-terminal RNA recognition motifs and a C-terminal protein interaction SPOC domain (Ariyoshi and Schwabe, 2003). We previously showed that FPA controls the site of cleavage and polyadenylation in some mRNAs, including autoregulation of *FPA* pre-mRNA (Duc et al., 2013; Hornyik et al., 2010; Lyons et al., 2013). These findings were extended to show that FPA can affect poly(A) site choice at genes with intronic heterochromatin, including *RPP7* (Deremetz et al., 2019). The poly(A) selection mechanism used by FPA remains unclear. FPA might mediate poly(A) choice either directly by recruiting the RNA 3′ end processing machinery to sensitive sites or indirectly, for example by influencing splicing, chromatin

107 modifications or the rate of transcription by RNA Polymerase II (Pol II). We previously used
108 Helicos BioSciences direct RNA sequencing (Helicos DRS) to map the 3′ ends of
109 Arabidopsis polyadenylated transcripts and identify genes affected by transcriptome-wide
110 loss of FPA function (Duc et al., 2013; Sherstnev et al., 2012). A limitation of this approach
111 was that it could only identify RNA 3′ end positions, and so could not resolve other potential
112 roles of FPA in gene expression.
113 In this study, we used two approaches to gain a clearer understanding of how FPA
114 functions. We first investigated which proteins FPA associates with inside living plant cells.
115 Next, we analysed the global impact of different levels of FPA activity on gene expression.
116 For this, we combined Helicos DRS with short-read Illumina RNA-Seq and Oxford Nanopore
117 Technologies (Nanopore) DRS, which can reveal the authentic processing and modification
118 of full-length mRNAs (Parker et al., 2020). Using these combined data together with new
119 computational approaches to study RNA processing, we found that the predominant role
120 of FPA is to promote poly(A) site choice. In addition, we uncovered an unusual degree of
121 complexity in the processing of NLR mRNAs, which is sensitive to FPA. The finding that
122 premature transcription termination functions as an additional layer of NLR expression
123 control has implications for understanding the dynamics of NLR regulation and evolution.
124

125 ## Results

126 ### FPA co-purifies with proteins that mediate mRNA 3′ end processing

127 In order to understand how FPA controls the site of mRNA 3′ end formation, we used *in vivo*
128 interaction proteomics–mass spectrometry (*IVI-MS*) to identify which proteins FPA
129 associates with inside living plant cells. First, we fixed molecular interactions using
130 formaldehyde infiltration of Arabidopsis seedlings expressing FPA fused to YFP
131 (*35S::FPA:YFP*). Wild-type Columbia-0 (Col-0) seedlings treated in the same way were used
132 as a negative control. We then purified nuclei and performed GFP-trap immunopurification
133 followed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) to identify
134 FPA-associated proteins. By comparing the proteins detected in three biological replicates
135 of *35S::FPA:YFP* and Col-0, we identified 203 FPA co-purifying proteins with a median $\log_2$
136 fold change in adundance of greater than two (Figure 1, Supplementary file 1). At least 56%
137 (113) of the enriched proteins are poly(A)+ mRNA-binding proteins as established by
138 orthogonal RNA-binding proteome analysis (Bach-Pages et al., 2020; Reichel et al., 2016).
139 Consistent with FPA control of mRNA 3′ end formation, 14 highly conserved
140 cleavage and polyadenylation factors (CPFs) co-purified with FPA (Figure 1A,
141 Supplementary file 1). These include members of the cleavage and polyadenylation
142 specificity factor (CPSF) complex, cleavage stimulating factor (CstF) complex, and cleavage
143 factor I and II (CFIm/CFIIm) complexes. The U2AF and U2 spliceosome components that
144 interact with CFIm–CPSF to mediate terminal exon definition were also detected (Kyburz et
145 al., 2006) (Figure 1B, Supplementary file 1). We additionally detected both subunits of
146 Pol II. Characteristically, Serine^5 of the Pol II C-terminal domain (CTD) heptad repeat is

147 phosphorylated when Pol II is at the 5′ end of genes, and Ser^2 is phosphorylated when Pol II
148 is at the 3′ end (Komarnitsky et al., 2000). The position-specific phosphorylation of these
149 sites alters the RNA processing factors which are recruited to the CTD at the different stages
150 of transcription. We found that the kinase CDKC;2, which phosphorylates Ser^2 (Wang et al.,
151 2014), and the phosphatase CPL1 (homolog of yeast Fcp1), which dephosphorylates Ser^5
152 (Koiwa et al., 2004), co-purified with FPA. We also detected the homolog of the human
153 exonuclease XRN2 (known as XRN3 in Arabidopsis), which mediates Pol II transcription
154 termination (Krzyszton et al., 2018).
155 A second major class of proteins that co-purified with FPA are components of the
156 autonomous flowering pathway (Andres and Coupland, 2012; Simpson, 2004)
157 (Figure 1C,Supplementary file 1). FPA functions in the autonomous pathway to limit
158 expression of the floral repressor *FLC*. FPA activity is associated with alternative
159 polyadenylation of long non-coding RNAs that are transcribed antisense to the *FLC* locus
160 (Andres and Coupland, 2012; Simpson, 2004). Consistent with this, conserved CPF proteins
161 such as FY (WDR33) (Hornyik et al., 2010; Liu et al., 2007), PCFS4 (Xing et al., 2008), CSTF64
162 and CSTF77 (Liu et al., 2010) were previously identified in late flowering mutant screens.
163 Other detected autonomous pathway factors are proteins with established roles in pre-
164 mRNA processing, including HLP1 (Zhang et al., 2015), FLK (Mockler et al., 2004) and
165 EMB1579/RSA1 (Zhang et al., 2020b). Notably, FLK has been found to associate with PEP,
166 HUA1 and HEN4 (Zhang et al., 2015), and we identified all four of these as FPA co-purifying
167 proteins. In addition to regulating *FLC*, the FLK–PEP complex has been shown to control
168 alternative polyadenylation within pre-mRNA encoding the floral homeotic transcription
169 factor AGAMOUS (Mockler et al., 2004). Their co-purification with FPA suggests that this
170 role may be more global and involve direct interactions at RNA 3′ ends.
171 A third group of proteins that co-purified with FPA are conserved members of the
172 mRNA N^6-methyladenosine (m^6A) writer complex (Ruzicka et al., 2017) (Figure 1D,
173 Supplementary file 1). The m^6A modification mediated by this complex is predominately
174 targeted to the 3′ untranslated region (UTR) of Arabidopsis protein-coding mRNAs (Parker
175 et al., 2020). The co-purification of FPA with m^6A writer complex components may be
176 explained by either a direct role for FPA in m^6A modification or, more simply, because both
177 CPF and m^6A writer proteins are found at RNA 3′ ends.
178 The picture that emerges from this analysis is that FPA is located in proximity to
179 proteins that promote cleavage, polyadenylation, transcription termination and RNA
180 modification at the 3′ end of Pol II-transcribed genes.
181
182 **FPA co-localises with RNA Pol II Ser^2 at the 3′ end of Arabidopsis genes**
183 We next used an orthogonal approach to investigate the association of FPA with proteins
184 that function at the 3′ end of Pol II-transcribed genes. We performed chromatin
185 immumunoprecipitation sequencing (ChIP-Seq) using antibodies against FPA and Pol II
186 phosphorylated at either Ser^5 or Ser^2 of the CTD heptad repeat (Yu et al., 2019). Our

187 metagene analysis revealed that FPA is enriched at the 3′ end of genes and co-localises
188 with Pol II phosphorylated at Ser^2 of the CTD (Figure 1E, Figure 1–figure supplement 1). The
189 close relationship between FPA and Pol II Ser^2 is reinforced by changes in the distribution
190 of Pol II isoforms in *fpa* mutants. For example, we previously showed that FPA is required
191 for 3′ end processing at *PIF5* (Duc et al., 2013). Pol II Ser^2 was enriched at the 3′ end of *PIF5*
192 in Col-0, but depleted from this region in *fpa-7* mutants (Figure 1–figure supplement 2).
193 Together, these orthogonal ChIP-Seq and *IVI-MS* analyses reveal the close association of
194 FPA with proteins involved in 3′ end processing and transcription termination at the 3′ end
195 of Arabidopsis genes.
196
197 **FPA predominantly promotes poly(A) site choice**
198 We next asked which RNA processing events are controlled by FPA. We used a
199 combination of Illumina RNA-Seq and Helicos and Nanopore DRS technologies to analyse
200 three different genetic backgrounds expressing different levels of FPA activity: wild-type
201 Col-0, loss-of-function *fpa-8* and a line overexpressing FPA fused to YFP (*35S::FPA:YFP*). In
202 combination, these orthogonal sequencing technologies can reveal different features of
203 transcriptomes: Helicos DRS short reads identify the 3′ ends of mRNAs, but cannot reveal
204 the full properties of the corresponding transcripts (Ozsolak et al., 2009); Illumina RNA-Seq
205 produces short reads derived from all expressed regions, meaning that changes in RNA 3′
206 end processing can only be detected by differences in coverage (Xia et al., 2014); and
207 Nanopore DRS long reads define the 3′ ends of mRNAs in the context of reads that can
208 correspond to full-length transcripts (Parker et al., 2020). For each genotype, we performed
209 three biological replicates with Helicos DRS, six with Illumina RNA-Seq and four with
210 Nanopore DRS. The resultant sequencing statistics are detailed in Supplementary file 1.
211 We first assessed the utility of the three sequencing technologies to map changes
212 in mRNA processing by focusing on the *FPA* locus. FPA autoregulates its expression by
213 promoting premature cleavage and polyadenylation within intron 1 of *FPA* pre-mRNA (Duc
214 et al., 2013; Hornyik et al., 2010). Consistent with this, a proximal poly(A) site in the first
215 intron and distal sites in the terminal intron and exon of *FPA* could be mapped in Col-0
216 using Nanopore and Helicos DRS (Figure 2A). Using all three data types, we detected a
217 quantitative shift towards selection of distal poly(A) sites in the loss-of-function *fpa-8* mutant
218 and a strong shift to proximal poly(A) site selection when FPA is overexpressed
219 (*35S::FPA:YFP*). Nanopore DRS provided the clearest picture of alternative polyadenylation
220 events because full-length reads reveal poly(A) site choice in the context of other RNA
221 processing events.
222 We next asked how transcriptome-wide RNA processing is affected by FPA activity.
223 Since mutations in FPA cause readthrough of annotated 3′UTRs (Duc et al., 2013), we
224 applied the software tool StringTie2 (Pertea et al., 2015) to create a bespoke reference
225 annotation with Nanopore DRS reads from Col-0, *fpa-8* and *35S::FPA:YFP*. We then
226 measured how changes in FPA expression altered the 3′ end distribution at each locus

227 using the earth mover's distance (EMD; also known as the Wasserstein distance). EMD
228 indicates the "work" required to transform one normalised distribution into another based
229 on the proportion of 3′ ends that would have to be moved and by what distance. We used
230 an EMD permutation test, in which reads are randomly shuffled between conditions, to
231 estimate *p*-values for each locus. Loci with an EMD greater than 25 and a false discovery
232 rate (FDR) less than 0.05 were considered differentially polyadenylated.

233 Using this approach on Nanopore DRS data, we identified 285 and 293 loci with
234 alternative polyadenylation events in *fpa-8* and *35S::FPA:YFP*, respectively (Figure 2B). In
235 all, 77.9% (222) of loci with alternative polyadenylation in *fpa-8* displayed a positive change
236 in the mean 3′ end position, indicating a predominant shift to distal poly(A) site selection
237 (Figure 2B, left panel). These loci also had greater effect sizes than those with shifts towards
238 proximal poly(A) sites (Figure 2C, left panel). In contrast, 56.7% (166) of loci with alternative
239 polyadenylation in *35S::FPA:YFP* displayed a negative change in the mean 3′ end position,
240 indicating a shift towards proximal poly(A) sites (Figure 2B, right panel). These loci had
241 greater effect sizes than those with positive changes in 3′ end profile (Figure 2C, right
242 panel). A total of 16 loci displayed a shift to distal poly(A) site selection in *fpa-8* and to
243 proximal poly(A) site selection in *35S::FPA:YFP* (hypergeometric test $p=3.9\times10^{-7}$),
244 demonstrating that loss of function versus overexpression of FPA can result in reciprocal
245 patterns of poly(A) site choice.

246 We used the same approach to identify loci with FPA-dependent alternative
247 polyadenylation in Helicos DRS data. We identified 319 and 299 genes with alternative
248 polyadenylation events in *fpa-8* and *35S::FPA:YFP*, respectively (Figure 2D and E).
249 Consistent with Nanopore DRS analysis, the predominant shifts in *fpa-8* and *35S::FPA:YFP*
250 were towards distal (79.0% or 252 loci) and proximal (75.3% or 225 loci) poly(A) sites,
251 respectively. In all, 44 loci displayed a shift to distal poly(A) sites in *fpa-8* and to proximal
252 poly(A) sites in *35S::FPA:YFP* (hypergeometric test $p=4.8\times10^{-30}$), again demonstrating
253 reciprocal poly(A) site selection depending on FPA activity. Of the 222 loci identified with
254 shifts to distal poly(A) sites in *fpa-8* using Nanopore DRS, 39.6% (88) were also detected
255 using Helicos DRS (Figure 2–figure supplement 1). Likewise, 44.0% of loci (73) with proximal
256 polyadenylation detected in *35S::FPA:YFP* using Nanopore DRS were also detected using
257 Helicos DRS. Across the DRS datasets, we identified 59 loci for which reciprocal poly(A) site
258 regulation by FPA could be detected by Nanopore DRS and/or Helicos DRS.

259 In order to analyse the Illumina RNA-Seq data, we developed annotation-agnostic
260 software for detecting alternative RNA 3′ end processing events, using a similar approach
261 to the existing tools DERfinder (Collado-Torres et al., 2017), RNAprof (Tran Vdu et al., 2016)
262 and DEXSeq (Anders et al., 2012). We segmented Illumina RNA-Seq data by coverage and
263 relative expression in *fpa*-8 or *35S::FPA:YFP* compared with Col-0. Segmented regions
264 were grouped into transcriptional loci using the annotations generated from Nanopore
265 DRS datasets. Differential usage of regions within each locus was then tested using
266 DEXSeq. Using this approach, we identified 2535 loci with differential RNA processing

events in *fpa-8*: 1792 were upregulated, 390 were downregulated, and 353 had both upregulated and downregulated regions (FDR<0.05, absolute logFC>1; Figure 2F, left panel). A total of 1747 loci with differential RNA processing events were identified in *35S::FPA:YFP*: 997 were upregulated, 532 were downregulated, and 218 had both upregulated and downregulated regions (Figure 2F, right panel). The median effect size for differentially processed regions was greater for upregulated regions than for downregulated regions in *fpa-8*. This is consistent with an increase in transcriptional readthrough events and elevated expression of intergenic regions and downstream genes. In contrast, the median effect size for differentially processed regions was similar for up- and downregulated regions in *35S::FPA:YFP*. This is consistent with an increase in the relative expression of proximal exonic and intronic regions, and loss of expression of distal exonic regions caused by preferential selection of proximal poly(A) sites. Similar results were seen for differential splice junction usage analysis (Figure 2G), suggesting that changes in splicing are the indirect effects of altered 3′ end processing in *fpa-8*, rather than direct effects of FPA on splice site choice. Evidence of this can be seen at the *PIF5* locus, where readthrough results in increased cryptic and canonical splicing of downstream *PAO3* (Figure 2–figure supplement 2).

We next asked whether FPA influences RNA modification. Our *IVI-MS* analysis had revealed that conserved members of the Arabidopsis m^6A writer complex co-purify with FPA (Figure 1D,Supplementary file file 1). The human proteins most closely related to FPA are RBM15/B, which co-purify with the human m^6A writer complex and are required for m^6A deposition (Patil et al., 2016). We used LC-MS/MS to analyse the m^6A/A (adenosine) ratio in mRNA purified from Col-0, *fpa-8*, *35S::FPA:YFP* and a mutant defective in the m^6A writer complex component VIR (*vir-1*). Consistent with previous reports, the level of mRNA m^6A in the hypomorphic *vir-1* allele was reduced to approximately 10% of wild-type levels (Parker et al., 2020; Ruzicka et al., 2017) (Figure 2–figure supplement 3). However, we detected no differences in the m^6A level between genotypes with altered FPA activity. Therefore, we conclude that FPA does not influence global levels of mRNA m^6A methylation.

Finally, we asked whether the FPA-dependent global changes in alternative polyadenylation result from an indirect effect on chromatin state. We previously showed that FPA controls the expression of histone demethylase IBM1 by promoting proximal polyadenylation within *IBM1* intron 7 (Duc et al., 2013). IBM1 functions to restrict $H3K9me^2$ levels, and *ibm1* mutants accumulate ectopic heterochromatic marks in gene bodies, which affects RNA processing at certain loci (Miura et al., 2009; Saze et al., 2008). When we analysed two independent ChIP-Seq datasets of $H3K9me^2$ in *ibm1–4* mutants (Inagaki et al., 2017; Lai et al., 2020), we found that only 10.6% of loci with altered poly(A) site choice in *35S::FPA:YFP* have altered $H3K9me^2$ in *ibm1* mutants compared with 14.2% of all loci tested (hypergeometric p=0.97; Figure 2–figure supplement 4). This result suggests that FPA-dependent regulation of poly(A) site choice is not an indirect consequence of FPA control of *IBM1*.

307 Overall, these analyses reveal that the primary function of FPA is to control poly(A)
308 site choice. FPA predominantly promotes poly(A) site selection; hence, *fpa* loss-of-function
309 backgrounds exhibit readthrough at sites used in the wild type, whereas FPA
310 overexpression results in increased selection of proximal poly(A) sites.
311
312 **NLRs are major targets of FPA-dependent poly(A) site selection**
313 We next asked which groups of genes are sensitive to FPA-dependent alternative
314 polyadenylation. We used InterPro annotations (Mitchell et al., 2019) to perform protein
315 family domain enrichment analysis of the loci affected by FPA (revealed by the Nanopore
316 and Helicos DRS analyses). We found that sequences encoding NB-ARC, Rx-like coiled coil
317 (CC), and/or LRR domains were enriched amongst the loci with increased proximal
318 polyadenylation in *35S::FPA:YFP* (Figure 3A and B). This combination of domains is
319 associated with NLR disease resistance proteins.
320 The Col-0 accession contains at least 206 genes encoding some combination of TIR,
321 CC, RPW8, NB-ARC and LRR domains, which might be classified as NLRs or partial NLRs
322 (Van de Weyer et al., 2019). In general, these can be grouped according to their encoded
323 N-terminal domain as TIR (TNLs), CC (CNLs) or RPW8 (RNLs) genes. We manually examined
324 these NLR genes to identify those with alternative polyadenylation. Reannotation of some
325 loci was required to interpret the effects of FPA regulation. For example, we found that the
326 TNL gene *AT5G46490,* located in the RPS6 cluster, is incorrectly annotated as two loci,
327 *AT5G46490* and *AT5G46500* (Figure 3–figure supplement 1). Nanopore DRS evidence
328 indicates that this is actually a single locus with a previously unrecognised 2.7 kb intron
329 containing a proximal poly(A) site, the use of which is regulated by FPA. This interpretation
330 is supported by nanoPARE data (Schon et al., 2018), which showed no evidence of capped
331 5′ ends originating from the annotated downstream gene. Use of the distal poly(A) site
332 introduces an additional ~400 amino acids to the C-terminus of the protein. This C-terminal
333 region has homology to other NLRs in the *RPS6* cluster, and is predicted to introduce
334 additional LRR repeats (Martin et al., 2020) (Figure 3–figure supplement 2).
335 Notably, we could also reannotate the chromosomal region around *RPS6* itself. The
336 extreme autoimmunity phenotypes of NMD mutants and mitogen-activated kinase pathway
337 mutants require RPS6 but the mechanisms involved are not understood (Gloggnitzer et al.,
338 2014; Takagi et al., 2020). Nanopore DRS indicates that the 3′UTR of *RPS6* is complex, with
339 multiple splicing events and poly(A) sites (Figure 3C). We also detected transcripts
340 expressed from this region that do not appear to be contiguous with *RPS6* 3′UTR reads.
341 Instead, these reads correspond to an independent unannotated gene that overlaps the
342 *RPS6* 3′UTR. This interpretation is supported by capped RNA 5′ ends detected in this region
343 by nanoPARE (Schon et al., 2018). In addition, Nanopore DRS analysis of the RNA exosome
344 mutant *hen2-2* (Parker et al., 2021) revealed that this unannotated gene is expressed at
345 relatively high levels, but that the transcripts are subject to degradation. Consequently,
346 steady-state levels of RNA expressed from this locus are relatively low in Col-0. The gene

347 encodes a TIR domain similar to that of RPS6 (Figure 3D). Therefore, use of the distal *RPS6*
348 poly(A) site constitutes readthrough into the downstream TIR-domain-only NLR. Based on
349 these analyses, we conclude that long-read Nanopore DRS data have the potential to
350 correct NLR gene annotation at complex loci that cannot be resolved by genome annotation
351 software or short-read Illumina RNA-Seq.
352
353 **Widespread premature transcription termination of NLRs includes frequent selection**
354 **of poly(A) sites in protein-coding exons**
355 Of the 206 NLR genes examined, 124 had a sufficient level of expression levels to identify
356 alternative polyadenylation in the Nanopore DRS data; of these 124, 62 (50.0%) were found
357 to have FPA-dependent alternative polyadenylation (Tables 1–3). Of the 74 expressed NLRs
358 located in major clusters, 44 (59.5%) were regulated by FPA (chi^2 p=0.02) (Lee and Chae,
359 2020). The localisation of NLRs to large genomic clusters is known to facilitate diversification
360 (Barragan and Weigel, 2020). Consistent with this, 20 (71.4%) of the 28 expressed NLRs
361 reported to be under high levels of diversifying selection were regulated by FPA (chi^2
362 p=0.02) (Prigozhin and Krasileva, 2021). In addition, FPA-regulated NLRs tended to be
363 located in regions with higher levels of synteny diversity (Jiao and Schneeberger, 2020),
364 although in this case the association was not significant (*t*-test p=0.09; Figure 4–figure
365 supplement 1). Overall, these findings suggest that FPA-dependent alternative
366 polyadenylation is associated with rapidly evolving NLRs.
367 The effects of FPA regulation can be broadly classified into three modes of control
368 involving (i) readthrough and chimeric RNAs, (ii) intronic poly(A) sites and (iii) poly(A) sites
369 within protein-coding exons. At certain complex loci, FPA can affect poly(A) site choice
370 using combinations of these different modes of regulation.
371 For 17 NLR genes, we found that a change in FPA activity altered the formation of
372 readthrough or chimeric RNAs containing one or more NLR loci (Table 1). The duplicated
373 *RPP7a/b*-like genes *AT1G58848* and *AT1G59218* (which form part of the *RPP7* cluster
374 containing five CNL-class NLRs) displayed increased readthrough into downstream
375 transposable elements (TEs) in *fpa-8* (Figure 4A). EMD tests could not be performed at these
376 loci due to the multi-mapping of reads at these duplicated genes (*AT1G58848* and
377 *AT1G59218*). Loss of FPA function can also lead to clusters of two or more NLR genes being
378 co-transcribed as a single transcriptional unit. For example, the TNL-class gene *AT1G63730*,
379 located in the *B4/RLM1* cluster, forms chimeric RNA with the downstream TNL-class gene
380 *AT1G63740* in *fpa-8* (Helicos EMD=1099, FDR=0.02; Figure 4–figure supplement 2).
381 We identified another 17 NLR genes with intronic polyadenylation regulated by FPA
382 (Table 2). Of these, four contained poly(A) sites in 5′UTR introns (which would result in non-
383 coding transcripts) and three contained alternative poly(A) sites after the stop codon (which
384 could alter potential regulatory sequences contained in 3′UTRs). The remainder contained
385 poly(A) sites in introns between protein-coding exons. Selection of these poly(A) sites
386 introduce premature stop codons that result in truncated open reading frames (ORFs). For

387 example, we identified a proximal poly(A) site within the third intron of *AT1G69550*, which
388 encodes a TNL-type singleton NLR (Figure 4B). Use of this poly(A) site results in mRNAs with
389 a premature stop codon; the encoded protein lacks most of the predicted LRR domain. In
390 *fpa-8*, readthrough at this poly(A) site is increased (Helicos EMD=1271, FDR=1.2×10^{-4}),
391 resulting in an increase in the relative number of full-length transcripts.

