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, controlling their functional expression and impacting immunity. Using long-read nanopore direct RNA sequencing we resolve 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 & Weigel, 2020; Jones *et al*, 2016; Tamborski & Krasileva, 2020; Wu *et al*, 2017). Plant genomes can encode large numbers (>1,000) of NLRs, which often occur in physical clusters (Wei *et al*, 2016). These features, coupled with the rapid birth and death of NLR genes and resulting intraspecific diversity, have made them relatively difficult to study. Indeed, it was not until recently that the near complete repertoire of Arabidopsis NLR genes was mapped (Van de Weyer *et al*, 2019). This advance depended on long read DNA sequencing that was able to read through complex NLR clusters in a diverse set of Arabidopsis accessions.

The benefit of NLR function to hosts is disease resistance. However, there are costs to increased NLR gene number or activity that can lead to detrimental autoimmunity (or hyperimmunity) (Rodriguez *et al*, 2016), restriction of association with beneficial microbes (Yang *et al*, 2010), and a general reduction in fitness (Tian *et al*, 2003). In some cases, epistatic interactions involving NLRs can result in autoimmunity that is detected as hybrid weakness or necrosis (Li *et al*, 2020). A key question, therefore, is how NLRs are regulated to enable limited expression for pathogen surveillance but enhanced expression during defence responses. This problem is compounded by the evolutionary dynamics of NLR genes because regulatory processes must not only enable, but keep pace with, the emergence of new NLR genes and the gain or loss of function of others. Consequently, the regulation of NLRs poses one of the most important and difficult challenges faced by plants.

NLR control measures occur at different stages of gene expression and at the protein level (Lai & Eulgem, 2018). The miR482/2118 superfamily of miRNAs limits the expression of many NLRs in legumes and Solanaceae by targeting conserved regions encoded in NLR mRNAs and triggering a cascade of phased siRNAs that suppress NLR

68 expression quite broadly (Shivaprasad *et al*, 2012; Zhai *et al*, 2011). The first example of
69 alternative splicing-dependent control having a consequence for NLR function came from
70 the study of the *N* gene, which promotes resistance to tobacco mosaic virus (Dinesh-Kumar
71 & Baker, 2000). Likewise, the function of RECOGNITION OF PSEUDOMONAS SYRINGAE 4
72 (RPS4, *AT5G45250*) in resistance to *Pseudomonas syringae* DC3000 depends on alternative
73 splicing of *RPS4* pre-mRNA and the co-ordinated expression of more than one *RPS4* mRNA
74 isoform (Zhang & Gassmann, 2007). The best characterised example of alternative
75 polyadenylation controlling NLR gene expression is at the Arabidopsis *RECOGNITION OF*
76 *PERONOSPORA PARASITICA 7* locus (*RPP7, AT1G58602*). Intragenic heterochromatin
77 associated with a retrotransposon located within the relatively long intron 1 of *RPP7*
78 influences the transcription of full-length mRNAs, with a functional consequence for
79 immunity (Tsuchiya & Eulgem, 2013). The analysis of Arabidopsis mutants that disrupt RNA
80 surveillance has uncovered other features of NLR control. For example, null mutants
81 defective in non-sense mediated RNA decay (NMD) are lethal in Arabidopsis because they
82 trigger autoimmunity dependent on the NLR gene *RPS6 (AT5G46470*) (Gloggnitzer *et al*,
83 2014). Most recently, mutations in the RNA exosome, which degrades RNAs in a 3′ to 5′
84 direction, were reported to suppress autoimmunity by disrupting the function of *RPS6*
85 (Takagi *et al*, 2020). The signalling and activity of NLR proteins is controlled by intra and
86 intermolecular protein interactions (Adachi *et al*, 2019). Specialization and diversification of
87 NLR function has led to the evolution of NLRs that either sense pathogen effectors directly
88 or help to signal effectors detected by other NLRs. Some NLRs control the activity of other
89 NLRs by direct protein interaction. As a result, networks of NLRs can control immune
90 responses. Despite this progress, our overall understanding of how NLRs evolve and are
91 controlled is limited and fragmented (Adachi *et al.*, 2019).
92 The RNA-binding protein FPA was first identified as a factor required for the control
93 of Arabidopsis flowering time (Koornneef *et al*, 1991). Loss of function *fpa* mutants flower
94 late due to elevated levels of the floral repressor, FLC (Schomburg *et al*, 2001). However,
95 this is not the exclusive function of FPA because it is much more widely conserved than *FLC*.
96 FPA is a member of the spen family of proteins, defined by three N-terminal RNA
97 recognition motifs and a C-terminal protein interaction SPOC domain (Ariyoshi & Schwabe,
98 2003). We previously showed that FPA could control the site of cleavage and
99 polyadenylation in some mRNAs, including autoregulation of expression at *FPA* pre-mRNA
100 (Duc *et al*, 2013; Hornyik *et al*, 2010). These findings have been extended to show that FPA
101 can affect poly(A) site choice at genes with intronic heterochromatin, including *RPP7*
102 (Deremetz *et al*, 2019). The mechanism by which FPA affects poly(A) site choice remains
103 unclear. FPA might mediate these effects directly by recruiting the RNA 3′ end processing
104 machinery to sensitive sites, or indirectly, by influencing splicing, chromatin modifications,
105 or the rate of transcription by RNA Polymerase II (Pol II), for example. We previously used
106 Helicos direct RNA sequencing (DRS) to map the 3′ ends of Arabidopsis polyadenylated
107 transcripts and to identify those genes affected by loss of FPA function transcriptome-wide

(Duc *et al.*, 2013; Sherstnev *et al*, 2012). A limitation of this approach was that it could only identify RNA 3′ end positions, and so we could not resolve a possible wider role for FPA in gene regulation.

In this study, we set out to gain a clearer understanding of how FPA functions using two approaches. First, we studied the proteins that FPA associates with inside living plant cells. Second, we analysed the global impact of different levels of FPA activity on gene expression. We combined Helicos DRS with short-read Illumina RNA-Seq and Oxford Nanopore Technology (nanopore) DRS, which can reveal the authentic processing and modification of full-length mRNAs (Parker *et al*, 2020b). Using this combination of data, and new computational approaches to study RNA processing, we provide evidence that the predominant role of FPA is to promote poly(A) site choice. In addition, we uncovered an unusual degree of complexity in the processing of NLR mRNAs that are sensitive to FPA. As a result, we reveal that premature transcription termination is a widespread layer of NLR gene regulation with implications for the dynamics of NLR evolution.