392 The most common form of FPA-dependent NLR regulation was premature
393 termination within exons (Table 3). We identified 45 NLRs controlled in this way: at 44 of
394 these loci, termination occurred within protein-coding exons. In most cases, this results in
395 stop-codonless transcripts that are predicted targets of non-stop decay (Szádeczky-Kardoss
396 et al., 2018). Many of these proximal exonic poly(A) sites could be identified at lower levels
397 in Col-0. For example, at *RPP28* (AT2G14080), which encodes a TNL-class singleton NLR,
398 we detected multiple exonic poly(A) sites located within the second and fourth exons, which
399 encode the NB-ARC and LRR domains, respectively (Figure 4C). Selection of these exonic
400 poly(A) sites was increased in *35S::FPA:YFP* (Helicos EMD=859, FDR=5.4×10^{-9}) and
401 decreased in *fpa-8* (Helicos EMD=912, FDR=7.6×10^{-9}). FPA was also found to promote
402 premature termination in the protein-coding sequence of single-exon, intronless NLR
403 genes. For example, at *RPP13* (AT3G46530), which encodes a CNL-class NLR protein, FPA
404 overexpression causes selection of proximal poly(A) sites located within the region
405 encoding the LRR domain (Helicos EMD=228, FDR=1.8×10^{-4}; Figure 4–figure
406 supplement 3).

407 Although the most frequent consequence of FPA selection of exonic poly(A) sites
408 was stop-codonless transcripts, we also identified examples where the protein-coding
409 potential was altered. For example, *AT5G40060* encodes a TNL-class NLR but has a
410 premature stop codon between the TIR and NB-ARC domains. Consequently, full-length
411 transcription results in an mRNA with an upstream ORF (uORF) encoding the TIR domain
412 and a larger downstream ORF encoding NB-ARC and LRR domains (Figure 4D). However,
413 transcripts with such large uORFs are targets of NMD in plants (Nyikó et al., 2009).
414 Therefore, FPA-regulated proximal polyadenylation in the region encoding the NB-ARC
415 domain results in a transcript containing only the uORF, which is not a predicted NMD target
416 and may be more efficiently translated into a TIR-only protein.

417 In seven of the identified genes, exonic proximal polyadenylation is associated with
418 retention of an upstream intron (Table 3). As a result, premature stop codons are
419 introduced, resulting in a truncated coding region. For example, the TNL-type NLR *RPS4*
420 was previously shown to be regulated by alternative splicing induced by the effector
421 AvrRps4 (Zhang and Gassmann, 2007). We identified an increase in *RPS4* intron 3 retention
422 in *35S::FPA:YFP* compared with Col-0 that was associated with proximal polyadenylation
423 events in exon 4 (Helicos EMD=34, not significant; Figure 4–figure supplement 4).
424 Therefore, inter-dependence between splicing and poly(A) site choice may explain *RPS4*
425 control.

426 FPA controlled NLR poly(A) site selection at 16 complex loci with combinations of
427 intronic, exonic and readthrough sites. One example is *RPP4* (AT4G16860), a TNL-class NLR
428 known to mediate Arabidopsis resistance to *Hpa* isolate Emoy2 (*Hpa*-Emoy2) (van der
429 Biezen et al., 2002). *RPP4* is part of the *RPP5* cluster, comprising seven TNL-class NLRs. In
430 agreement with a previous study (Wang and Warren, 2010), we found that in wild-type Col-
431 0, *RPP4* can be transcribed as a chimeric RNA together with the downstream *AtCOPIA4* TE
432 (AT4G16870) through selection of one of the two distal poly(A) sites located within the TE
433 (Figure 5) (Wang and Warren, 2010) or selection of a third poly(A) site in the downstream
434 gene *AT4G16857*. Use of the proximal poly(A) site within the TE is associated with an
435 approximately 8-kb cryptic splicing event between the 5′ splice site of the first exon of *RPP4*
436 and a 3′ splice site located within the TE. Both Nanopore DRS and Illumina RNA-Seq data
437 provide evidence for this cryptic splicing event, which skips all *RPP4* exons downstream of
438 exon 1, removing most of the *RPP4* coding sequence and introducing a stop codon
439 (Figure 5, Inset 1). The resulting transcript is predicted to encode a TIR-domain-only protein.
440 Loss of FPA function decreases chimeric RNA formation by shifting poly(A) site selection
441 towards a proximal poly(A) site located within the protein-coding region of the final exon
442 (Figure 5–figure supplement 1). This results in the production of *RPP4* transcripts lacking in-
443 frame stop codons (Figure 5, Inset 2). Furthermore, in *35S::FPA:YFP*, we observed increased
444 selection of a proximal poly(A) site located within the first intron of *RPP4*, which would also
445 encode a truncated RPP4 protein. We conclude that FPA-regulated alternative
446 polyadenylation at *RPP4* produces transcripts with unusually long 3′UTRs, alternative
447 protein isoforms and transcripts that cannot be efficiently translated.
448
449 **FPA controls *RPP7* by promoting premature termination within protein-coding exon**
450 **6**
451 To examine the functional impact of FPA on the regulation of NLRs, we focused on *RPP7*.
452 *RPP7* encodes a CNL-class NLR protein crucial for strain-specific resistance to *Hpa* isolate
453 Hiks1 (*Hpa*-Hiks1) (Tsuchiya and Eulgem, 2013). The full-length expression of *RPP7* is
454 controlled by elongation factors that interact with H3K9me^2, which is associated with the
455 COPIA-type retrotransposon (*COPIA-R7*) located in *RPP7* intron 1 (Saze et al., 2013). Using
456 Nanopore and Helicos DRS data, we identified at least two poly(A) sites within the *COPIA-*
457 *R7* element, both of which were selected more frequently in *fpa-8* (Figure 6A, Figure 6–
458 figure supplement 1). We also identified two poly(A) sites within the second intron of *RPP7*.
459 The use of both sites is reciprocally regulated by FPA, with a moderate decrease in *fpa-8*
460 and an increase in *35S::FPA:YFP*. All of these intronic proximal poly(A) sites are located
461 before the start of the *RPP7* ORF and generate transcripts that do not encode RPP7 protein.
462 At the 3′ end of *RPP7*, we found three alternative poly(A) sites located in the terminal intron,
463 in addition to the previously reported most distal and most commonly used poly(A) site in
464 the terminal exon (Figure 6A, Inset 1) (Tsuchiya and Eulgem, 2013). Selection of each of
465 these poly(A) sites is associated with alternative splicing events that lead to the generation

466 of four possible 3′UTR sequences. Termination at the 3′UTR intronic poly(A) sites is
467 suppressed by FPA: their usage is increased in *fpa-8* and decreased in *35S::FPA:YFP*. These
468 data indicate that FPA influences *RPP7* intronic polyadenylation at a larger number of
469 poly(A) sites than previously supposed.
470 The major effect of FPA on *RPP7* is within protein-coding exon 6, where we identified
471 three poly(A) sites (Figure 6A, Inset 2): two at the end of the region encoding the NB-ARC
472 domain and one within the region encoding the LRR repeats. Cleavage and polyadenylation
473 at these sites result in transcripts without in-frame stop codons, thereby disrupting the
474 coding potential of *RPP7* mRNA. These poly(A) sites were identified in both Helicos and
475 Nanopore DRS data, indicating that they are unlikely to be caused by alignment errors. The
476 relative selection of exon 6 poly(A) sites depends on FPA expression: in Col-0, 25% of *RPP7*
477 Nanopore DRS reads terminate at one of these exon 6 poly(A) sites; and when FPA is
478 overexpressed, this figure increases to 63%. Consistent with this, a relative drop in coverage
479 at exon 6 was also observed in *35S::FPA:YFP* Illumina RNA-Seq data. Consequently, only
480 23% of *RPP7* transcripts are expected to encode RPP7 protein in the FPA-overexpressing
481 line. In contrast, 4% of *RPP7* Nanopore DRS reads identified in *fpa-8* terminate in exon 6,
482 and 79% of transcripts are expected to be protein coding. In an orthogonal approach, we
483 used RNA gel blot analysis to visualise *RPP7* mRNAs in Col-0, *fpa-8* and *35S::FPA:YFP*
484 backgrounds and detected a clear decrease in signal corresponding to full-length *RPP7*
485 transcripts in *35S::FPA:YFP* (Figure 6B). These data support previous evidence of FPA-
486 dependent control of *RPP7* (Deremetz et al., 2019) but reveal that the predominant
487 mechanism is via exonic transcription termination.
488
489 **FPA modulates *RPP7*-dependent, race-specific pathogen susceptibility**
490 We next asked whether FPA-dependent premature transcription termination at *RPP7* exon
491 6 has a functional consequence. Since FPA reduced the level of full-length protein-coding
492 *RPP7* transcripts, we asked whether increased FPA activity might compromise RPP7-
493 dependent immunity. To test this hypothesis, we carried out pathogenesis assays using the
494 oomycete strain *Hpa*-Hiks1. RPP7 provides race-specific immunity to *Hpa*-Hiks1 in Col-0.
495 The Keswick (Ksk-1) accession is susceptible to *Hpa*-Hiks1 (Lai et al., 2019) and we used it
496 as a control in these studies.
497 We inoculated Arabidopsis seedlings with *Hpa*-Hiks1 spores in three independent
498 experiments. Four days after inoculation, we checked susceptibility by counting the number
499 of sporangiophores. With the exception of Ksk-1, all of the lines we tested were in a Col-0
500 background. As expected, Col-0 plants were resistant to infection
501 (median: 0 sporangiophores per plant), and Ksk-1 plants were sensitive to infection
502 (median: 5 sporangiophores per plant; $p=1.7\times10^{-32}$; Figure 6C). *fpa-7* mutants were as
503 resistant to infection as Col-0 (median: 0 sporangiophores per plant, $p=0.19$). This is
504 consistent with our finding that full-length *RPP7* transcript expression is not reduced in the
505 absence of FPA. *fpa-8* mutants were also resistant to infection (median: 0 sporangiophores

506 per plant), however there was slight variability in their resistance compared to *fpa-7*
507 (p=2.4×10^{-12}). This variability was not restored by complementation with a *pFPA::FPA*
508 transgene (p=0.23) indicating that it is not caused by loss of FPA function, and is likely to
509 result from other mutations in the *fpa-8* background. In contrast, *35S::FPA:YFP* plants were
510 significantly more sensitive to *Hpa*-Hiks1 than *pFPA::FPA* (median: 3 sporangiophores per
511 plant; p=3.8×10^{-9}), indicating that overexpression of FPA compromises immunity. We
512 conclude that FPA-mediated transcription termination of *RPP7* pre-mRNA has a functional
513 consequence for race-specific immunity. Therefore, FPA control of poly(A) site selection can
514 modulate NLR function.
515

516 ## Discussion

517 We have identified a novel role for the RNA-binding protein FPA in the control of plant
518 innate immunity. Using *IVI-MS* proteomics and ChIP-Seq, we showed that FPA is closely
519 associated with proteins involved in RNA 3′ processing and co-localises with Ser^2
520 phosphorylated Pol II at the 3′ ends of genes. Integrative analysis using three RNA
521 sequencing technologies confirmed that the major effect of modulating FPA activity is to
522 alter poly(A) site selection. An unexpected finding was that half of expressed NLR loci were
523 sensitive to FPA activity. In most cases, FPA promoted the use of poly(A) sites within protein-
524 coding exons of NLR genes. At *RPP7*, an increase in exonic polyadenylation caused by FPA
525 overexpression was shown to compromise immunity to *Hpa*-Hiks1. The widespread nature
526 of this control mechanism suggests that transcription termination plays an important role in
527 the regulatory and evolutionary dynamics of NLR genes.
528

529 **Uncovering protein assemblies that mediate 3′ end processing in living plant cells**

530 We used an *in vivo* formaldehyde cross-linking approach to identify proteins that co-
531 localise with FPA inside living plant cells. These data provide in-depth knowledge of the
532 proteins involved in Arabidopsis RNA 3′ end processing and clues to the function of the
533 uncharacterised proteins identified here. Components of the m^{6}A writer complex also co-
534 purify with FPA. However, unlike related proteins in human and Drosophila (Knuckles et al.,
535 2018; Patil et al., 2016), we found that FPA is not required to maintain global levels of m^{6}A
536 modification in Arabidopsis.
537 Two Arabidopsis PCF11 paralogs with Pol II CTD-interacting domains (CIDs), PCFS2
538 and PCFS4, co-purified with FPA, but two paralogs lacking CIDs, PCFS1 and PCFS5, did
539 not. PCF11 was previously shown to have functionally separable roles in transcription
540 termination and cleavage and polyadenylation (Sadowski et al., 2003): the N-terminal
541 PCF11 CID is required for transcription termination, whereas the C-terminal domains are
542 required for cleavage and polyadenylation. The specific interaction of FPA with CID-
543 containing PCF11 paralogs suggests that FPA regulates alternative polyadenylation by
544 altering Pol II speed and transcription termination. The human SPOC domain protein PHF3
545 can bind to two adjacent Ser^2 phosphorylated heptads of the CTD of Pol II via two

546 electropositive patches on the surface of its SPOC domain (Appel, Franke et al. 2020). One
547 of these patches, and the key amino acid residues within it, is conserved in the structure of
548 the FPA SPOC domain (Zhang et al., 2016). Consequently, FPA might also interact with the
549 CTD, possibly in conjunction with CID domains of PCFS2 and PCFS4. Such interactions
550 could explain the specific co-association of FPA and Pol II Ser^2 at the 3′ end of genes.
551
552 **Widespread control of NLR transcription termination by FPA**
553 An unanticipated finding of this study is that Arabidopsis NLR loci were the most enriched
554 group of FPA targets. NLRs function in the immune response and, consistent with this
555 crucial role, they are under powerful and dynamic selective pressure. Defining the inventory
556 of Arabidopsis NLRs depended on long-range DNA sequencing of diverse accessions (Van
557 de Weyer et al., 2019). Here we show that long-read Nanopore DRS provides insight into
558 the authentic complexity of NLR mRNA processing and enables the accurate annotation of
559 NLR genes. For example, our reannotation of the *RPS6* locus is essential to understand the
560 recurring role of RPS6 in autoimmunity. The autoimmune phenotypes of mutants defective
561 in NMD or the mitogen-activated kinase pathway are RPS6 dependent, but the mechanisms
562 involved are unclear (Gloggnitzer et al., 2014; Takagi et al., 2020). We found that *RPS6* is
563 transcribed through a previously unrecognised downstream gene that encodes an RPS6-
564 like TIR domain. We showed that expression of the downstream gene is dependent on the
565 RNA exosome component HEN2. In addition, mutations in *HEN2* were recently identified
566 as suppressors of RPS6-dependent autoimmune phenotypes (Takagi et al., 2020). It is clear
567 that accurate annotation of complex NLR loci facilitates the interpretation of basic features
568 of NLR function.
569 Of the 124 NLRs with detectable expression in Nanopore DRS data, 62 were
570 sensitive to FPA activity. FPA controls 3′ end formation of NLR genes in three different
571 transcript locations (Figure 7): (i) 3′UTRs, where it can prevent readthrough and chimeric
572 RNA formation; (ii) introns, where it promotes proximal polyadenylation; and (iii) protein-
573 coding exons, where it promotes stop-codonless transcript formation. The consequences
574 of such complex control of RNA 3′ end formation are wide-ranging and likely to be context
575 dependent (Mayr, 2019).
576 Where FPA controls readthrough and chimeric RNA formation, it affects 3′UTR
577 length, sequence composition and cryptic splice site usage. Long or intron-containing
578 3′UTRs are targeted by NMD, leading to RNA decay or suppressed translatability. Long,
579 unstructured 3′UTRs influence intermolecular RNA interactions and phase separation,
580 changing the subcellular localisation of mRNAs (Ma et al., 2021). The close proximity of
581 mRNAs in the resulting granules may enable co-translational protein complex formation.
582 Readthrough transcription may also disrupt the expression of downstream genes by
583 transcription interference (Proudfoot, 1986).
584 FPA-dependent premature transcription termination at intronic poly(A) sites can
585 introduce novel stop codons, resulting in transcripts that encode truncated NLR proteins

586 with altered functions. For example, some TIR domain-only proteins are known to be active
587 in NLR regulation, resulting in constitutive signalling activity (Zhang et al., 2004) or act as
588 competitive inhibitors by titrating full-length NLR protein (Williams et al., 2014). In other
589 cases, TIR-domain-only proteins are sufficient for pathogen recognition (Nishimura et al.,
590 2017). The TE-containing 3′UTR of *RPP4* appears to be required for resistance to the
591 pathogen *Hpa*-Emoy2, although the mechanism involved is unclear (Wang and Warren,
592 2010). We discovered that cryptic splicing of *RPP4* exon 1 to a novel 3′ splice site within the
593 TE can produce a unique transcript that encodes only the RPP4 TIR domain. It will be
594 interesting to examine whether the TIR-only RPP4 isoform is required for full pathogen
595 resistance. We also found that intron retention at *RPS4,* which is essential for RPS4-
596 dependent resistance against *P. syringae* DC3000 (Zhang and Gassmann, 2007), is linked
597 to exonic proximal polyadenylation. Intron retention without accompanying proximal
598 polyadenylation will result in transcripts with long 3′UTRs that are likely to be sensitive to
599 NMD, whereas proximally polyadenylated transcripts could be translated into truncated
600 protein. Therefore, a combination of alternative polyadenylation and splicing probably
601 underpins *RPS4* control. In future, sensitive proteomic analyses will be important to
602 determine the impact of alternative polyadenylation on NLR protein isoform expression.
603 A remarkable finding was that FPA mostly targets the protein-coding exons of NLR
604 genes and, even controls premature transcription termination within the ORF of single-exon
605 NLR genes such as *RPP13*. Premature transcription termination in protein-coding exons
606 results in the production of stop-codonless transcripts that cannot be efficiently translated
607 into protein. These truncated transcripts may be subject to decay by RNA surveillance
608 mechanisms (e.g. the non-stop decay pathway) or act as non-coding RNA decoys to titrate
609 the levels of regulatory microRNAs (Shivaprasad et al., 2012). Increased rates of NLR
610 transcription in plants under pathogen attack could promote elongation through such
611 "regulatory" poly(A) sites. In this way, the expression of NLR proteins might be restricted
612 during pathogen surveillance but kept poised for rapid activation during infection.
613 Since the evolution of *cis*-regulatory elements controlling poly(A) site choice within
614 introns or 3′UTRs is free from the constraints of protein-coding functionality, why should
615 protein-coding exons be targeted so frequently? One possibility is that this enables the
616 expression of newly created NLR genes to be kept under tight control, thereby facilitating
617 rapid evolution whilst reducing the chances of autoimmunity (Figure 7). This hypothesis is
618 strengthened by the finding that many NLRs with high allelic diversity (Prigozhin and
619 Krasileva, 2021) are regulated by FPA. Alternative polyadenylation might also function to
620 hide NLR genes from negative selection and contribute to cryptic genetic variation in a
621 similar way to the mechanism proposed for NMD- and microRNA-mediated NLR control
622 (Raxwal and Riha, 2016; Shivaprasad et al., 2012). Cryptically spliced chimeric RNAs, with
623 subsequent retrotransposition, can be a source of new genes (Akiva et al., 2006). Therefore,
624 the control of transcription termination could directly facilitate the neofunctionalisation of
625 NLRs. In the future, it will be important to compare patterns of transcription termination at

626 NLRs across Arabidopsis accessions. For example, analysis of transcriptomic data will
627 determine whether proximal polyadenylation is conserved in NLRs with high allelic
628 diversity, whilst an integrative analysis of transcriptomic and genomic data will establish
629 whether chimeric NLR transcripts identified in some accessions are found as
630 retrotransposed genes in others.

631 At least two distinct patterns of alternative polyadenylation mediate *RPP7*
632 regulation, one involving intronic heterochromatin (Tsuchiya and Eulgem, 2013) and
633 another involving FPA-dependent termination in exon 6. The latter mechanism is conserved
634 across all NLR genes of the Col-0 *RPP7* locus (Table 3). Alleles of these RPP7-like NLR genes
635 have been identified as the causes of specific cases of hybrid necrosis (Bomblies and
636 Weigel, 2007). In these cases, autoimmunity is explained by allele-specific physical
637 interactions between RPP7 protein and the RPW8-only protein HR4 (Li et al., 2020). We
638 found that not only are *RPP7*-like genes targeted by FPA-dependent premature transcript
639 termination, but so too is *HR4* (Table 2). This raises the possibility that FPA could rescue
640 hybrid necrosis by limiting the expression of these proteins. FPA also appears to regulate
641 the proximal polyadenylation of *DANGEROUS MIX 10 (DM10)*, producing transcripts that
642 could encode a protein with truncated LRR repeats. *DM10* alleles with LRR truncations have
643 been demonstrated to cause autoimmunity in specific crosses (Barragan et al., 2021),
644 suggesting that in other cases FPA overexpression could trigger or enhance autoimmune
645 phenotypes. Consequently, modulation of transcription termination may shift the balance
646 of costs and benefits associated with NLR gene expression. This phenomenon is not likely
647 to be restricted to FPA because mutations in the RNA 3′ processing factor CPSF30 can also
648 suppress autoimmunity (Bruggeman et al., 2014).