Results

FPA co-purifies with the mRNA 3′ end processing machinery

In order to understand how FPA controls the site of mRNA 3′ end formation, we first asked which proteins FPA associates with inside living plant cells. To address this question, we used *in vivo* interaction proteomics-mass spectrometry (*IVI-MS*). We fixed molecular interactions using formaldehyde (FA) infiltration of Arabidopsis seedlings expressing FPA fused to YFP (*35S::FPA:YFP*) and wild type (Col-0) plants as a control. We then purified nuclei and performed GFP-trap immunoprecipitation followed by liquid chromatography-tandem MS (LC-MS/MS) to identify FPA-associated proteins.

By comparing the proteins detected in three biological replicates of *35S::FPA:YFP* and Col-0, we identified 203 proteins enriched with FPA at a median log fold change greater than two (Figure 1, Supplementary file 1). 56% (113) of these are established RNA-binding proteins identified using RNA-binding proteome procedures (Bach-Pages *et al*, 2020; Reichel *et al*, 2016). At least three functionally relevant associations emerge from this analysis. First, consistent with the role of FPA in controlling mRNA 3′ end formation, most of the highly conserved cleavage and polyadenylation factors (CPF) (Figure 1A, Supplementary file 1) and proteins functioning in transcription termination, co-purify with FPA (Figure 1B, Supplementary file 1). Second, in agreement with the role of FPA in flowering, many of the FPA-associated proteins function in the same autonomous pathway of flowering time control (Figure 1C, Supplementary file 1). Third, components of the m^6A writer complex, which deposits m^6A in the 3′UTRs of protein-coding mRNAs, also co-purify with FPA (Figure 1D, Supplementary file 1).

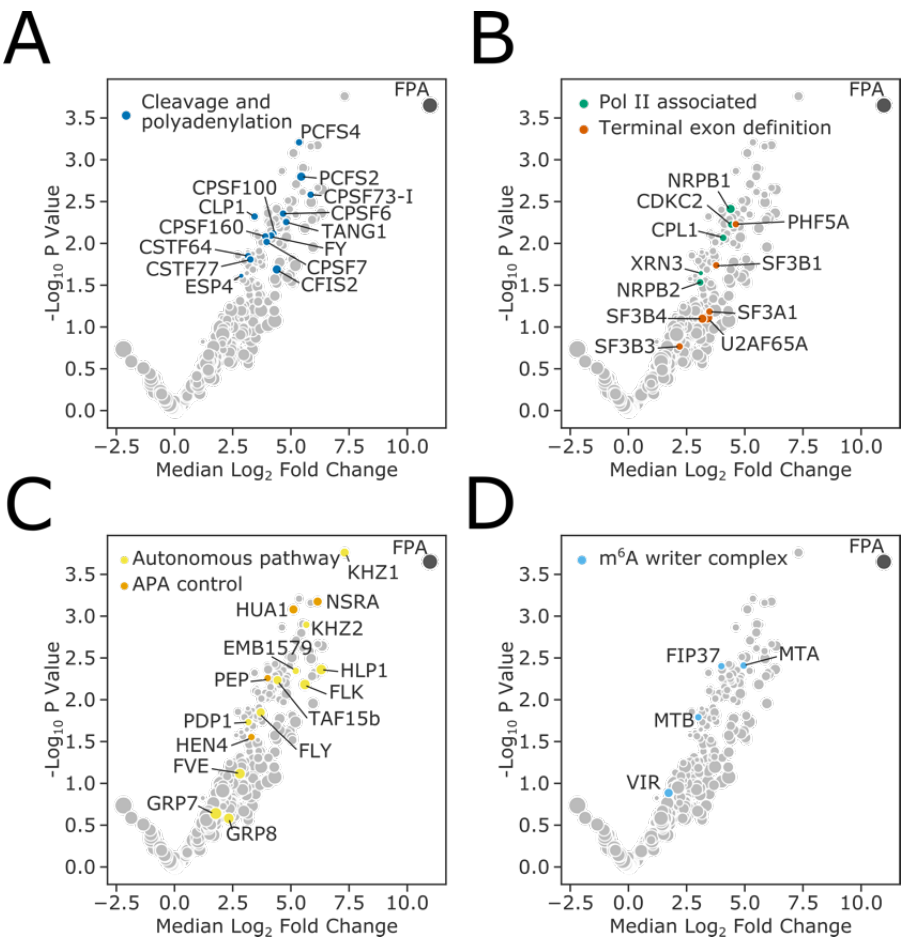


Figure 1: **FPA associates with proteins that function to process the 3' ends of Pol II transcribed RNAs and promote transcription termination**

(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)** cleavage and polyadenylation factors, 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 (APA), in light orange; and **(D)** m^6A writer complex components, in light blue.