649 The impact of FPA overexpression on gene expression and autoimmunity revealed
650 here derives from artificial transgene expression. However, pathogens could similarly
651 modulate NLR activity by evolving effectors that target the expression or activity of factors
652 controlling NLR poly(A) site choice. Consistent with this idea, the HopU1 effector of
653 *P. syringae* targets the RNA-binding protein AtGRP7 (Fu et al., 2007), which co-purified with
654 FPA. In addition, the Pi4089 effector of the oomycete pathogen *Phytophthora infestans*
655 targets the KH-domain RNA-binding protein StKRBP1 in potato; as a result, the abundance
656 of StKRBP1 increases and infection by *P. infestans* is enhanced (Wang et al., 2015). This
657 precedent reveals that effector-mediated increases in RNA-binding protein abundance can
658 transform host RNA-binding proteins into susceptibility factors. Phylogenetic analysis of
659 StKRBP1 suggests that a direct homolog is absent in Brassicaceae. However, the most
660 closely related Arabidopsis proteins are FLK and PEP (Zhang et al., 2020a), both of which
661 co-purify with FPA and have been shown to regulate poly(A) site choice (Rodriguez-Cazorla
662 et al., 2015). FPA, GRP7, FLK and PEP, along with other RNA-binding proteins, act in concert
663 to fine-tune the timing of flowering through the regulation of *FLC*. In a similar way, RNA-
664 binding protein-dependent modulation of NLR expression might explain how quantitative
665 disease resistance occurs (Corwin and Kliebenstein, 2017).

666
667 **New ways to analyse RNA processing**
668 An essential feature of our study was the introduction of new approaches to study RNA
669 processing and 3′ end formation. The use of long-read Nanopore DRS transformed our
670 understanding of the complexity of NLR gene expression by providing insight that short-
671 read Illumina RNA-Seq and Helicos DRS could not. We recently showed that Nanopore DRS
672 mapping of RNA 3′ ends closely agrees with short-read Helicos DRS, and that Nanopore
673 DRS is not compromised by internal priming artefacts (Parker et al., 2020). Consequently,
674 we used Nanopore DRS to quantify alternative patterns of cleavage and polyadenylation.
675 We also introduced a new approach to analyse alternative polyadenylation by applying the
676 EMD metric. EMD incorporates information on the both the relative abundance and the
677 genomic distance between alternative poly(A) sites. This is valuable because large
678 distances between poly(A) sites are more likely to impact the mRNA coding potential or
679 trigger mRNA surveillance compared with subtle changes in 3′UTR length.
680 A limitation of short-read analyses of RNA processing is their dependence upon
681 reference transcript annotations because these may be incomplete. For example, in disease
682 or mutant conditions, RNA processing often occurs at novel sites that are not present in
683 reference transcriptomes (as was the case here for NLR genes). For this reason, using long-
684 read sequencing data to generate bespoke reference transcriptomes for the genotypes
685 under analysis can increase the value of short-read sequencing data. Until the throughput
686 of long-read sequencing matches that of short-read technologies, a combined approach is
687 likely to be generally useful in interpreting transcriptomes.
688
689 **Concluding remarks**
690 It is difficult to identify alternative polyadenylation from conventional short-read RNA-Seq
691 data. As a result, the impact of alternative polyadenylation is probably under-reported.
692 Here we show that premature transcription termination of NLR genes is widespread. Using
693 Nanopore DRS, we could improve the accuracy of NLR annotation and revealed a layer of
694 NLR gene regulation that may also influence the dynamics of NLR evolution. The continued
695 development of approaches that reveal full-length native RNA molecules is likely to provide
696 new insight into other important, but previously unrecognised, aspects of biology.
697

698 # Materials and methods

699 ## Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain (*Arabidopsis thaliana*)	Columbia (Col-0)	NA	ABRC: CS22625	Country of Origin: USA
Strain (Arabidopsis thaliana)	Keswick (Ksk-1)	Lai et al. 2018	ABRC: CS1634	Country of Origin: UK
Gene (*Arabidopsis thaliana)*	*FPA*	NA	TAIR/ABRC: AT2G43410	-
Gene (Arabidopsis thaliana)	*RPP7*	NA	TAIR/ABRC: AT1G58602	-
Genetic reagent (*Arabidopsis thaliana*)	*fpa-7*	Duc et al., 2013	ABRC: SALK_021959C	T-DNA insertion mutant in Col-0 background. Gifted by R. Amasino, UW-Madison.
Genetic reagent (*Arabidopsis thaliana*)	*fpa-8*	Baurle et al., 2007	TAIR: 4515120225	EMS point mutation in Col-0 background. Gifted by C. Dean, John Innes Centre
Genetic reagent (*Arabidopsis thaliana*)	*35S::FPA:YFP fpa-8*	Baurle et al., 2007	NA	Transgenic line in *fpa-8* background, gifted by C. Dean, John Innes Centre
Genetic reagent (*Arabidopsis thaliana*)	*pFPA::FPA fpa-8*	Zhang et al., 2016	NA	Transgenic line in *fpa-8* background.
Genetic reagent (*Arabidopsis thaliana*)	*vir-1*	Růžička et al., 2017	TAIR: 6532672723	EMS point mutant in Col-0 background. Gifted by K. Růžička, Brno.
Commercial assay, kit	Rneasy Plant Mini kit	QIAGEN	Cat#: 74904	-
Commercial assay, kit	SuperScript™ III Reverse Transcriptase	Thermo Fisher Scientific	Cat#: 18080044	-
Commercial assay, kit	NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®	New England Biolabs	Cat#: E7420	-
Commercial assay, kit	Dynabeads™ mRNA Purification Kit	Thermo Fisher Scientific	Cat#: 61006	-
Commercial assay, kit	Nanopore Direct RNA sequencing kit	Oxford Nanopore Technologies	Cat#: SQK-RNA001	-
Commercial assay, kit	MinION Flow cell r9.4	Oxford Nanopore Technologies	Cat#: FLO-MIN106	-
Peptide, recombinant protein	T4 DNA ligase	New England Biolabs	Cat#: M0202	-

Commercial assay, kit	Quick Ligase reaction buffer	New England Biolabs	Cat#: B6058S	-
Commercial assay, kit	Agencourt RNAclean XP magnetic beads	Beckman Coulter	Cat#: A63987	-
Commercial assay, kit	Qubit RNA BR Assay Kit	Thermo Fisher Scientific	Cat#: Q10210	-
Commercial assay or kit	RNA ScreenTape System	Agilent	Cat#: 5067-5576 - 5067-5578	-
Antibody	FPA antibody	Covance	NA	Rabbit polyclonal antibody. Raised against FPA amino acids 536-901.
Chemical compound	[γ-32P]-ATP	Perkin Elmer	Cat#: BLU012H250UC	-
Commercial assay or kit	DECAprime II DNA labelling kit	Thermo Fisher Scientific	Cat#: AM1455	-
Commercial assay or kit	Illustra MicroSpin G-50 Columns	GE Healthcare	Cat#: 27-5330-01	-
Commercial assay or kit	RiboRuler High Range RNA Ladder	Thermo Fisher Scientific	Cat#: SM1821	-
Peptide, recombinant protein	FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific	Cat#: EF0651	-
Peptide, recombinant protein	T4 Polynucleotide Kinase	Thermo Fisher Scientific	Cat#: EK0031	-
Peptide, recombinant protein	Nuclease P1	Merck	Cat#: N8630-1VL	-
Peptide, recombinant protein	Calf Intestinal Alkaline Phosphatase	New England Biolabs	Cat#: M0290S	-
Chemical compound	N6-Methyladenosine (m6A), Modified adenosine analog	Abcam	Cat#: ab145715	-
Chemical compound	Adenosine, Endogenous P1 receptor agonist	Abcam	Cat#: ab120498	-
Commercial assay or kit	GFP-Trap Agarose	Chromotek	Cat#: gta-20	-
Software, algorithm	d3pendr	10.5281/zenodo.4319112	NA	Scripts to perform differential 3' end analysis using Nanopore DRS or Helicos DRS data
Software, algorithm	Simpson_Barton_FPA_NLRs	10.5281/zenodo.4319108	NA	All pipelines, scripts and notebooks used for analyses in this manuscript.

700

701 **Plants**

702 *Plant material and growth conditions*

703 The wild-type Col-0 accession and *fpa-7* were obtained from the Nottingham Arabidopsis
704 Stock Centre. The *fpa-8* mutant (Col-0 background) and *35S::FPA:YFP* in *fpa-8* (Baurle et
705 al., 2007) were provided by C. Dean (John Innes Centre). Generation of the *pFPA::FPA* line

706 was previously described (Zhang et al., 2016). Surface-sterilised seeds were sown on MS10
707 medium plates containing 2% agar, stratified at 4°C for 2 days, germinated in a controlled
708 environment at 20°C under 16-h light/8-h dark conditions and harvested 14 days after
709 transfer to 20°C.
710
711 ***IVI-MS***
712 *Preparation of IVI-MS samples*
713 Seedlings were harvested 14 days after germination and cross-linked with 1 % (v/v)
714 formaldehyde under vacuum. The cross-linking reaction was stopped after 15 min by the
715 addition of glycine to a final concentration of 0.125 M and returned to vacuum for a further
716 5 min. Nuclei were isolated from frozen ground plant tissue using Honda buffer (20 mM
717 Hepes-KOH pH 7.4, 10 mM $MgCl_2$, 440 mM sucrose, 1.25 % (w/v) Ficoll, 2.5 % (w/v) Dextran
718 T40, 0.5 % (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF, 1% (v/v) Protease Inhibitor Cocktail;
719 (Sigma)) and collected by centrifugation at 2000*g* for 17 min at 4°C. Nuclei were washed
720 twice with Honda buffer (centrifugation at 1500*g* for 15 min at 4°C between washes) and
721 lysed in nuclear lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1 % (w/v) SDS, 1 mM PMSF,
722 1 % (v/v) Protease Inhibitor Cocktail) by sonication for four cycles of 30 s pulses with low
723 power and 60 s cooling between pulses using a Bioruptor UCD-200 (Diagenode).
724 Following centrifugation (16,100*g* for 10 min at 4°C), the supernatant was diluted 10-fold
725 with sample dilution buffer (16.7 mM Tris-HCl pH 8, 167 mM NaCl, 1.1 % (v/v) Triton X-100,
726 1 % (v/v) Protease Inhibitor Cocktail). Cross-linked protein complexes were isolated with
727 GFP-trap agarose beads (Chromotek) and incubated at 4°C with constant rotation for 5 h,
728 followed by centrifugation (141*g* for 3 min at 4°C). Beads were washed three times with
729 washing buffer (150 mM NaCl, 20 mM Tris-HCl pH 8, 2 mM EDTA pH 8, 1 % (v/v) Triton X-
730 100, 0.1 % (w/v) SDS, 1 mM PMSF) by centrifugations between washes (400*g* for 3 min at
731 4°C). Samples were incubated at 90°C for 30 min to reverse the cross-linking prior to SDS-
732 PAGE. Each biological replicate was separated into five fractions following SDS-PAGE,
733 subjected to in-gel digestion with trypsin and submitted for LC-MS/MS analysis (LTQ
734 Orbitrap Velos Pro mass spectrometer; Thermo Fisher Scientific). Three biological
735 replicates were performed for each genotype.
736
737 *IVI-MS data analysis*
738 Raw peptide data files from *IVI-MS* were analysed by MaxQuant software (version 1.6.10.43)
739 (Cox and Mann, 2008). Peptide tables were then loaded using Proteus (version 0.2.14)
740 (Gierlinski et al., 2018) and summarised to protein level counts using the hi-flyer method
741 (mean of the top three most abundant peptides). Because wild-type plants lacking GFP
742 were used as controls, a large number of the proteins enriched by immunoprecipitation
743 were below the detection threshold in the control. This group of proteins can be classified
744 as "missing not at random" (MNAR). In all proteomics experiments, there will also be a
745 number of proteins which are not detected purely by chance: these are referred to as

"missing at random" (MAR). We treated proteins that were missing from all replicates of a condition as MNAR, and proteins that were missing only from a subset of replicates as MAR. Using the imputeLCMD package (version 2.0) (Lazar, 2015), a K nearest neighbours' strategy was used to impute MAR examples, and a quantile regression imputation of left centred data (QRILC) approach was used to impute MNAR examples. Differential expression analysis was performed on imputed data using limma (version 3.40.0) (Ritchie et al., 2015). Because imputation is not deterministic (i.e. will lead to different outcomes every time), we improved the robustness of the analysis by performing 999 bootstraps of the imputation and differential expression, and summarising the results using the median $\log_2$ fold change and harmonic mean *p* value.

ChIP-Seq

Preparation of libraries for ChIP-Seq

ChIP against FPA and Pol II phosphorylated at either Ser^5 or Ser^2 of the CTD heptad repeat was performed as previously described (Yu et al., 2019). Polyclonal antibodies against FPA amino acids 536–901 were raised in rabbit by Covance.

ChIP-Seq data processing

FPA and Pol II ChIP-Seq data are available at ENA accession PRJNA449914. $H3K9me^2$ ChIP-Seq data were downloaded from ENA accessions PRJDB5192 (Inagaki et al., 2017) and PRJNA427432 (Lai et al., 2020). Reads were aligned to the TAIR10 reference genome using Bowtie2 (version 2.3.5.1) (Langmead and Salzberg, 2012) with the parameters –mm –very-sensitive –maxins 800 –no-mixed –no-discordant. Counts per million normalised coverage profiles were generated using deepTools (version 3.4.3) (Ramirez et al., 2014). For 3′ end centred metagene profiles, we determined the major 3′ position per gene using the Araport11 annotation and existing Col-0 Helicos DRS data (Sherstnev et al., 2012). Metagenes centred on these positions were then generated in Python 3.6 using pyBigWig (version 0.3.17) (Ramirez et al., 2014), Numpy (version 1.18.1) (Harris et al., 2020) and Matplotlib (version 3.1.3) (Hunter, 2007). For differential $H3K9me^2$ analysis, read counts per gene (including intronic regions) were generated using pysam (version 0.16.0), and differential expression analysis was performed using edgeR (version 3.22.5) (Robinson et al., 2010).

RNA

Total RNA isolation

Total RNA was isolated using RNeasy Plant Mini kit (QIAGEN) and treated with TURBO DNase (Thermo Fisher Scientific) according to the manufacturers' instructions. The total RNA concentration was measured using a Qubit 1.0 Fluorometer and Qubit RNA BR Assay Kit (Thermo Fisher Scientific), whilst RNA quality and integrity was assessed using a

785 NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific) and Agilent 2200
786 TapeStation System (Agilent).
787
788 **Nanopore DRS**
789 *Preparation of libraries for DRS using nanopores*
790 Total RNA was isolated from Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as described above.
791 mRNA was isolated and Nanopore DRS libraries prepared (using the SQK-RNA001
792 Nanopore DRS Kit; Oxford Nanopore Technologies) as previously described (Parker et al.,
793 2020). Libraries were loaded onto R9.4 flow cells (Oxford Nanopore Technologies) and
794 sequenced using a 48-h runtime. Four biological replicates were performed for each
795 genotype.
796
797 *Nanopore DRS data processing*
798 Nanopore DRS reads were basecalled using the Guppy (version 3.6.0) high accuracy
799 model. Reads were mapped to the Arabidopsis *TAIR10* genome (The Arabidopsis
800 Genome, 2000) using minimap2 (version 2.17) with the parameters -a -L --cs=short -x splice
801 -G20000 --end-seed-pen=12 --junc-bonus=12 -uf. Spliced alignment was guided using
802 junctions from the Araport11 annotation (Cheng et al., 2017). Nanopore DRS reads can
803 suffer from "oversplitting" – where the signal originating from a single RNA molecule is
804 incorrectly interpreted as two or more reads (Parker et al., 2020). These errors can be
805 systematic and result in false positive 3′ ends. To filter these errors, we identified reads that
806 were sequenced consecutively through the same pore and also mapped contiguously on
807 the genome (within 1 kb of each other). In this way, we filtered all except the most 3′ reads,
808 which should contain the genuine RNA 3′ end. Pipelines for processing Nanopore DRS data
809 were built using Snakemake (Koster and Rahmann, 2012).
810
811 **Helicos DRS**
812 *Preparation of samples for Helicos DRS*
813 Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as described
814 above. Samples were processed by Helicos BioSciences as previously described (Ozsolak
815 et al., 2009; Sherstnev et al., 2012). Three biological replicates were performed for each
816 genotype.
817
818 *Helicos DRS data processing*
819 Helicos DRS reads were mapped to the *Arabidopsis* TAIR10 genome using Heliosphere
820 (version 1.1.498.63) as previously described (Sherstnev et al., 2012). Reads were filtered to
821 remove those with insertions or deletions of >4 nt and to mask regions with low complexity,
822 as determined using DustMasker (Camacho et al., 2009) (from BLAST+ suite version 2.10.1)
823 set at DUST level 15 (Sherstnev et al., 2012).
824

825 **Differential 3′ end analysis of Nanopore and Helicos DRS datasets**
826 Transcriptional loci were first identified in Col-0, *fpa*-8 and *35S::FPA:YFP* Nanopore DRS
827 reads using the long-read transcript assembly tool StringTie2 version 2.1.1 (Pertea et al.,
828 2015). Novel transcriptional loci were merged with annotated loci from the Araport11
829 reference (Cheng et al., 2017). To detect sites with altered 3′ end distributions in *fpa-8* and
830 *35S::FPA:YFP* mutants, we pooled the replicates of either Nanopore or Helicos DRS data
831 and identified reads overlapping each transcriptional locus. These reads were used to build
832 distributions of 3′ end locations. The difference in 3′ end distributions between the
833 treatment and control (Col-0) was measured using EMD. To identify loci with statistically
834 significant differences in 3′ distributions, we performed an EMD permutation test using 999
835 bootstraps: for this, reads for each locus were randomly shuffled between the treatment
836 and control samples to create null distributions, and the EMD recalculated. The histogram
837 of null EMDs was fitted using a gamma distribution, and the *p*-value (probability of
838 achieving the observed EMD or greater by chance) was calculated from the distribution. *P*-
839 values were corrected for multiple testing using the Benjamini–Hochberg method. Genes
840 with an EMD>25 and an FDR<0.05 were considered to be differentially alternatively
841 polyadenylated, and the directionality of change was identified using the difference in
842 mean 3′ position. Software developed to perform differential 3′ analysis is available on
843 GitHub at https://github.com/bartongroup/d3pendr and 10.5281/zenodo.4319113, and
844 can be used with Nanopore DRS, Helicos DRS, or Illumina 3′ tag-based datasets.
845
846 **Illumina RNA sequencing**
847 *Preparation of libraries for Illumina RNA sequencing*
848 Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as described
849 above. mRNA was isolated and sequencing libraries prepared using the NEBNext Ultra
850 Directional RNA Library Prep Kit for Illumina (New England Biolabs) by the Centre for
851 Genomic Research (University of Liverpool). 150-bp paired-end sequencing was carried out
852 on Illumina HiSeq 4000. Six biological replicates were performed for each genotype.
853
854 *Illumina RNA sequencing data processing*
855 Illumina RNA-Seq data were assessed for quality using FastQC (version 0.11.9) and MultiQC
856 (version 1.8) (Andrews, 2010; Ewels et al., 2016). Reads were mapped to the TAIR10
857 genome using STAR (version 2.7.3a) (Dobin et al., 2013) with a splice junction database
858 generated from the Araport11 reference annotation (Cheng et al., 2017). Counts per million
859 normalised coverage tracks were created using samtools (version 1.10) and deepTools
860 (version 3.4.3) (Ramirez et al., 2014). To identify expressed regions in each locus, the
861 coverage profiles of each treatment and control replicate were first extracted using
862 pyBigWig (version 0.3.17) (Ramirez et al., 2014). These were normalised such that the area
863 under each profile was equal to the mean area under the profiles. A normalised coverage
864 threshold of 1 was used to identify expressed regions of the loci. These regions were further

segmented when at least two-fold differences in expression within a 25-nt window were found between control and treatment conditions (and then regions smaller than 50 nt removed). Expression of the segmented regions was then calculated using featureCounts (version 2.0.0) (Liao et al., 2013). Each read pair was counted as one fragment, and only properly paired, concordant and primary read pairs were considered. Differential usage within transcriptional loci was assessed using DEXSeq (version 1.32.0) (Reyes et al., 2013). Loci were considered to be differentially processed if they had a locus-level FDR<0.05 and at least one region with an absolute logFC>1 and FDR<0.05. For differential splice junction usage analysis, counts of splice junctions annotated in the bespoke Nanopore DRS-derived annotation, plus Araport11 annotation, were generated for each locus using pysam (version 0.16.0). Differential splice junction usage was assessed using DEXSeq (version 1.32.0) (Reyes et al., 2013). Loci were considered to be differentially spliced if they had a locus-level FDR<0.05 and at least one junction with an absolute logFC>1 and FDR<0.05.

Gene tracks

Gene track figures were generated in Python 3.6 using Matplotlib (version 3.1.3) (Hunter, 2007). For gene tracks where any condition had >200 Nanopore DRS read alignments, 200 representative alignments were selected by random sampling without replacement (except for the *FPA* gene track figure, where 500 read alignments were sampled). NanoPARE data (Schon et al., 2018) were processed as previously described (Parker et al., 2020). For reannotated gene loci, domains were predicted using the InterproScan web client (Mitchell et al., 2019) and LRRs were predicted using LRRpredictor web client (Martin et al., 2020). Protein alignments were created and visualised in Jalview (version 2.11) (Waterhouse et al., 2009) using T-Coffee (Notredame et al., 2000).