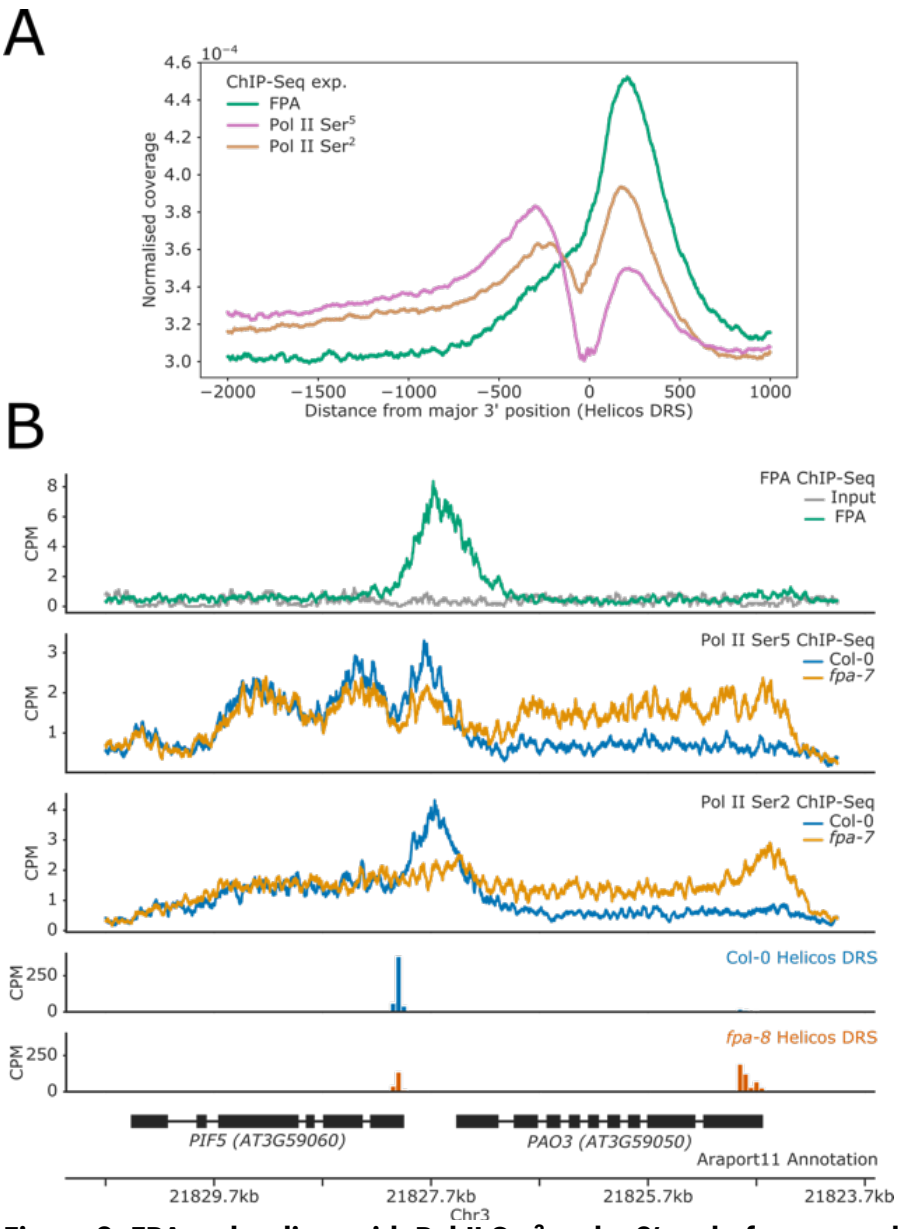
We detected almost all known components of the conserved CPF machinery (Tian & Manley, 2017) co-purifying with FPA (Figure 1A, Supplementary file 1). For example, we identified homologues of the human cleavage and polyadenylation specificity factor (CPSF) complex proteins CPSF160, CPSF100, CPSF73-I and FY (WDR33); the cleavage stimulating factor (CstF) complex proteins CSTF77, CSTF64; and paralogs of symplekin-like proteins TANG1 and ESP4. We detected the cleavage factor I (CFIm) components CFIS2, CPSF6 and CPSF7; the CFIIm components CLP1 and CLPS3; and two paralogs of PCF11 (PCFS2 and PCFS4). The only established core CPF proteins that we did not detect were CPSF30, FIP1, CSTF50 and poly (A) polymerases.

Consistent with the coupling between splicing and 3' end formation that mediates definition of the terminal exon of transcripts, we found several components of the U2 snRNP complex that are involved in this process (Figure 1B, Supplementary file 1). For example,

171 we detected SF3a (SF3A1) and SF3b (SF3B1, B3, B4 and PHF5A) proteins, together with
172 U2AF65A, which make specific interactions with CPSF during terminal exon definition
173 (Kyburz *et al*, 2006). Cleavage and polyadenylation is also intimately linked to transcription
174 termination. Consistent with this, we identified the Pol II subunits NRPB1 and NRPB2, as well
175 as proteins that modify the phosphorylation status of the heptad repeat of the Pol II C-
176 terminal domain (CTD) (Figure 1B, Supplementary file 1). Characteristically, Serine^5 of the
177 CTD is phosphorylated when Pol II is at the 5′ end of genes, and Ser^2 is phosphorylated
178 when Pol II is at 3′ ends (Komarnitsky *et al*, 2000). We found the kinase CDKC;2, which
179 phosphorylates Ser^2 (Wang *et al*, 2014) and the phosphatase CPL1 (homolog of yeast
180 Fcp1), which dephosphorylates Ser^5 (Koiwa *et al*, 2004), co-purified with FPA (Figure 1B,
181 Supplementary file 1). Finally, we detected the homolog of human XRN2 5′ to 3′
182 exonuclease (known as XRN3 in Arabidopsis), which functions in transcription termination
183 (Figure 1B, Supplementary file 1) (Krzyszton *et al*, 2018). These data suggest that we have
184 captured FPA at the 3′ end of actively transcribed genes in close proximity to factors that
185 mediate multiple features of RNA 3′ end processing and Pol II transcription termination.
186 FPA is a component of the genetically defined autonomous pathway of flowering
187 time control (Andres & Coupland, 2012; Simpson, 2004). Consistent with this function, a
188 second major class of proteins co-purifying with FPA are components of this flowering
189 pathway (Figure 1C, Supplementary file 1). Autonomous pathway components limit the
190 expression of the floral repressor *FLC*. Regulation of *FLC* expression is complex, and not
191 fully understood, but it is associated with long non-coding antisense RNAs transcribed at
192 this locus that are alternatively polyadenylated (Hornyik *et al.*, 2010; Liu *et al*, 2007).
193 Components of the autonomous pathway, identified through mutant screens, include
194 conserved CPF proteins such as FY (WDR33) (Simpson *et al*, 2003), CSTF64 and CSTF77
195 (Liu *et al*, 2010) that co-purified here with FPA. In addition, the PCF11 homolog PCFS4,
196 which also co-purifies with FPA, was identified as a component of the autonomous pathway
197 by reverse genetics (Xing *et al*, 2008). Consequently, there is an overlap in classification of
198 these proteins because of their biochemical role as CPF proteins and the biological impact
199 that they have on flowering time control at *FLC*. We found that other autonomous pathway
200 RNA binding proteins, which are not core CPF proteins but do affect RNA 3′ end formation,
201 also co-purify with FPA (Figure 1C, Supplementary file 1). These include HLP1, a
202 heterogenous ribonucleoprotein A (hnRNP A)-related protein (Zhang *et al*, 2015).
203 Additional RNA binding proteins of the autonomous pathway, such as FLK (Mockler *et al*,
204 2004), KHZ1/KHZ2 (Yan *et al*, 2019), FLY (Dai *et al*, 2020) and GRP7/8 (Streitner *et al*, 2008)
205 were also detected. Not all members of the autonomous pathway are RNA binding
206 proteins, and we detected FVE co-purifying with FPA. FVE homologs function in diverse
207 chromatin remodelling complexes in other organisms (Ausin *et al*, 2004). Notably, we also
208 detected PDP1 and EMB1579/RSA1, which have recently been shown to physically interact
209 with FVE (Zhang *et al*, 2020; Zhou *et al*, 2018). EMB1579/RSA1 controls nuclear condensate
210 formation and global patterns of transcription and splicing, and *emb1579/rsa1* mutants

211 flower late because of elevated levels of *FLC* (Zhang *et al.*, 2020). In addition, we detected
212 TAF15b which represses *FLC* transcription (Eom *et al*, 2018). The autonomous pathway
213 components FLD, LD and SDG26, which were recently found to co-purify together (Fang *et*
214 *al*, 2020), were not detected here. We identified NSRA, which affects alternative
215 polyadenylation of antisense RNAs at *FLC*, but functions antagonistically to the autonomous
216 pathway by promoting *FLC* expression (Bazin *et al*, 2018). FLK has been found to associate
217 with PEP, HUA1 and HEN4 (Rodriguez-Cazorla *et al*, 2015), and here, we found all four of
218 these proteins co-purifying with FPA (Figure 1C, Supplementary file 1). In addition to
219 regulating *FLC* expression, these four proteins have been shown to interact to control
220 alternative polyadenylation within pre-mRNA encoding the floral homeotic transcription
221 factor AGAMOUS (Rodriguez-Cazorla *et al.*, 2015). Their co-purification here suggests that
222 this role may be more global and involve direct interactions at RNA 3′ ends.
223 The third major group of proteins co-purifying with FPA are conserved members of
224 an mRNA m^6A writer complex; MTA, MTB, VIR and FIP37 (Ruzicka *et al*, 2017) (Figure 1D,
225 Supplementary file 1). The m^6A modification dependent on this writer complex is
226 predominately targeted to the 3′ UTR of Arabidopsis protein coding mRNAs (Parker *et al.*,
227 2020b). The co-purification of FPA with the m^6A writer complex components could
228 therefore be explained by a direct role for FPA in m^6A modification, or simply because both
229 CPF and m^6A writer proteins are found at RNA 3′ ends.
230 The overall picture that emerges from this analysis is that FPA associates in close
231 proximity to diverse proteins that function to promote cleavage, polyadenylation,
232 transcription termination and RNA modification at the 3′ end of Pol II-transcribed genes.
233
234 **FPA co-localises with Pol II phosphorylated at Ser^2 of the CTD heptad repeat at the 3′**
235 **end of Arabidopsis genes**
236 We next asked whether FPA was recruited specifically to the 3′ ends of genes or was
237 associated with Pol II throughout transcription. To distinguish between these possibilities,
238 we used antibodies against FPA and Pol II phosphorylated at either Ser^5 or Ser^2 of the CTD
239 heptad repeat to perform ChIP-Seq analysis (Yu *et al*, 2019). Our metagene analysis reveals
240 that FPA is enriched at the 3′ end of genes and co-localises with Pol II phosphorylated at
241 Ser^2 of the CTD (Figure 2A, Figure 2 – supplement 1). These findings are consistent with the
242 idea that FPA is specifically recruited to the 3′ ends of genes and is not associated with Pol
243 II during elongation.
244 Next, we investigated the relationship between FPA and the different
245 phosphorylated forms of Pol II at individual genes. We have previously identified specific
246 loci at which loss of FPA function causes defective 3′ end formation and readthrough into
247 downstream genes resulting in the formation of chimeric RNAs (Duc *et al.*, 2013). For
248 example, at *PIF5 (AT3G59060)*, chimeric RNAs are formed with the downstream gene *PAO3*
249 *(AT3G59050)* in loss-of-function *fpa-7 and fpa-8* mutant backgrounds (Duc *et al.*, 2013).
250 Here we found that, in wild type Col-0 plants, FPA and Ser^2 phosphorylated Pol II co-localise

251 at the 3′ end of *PIF5,* with low levels of Ser$^{5/2}$ phosphorylated Pol II detected at *PAO3* (Figure
252 2B). In *fpa-7* mutants, however, Pol II phosphorylated at Ser^{2} of the CTD is lost from the 3′
253 end of *PIF5,* with increased levels of Ser^{5} and Ser^{2} phosphorylated Pol II detected at *PAO3*
254 downstream. These data are consistent with FPA influencing 3′ end formation and the
255 termination of transcription by Pol II.
256



257
258 **Figure 2: FPA co-localises with Pol II Ser^{2} at the 3′ end of genes and promotes transcription**
259 **termination**
260 **(A)** ChIP-Seq metagene profile showing the normalised occupancy of FPA (green) and Pol II
261 phosphorylated at Ser^{5} (pink) and Ser^{2} (brown) of the CTD (Yu *et al.*, 2019) relative to the major 3′
262 position of each gene as measured using Helicos DRS. Only long genes (> 2.5 kb) are included
263 (n=10,215). **(B)** ChIP-Seq occupancy in counts per million (CPM) of FPA (input shown in gray; FPA,
264 in green) and Pol II phosphorylated at Ser^{5} or Ser^{2} (Col-0 is shown in blue; *fpa-7*, in orange) at the
265 *PIF5* (*AT3G59060*) and *PAO3* (*AT3G59050*) loci (top panels); Helicos DRS (middle panels,
266 normalised coverage density) and nanopore DRS (bottom panels, read alignments) - Col-0 is shown
267 in blue; *fpa-8,* in orange; and Araport11 annotation, in black.

268 **FPA affects RNA processing by promoting poly(A) site choice**

269 We next asked which RNA processing events were controlled by FPA. We used a
270 combination of Illumina RNA-Seq, Helicos and nanopore DRS technologies, to analyse
271 three different genetic backgrounds expressing different levels of FPA: wild-type Col-0, loss
272 of function *fpa-8* and a line overexpressing FPA fused to YFP (*35S::FPA:YFP*).