Protein domain family enrichment analysis

To conduct protein domain enrichment analysis, InterPro domain annotations of Arabidopsis proteins were downloaded from BioMart (Smedley et al., 2009) and converted to genomic co-ordinates using the Araport11 annotation (Cheng et al., 2017). Domain families overlapping each locus tested for alternative polyadenylation using either Nanopore or Helicos DRS were identified using pybedtools (version 0.8.1) (Dale et al., 2011). To identify enriched domain families, domains were randomly shuffled between tested loci in 19,999 bootstraps, and the number of times that each domain class overlapped by chance with significantly alternatively polyadenylated loci was recorded. This was compared with the observed overlap of each domain family with alternatively polyadenylated loci to calculate *p*-values, which were corrected for multiple testing using the Benjamini–Hochberg method.

903 **Manual annotation of alternatively polyadenylated NLR genes**
904 To identify which of the 206 previously annotated NLR genes present in Col-0 were
905 alternatively polyadenylated in *fpa-8* and *35S::FPA*, we devised a standard operating
906 procedure for visual inspection. Genes that had Nanopore DRS read coverage in at least
907 two conditions were considered to be expressed. Genes were considered to be
908 alternatively polyadenylated if they had multiple 3′ end locations with each supported by
909 at least four Nanopore DRS reads, and if there was a clear difference in Nanopore DRS
910 coverage in the treatment condition compared with Col-0. Helicos and Illumina
911 corroboration of poly(A) sites and coverage changes was also taken into consideration.
912
913 **Genomic organisation of alternatively polyadenylated NLR genes**
914 To test whether expressed NLR genes with FPA-dependent alternative polyadenylation
915 were associated with NLR gene clusters, we used previously produced cluster assignments
916 for Col-0 NLR genes (Lee and Chae, 2020). We also tested the association of FPA-
917 dependent alternative polyadenylation with previously produced hypervariable NLR
918 classifications (Prigozhin and Krasileva, 2021). The association of alternatively
919 polyadenylated genes with both major NLR gene clusters and hypervariable NLRs was
920 assessed using a Chi squared test. To test whether FPA-regulated NLRs are found in regions
921 with high synteny diversity, we used 5-kb sliding window estimates of synteny diversity
922 calculated from seven diverse Arabidopsis ecotypes (Jiao and Schneeberger, 2020). For
923 each expressed NLR gene, the window with the largest overlap was used as the estimate of
924 synteny diversity. The association with alternatively polyadenylated genes was assessed
925 using a *t*-test.
926
927 **RNA gel blot analysis of *RPP7* mRNAs**
928 RNA gel blot analysis was carried out as previously described (Quesada et al., 2003) with
929 minor modifications. *RPP7* mRNA was detected using a probe annealing to the second
930 exon of the *RPP7* (*AT1G58602*) gene (200 bp PCR product amplified with the following
931 primers: Forward: 5′-TCGGGGACTACTACTACTCAAGA-3′ and Reverse: 5′-
932 TCTTGATGGTGTGAAAGAATCTAGT-3′). *Beta-TUBULIN* mRNA was used as a loading
933 control and visualised by a probe annealing to the third exon of the *Beta-TUBULIN*
934 *(AT1G20010) gene* (550 bp PCR product amplified with the following primers: Forward: 5′-
935 CTGACCTCAGGAAACTCGCG-3′ and Reverse: 5′- CATCAGCAGTAGCATCTTGG-3′). The
936 probes were 5′ labelled using [γ-^{32}P]-ATP (Perkin Elmer) and DECAprime™ II DNA labelling
937 kit (Thermo Fisher Scientific) and purified on illustra G-50 columns (GE Healthcare Life
938 Sciences). mRNA isoforms were visualised and quantified using an Amersham Typhoon Gel
939 and Blot Imaging System (GE Healthcare Bio-Sciences AB). The RiboRuler High Range RNA
940 Ladder (Thermo Fisher Scientific), used to identify the approximate size of RNA bands, was
941 first dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher

942 Scientific) and then labelled with [γ-^{32}P]-ATP (Perkin Elmer) using T4 Polynucleotide Kinase
943 (Thermo Fisher Scientific) before gel loading.
944
945 **m^{6}A LC-MS/MS**
946 Total RNA was isolated and checked as described above. mRNA was extracted twice from
947 approximately 75 µg of total RNA using the Dynabeads mRNA Purification Kit (Thermo
948 Fisher Scientific) according to the manufacturer's instructions. The quality and quantity of
949 mRNA was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific)
950 and Agilent 2200 TapeStation System (Agilent). Samples for m^{6}A LC-MS/MS were prepared
951 as previously described (Huang et al., 2018) with several modifications. First, 100 ng mRNA
952 was diluted in a total volume of 14 ml nuclease-free water (Thermo Fisher Scientific) and
953 digested by nuclease P1 (1 U, Merck) in 25 µl buffer containing 20 mM NH_4Oac (pH 5.3) at
954 42°C for 2 h. Next, 3 µl freshly made 1 M NH_4HCO_3 and calf intestinal alkaline phosphatase
955 (1 U, New England Biolabs) were added, and samples were incubated at 37°C for 2 h. The
956 samples were then diluted to 50 µl with nuclease-free water and filtered (0.22 µm pore size,
957 4 mm diameter; Millipore). LC-MS/MS was carried out by the FingerPrints Proteomics
958 facility at the University of Dundee. m^{6}A/A ratio quantification was performed in
959 comparison with the curves obtained from pure adenosine (endogenous P1 receptor
960 agonist, Abcam) and m^{6}A (modified adenosine analog, Abcam) nucleoside standards.
961 Statistical analysis was performed using a 2-way *t*-test.
962
963 **Pathogenesis assays**
964 Pathogenesis assays were carried out as previously described (Tome et al., 2014). The *Hpa*-
965 Hiks1 isolate was maintained by weekly sub-culturing on Ksk-1 plants. A solution containing
966 *Hpa*-Hiks1 spores was used to inoculate 14-day-old Col-0, Ksk-1, *fpa-7, fpa-8, pFPA::FPA*
967 and *35S::FPA:YFP* seedlings. Sporangiophores were counted 4 days after inoculation. The
968 experiment was repeated three times using up to 45 plants per genotype per each repeat.
969 Statistical analysis was performed with negative binomial regression using Statsmodels
970 (version 0.11.0) (Seabold and Perktold, 2010), plants were grouped by experimental repeat
971 during testing to control for variation between repeats.
972
973 **Code availability**
974 All pipelines, scripts and notebooks used to generate figures are available from GitHub at
975 https://github.com/bartongroup/Simpson_Barton_FPA_NLRs and Zenodo at
976 10.5281/zenodo.4319108. The software tool developed for detecting changes in poly(A)
977 site choice in Nanopore and Helicos DRS data are available from GitHub at
978 https://github.com/bartongroup/d3pendr and Zenodo at 10.5281/zenodo.4319112.
979

980 **Data availability**

981 *IVI-MS* data are available from the ProteomeXchange Consortium via the PRIDE (Perez-
982 Riverol et al., 2019) partner repository, with the dataset identifier PXD022684. FPA and Pol II
983 ChIP-Seq data are available from ENA accession PRJNA449914 (Yu et al., 2019). Col-0
984 Nanopore DRS data are available from ENA accession PRJEB32782 (Parker et al., 2020).
985 *fpa-8* and *35S::FPA:YFP* Nanopore DRS data are available from ENA accession
986 PRJEB41451. *Hen2-2* Nanopore DRS data are available from ENA accession PRJEB41381.
987 Col-0, *fpa-8* and *35S::FPA:YFP* Helicos DRS data are available from Zenodo
988 DOI 10.5281/zenodo.4309752. Col-0, *fpa-8* and *35S::FPA:YFP* Illumina RNA-Seq data are
989 available from ENA accession PRJEB41455.

990

991 **Acknowledgments**

992 We thank Paul Birch and Ingo Hein for comments on the manuscript and David Baulcombe,
993 Ian Henderson and Wenbo Ma for helpful NLR discussions. We thank Abdelmadjid Atrih
994 (Centre for Advanced Scientific Technologies, School of Life Sciences) for the m^6A LC-
995 MS/MS analysis.

996

997 **Competing interests**

998 The authors declare no competing interests.

999

1000 **Funding**

1001 This work was supported by awards from the BBSRC (BB/M010066/1; BB/J00247X/1;
1002 BB/M004155/1), the University of Dundee Global Challenges Research Fund, a University
1003 of Dundee PhD studentship to V.Z. and a European Union Horizon 2020 research and
1004 innovation programme under Marie Skłodowska-Curie grant agreement No. 799300 to K.K.
1005 The FingerPrints Proteomics Facility of the University of Dundee is supported by a
1006 Wellcome Trust Technology Platform Award (097945/B/11/Z).

1007

1008 ## References

1009

1010 Adachi, H., Derevnina, L., and Kamoun, S. (2019). NLR singletons, pairs, and networks:
1011 evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants.
1012 Curr Opin Plant Biol *50*, 121-131.
1013 Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., and Sorek, R.
1014 (2006). Transcription-mediated gene fusion in the human genome. Genome Res *16*, 30-36.
1015 Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-
1016 seq data. Genome Res *22*, 2008-2017.
1017 Andres, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal
1018 cues. Nat Rev Genet *13*, 627-639.
1019 Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

1020 Ariyoshi, M., and Schwabe, J.W. (2003). A conserved structural motif reveals the essential
1021 transcriptional repression function of Spen proteins and their role in developmental
1022 signaling. Genes Dev *17*, 1909-1920.
1023 Bach-Pages, M., Homma, F., Kourelis, J., Kaschani, F., Mohammed, S., Kaiser, M., van der
1024 Hoorn, R., Castello, A., and Preston, G. (2020). Discovering the RNA-Binding Proteome of
1025 Plant Leaves with an Improved RNA Interactome Capture Method. Biomolecules *10, 661*.
1026 Barragan, A.C., Collenberg, M., Wang, J., Lee, R.R.Q., Cher, W.Y., Rabanal, F.A., Ashkenazy,
1027 H., Weigel, D., Chae, E., and Patricia, W. (2021). A Truncated Singleton NLR Causes Hybrid
1028 Necrosis in Arabidopsis thaliana. Molecular Biology and Evolution *38*, 557-574.
1029 Barragan, A.C., and Weigel, D. (2020). Plant NLR Diversity: The Known Unknowns of Pan-
1030 NLRomes. The Plant Cell, koaa002.
1031 Baurle, I., Smith, L., Baulcombe, D.C., and Dean, C. (2007). Widespread role for the
1032 flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. Science *318*,
1033 109-112.
1034 Bomblies, K., and Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow
1035 barrier in plant species. Nat Rev Genet *8*, 382-393.
1036 Bruggeman, Q., Garmier, M., de Bont, L., Soubigou-Taconnat, L., Mazubert, C., Benhamed,
1037 M., Raynaud, C., Bergounioux, C., and Delarue, M. (2014). The Polyadenylation Factor
1038 Subunit CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR30: A Key Factor of
1039 Programmed Cell Death and a Regulator of Immunity in Arabidopsis. Plant Physiology *165*,
1040 732-746.
1041 Cai, Q., Liang, C., Wang, S., Hou, Y., Gao, L., Liu, L., He, W., Ma, W., Mo, B., and Chen, X.
1042 (2019). Author Correction: The disease resistance protein SNC1 represses the biogenesis of
1043 microRNAs and phased siRNAs. Nat Commun *10*, 642.
1044 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden,
1045 T.L. (2009). BLAST+: architecture and applications. BMC bioinformatics *10*, 421-421.
1046 Canto-Pastor, A., Santos, B., Valli, A.A., Summers, W., Schornack, S., and Baulcombe, D.C.
1047 (2019). Enhanced resistance to bacterial and oomycete pathogens by short tandem target
1048 mimic RNAs in tomato. Proc Natl Acad Sci U S A *116*, 2755-2760.
1049 Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D.
1050 (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome.
1051 The Plant Journal *89*, 789-804.
1052 Collado-Torres, L., Nellore, A., Frazee, A.C., Wilks, C., Love, M.I., Langmead, B., Irizarry, R.A.,
1053 Leek, J.T., and Jaffe, A.E. (2017). Flexible expressed region analysis for RNA-seq with
1054 derfinder. Nucleic Acids Res *45*, e9.
1055 Corwin, J.A., and Kliebenstein, D.J. (2017). Quantitative Resistance: More Than Just
1056 Perception of a Pathogen. Plant Cell *29*, 655-665.
1057 Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates,
1058 individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat
1059 Biotechnol *26*, 1367-1372.
1060 Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011). Pybedtools: a flexible Python library for
1061 manipulating genomic datasets and annotations. Bioinformatics *27*, 3423-3424.
1062 Deremetz, A., Le Roux, C., Idir, Y., Brousse, C., Agorio, A., Gy, I., Parker, J.E., and Bouche, N.
1063 (2019). Antagonistic Actions of FPA and IBM2 Regulate Transcript Processing from Genes
1064 Containing Heterochromatin. Plant Physiol *180*, 392-403.

1065 Dinesh-Kumar, S.P., and Baker, B.J. (2000). Alternatively spliced N resistance gene
1066 transcripts: their possible role in tobacco mosaic virus resistance. Proc Natl Acad Sci U S A
1067 *97*, 1908-1913.
1068 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
1069 and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-
1070 21.
1071 Duc, C., Sherstnev, A., Cole, C., Barton, G.J., and Simpson, G.G. (2013). Transcription
1072 termination and chimeric RNA formation controlled by *Arabidopsis thaliana* FPA. PLoS
1073 Genet *9*, e1003867.
1074 Ewels, P., Magnusson, M., Käller, M., and Lundin, S. (2016). MultiQC: summarize analysis
1075 results for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048.
1076 Fu, Z.Q., Guo, M., Jeong, B.R., Tian, F., Elthon, T.E., Cerny, R.L., Staiger, D., and Alfano, J.R.
1077 (2007). A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity.
1078 Nature *447*, 284-288.
1079 Gierlinski, M., Gastaldello, F., and Barton, G.J. (2018). Proteus: an R package for downstream
1080 analysis of MaxQuant output. bioRxiv, 416511.
1081 Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor, J.,
1082 Dekrout, B., Jonak, C., Jimenez-Gomez, J.M.*, et al.* (2014). Nonsense-mediated mRNA decay
1083 modulates immune receptor levels to regulate plant antibacterial defense. Cell Host
1084 Microbe *16*, 376-390.
1085 Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D.,
1086 Wieser, E., Taylor, J., Berg, S., Smith, N.J.*, et al.* (2020). Array programming with NumPy.
1087 Nature *585*, 357-362.
1088 Hornyik, C., Terzi, L.C., and Simpson, G.G. (2010). The spen family protein FPA controls
1089 alternative cleavage and polyadenylation of RNA. Dev Cell *18*, 203-213.
1090 Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B.S., Mesquita, A., Liu, C., Yuan,
1091 C.L.*, et al.* (2018). Recognition of RNA N^6-methyladenosine by IGF2BP proteins enhances
1092 mRNA stability and translation. Nature Cell Biology *20*, 285-295.
1093 Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science &
1094 Engineering *9*, 90-95.
1095 Inagaki, S., Takahashi, M., Hosaka, A., Ito, T., Toyoda, A., Fujiyama, A., Tarutani, Y., and
1096 Kakutani, T. (2017). Gene-body chromatin modification dynamics mediate epigenome
1097 differentiation in Arabidopsis. EMBO J *36*, 970-980.
1098 Jiao, W.B., and Schneeberger, K. (2020). Chromosome-level assemblies of multiple
1099 Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary
1100 dynamics. Nat Commun *11*, 989.
1101 Jones, J.D.G., Vance, R.E., and Dangl, J.L. (2016). Intracellular innate immune surveillance
1102 devices in plants and animals. Science *354*, aaf6395-aaf6395.
1103 Knuckles, P., Lence, T., Haussmann, I.U., Jacob, D., Kreim, N., Carl, S.H., Masiello, I., Hares,
1104 T., Villasenor, R., Hess, D.*, et al.* (2018). Zc3h13/Flacc is required for adenosine methylation
1105 by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component
1106 Wtap/Fl(2)d. Gene Dev *32*, 415-429.
1107 Koiwa, H., Hausmann, S., Bang, W.Y., Ueda, A., Kondo, N., Hiraguri, A., Fukuhara, T., Bahk,
1108 J.D., Yun, D.J., Bressan, R.A.*, et al.* (2004). Arabidopsis C-terminal domain phosphatase-like 1
1109 and 2 are essential Ser-5-specific C-terminal domain phosphatases. Proc Natl Acad Sci U S A
1110 *101*, 14539-14544.

1111 Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA
1112 polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*,
1113 2452-2460.
1114 Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological
1115 analysis of late flowering mutants in Arabidopsis thaliana. Mol Gen Genet *229*, 57-66.
1116 Koster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics workflow engine.
1117 Bioinformatics *28*, 2520-2522.
1118 Krzyszton, M., Zakrzewska-Placzek, M., Kwasnik, A., Dojer, N., Karlowski, W., and Kufel, J.
1119 (2018). Defective XRN3-mediated transcription termination in Arabidopsis affects the
1120 expression of protein-coding genes. Plant J *93*, 1017-1031.
1121 Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between
1122 subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end
1123 processing and splicing. Mol Cell *23*, 195-205.
1124 Lai, Y., Cuzick, A., Lu, X.M., Wang, J., Katiyar, N., Tsuchiya, T., Le Roch, K., McDowell, J.M.,
1125 Holub, E., and Eulgem, T. (2019). The Arabidopsis RRM domain protein EDM3 mediates race-
1126 specific disease resistance by controlling H3K9me2-dependent alternative polyadenylation
1127 of RPP7 immune receptor transcripts. Plant J *97*, 646-660.
1128 Lai, Y., and Eulgem, T. (2018). Transcript-level expression control of plant NLR genes. Mol
1129 Plant Pathol *19*, 1267-1281.
1130 Lai, Y., Lu, X.M., Daron, J., Pan, S., Wang, J., Wang, W., Tsuchiya, T., Holub, E., McDowell,
1131 J.M., Slotkin, R.K.*, et al.* (2020). The Arabidopsis PHD-finger protein EDM2 has multiple roles
1132 in balancing NLR immune receptor gene expression. PLoS Genet *16*, e1008993.
1133 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat
1134 Methods *9*, 357-359.
1135 Lazar, C. (2015). imputeLCMD: a collection of methods for left-censored missing data
1136 imputation. R package, version *2*.
1137 Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Tremousaygue, D., Kraut, A., Zhou, B.,
1138 Levaillant, M., Adachi, H., Yoshioka, H.*, et al.* (2015). A receptor pair with an integrated
1139 decoy converts pathogen disabling of transcription factors to immunity. Cell *161*, 1074-
1140 1088.
1141 Lee, R.R.Q., and Chae, E. (2020). Variation Patterns of NLR Clusters in Arabidopsis thaliana
1142 Genomes. Plant Communications *1,* 100089.
1143 Li, L., Habring, A., Wang, K., and Weigel, D. (2020). Atypical Resistance Protein RPW8/HR
1144 Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity. Cell Host
1145 Microbe *27*, 405-417.
1146 Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable
1147 read mapping by seed-and-vote. Nucleic Acids Res *41*, e108-e108.
1148 Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S., and Dean, C. (2007). The
1149 Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to
1150 downregulate FLC. Mol Cell *28*, 398-407.
1151 Liu, F.Q., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3 '
1152 Processing of Antisense Transcripts Triggers Arabidopsis FLC Chromatin Silencing. Science
1153 *327*, 94-97.
1154 Lyons, R., Iwase, A., Gansewig, T., Sherstnev, A., Duc, C., Barton, G.J., Hanada, K., Higuchi-
1155 Takeuchi, M., Matsui, M., Sugimoto, K.*, et al.* (2013). The RNA-binding protein FPA regulates
1156 flg22-triggered defense responses and transcription factor activity by alternative
1157 polyadenylation. Sci Rep *3*, 2866.

1158 Ma, W., Zhen, G., Xie, W., and Mayr, C. (2020). Unstructured mRNAs form multivalent RNA-
1159 RNA interactions to generate TIS granule networks. bioRxiv, 2020.2002.2014.949503.
1160 Martin, E.C., Sukarta, O.C.A., Spiridon, L., Grigore, L.G., Constantinescu, V., Tacutu, R.,
1161 Goverse, A., and Petrescu, A.-J. (2020). LRRpredictor—A New LRR Motif Detection Method
1162 for Irregular Motifs of Plant NLR Proteins Using an Ensemble of Classifiers. Genes-Basel *11,*
1163 286.
1164 Mayr, C. (2019). What Are 3' UTRs Doing? Csh Perspect Biol *11*.a034728.
1165 Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang,
1166 H.Y., El-Gebali, S., Fraser, M.I.*, et al.* (2019). InterPro in 2019: improving coverage,
1167 classification and access to protein sequence annotations. Nucleic Acids Res *47*, D351-D360.
1168 Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An
1169 Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG
1170 sites. EMBO J *28*, 1078-1086.
1171 Mockler, T.C., Yu, X., Shalitin, D., Parikh, D., Michael, T.P., Liou, J., Huang, J., Smith, Z.,
1172 Alonso, J.M., Ecker, J.R.*, et al.* (2004). Regulation of flowering time in Arabidopsis by K
1173 homology domain proteins. Proc Natl Acad Sci U S A *101*, 12759-12764.
1174 Nishimura, M.T., Anderson, R.G., Cherkis, K.A., Law, T.F., Liu, Q.L., Machius, M., Nimchuk,
1175 Z.L., Yang, L., Chung, E.-H., El Kasmi, F.*, et al.* (2017). TIR-only protein RBA1 recognizes a
1176 pathogen effector to regulate cell death inArabidopsis. Proc Natl Acad Sci U S A *114*, E2053-
1177 E2062.
1178 Nishimura, M.T., Monteiro, F., and Dangl, J.L. (2015). Treasure your exceptions: unusual
1179 domains in immune receptors reveal host virulence targets. Cell *161*, 957-960.
1180 Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D. (2009). Plant upstream ORFs
1181 can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol Biol *71*,
1182 367-378.
1183 Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson,
1184 J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature *461*, 814-
1185 818.
1186 Parker, M.T., Knop, K., Barton, G.J., and Simpson, G.G. (2021). 2passtools: two-pass
1187 alignment using machine-learning-filtered splice junctions increases the accuracy of intron
1188 detection in long-read RNA sequencing. Genome Biology *22*.1-24
1189 Parker, M.T., Knop, K., Sherwood, A.V., Schurch, N.J., Mackinnon, K., Gould, P.D., Hall,
1190 A.J.W., Barton, G.J., and Simpson, G.G. (2020). Nanopore direct RNA sequencing maps the
1191 complexity of Arabidopsis mRNA processing and m6A modification. eLife *9*, e49658.
1192 Patil, D.P., Chen, C.K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R.
1193 (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature
1194 *537*, 369-373.
1195 Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J.,
1196 Inuganti, A., Griss, J., Mayer, G., Eisenacher, M.*, et al.* (2019). The PRIDE database and
1197 related tools and resources in 2019: improving support for quantification data. Nucleic Acids
1198 Res *47*, D442-D450.
1199 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L.
1200 (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.
1201 Nat Biotechnol *33*, 290-295.
1202 Prigozhin, D.M., and Krasileva, K.V. (2021). Analysis of intraspecies diversity reveals a subset
1203 of highly variable plant immune receptors and predicts their binding sites. The Plant Cell,
1204 koab013.