273 Illumina RNA-Seq short reads map to whole transcripts, but changes in RNA 3′ end
274 processing can only be detected by changes in coverage (Xia *et al*, 2014). The relatively
275 high depth of Illumina RNA-Seq (approx. 20-100 million reads per replicate) gives statistical
276 power to detect such changes. In Helicos DRS, short read tags represent the true 3′ ends of
277 mRNAs (at 2 nt resolution), and so can be used to identify cleavage and polyadenylation
278 sites, but not the full properties of the corresponding transcripts (Ozsolak *et al*, 2009). In
279 nanopore DRS, long reads represent the true 3′ ends of mRNAs in the context of reads that
280 can correspond to full-length transcripts (Parker *et al.*, 2020b). The Helicos (approx. 3-5
281 million reads per replicate) and nanopore (approx. 0.5-2 million reads per replicate) DRS
282 datasets have less coverage than Illumina RNA-Seq. We performed three biological
283 replicates of each genotype with Helicos DRS, six biological replicates with Illumina RNA-
284 Seq and four biological replicates with nanopore DRS. The corresponding read statistics
285 for the sequencing data are detailed in Supplementary file 2.

286 We first assessed the utility of the three sequencing technologies to map changes
287 in RNA 3′ end processing by focusing on the well characterised *FPA* locus. FPA
288 autoregulates its expression by promoting premature cleavage and polyadenylation within
289 intron 1 of *FPA* pre-mRNA (Duc *et al.*, 2013; Hornyik *et al.*, 2010). Consistent with this, a
290 proximal poly(A) site in the first intron of *FPA* and distal sites in the terminal intron and exon
291 can be mapped precisely in Col-0 using nanopore and Helicos DRS (Figure 3). Using all
292 three data types we detected a quantitative shift towards selection of distal poly(A) sites in
293 the loss-of-function *fpa-8* mutant and a strong shift to proximal poly(A) site selection when
294 FPA is overexpressed (*35S::FPA:YFP*). Nanopore DRS provides the clearest picture of
295 alternative polyadenylation events because the context in full-length reads is revealed.

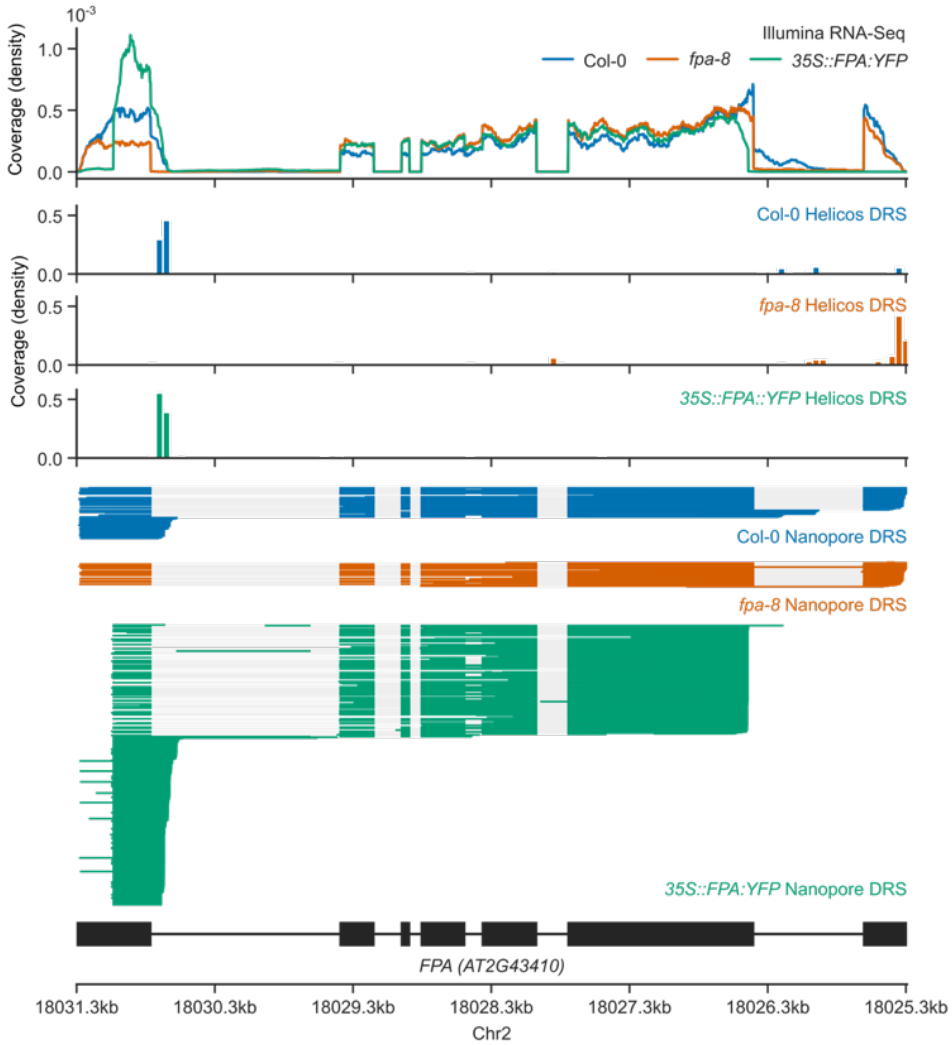


Figure 3: Changes in RNA processing at the *FPA* locus can be precisely mapped using a combination of three RNA sequencing technologies

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments) reveal FPA-dependent RNA 3′ end processing changes at the *FPA* (*AT2G43410*) locus. Col-0 is shown in blue; *fpa-8*, in orange; *35S::FPA:YFP*, in green; and Araport11 annotation, in black.

We next asked how RNA processing was affected by FPA activity transcriptome-wide. Since mutations in FPA cause readthrough of annotated 3′UTRs (Duc *et al.*, 2013) we applied the software tool StringTie2 (Pertea *et al*, 2015) to create a new bespoke reference annotation with nanopore DRS reads from Col-0, *fpa-8* and *35S::FPA:YFP*. We then used this new reference annotation to build RNA 3′ end position distributions of nanopore or Helicos DRS reads overlapping each transcriptional locus. To determine if poly(A) site choice was altered in different backgrounds, we measured differences in the 3′ end distributions using a metric called the earth-mover distance (EMD). This metric indicates the "work" required to transform one normalised distribution into another, considering the amount of 3′ ends which would have to be moved and the distance by which they would be moved. We used an EMD permutation test, where reads are randomly shuffled between

conditions, to estimate p-values for each gene. Using this approach on nanopore DRS data, we identified 285 and 293 loci (EMD > 25, FDR < 0.05) with alternative polyadenylation events in *fpa-8* and *35S::FPA:YFP*, respectively (Figure 4A). Out of all alternatively polyadenylated loci identified in *fpa-8*, 77.9% (222) displayed a positive change in mean 3′ end position, indicating a predominant shift to distal poly(A) site selection (Figure 4A, left panel). These loci also had greater effect sizes than those shifting towards proximal poly(A) sites (Figure 4B, left panel). In contrast, 56.7% (166) of loci with identified alternative polyadenylation in *35S::FPA:YFP* displayed a negative change in mean 3′ end position, indicating a shift towards proximal poly(A) sites (Figure 4A, right panel). These loci had greater effect sizes than those with positive changes in 3′ end profile (Figure 4B, right panel). 16 loci displayed both a shift to distal poly(A) site selection in *fpa-8* and to proximal poly(A) site choice in *35S::FPA:YFP* (hypergeometric test $p=3.9\times10^{-7}$) (Figure 4A-B), demonstrating that loss-of-function or overexpression of FPA can result in reciprocal patterns of poly(A) site choice.