1205 Proudfoot, N.J. (1986). Transcriptional interference and termination between duplicated
1206 alpha-globin gene constructs suggests a novel mechanism for gene regulation. Nature *322*,
1207 562-565.
1208 Quesada, V., Macknight, R., Dean, C., and Simpson, G.G. (2003). Autoregulation of *FCA* pre-
1209 mRNA processing controls *Arabidopsis* flowering time. EMBO J *22*, 3142-3152.
1210 Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible
1211 platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.
1212 Raxwal, V.K., and Riha, K. (2016). Nonsense mediated RNA decay and evolutionary
1213 capacitance. Biochim Biophys Acta *1859*, 1538-1543.
1214 Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A.M., Horos, R., Hentze,
1215 M.W., Preiss, T., and Millar, A.A. (2016). *In planta* determination of the mRNA-binding
1216 proteome of Arabidopsis etiolated seedlings. Plant Cell *28*, 2435-2452.
1217 Reyes, A., Anders, S., Weatheritt, R.J., Gibson, T.J., Steinmetz, L.M., and Huber, W. (2013).
1218 Drift and conservation of differential exon usage across tissues in primate species. Proc Natl
1219 Acad Sci U S A *110*, 15377-15382.
1220 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma
1221 powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic
1222 Acids Res *43*, e47.
1223 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for
1224 differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
1225 Rodriguez, E., El Ghoul, H., Mundy, J., and Petersen, M. (2016). Making sense of plant
1226 autoimmunity and 'negative regulators'. FEBS J *283*, 1385-1391.
1227 Rodriguez-Cazorla, E., Ripoll, J.J., Andujar, A., Bailey, L.J., Martinez-Laborda, A., Yanofsky,
1228 M.F., and Vera, A. (2015). K-homology nuclear ribonucleoproteins regulate floral organ
1229 identity and determinacy in arabidopsis. PLoS Genet *11*, e1004983.
1230 Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., Eeckhout, D., El-Showk,
1231 S., Li, H., Zhong, S*., et al.* (2017). Identification of factors required for m^6A mRNA
1232 methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. New
1233 Phytol *215*, 157-172.
1234 Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. (2003). Independent functions of yeast
1235 Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J *22*, 2167-
1236 2177.
1237 Saze, H., Kitayama, J., Takashima, K., Miura, S., Harukawa, Y., Ito, T., and Kakutani, T. (2013).
1238 Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic
1239 heterochromatin. Nat Commun *4*, 2301.
1240 Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. (2008). Control of genic DNA methylation
1241 by a jmjC domain-containing protein in Arabidopsis thaliana. Science *319*, 462-465.
1242 Schomburg, F.M., Patton, D.A., Meinke, D.W., and Amasino, R.M. (2001). FPA, a gene
1243 involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition
1244 motifs. Plant Cell *13*, 1427-1436.
1245 Schon, M.A., Kellner, M.J., Plotnikova, A., Hofmann, F., and Nodine, M.D. (2018). NanoPARE:
1246 parallel analysis of RNA 5' ends from low-input RNA. Genome Res *28*, 1931-1942.
1247 Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with
1248 Python. In Proceedings of the 9th Python in Science Conference, pp. 92-96.
1249 Sherstnev, A., Duc, C., Cole, C., Zacharaki, V., Hornyik, C., Ozsolak, F., Milos, P.M., Barton,
1250 G.J., and Simpson, G.G. (2012). Direct sequencing of *Arabidopsis thaliana* RNA reveals
1251 patterns of cleavage and polyadenylation. Nat Struct Mol Biol *19*, 845-852.

1252 Shivaprasad, P.V., Chen, H.M., Patel, K., Bond, D.M., Santos, B.A., and Baulcombe, D.C.
1253 (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and
1254 other mRNAs. Plant Cell *24*, 859-874.
1255 Simpson, G.G. (2004). The autonomous pathway: epigenetic and post-transcriptional gene
1256 regulation in the control of Arabidopsis flowering time. Curr Opin Plant Biol *7*, 570-574.
1257 Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A.
1258 (2009). BioMart--biological queries made easy. BMC Genomics *10*, 22.
1259 Swiderski, M.R., Birker, D., and Jones, J.D.G. (2009). The TIR Domain of TIR-NB-LRR
1260 Resistance Proteins Is a Signaling Domain Involved in Cell Death Induction. Molecular Plant-
1261 Microbe Interactions *22*, 157-165.
1262 Szádeczky-Kardoss, I., Csorba, T., Auber, A., Schamberger, A., Nyikó, T., Taller, J., Orbán, T.I.,
1263 Burgyán, J., and Silhavy, D. (2018). The nonstop decay and the RNA silencing systems
1264 operate cooperatively in plants. Nucleic Acids Res *46*, 4632-4648.
1265 Takagi, M., Iwamoto, N., Kubo, Y., Morimoto, T., Takagi, H., Takahashi, F., Nishiuchi, T.,
1266 Tanaka, K., Taji, T., Kaminaka, H.*, et al.* (2020). Arabidopsis SMN2/HEN2, Encoding DEAD-Box
1267 RNA Helicase, Governs Proper Expression of the Resistance Gene SMN1/RPS6 and Is
1268 Involved in Dwarf, Autoimmune Phenotypes of mekk1 and mpk4 Mutants. Plant Cell Physiol
1269 *61*, 1507-1516.
1270 Tamborski, J., and Krasileva, K.V. (2020). Evolution of Plant NLRs: From Natural History to
1271 Precise Modifications. Annu Rev Plant Biol *71*, 355-378.
1272 The Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant
1273 *Arabidopsis thaliana*. Nature *408*, 796-815.
1274 Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-
1275 gene-mediated resistance in Arabidopsis thaliana. Nature *423*, 74-77.
1276 Tome, D.F., Steinbrenner, J., and Beynon, J.L. (2014). A growth quantification assay for
1277 Hyaloperonospora arabidopsidis isolates in Arabidopsis thaliana. Methods Mol Biol *1127*,
1278 145-158.
1279 Tran Vdu, T., Souiai, O., Romero-Barrios, N., Crespi, M., and Gautheret, D. (2016). Detection
1280 of generic differential RNA processing events from RNA-seq data. RNA Biol *13*, 59-67.
1281 Tsuchiya, T., and Eulgem, T. (2013). An alternative polyadenylation mechanism coopted to
1282 the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad
1283 Sci U S A *110*, E3535-3543.
1284 Van de Weyer, A.L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K., Jones,
1285 J.D.G., Dangl, J.L., Weigel, D., and Bemm, F. (2019). A Species-Wide Inventory of NLR Genes
1286 and Alleles in Arabidopsis thaliana. Cell *178*, 1260-1272 e1214.
1287 van der Biezen, E.A., Freddie, C.T., Kahn, K., Parker, J.E., and Jones, J.D. (2002). Arabidopsis
1288 RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy
1289 mildew resistance through multiple signalling components. Plant J *29*, 439-451.
1290 Wang, X., Boevink, P., McLellan, H., Armstrong, M., Bukharova, T., Qin, Z., and Birch, P.R.
1291 (2015). A Host KH RNA-Binding Protein Is a Susceptibility Factor Targeted by an RXLR
1292 Effector to Promote Late Blight Disease. Mol Plant *8*, 1385-1395.
1293 Wang, Y.H., and Warren, J.T., Jr. (2010). Mutations in retrotransposon AtCOPIA4
1294 compromises resistance to Hyaloperonospora parasitica in Arabidopsis thaliana. Genet Mol
1295 Biol *33*, 135-140.
1296 Wang, Z.W., Wu, Z., Raitskin, O., Sun, Q., and Dean, C. (2014). Antisense-mediated FLC
1297 transcriptional repression requires the P-TEFb transcription elongation factor. Proc Natl
1298 Acad Sci U S A *111*, 7468-7473.

1299 Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview
1300 Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*,
1301 1189-1191.
1302 Wei, C., Chen, J., and Kuang, H. (2016). Dramatic Number Variation of R Genes in Solanaceae
1303 Species Accounted for by a Few R Gene Subfamilies. Plos One *11*, e0148708.
1304 Williams, S.J., Sohn, K.H., Wan, L., Bernoux, M., Sarris, P.F., Segonzac, C., Ve, T., Ma, Y.,
1305 Saucet, S.B., Ericsson, D.J.*, et al.* (2014). Structural basis for assembly and function of a
1306 heterodimeric plant immune receptor. Science *344*, 299-303.
1307 Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W.
1308 (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR
1309 landscape across seven tumour types. Nat Commun *5*, 5274.
1310 Xing, D.H., Zhao, H.W., Xu, R.Q., and Li, Q.S.Q. (2008). Arabidopsis PCFS4, a homologue of
1311 yeast polyadenylation factor Pcf11p, regulates FCA alternative processing and promotes
1312 flowering time. Plant Journal *54*, 899-910.
1313 Yang, S., Tang, F., Gao, M., Krishnan, H.B., and Zhu, H. (2010). R gene-controlled host
1314 specificity in the legume-rhizobia symbiosis. Proc Natl Acad Sci U S A *107*, 18735-18740.
1315 Yu, X., Martin, P.G.P., and Michaels, S.D. (2019). BORDER proteins protect expression of
1316 neighboring genes by promoting 3' Pol II pausing in plants. Nat Commun *10*, 4359.
1317 Zhai, J., Jeong, D.H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., Gonzalez, A.J., Yan, Z., Kitto, S.L.,
1318 Grusak, M.A.*, et al.* (2011). MicroRNAs as master regulators of the plant NB-LRR defense
1319 gene family via the production of phased, trans-acting siRNAs. Genes Dev *25*, 2540-2553.
1320 Zhang, P., Berardini, T.Z., Ebert, D., Li, Q., Mi, H., Muruganujan, A., Prithvi, T., Reiser, L.,
1321 Sawant, S., Thomas, P.D.*, et al.* (2020a). PhyloGenes: An online phylogenetics and functional
1322 genomics resource for plant gene function inference. Plant Direct *4*.
1323 Zhang, X.C., and Gassmann, W. (2007). Alternative splicing and mRNA levels of the disease
1324 resistance gene RPS4 are induced during defense responses. Plant Physiol *145*, 1577-1587.
1325 Zhang, Y., Dorey, S., Swiderski, M., and Jones, J.D. (2004). Expression of RPS4 in tobacco
1326 induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. Plant J *40*, 213-
1327 224.
1328 Zhang, Y., Gu, L., Hou, Y., Wang, L., Deng, X., Hang, R., Chen, D., Zhang, X., Zhang, Y., Liu, C.*,*
1329 *et al.* (2015). Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein,
1330 regulates plant flowering by targeting alternative polyadenylation. Cell Res *25*, 864-876.
1331 Zhang, Y., Li, Z., Chen, N., Huang, Y., and Huang, S. (2020b). Phase separation of Arabidopsis
1332 EMB1579 controls transcription, mRNA splicing, and development. Plos Biol *18*, e3000782.
1333 Zhang, Y., Rataj, K., Simpson, G.G., and Tong, L. (2016). Crystal Structure of the SPOC
1334 Domain of the Arabidopsis Flowering Regulator FPA. Plos One *11*, e0160694.
1335
1336

Figures

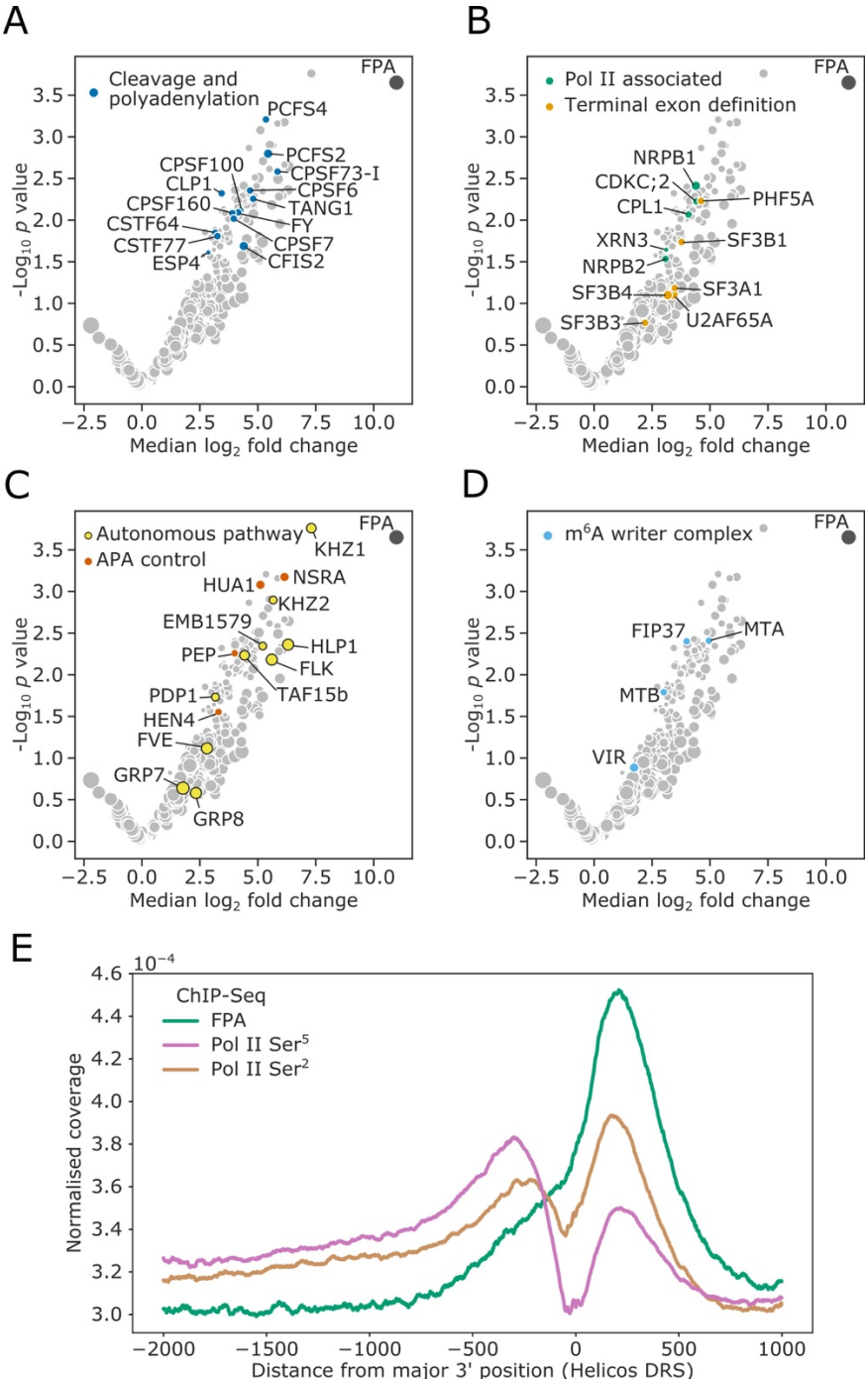


Figure 1: **FPA associates with proteins that function to process the 3′ ends of Pol II-transcribed RNAs and promote transcription termination.**

(A-D) Volcano plots representing proteins co-purifying with FPA using *IVI-MS*. Only proteins detected in all three biological replicates of the *35S::FPA:YFP* line are shown (light grey). The following classes are highlighted: **(A)** CPFs in dark blue; **(B)** Pol II-associated factors in green; terminal exon definition factors in dark orange; **(C)** autonomous pathway components in yellow and factors controlling alternative polyadenylation in light orange; and **(D)** m^{6}A writer complex components in light blue. **(E)** ChIP-Seq metagene profile showing the normalised occupancy of FPA (green) and Pol II phosphorylated at Ser^5 (pink) and Ser^2 (brown) of the CTD (Yu et al., 2019) relative to the major 3′ position of each gene, as measured using Helicos DRS. Only long genes (> 2.5 kb) are included (n=10,215).

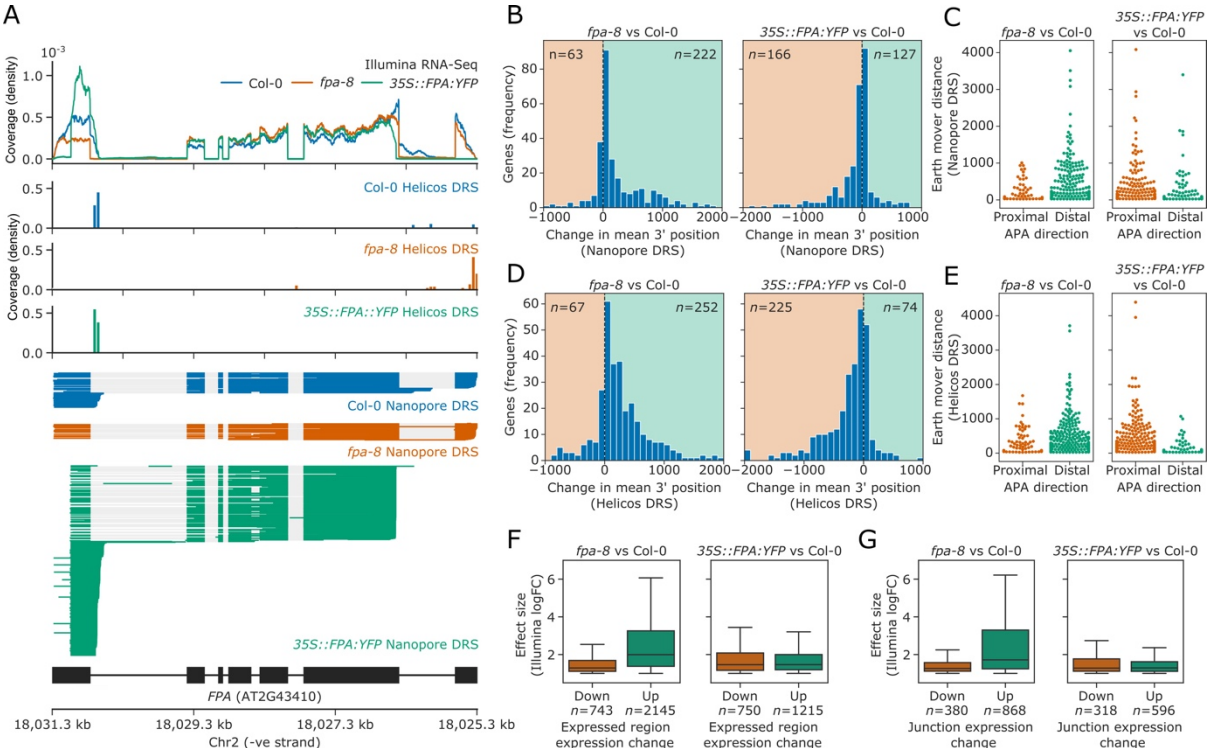


Figure 2: FPA regulates poly(A) site selection.

Loss of FPA function is associated with the preferential selection of distal poly(A) sites, whereas FPA overexpression leads to the preferential selection of proximal poly(A) sites. **(A)** Illumina RNA-Seq, Helicos DRS and Nanopore DRS reveal FPA-dependent RNA 3′ end processing changes at the *FPA* (*AT2G43410*) locus. The *35S::FPA:YFP* construct has alternative transgene-derived untranslated regions, so mRNAs derived from the transgene do not align to the native FPA 5′UTR and 3′UTR. **(B)** Histograms showing change in mean RNA 3′ end position for significantly alternatively polyadenylated loci (EMD>25, FDR<0.05) in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, as detected using Nanopore DRS. Orange and green shaded regions indicate sites with negative and positive RNA 3′ end position changes, respectively. **(C)** Effect size of significant proximal (orange) and distal (green) alternative polyadenylation events in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, as measured using the EMD. **(D)** Histograms showing change in mean RNA 3′ end position for significantly alternatively polyadenylated loci (EMD>25, FDR<0.05) in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, as detected using Nanopore DRS. Orange and green shaded regions indicate sites with negative and positive RNA 3′ end position changes, respectively. **(E)** Effect size of significant proximal (orange) and distal (green) alternative polyadenylation events in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, as measured using the EMD. **(F)** Boxplots showing the effect size (absolute $\log_2$ fold change (logFC)) of alternatively processed loci identified using Illumina RNA-Seq in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) respectively. Down- and upregulated loci are shown in orange and green, respectively. For each locus, the region with the largest logFC was selected to represent the locus. Loci with both up- and downregulated regions contribute to both boxes. **(G)** Boxplots showing the effect size (absolute logFC) of loci with alternative splice junction usage identified using Illumina RNA-Seq in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) respectively. Down- and upregulated loci are shown in orange and green, respectively. For each locus, the junction with the largest logFC was selected to represent the locus. Loci with both up- and downregulated junctions contribute to both boxes.