We next used the same approach to identify loci with FPA-dependent alternative polyadenylation in Helicos DRS data. We identified 319 and 299 genes (EMD > 25, FDR < 0.05) with alternative polyadenylation events in *fpa-8* and *35S::FPA:YFP*, respectively (Figure 4C). 79.0% (252) of the loci with alternative polyadenylation in *fpa-8* displayed a positive change in mean 3′ end position, indicating a shift to distal poly(A) site selection (Figure 4C, left panel). These loci are also associated with greater effect sizes than those shifting towards proximal poly(A) sites (Figure 4D, left panel). 75.3% (225) of genes with detectable alternative polyadenylation in *35S::FPA:YFP* displayed a negative shift in mean 3′ end position, indicating proximal poly(A) site selection (Figure 4C, right panel). These loci had stronger effect sizes than genes with increased distal poly(A) site selection (Figure 4D, right panel). 44 genes displayed a shift to distal poly(A) sites in *fpa-8* and to proximal poly(A) sites in *35S::FPA:YFP* (hypergeometric test $p=4.8\times10^{-30}$), again demonstrating reciprocal poly(A) site selection according to FPA activity. Of the loci identified as distally alternatively polyadenylated in *fpa-8* using nanopore DRS, 39.6% (88) were also detected using Helicos DRS (Figure 4 – supplement 1). Likewise, 44% of loci (73) with proximal polyadenylation detected in *35S::FPA:YFP* using nanopore DRS, were also detected using Helicos DRS. Across the DRS datasets, we identified 59 loci where reciprocal poly(A) site regulation by FPA could be detected by nanopore DRS, Helicos DRS, or a mixture of both technologies.

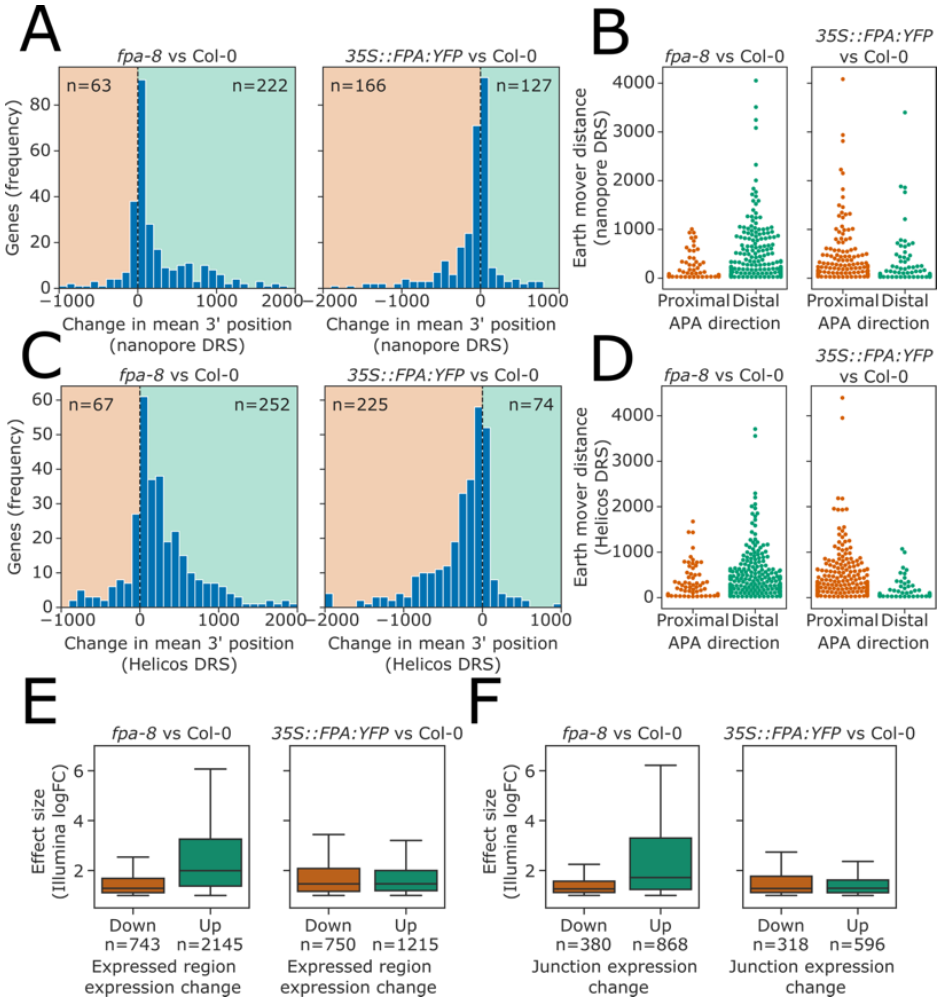


Figure 4: FPA promotes poly(A) site selection

Loss of FPA function is associated with the preferential selection of distal poly(A) sites, whilst FPA overexpression leads to the preferential selection of proximal poly(A) sites. **(A)** 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, detected using nanopore DRS. Orange and green shaded regions indicate sites with negative and positive RNA 3′ end position changes, respectively. **(B)** Effect size of significant proximal (orange) and distal (green) APA events in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, measured using the EMD. **(C)** 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, detected using nanopore DRS. Orange and green shaded regions indicate sites with negative and positive RNA 3′ end position changes, respectively. **(D)** Effect size of significant proximal (orange) and distal (green) APA events in *fpa-8* (left panel) and *35S::FPA:YFP* (right panel) compared with Col-0, measured using the EMD. **(E)** Boxplots showing the effect size (absolute log 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 upregulated and downregulated regions contribute to both boxes. **(F)** 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 upregulated and downregulated junctions contribute to both boxes.