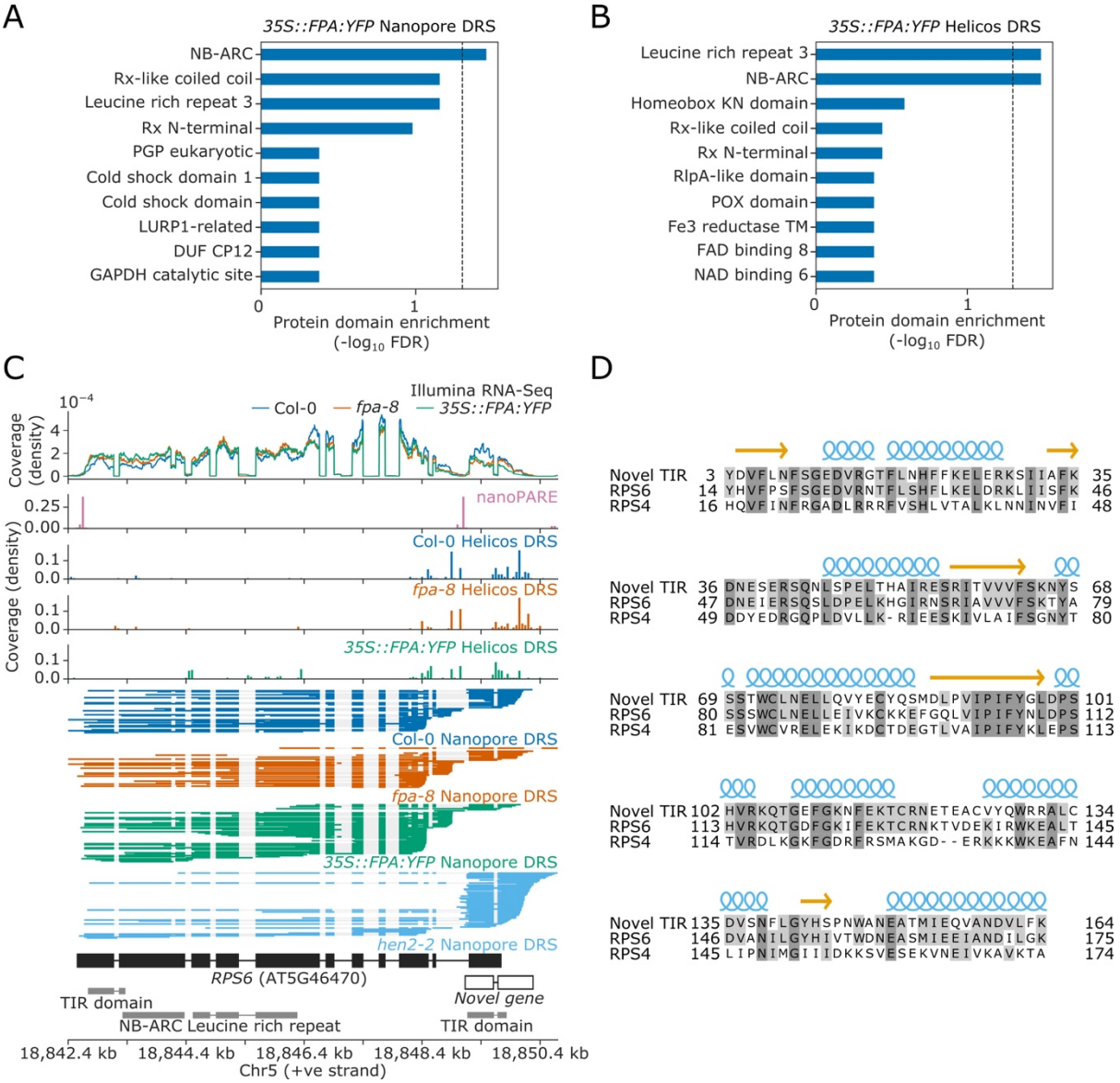


Figure 3: Nanopore and Helicos DRS identify NLR genes regulated by alternative polyadenylation.

(A-B) Protein domain enrichment analysis for loci with increased proximal poly(A) site selection in *35S::FPA:YFP* line, as detected using **(A)** Nanopore DRS or **(B)** Helicos DRS. **(C)** Nanopore DRS reveals the complexity of RNA processing at *RPS6*. Protein domain locations (shown in grey) represent collapsed InterPro annotations. The novel TIR domain was annotated using InterProScan (Mitchell et al., 2019). **(D)** Protein alignment of the predicted TIR domain from the novel gene downstream of RPS6, with the sequence of the TIR domains from RPS6 and RPS4. Helix and strand secondary structures (from UniProt: RPS4, Q9XGM3) are shown in blue and yellow, respectively. Residues are shaded according to the degree of conservation.

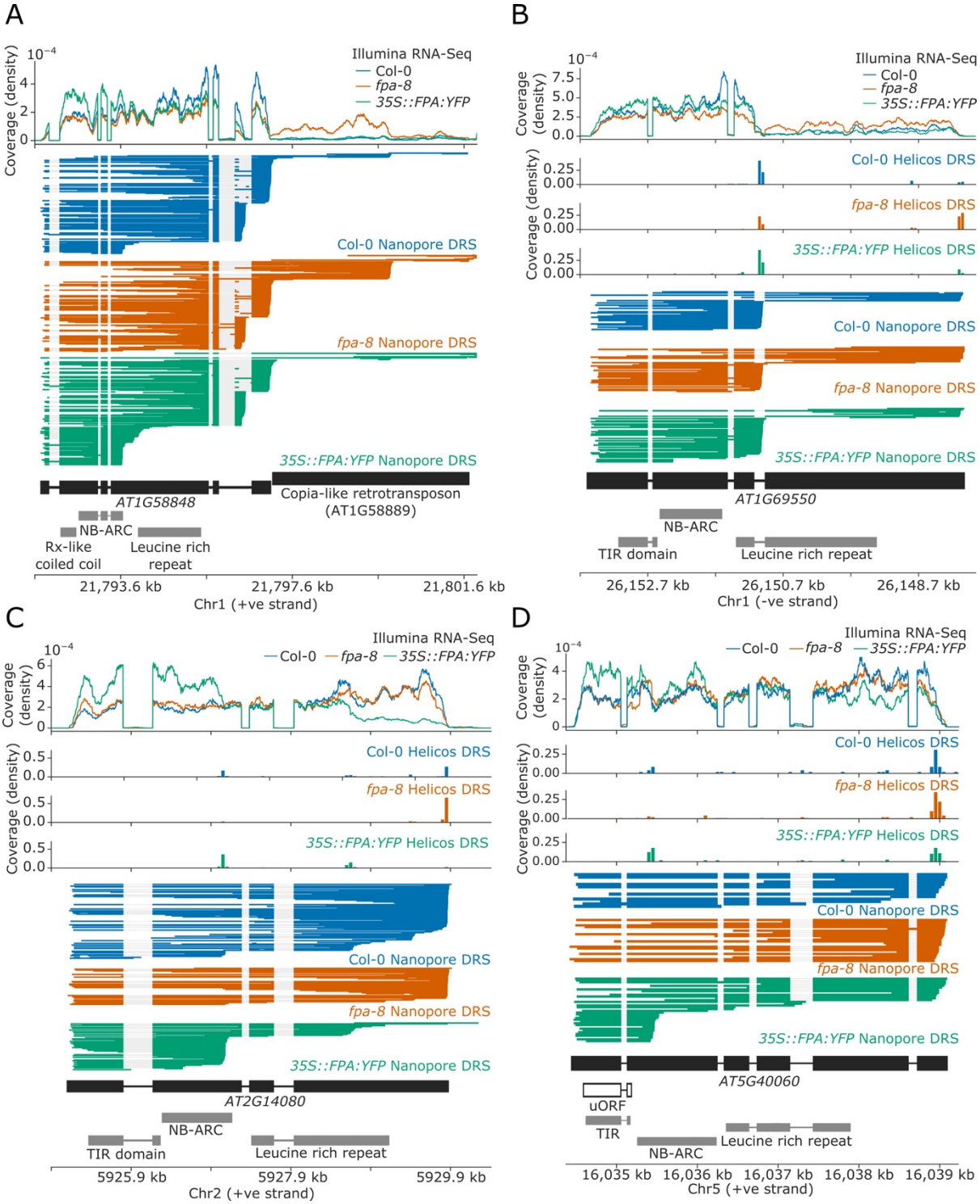


Figure 4: FPA-dependent alternative polyadenylation of NLR transcripts.
FPA regulates **(A)** readthrough and chimeric RNA formation at *AT1G58848* (unique mapping of short Helicos DRS reads was not possible due to the high homology of *AT1G58848* to tandemly duplicated NLR loci in the same cluster); **(B)** intronic polyadenylation at *AT1G69550*, resulting in transcripts encoding a protein with a truncated LRR domain; **(C)** exonic polyadenylation at *AT2G14080*, resulting in stop-codonless transcripts; and **(D)** exonic polyadenylation at *AT5G40060*, resulting in transcripts encoding a TIR-domain-only protein due to an upstream ORF.

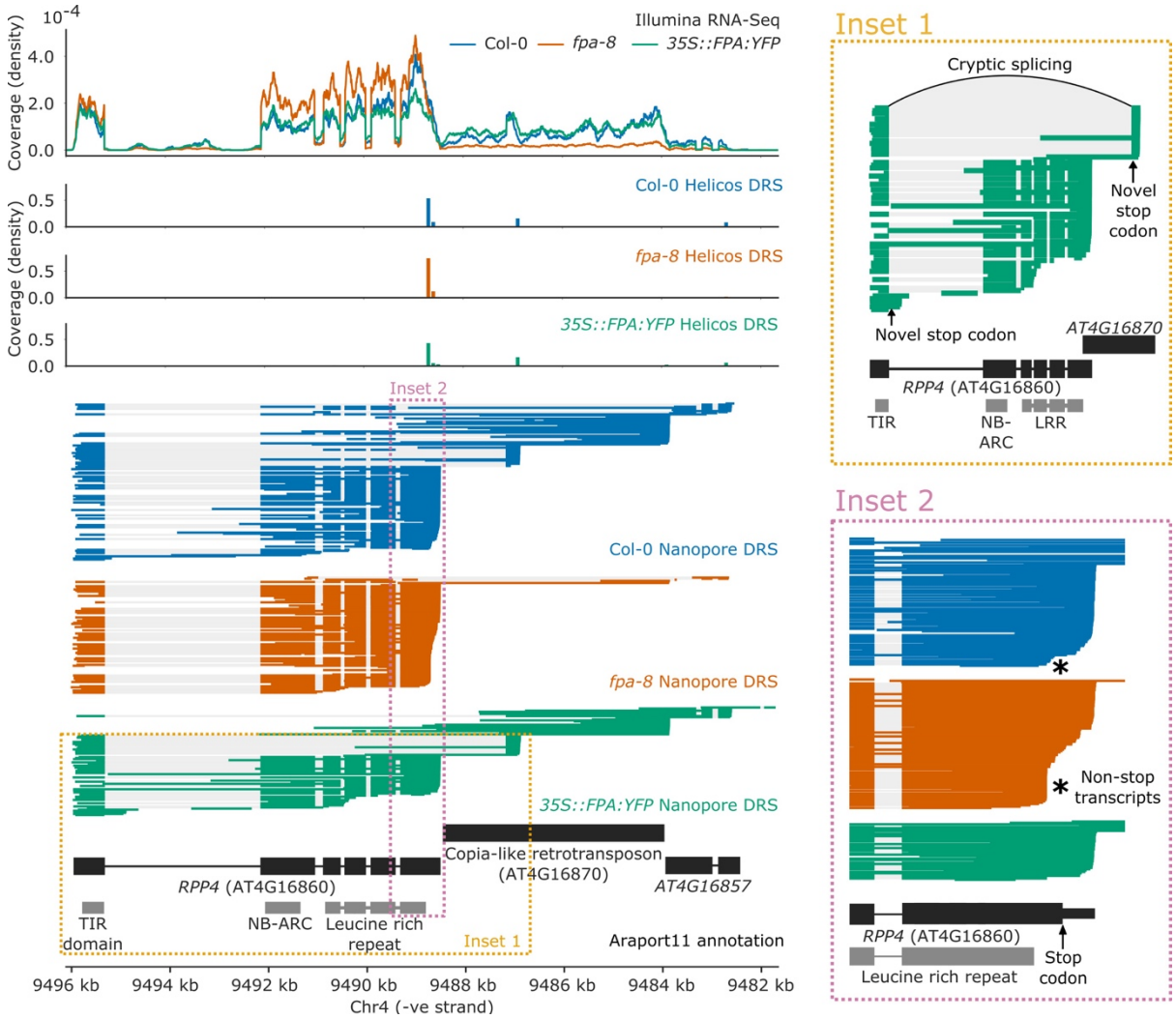


Figure 5: Complex FPA-dependent patterns of alternative polyadenylation at *RPP4*. FPA regulates intronic, exonic and readthrough poly(A) site selection in *RPP4*. **(Inset 1)** A magnified view of TIR-domain-only *RPP4* transcripts detected in *35S::FPA:YFP* caused by proximal polyadenylation in intron 1, and distal polyadenylation within the TE associated with cryptic splicing. **(Inset 2)** A magnified view of the stop-codonless transcripts produced within the protein-coding *RPP4* region in *fpa-8*.

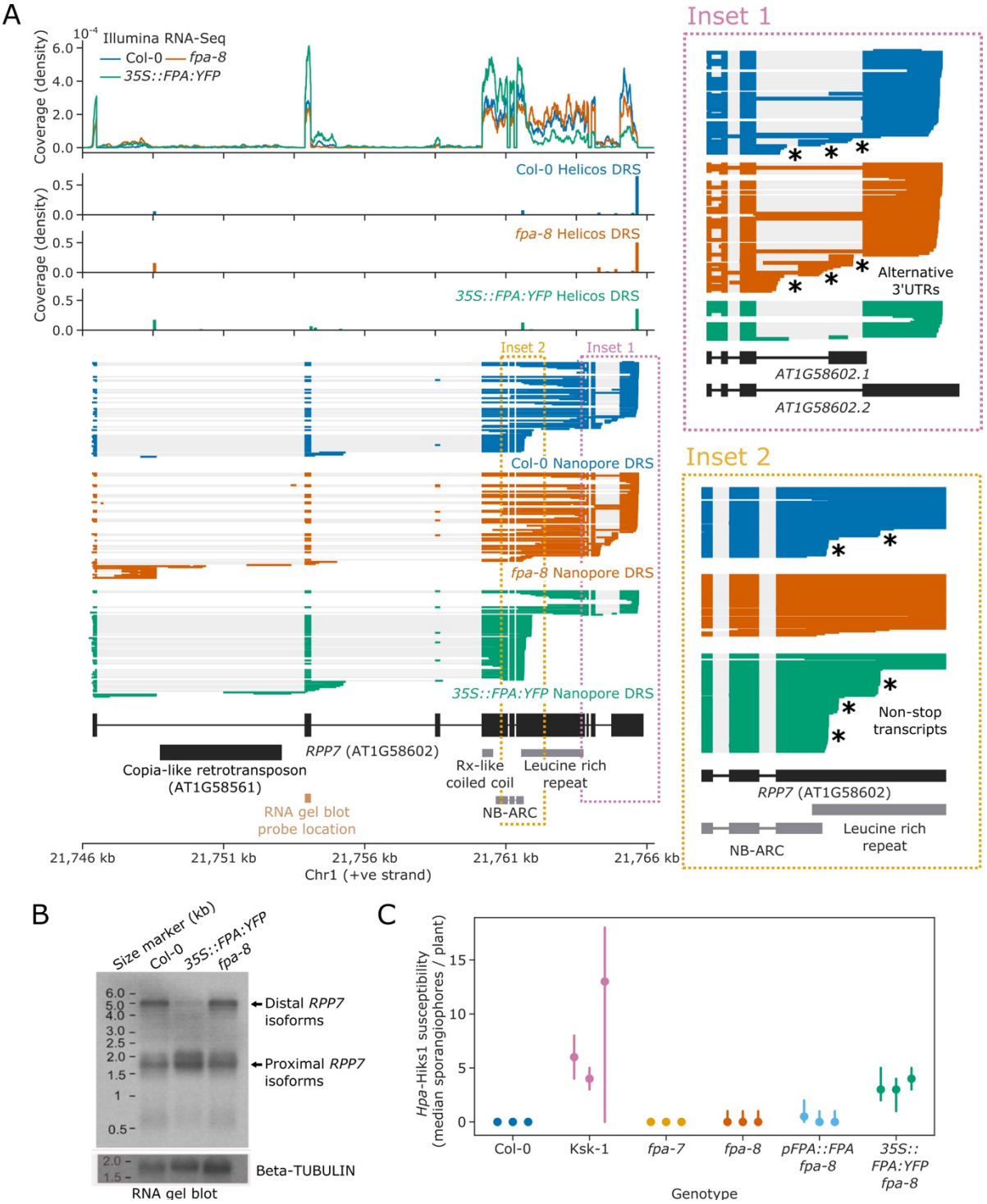


Figure 6: FPA promotes premature cleavage and polyadenylation within *RPP7* protein-coding exon 6 that compromises plant immunity against *Hyaloperonospora arabidopsidis* isolate Hiks1(*Hpa*-Hiks1).

(A) FPA-dependent RNA 3′ end formation changes at the *RPP7 (AT1G58602)* locus. **(Inset 1)** Magnified view of the *RPP7* 3′UTR region with alternative RNA 3′ ends. **(Inset 2)** Magnified view of the stop-codonless transcripts produced in protein-coding *RPP7* exon 6. **(B)** RNA gel blot visualising *RPP7* transcripts in Col-0, *fpa-8* and *35S::FPA:YFP*. Probe location in second exon is shown on (A) (light brown). Beta-TUBULIN was used as an internal control. **(C)** FPA-dependent premature exonic termination of *RPP7* compromises immunity against *Hpa*-Hiks1. Point plot showing median number of sporangiophores per plant calculated 4 days after *Hpa*-Hiks1 inoculation. Error bars are 95% confidence intervals. Each experimental replicate was generated from 7–45 plants per genotype.

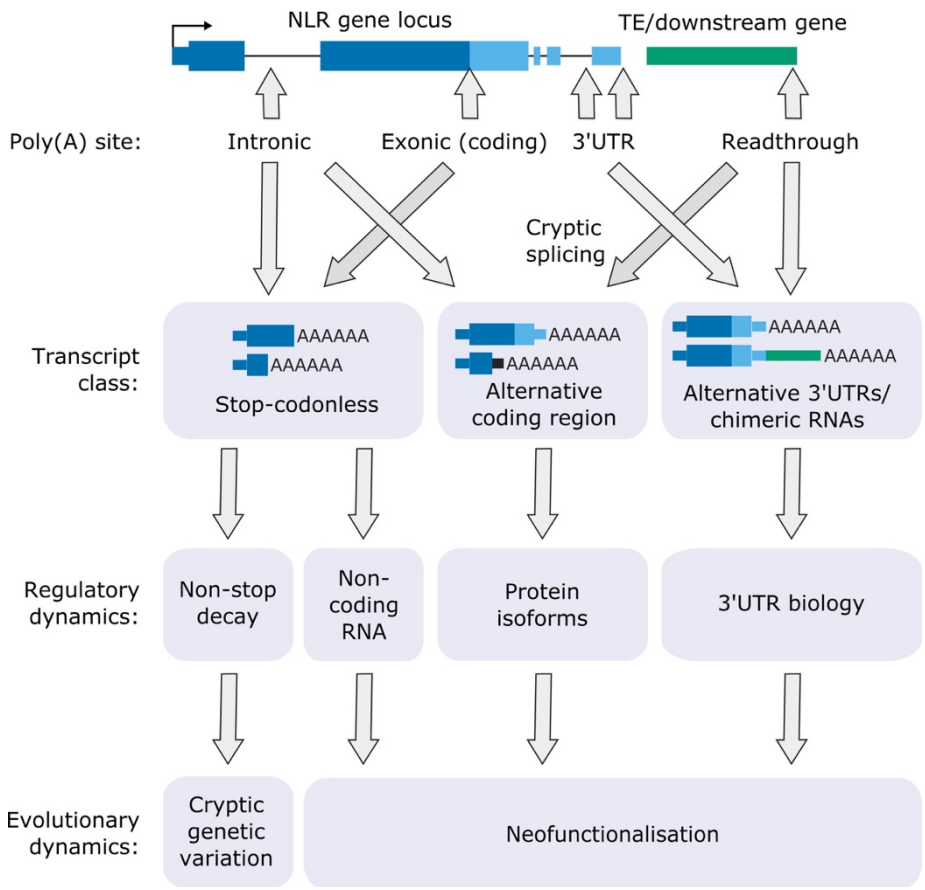


Figure 7: Functional consequences of FPA-dependent alternative polyadenylation at NLR loci. Model diagram showing how FPA-regulated alternative polyadenylation at NLR loci might affect the regulatory and evolutionary dynamics of plant disease resistance.