373 In order to analyse the Illumina RNA-Seq data, we developed annotation-agnostic
374 software for detecting alternative RNA 3′ end processing events, using a similar approach
375 to the existing tools DERfinder (Collado-Torres *et al*, 2017), RNAprof (Tran Vdu *et al*, 2016)
376 and DEXSeq (Anders *et al*, 2012). We segmented the Illumina RNA-Seq data using
377 coverage thresholding and relative changes in expression in *fpa*-8 or *35S::FPA:YFP*
378 compared to Col-0. Segmented regions were grouped into transcriptional loci using the
379 annotations generated from Col-0, *fpa-8*, and *35S::FPA:YFP* nanopore DRS datasets.
380 Differential usage of regions within each locus was then tested using DEXSeq (Figure 4E).
381 Using this approach, we identified 2535 loci with differential RNA processing events in *fpa-*
382 *8*: 1792 upregulated, 390 downregulated, and 353 with both upregulated and
383 downregulated regions (FDR < 0.05, absolute logFC > 1) (Figure 4E, left panel). 1747 loci
384 with differential RNA processing events were identified in *35S::FPA:YFP*: 997 upregulated,
385 532 downregulated, and 218 with both upregulation and downregulation (Figure 4E, right
386 panel). In the *fpa-8* mutant, we found that the median effect size for differentially processed
387 regions was greater for upregulated regions than downregulated regions (Figure 4E, left
388 panel). This is consistent with the major effect of loss of FPA function resulting in
389 transcriptional readthrough events and increased expression of intergenic regions,
390 downstream genes and/or transposable elements (TE). In contrast, the median effect size
391 for differentially processed regions was similar for up- and downregulated regions in
392 *35S::FPA:YFP* (Figure 4E, right panel). This is consistent with an increase in the relative
393 expression of proximal exonic and intronic regions, and a loss of expression of distal exonic
394 regions caused by preferential selection of proximal poly(A) sites.
395 To test whether changes in FPA expression result in alternative splicing, we
396 quantified the number of reads supporting each annotated splice junction, grouped by
397 transcriptional locus, and performed differential junction usage analysis with DEXSeq
398 (Figure 4F). We found that the clearest alterations in splice site usage occurred in *fpa-8*,
399 where 868 loci have splice junctions that are strongly upregulated (Figure 4F, left panel).
400 The effect sizes for these loci tended to be greater than those of differentially spliced loci
401 detected in *35S::FPA:YFP* (Figure 4F, right panel). Closer inspection of the junctions
402 affected in *fpa-8* indicated that they are the result of increased cryptic splicing events
403 associated with altered poly(A) site choice and the increased splicing of downstream genes
404 in readthrough and chimeric RNAs. These changes in processing can be illustrated by
405 examining the chimeric RNAs formed at the *PIF5* locus. Here, readthrough in *fpa-8* mutants
406 is associated with cryptic splicing events between the annotated terminal exon of *PIF5* and
407 the first exon of *PAO3* (Figure 4 – supplement 2). In addition, because the detectable
408 expression of the downstream gene *PAO3* is increased, there is a corresponding increase
409 in reads spanning annotated *PAO3* splice junctions. These changes in splicing can be
410 explained by the effect of FPA on 3′ end formation at *PIF5*. We conclude that changes in
411 splicing are indirect effects of the altered 3′ end processing which occurs in *fpa-8* mutants,
412 rather than direct effects of FPA on splice site choice.

413 Our *IVI-MS* analysis revealed that conserved members of the Arabidopsis m^{6}A writer
414 complex co-purify with FPA (Figure 1D, Supplementary file 1). The most related proteins to
415 FPA in humans are RBM15/B, which co-purify with the human m^{6}A writer complex and are
416 required for m^{6}A deposition (Patil *et al*, 2016). We therefore asked if FPA influenced the
417 level of m^{6}A in Arabidopsis mRNA. We used LC-MS/MS to analyse the ratio of m^{6}A/A in
418 mRNA purified from Col-0, *fpa-8*, *35S::FPA:YFP* and a mutant defective in the m^{6}A writer
419 complex component VIR (*vir-1*). Consistent with previous reports, the level of mRNA m^{6}A in
420 the hypomorphic *vir-1* allele was reduced to approximately 10% of wild-type levels (Parker
421 *et al.*, 2020b; Ruzicka *et al.*, 2017) (Figure 4 – supplement 3). However, we did not detect
422 any differences in the m^{6}A level between genotypes with altered FPA activity. We conclude
423 that FPA does not influence global levels of mRNA m^{6}A methylation.
424 Overall, these analyses reveal that the primary function of FPA is to control poly(A)
425 site choice. FPA predominantly promotes poly(A) site selection, hence *fpa* loss-of-function
426 backgrounds exhibit readthrough at sites used in wild-type, while FPA overexpression
427 results in increased selection of proximal poly(A) sites. Changes in splicing patterns in
428 backgrounds with different FPA activities can be explained as consequences of alternative
429 polyadenylation. Moreover, FPA is not required for the global levels of mRNA m^{6}A, despite
430 co-purifying with components of the m^{6}A writer complex.
431
432 **FPA controls the processing of NLR transcripts**
433 We next asked what groups of genes were sensitive to FPA-dependent alternative
434 polyadenylation. We found that genes encoding proteins involved in cleavage,
435 polyadenylation and transcription termination, or which co-purified with FPA, were affected
436 by FPA-mediated alternative polyadenylation. These included *CFIS2*, *CPSF160, CPSF73-1*,
437 *EMB1579/RSA1* and *ATSPEN2* (Figure 4 – supplement 4-8). FPA was also found to promote
438 exonic proximal polyadenylation within the coding regions of transcripts encoding PCF11-
439 related proteins *PCFS1* and *PCFS5,* despite both genes being encoded by a single exon
440 (Figure 4 – supplement 9-10). Moreover, we found that the *BORDER1* and *BORDER2* genes,
441 which encode proteins that protect expression of adjacent Arabidopsis genes by 3' pausing
442 of Pol II (Yu *et al.*, 2019), were controlled by FPA-dependent alternative polyadenylation
443 (Figure 4 – supplement 11-12).
444 We next used Interpro annotations (Mitchell *et al*, 2019) to perform protein family
445 domain enrichment analysis of the genes affected by FPA as revealed by the nanopore and
446 Helicos DRS analyses. This approach is similar to gene ontology analysis but relies on the
447 automatic annotation of domains using profile hidden Markov models, rather than on a
448 mixture of curated automatic and manual annotation. Consequently, annotations of poorly
449 studied genes should be similarly detailed to well-studied ones. We found that genes
450 encoding nucleotide binding NB-ARC, Rx-like coiled coil (CC), and Leucine Rich Repeat
451 (LRR) domains were most enriched amongst those with increased proximal polyadenylation