1421 # Tables

1422

Gene ID	Gene name	NLR class	Chimeric pair (upstream–downstream)
AT1G12220	RPS5	CNL	AT1G12220–AT1G12230
AT1G58848	RPP7a/b	TNL	AT1G58848–AT1G58889
AT1G59218	RPP7a/b	TNL	AT1G59218–AT1G59265
AT1G61190	-	CNL	ncRNA–AT1G61190
AT1G63730	-	TNL	AT1G63730–AT1G63740
AT1G63740	-	TNL	AT1G63730–AT1G63740
AT3G46730	-	CNL	AT3G46740–AT3G46730
AT4G16860	RPP4	TNL	AT4G16860–AT4G16870–AT4G16857
AT4G16960	SIKIC3	TNL	AT4G16970–AT4G16960–AT4G16957
AT4G19060	-	NB only	AT4G19070–AT4G19060
AT4G19530	-	TNL	AT4G19530–AT4G19540
AT5G38850	-	TNL	AT5G38850–AT5G38860
AT5G40090	CHL1	TNL	ncRNA–AT5G40090
AT5G44510	TAO1	TNL	AT5G44520–AT5G44510
AT5G45490	-	CNL	AT5G45472–AT5G45490
AT5G46470	RPS6	TNL	AT5G46470–TIR gene
AT5G48780	-	TNL	AT5G48775–AT5G48780

1423 **Table 1: Readthrough and chimeric RNA formation events at NLR genes regulated by FPA.**

1424

Gene ID	Gene name	NLR class	Predicted function	Protein isoform
AT1G12210	RFL1	CNL	non-coding (5′UTR)	-
AT1G58602	RPP7	CNL	non-coding (5′UTR); alternative 3′UTR	-
AT1G63750	WRR9	TNL	protein coding	TIR only
AT1G63880	RLM1B	TNL	protein coding; non-stop	TIR only
AT1G69550	-	TNL	protein coding	LRR truncation
AT3G44480	RPP1	TNL	protein coding	LRR truncation
AT3G50480	HR4	RPW8	protein coding	RPW8 truncation
AT4G16860	RPP4	TNL	protein coding	TIR only
AT4G16900	-	TNL	protein coding	LRR truncation
AT4G19510	RPP2B	TNL	alternative 3′UTR	-
AT5G17890	DAR4/CHS3	TNL	protein coding	TIR only
AT5G40910	-	TNL	protein coding	TIR only
AT5G43730	RSG2	CNL	non-coding (5′UTR)	-
AT5G43740	-	CNL	non-coding (5′UTR)	-
AT5G46270	-	TNL	protein coding	TIR/NB-ARC only; LRR truncation
AT5G46470	RPS6	TNL	alternative 3′UTR	
AT5G46490	-	TNL	protein coding; non-stop	TIR/NB-ARC only; LRR truncation

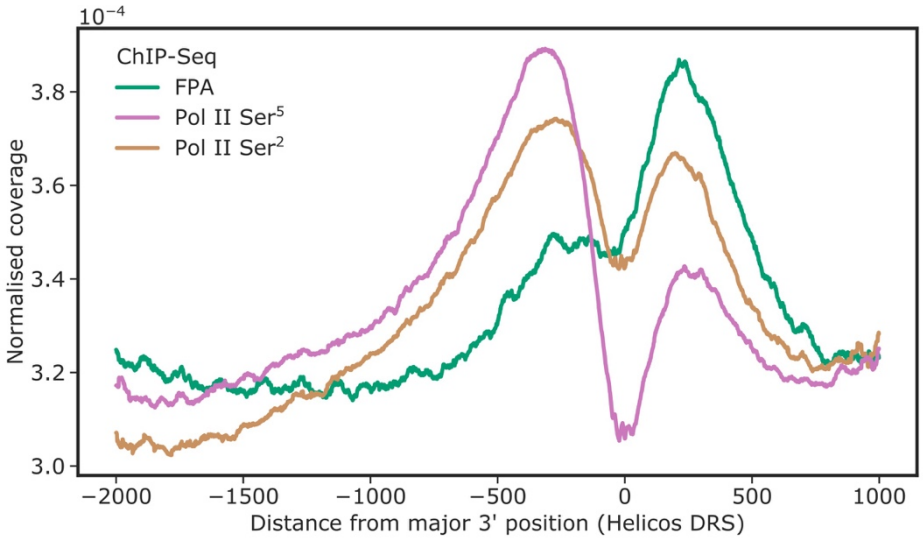
1425 **Table 2: Intronic proximal polyadenylation events at FPA-regulated NLR genes**

1426

Gene ID	Gene name	NLR class	Predicted function	Protein isoform
AT1G10920	LOV1	CNL	protein coding*	CC-only*
AT1G27180	-	TNL	non-stop	-
AT1G31540	RAC1	TNL	non-stop; protein coding^	LRR truncation^
AT1G33560	ADR1	RNL	non-stop	-
AT1G53350	-	CNL	non-stop	-
AT1G56510	WRR4A	TNL	non-stop	-
AT1G56520	-	TNL	non-stop	-
AT1G58602	RPP7	CNL	non-stop	-
AT1G58807	RF45	CNL	non-stop	-
AT1G58848	RPP7a/b	CNL	non-stop	-
AT1G59124	RDL5	CNL	non-stop	-
AT1G59218	RPP7a/b	CNL	non-stop	-
AT1G61300	-	CNL	non-stop	-
AT1G62630	-	CNL	non-stop	-
AT1G63360	-	CNL	non-stop	-
AT1G63730	-	TNL	non-stop	-
AT1G63860	-	TNL	non-stop	-
AT1G63880	RLM1B	TNL	non-stop	-
AT1G72840	-	TNL	non-coding (5′UTR)	-
AT2G14080	RPP28	TNL	non-stop	-
AT3G44480	RPP1	TNL	non-stop; protein coding^	LRR truncation^
AT3G44630	-	TNL	non-stop	-
AT3G44670	-	TNL	non-stop; protein coding^	TIR only^
AT3G46530	RPP13	CNL	non-stop	-
AT4G16860	RPP4	TNL	non-stop	-
AT4G16890	SNC1	TNL	non-stop	-
AT4G16900	-	TNL	non-stop	-
AT4G19520	-	TNL	non-stop	-
AT4G19530	-	TNL	non-stop	-
AT4G36140	-	TNL	non-stop	-
AT5G17890	DAR4/CHS3	TNL	non-stop	-
AT5G35450	-	CNL	non-stop	-
AT5G38850	-	TNL	non-stop	-
AT5G40060	-	TNL	protein coding*	TIR only*
AT5G40910	-	TNL	non-stop	-
AT5G43470	RPP8	CNL	non-stop	-
AT5G43740	-	CNL	non-stop	-
AT5G44510	TAO1	TNL	non-stop; protein coding^	LRR truncation^
AT5G44870	LAZ5	TNL	non-stop	-
AT5G45050	RRS1B	TNL	non-stop	-
AT5G45250	RPS4	TNL	protein coding^	LRR truncation^
AT5G45260	RRS1	TNL	non-stop	-
AT5G46270	-	TNL	non-stop; protein coding^	LRR truncation^
AT5G48620	-	CNL	non-stop	-
AT5G58120	DM10	TNL	non-stop; protein coding^	LRR truncation^

1427 **Table 3: Exonic proximal polyadenylation events at NLR genes regulated by FPA.**
1428 * indicates loci where exonic proximal polyadenylation generates transcripts that may be protein
1429 coding due to upstream ORFs. ^ indicates loci where exonic proximal polyadenylation coupled with
1430 intron retention results in a protein-coding ORF.
1431

1432 # Figure supplement legends



1433
1434 **Figure 1-figure supplement 1: FPA co-localises with Pol II Ser^2 at the 3′ end of genes.**
1435 ChIP-Seq metagene profile showing the normalised occupancy of FPA (green) and Pol II
1436 phosphorylated at Ser^5 and Ser^2 of the CTD relative to the major 3′ position of each gene, as
1437 measured using Helicos DRS. Only short genes (<2.5 kb) are included (*n*=17,440).
1438

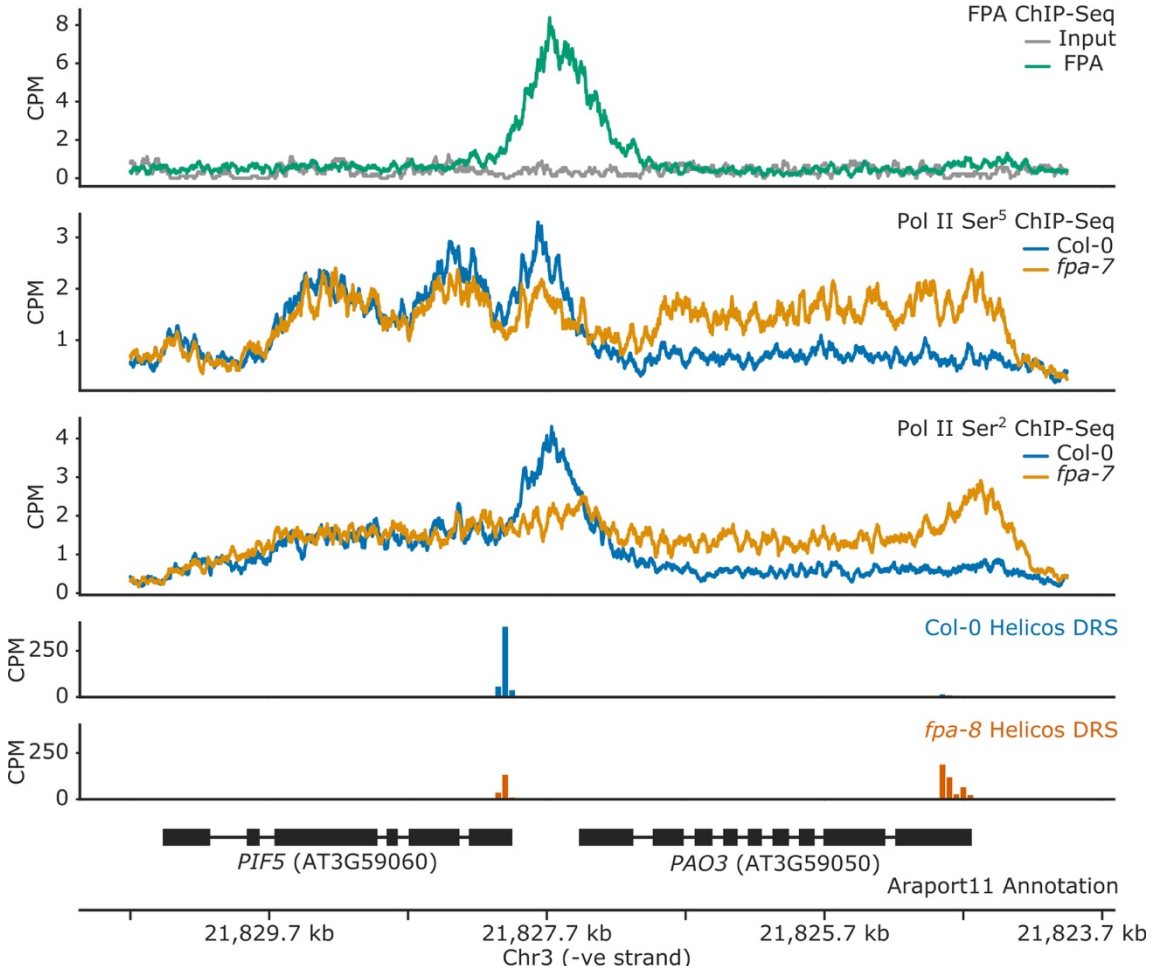


Figure 1-figure supplement 2: FPA controls Pol II occupancy and chimeric RNA formation at *PIF5*.

ChIP-Seq occupancy in counts per million (CPM) of FPA and Pol II phosphorylated at Ser^5 or Ser^2 at the *PIF5* and *PAO3* loci. *fpa* mutants display readthrough of the canonical *PIF5* poly(A) site, with a concomitant loss of Ser^2 at the poly(A) site, and an increase in Ser^5 in downstream *PAO3*.

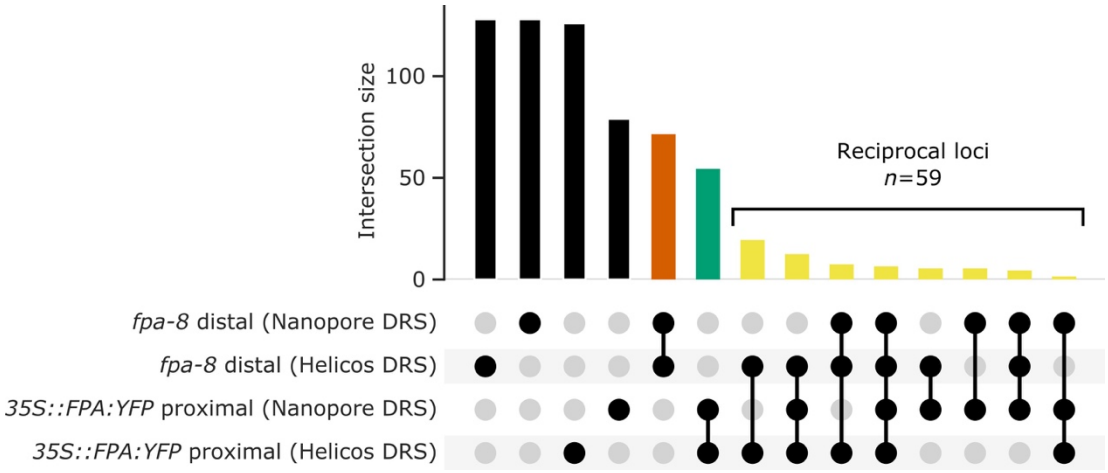


Figure 2-figure supplement 1: Nanopore and Helicos DRS reveal FPA-dependent RNA 3′ end processing changes.

(A) Comparison of RNA 3′ ends identified in Nanopore and Helicos DRS datasets in *fpa-8* and *35S::FPA:YFP* (compared with Col-0). Bar size indicates the number of alternatively polyadenylated loci common to an intersection (highlighted using circles below). Bars indicating loci that are identified as alternatively polyadenylated in a single condition (*fpa-8* or *35S::FPA:YFP*) using a single technique (Nanopore or Helicos DRS) are presented in black; bars indicating loci identified as distally polyadenylated in *fpa-8* using both Nanopore and Helicos DRS, in orange; bars indicating loci identified as proximally polyadenylated in *35S::FPA:YFP* using both Nanopore and Helicos DRS, in green; and bars indicating loci identified as reciprocally regulated by FPA (distal polyadenylation in *fpa-8,* proximal in *35S::FPA:YFP*) using at least one technique, in yellow.

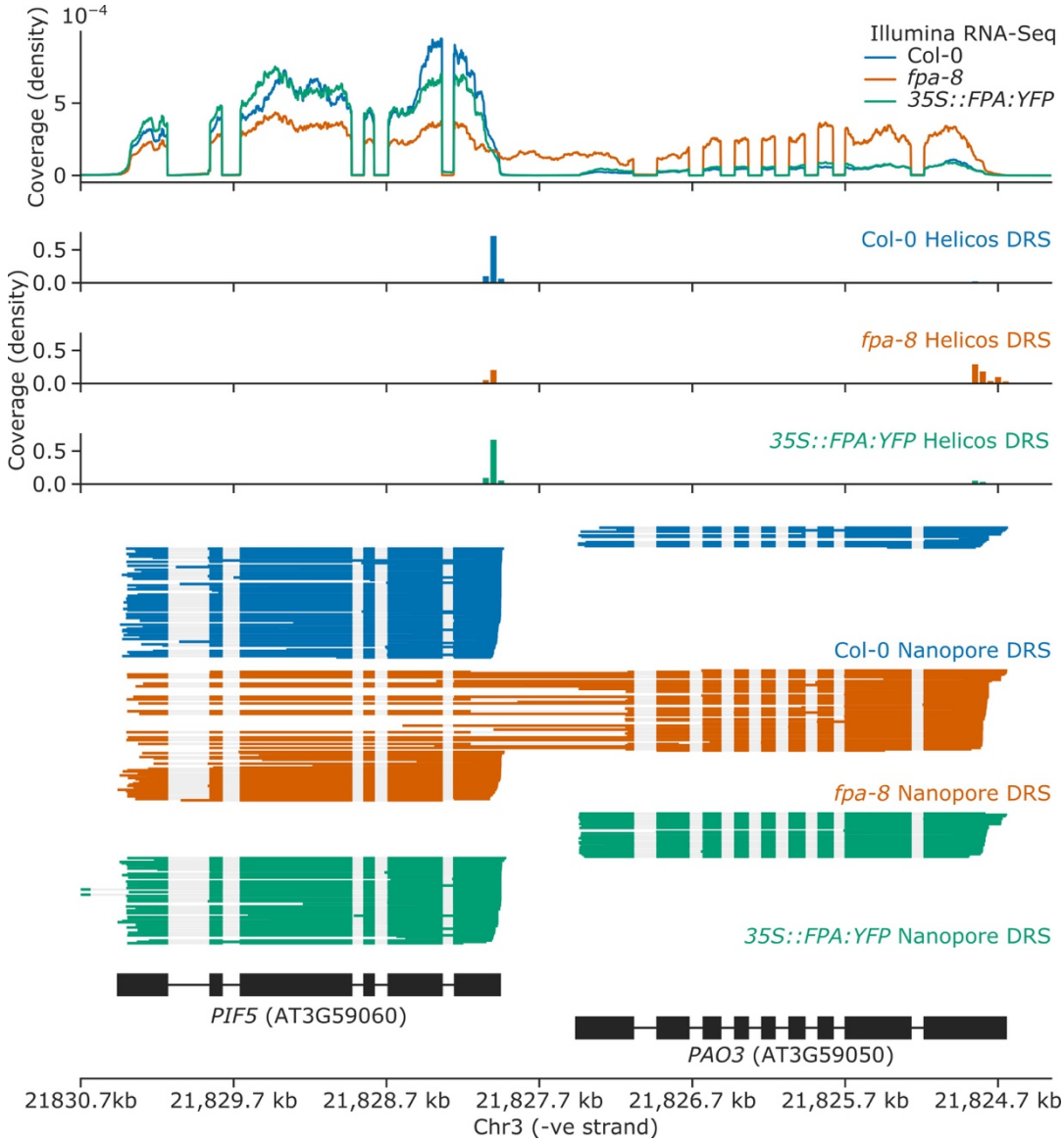


Figure 2-figure supplement 2: Splicing alterations in *fpa-8* can be explained by changes in RNA 3′ end formation.

Gene track showing chimeric RNA formation at the *PIF5* gene locus, as detected with Illumina RNA-Seq, Helicos DRS and Nanopore DRS.

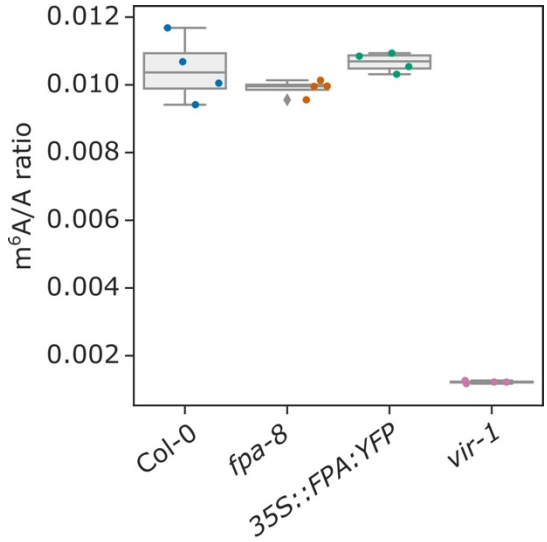


Figure 2-figure supplement 3: FPA does not affect global mRNA m^{6}A methylation.

Box plot showing the m^{6}A/A ratio, as analysed using LC-MS/MS.

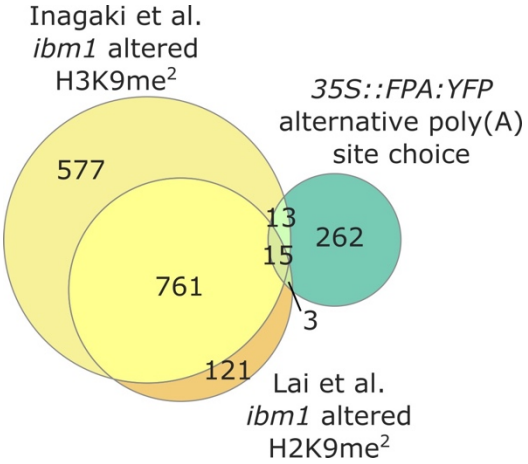


Figure 2-figure supplement 4: FPA-dependent regulation of NLR expression is independent of IBM1.

Venn diagram showing genes with altered H3K9me^2 levels in *ibm1–4* mutants, in yellow (Inagaki et al., 2017) and orange (Lai et al., 2020); and genes with altered poly(A) site choice in *35S::FPA:YFP,* in green.

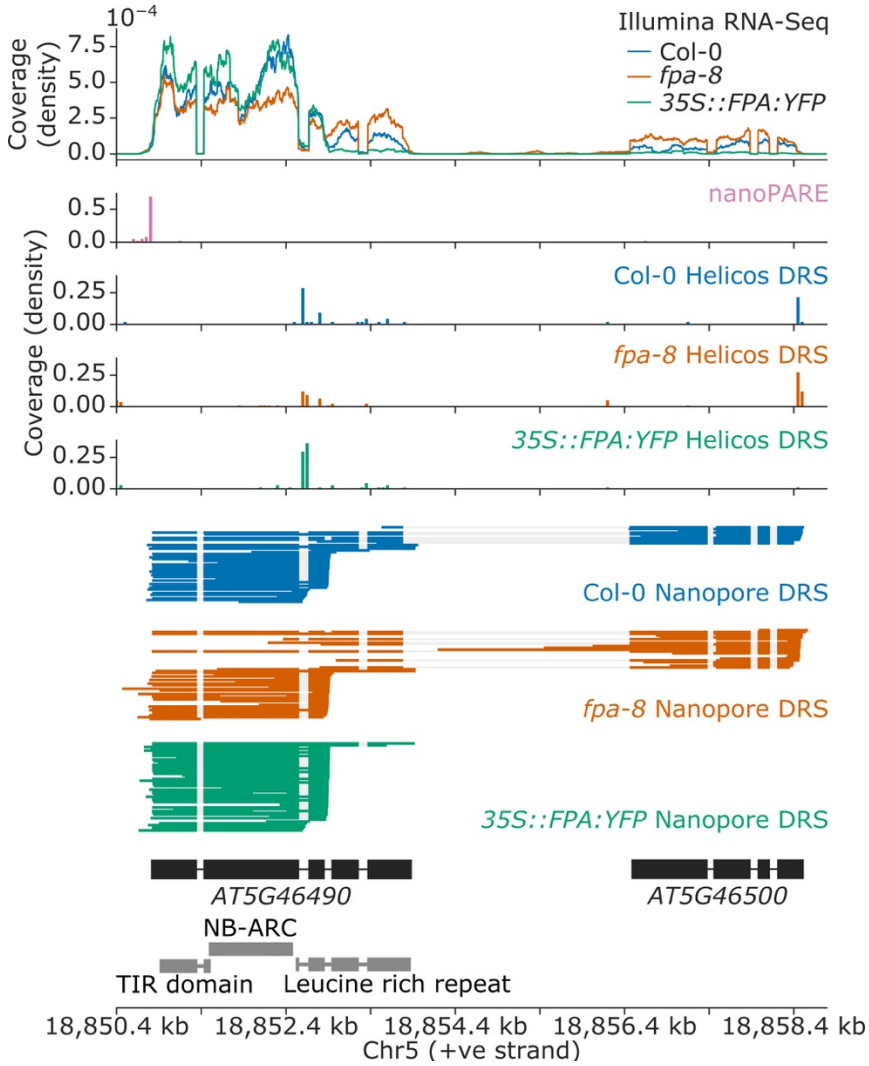


Figure 3-figure supplement 1: Nanopore DRS informs reannotation of the complex NLR locus encompassing the *AT5G46490* and *AT5G46500* annotations.

Gene track showing alternative polyadenylation at the *AT5G46490* gene locus, as detected with Illumina RNA-Seq, nanoPARE, Helicos DRS and Nanopore DRS.

```
AT5G46500   11 CSRLKCVSLHISKLKHLEDALFPACGALNRVELSGSSS   48
AT5G46260  875 CSRLKCVSLHISKLKRLGKVDFKDCGALTIVDLCGCPI  912
AT5G46520  873 CRELKCVSLNIFKLKHLGEVSFSNCGALTRVDLSCYPS  910

AT5G46500   49 GM---KADNIDTAS----SSLP---QVELDFRECFNLD   76
AT5G46260  913 GME-MEANNIDTVS-----------KVKLDFRDCFNLD  938
AT5G46520  911 GVEMMKADNADIVSEETTSSLPDSCVLNVNFMDCVNLD  948

AT5G46500   77 PETVLHQESIIFKYMLFPGKEEVPSYFTYRTTGV----  110
AT5G46260  939 PETVLHQESIIFKYMLFPGKEEMPSYFTYRTTGS----  972
AT5G46520  949 REPVLHQQSIIFNSMILPG-EEVPSYFTYRTSDSQPFG  985

AT5G46500  111 --SSLTIPLLHLPLSQPFFRFRVGALVTNVKHGKNIKV  146
AT5G46260  973 --SSLTIPLLHLPLSQPFFRFRVGALVTNVKHGKNIKV 1008
AT5G46520  986 TSSSLPIPLLPTQLSQPFFRFRVCAVV-SASNGVYIGV 1022

AT5G46500  147 KCEFKDRFGNSFHVGSDDFYVYLLFTKSQKGSQMLTIL  184
AT5G46260 1009 KCEFKDRFGNSFHVGSDDFYVYLLFTKSQKGSQMLTIL 1046
AT5G46520 1023 YSRFKGRIGNKFD--S--FGEVHNFMEIEKG-IHLCIF 1055

AT5G46500  185 DCCIPLNEGNASLAQGNYYDHVDINIHITESLGSFGST  222
AT5G46260 1047 DCCIPLNEGNASLAQGNYYDHVDINIHISS--GGWRST 1082
AT5G46520 1056 DCRIRLYKDNVPLSQLN-YDHVDINIHITS--GDWRST 1090

AT5G46500  223 SELKEWGIRLLEEDSSSAENQLGNPNSTLPHVSEAEEG  260
AT5G46260 1083 FELKEWGIRLLEEDSSSAENGLGNPNSTLPHVSEAEEG 1120
AT5G46520 1091 VVLKEWGIRLL-ETGSSAENRLGNPNSTLPHVSQAEEG 1127

AT5G46500  261 NMGYYTPVQ-GLVNEIEHNGESGDNNVETERSTKHAA   296
AT5G46260 1121 NMGYYTPLQEGLVNEIEHSEESGDINVGTKRSKKRMR  1157
AT5G46520 1128 NMGYYTHVQ-GLVNEIENSEDSGDNNVETERSKKRMR  1163
```

Figure 3-figure supplement 2: Nanopore DRS informs reannotation of the complex NLR locus encompassing the *AT5G46490* and *AT5G46500* annotations.

Protein alignment showing similarity between the *AT5G46500* protein sequence (which forms the C-terminal portion of distally polyadenylation *AT5G46490–AT5G46500* mRNAs) and other NLR protein sequences in the *RPS6* cluster. LRR predictions, generated with LRRpredictor (Martin et al., 2020), are shown in orange.

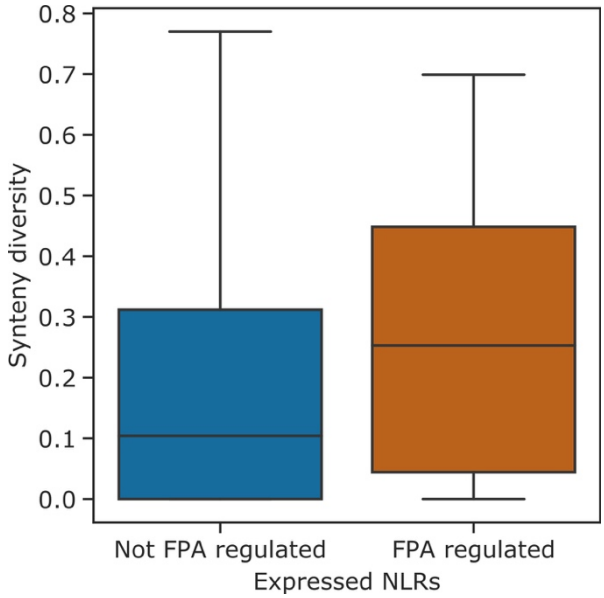


Figure 4–figure supplement 1: NLR genes with FPA-regulated alternative polyadenylation are found in hotspots of rearrangements.

Boxplot showing the synteny diversity, calculated from seven diverse *A. thaliana* accessions (Jiao and Schneeberger, 2020), of expressed NLR genes with and without FPA-regulated alternative polyadenylation.

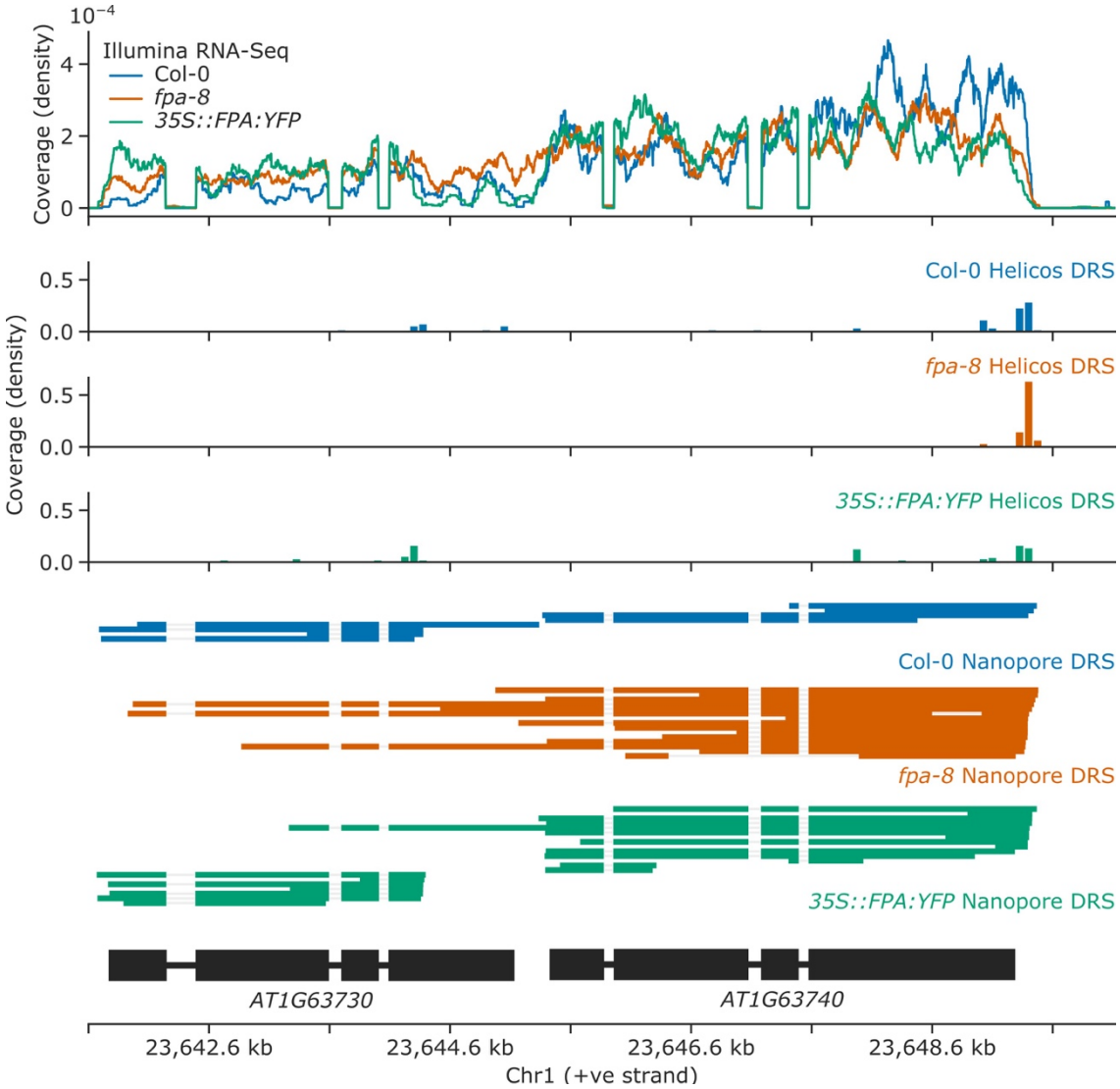


Figure 4-figure supplement 2: Loss of FPA function causes chimeric RNA formation at *AT1G63730* and *AT1G63740* NLR loci.

Gene track showing chimeric RNA formation at the *AT1G63730* gene locus, as detected with Illumina RNA-Seq, Helicos DRS and Nanopore DRS.

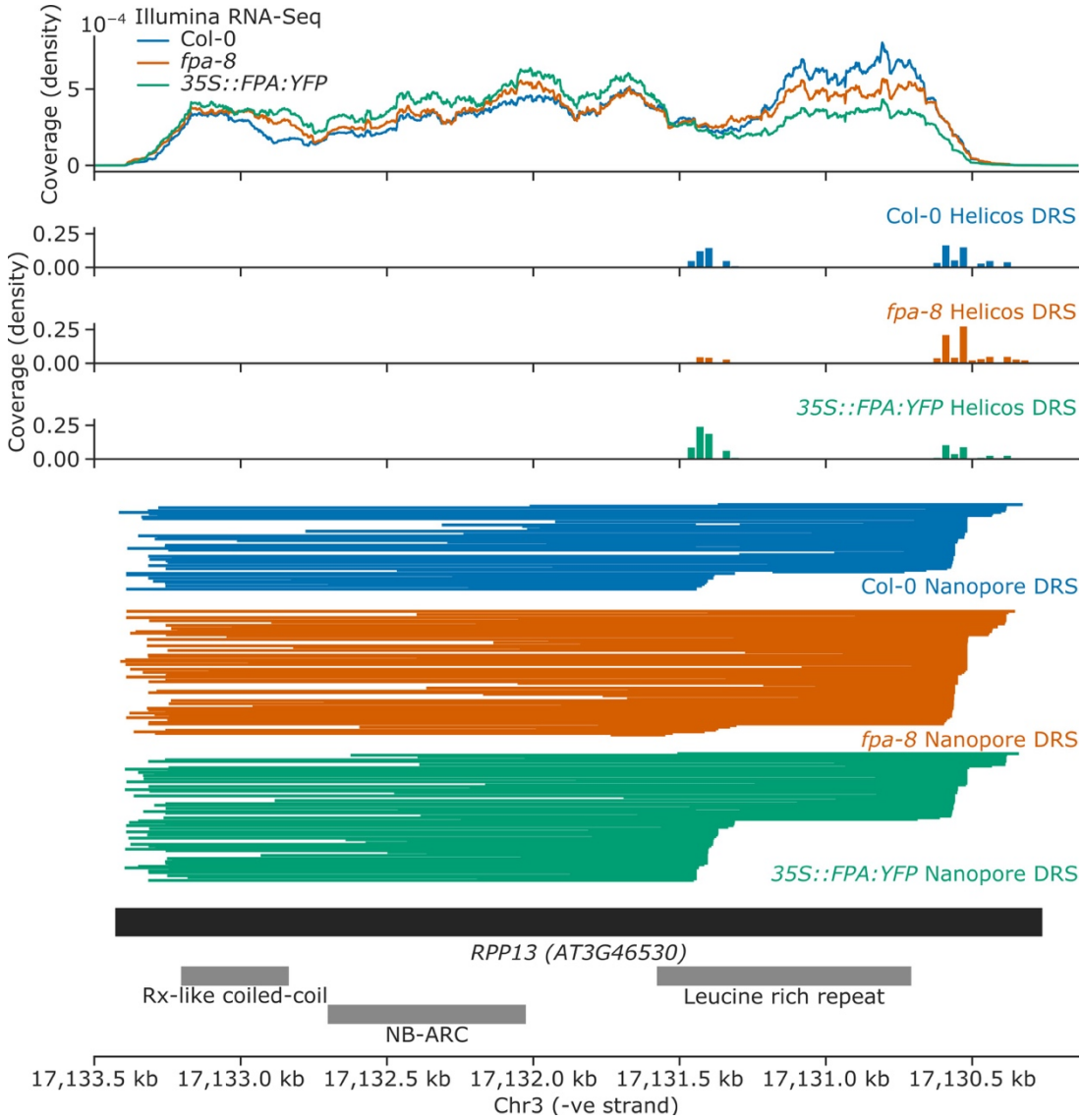


Figure 4-figure supplement 3: FPA overexpression increases exonic proximal polyadenylation of *RPP13*.

Gene track showing proximal polyadenylation at the *RPP13* gene locus, as detected with Illumina RNA-Seq, Helicos DRS and Nanopore DRS.

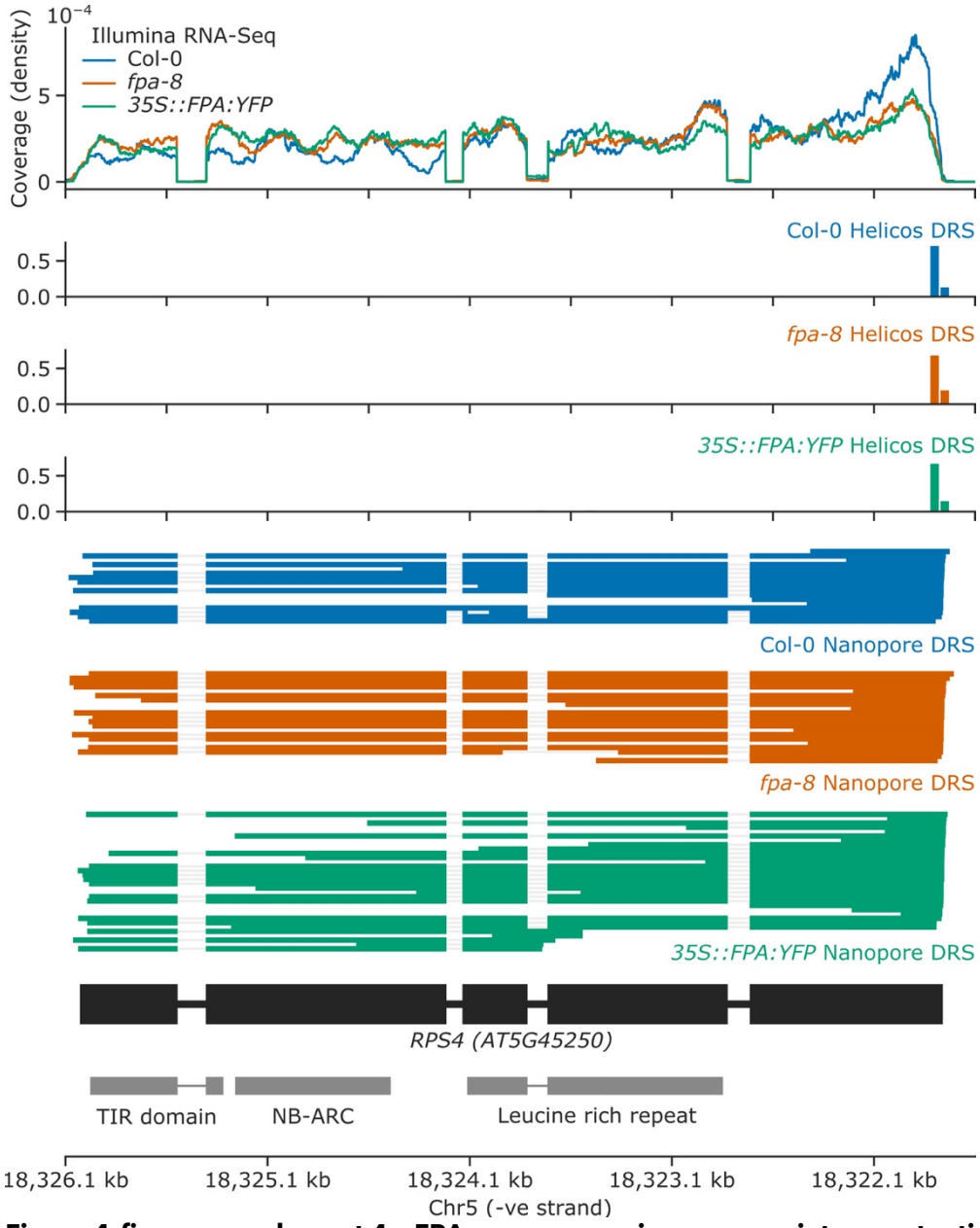


Figure 4-figure supplement 4: FPA overexpression causes intron retention and exonic proximal polyadenylation at intron 3 of *RPS4*.

Gene track showing proximal polyadenylation at the *RPS4* gene locus, as detected with Illumina RNA-Seq, Helicos DRS and Nanopore DRS.

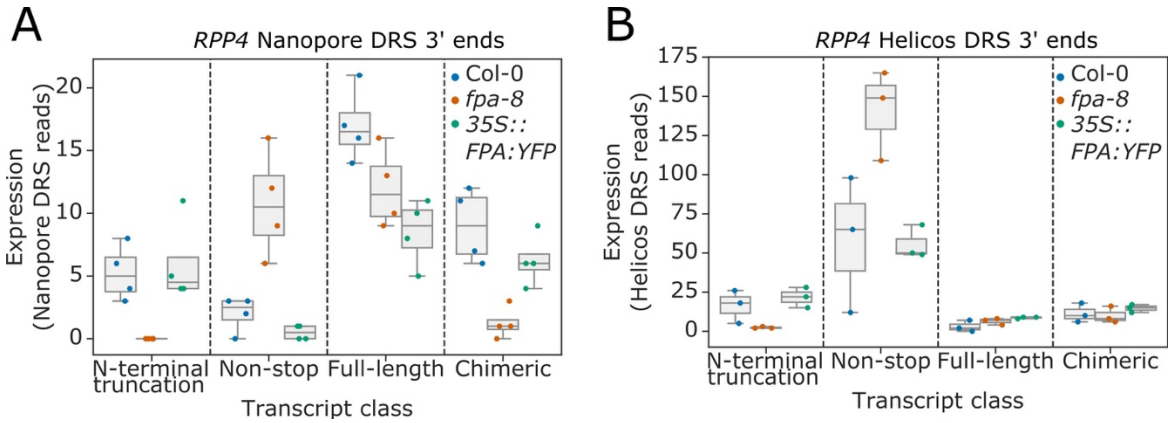


Figure 5-figure supplement 1: Complex FPA-dependent patterns of alternative polyadenylation at the *RPP4* locus.

Comparison of the expression of four classes of *RPP4* (AT4G16860) transcripts detected using **(A)** Nanopore DRS or **(B)** Helicos DRS. *N-terminal truncation*, TIR-domain-only transcripts generated by proximal intronic polyadenylation or distal polyadenylation and cryptic splicing; *Non-stop*, mRNAs lacking in-frame stop codons; *Full-length*, full-length protein-coding mRNAs; and Chimeric, mRNAs containing *RPP4*, COPIA-like retrotransposon (AT4G16870) and/or downstream *AT4G16857*.

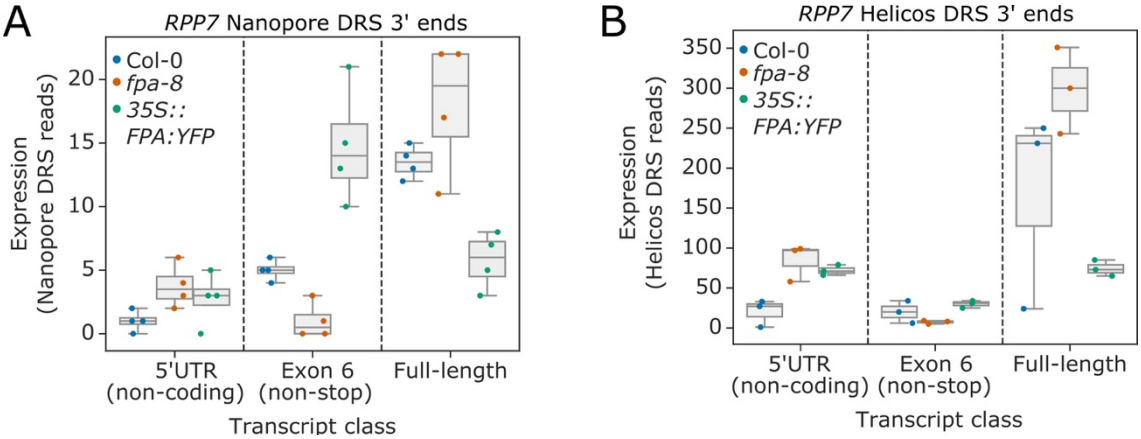


Figure 6-figure supplement 1: Complex FPA-dependent patterns of alternative polyadenylation at the *RPP7* locus.

Comparison of the expression of three classes of *RPP7* transcripts detected using **(A)** nanopore DRS or **(B)** Helicos DRS. *5'UTR (non-coding)*, mRNAs prematurely terminated within the 5'UTR; *exon 6 (non-stop)*, stop-codonless transcripts terminated at proximal poly(A) sites in exon 6; and *full-length*, protein-coding mRNAs terminated at distal poly(A) sites within the 3'UTR.

List of supplementary files:

Supplementary file 1: Proteins co-purifying with FPA, as identified by *IVI-MS* **[Linked to Figure 1].**

Supplementary file 2: Properties of the sequencing datasets produced using Nanopore DRS, Helicos DRS and Illumina RNA-Seq **[Linked to Figure 2]**.

List of source datasets:

Figure 2 source data 1: Nanopore StringTie assembly **[Linked to Figure 2A-B]**.

Figure 2 source data 2: Differential 3′ processing results for *fpa-8* vs Col-0, as identified by Nanopore DRS **[Linked to Figure 2B–C]**.

Figure 2 source data 3: Differential 3′ processing results for *35S::FPA:YFP* vs Col-0, as identified by Nanopore DRS **[Linked to Figure 2B–C]**.

Figure 2 source data 4: Differential 3′ processing results for *fpa-8* vs Col-0, as identified by Helicos DRS **[Linked to Figure 2D–E]**.

Figure 2 source data 5: Differential 3′ processing results for *35S::FPA:YFP* vs Col-0, as identified by Helicos DRS **[Linked to Figure 2D–E]**.

Figure 2 source data 6: Differentially expressed regions results for *fpa-8* vs Col-0, as identified by Illumina RNA-Seq **[Linked to Figure 2F]**.

Figure 2 source data 7: Differentially expressed regions results for *35S::FPA:YFP* vs Col-0, as identified by Illumina RNA-Seq **[Linked to Figure 2F]**.

Figure 2 source data 8: Differential splice junction usage results for *fpa-8* vs Col-0, as identified by Illumina RNA-Seq **[Linked to Figure 2G]**.

Figure 2 source data 9: Differential splice junction usage results for *35S::FPA:YFP* vs Col-0, as identified by Illumina RNA-Seq **[Linked to Figure 2G]**.

Figure 2 source data 10: m^6A/A ratios for Col-0, *fpa-8*, *35S::FPA:YFP* and *vir-1*, as detected by LC-MS/MS **[Linked to Figure 2-figure supplement 3]**.

Figure 2 source data 11: Differential $H3K9me^2$ results for *ibm1–4* vs Col-0 **[Linked to Figure 2-figure supplement 4]**.

Figure 6 source data 1: *Hpa*-Hiks1 susceptibility results for the Col-0, Ksk-1, *fpa-7, fpa-8, pFPA::FPA* and *35S::FPA:YFP* lines **[Linked to Figure 6C]**.