452 in *35S::FPA:YFP* (Figure 5A-B). This combination of protein domains is associated with the
453 family of disease resistance factors known as NLR proteins.
454 According to the recently published pan-NLRome of Arabidopsis, the Col-0
455 accession contains at least 206 genes with some combination of TIR (Toll-Interleukin-1
456 Receptor), CC, RPW8-like CC, NB-ARC, or LRR domains, which might be classified as NLRs
457 or partial NLRs (Van de Weyer *et al.*, 2019). In general, these can be grouped into classes
458 by their N-terminal domain as either TIR (TNLs), CC (CNLs), RPW8-like (RNLs), or NB-ARC
459 and LRR only (NLs) NLR genes. There are also many genes that comprise only TIR, CC, or
460 CCR domains and NLRs that incorporate other domains that function as integrated decoys
461 (Le Roux *et al*, 2015). We manually examined these genes to identify those with alternative
462 polyadenylation dependent on FPA. Nanopore DRS reveals that pre-mRNA processing at
463 many NLR loci is more complicated than current annotation would indicate. As a result, we
464 revised the annotation of some NLR genes in this analysis. For example, we find that the TNL
465 gene *AT5G46490* is incorrectly split into two loci, *AT5G46490* and *AT5G46500* (Figure 5 –
466 supplement 1, Supplementary file 3). This is in fact a single locus with a previously
467 unrecognised 2.7 kb intron containing a proximal poly(A) site, use of which is regulated by
468 FPA. This finding was supported by nanoPARE data (Schon *et al*, 2018), which shows that
469 there are no capped 5′ ends originating from the annotated downstream gene. Use of the
470 distal poly(A) site introduces an additional ~400 amino acids to the C-terminus of the
471 protein with predicted homology to RPP5 and lengthens the LRR domain. At the divergent
472 TNL pair *AT1G72840* and *AT1G72850*, an annotated natural antisense RNA opposite
473 *AT1G72850* (*AT1G72852)* was found to be transcribed as a single RNA with *AT1G72840*
474 (Figure 5 – supplement 2, Supplementary file 3). FPA-regulated proximal polyadenylation
475 can occur at the annotated *AT1G72852* 3′ end, but we saw no evidence in nanoPARE data
476 that downstream *AT1G72840* can be transcribed on its own.
477 Clearly, correct annotation is a necessary part of understanding how NLRs are
478 expressed and controlled. Notably, we could also reannotate the chromosomal region
479 around *RPS6* (*AT5G46470*). The extreme autoimmunity phenotypes of NMD mutants and
480 mitogen activated kinase pathway mutants are caused by the dysregulation of RPS6 but the
481 mechanisms involved are not understood (Gloggnitzer *et al.*, 2014; Takagi *et al.*, 2020). Our
482 nanopore DRS analysis reveals a long and complex *RPS6* 3′UTR, with multiple exons and
483 alternative poly(A) sites (Figure 5 – supplement 3). We also detect evidence of rare
484 transcripts in this region that do not appear to be contiguous with *RPS6* 3′UTR reads.
485 Instead, these reads correspond to an independent unannotated downstream gene that
486 overlaps with the *RPS6* 3′UTR. Consistent with this idea, capped 5′ ends identified by
487 nanoPARE map to this region (Schon *et al.*, 2018) and our nanopore DRS analysis of the RNA
488 exosome mutant *hen2-2* (Parker *et al*, 2020a) reveals relatively high levels of independently
489 transcribed RNAs in the corresponding position (Figure 5 – supplement 3). The downstream
490 gene encodes a TIR domain similar to the RPS6 TIR domain (Figure 5 – supplement 4). Use
491 of the distal *RPS6* poly(A) site therefore constitutes readthrough into the downstream TIR-

492 domain-only NLR (Figure 5 – supplement 3). We conclude that long read nanopore DRS
493 data has the potential to resolve the complexity of NLR gene annotation, that long read DNA
494 sequence data and short read Illumina RNA-Seq cannot.
495 Of the 206 NLR genes examined, 124 were sufficiently expressed to identify
496 alternative polyadenylation in the nanopore DRS data, and of these, 62 (50%) were found
497 to have FPA-dependent alternative polyadenylation (Supplementary file 3). These effects
498 could be classified into three modes of regulation: (i) FPA control of readthrough and
499 chimeric RNAs, usually by repressing their formation; (ii) FPA regulation of intronic poly(A)
500 site selection; and (iii) FPA control of poly(A) site selection within protein-coding exons. FPA
501 can affect poly(A) site choice using a combination of all three modes of regulation at a
502 number of complex loci.
503 We detected 16 NLR genes where a change in FPA levels altered formation of
504 readthrough or chimeric RNAs containing one or more NLR loci (Supplementary file 3).
505 *AT1G58848*, which is part of the *RPP7* cluster containing five CNL-class NLRs, displayed
506 increased readthrough into the downstream TE *AT1G58889* in the *fpa-8* mutant (Figure 5C,
507 Supplementary file 3). Loss of FPA function can also lead to the generation of clusters
508 containing two or more NLR genes transcribed together as one transcriptional unit. For
509 example, the TNL-class gene *AT1G63730* forms chimeric RNA with the downstream TNL-
510 class gene *AT1G63740* in *fpa-8* (Figure 5 – supplement 5, Supplementary file 3).
511 We identified 18 NLR genes with intronic polyadenylation regulated by FPA
512 (Supplementary file 3). At *AT1G69550*, encoding a TNL-type NLR, we identified a proximal
513 poly(A) site within the third intron (Figure 5D, Supplementary file 3). Use of this poly(A) site
514 results in mRNAs with a premature stop codon, which encodes a form of the AT1G69550
515 protein lacking the majority of the predicted LRR domain. In *fpa-8* mutants, readthrough of
516 this poly(A) site increased, resulting in higher levels of full-length transcripts. At the TNL-
517 type NLR *RPS4,* which has previously been shown to be regulated by alternative splicing
518 (Zhang & Gassmann, 2007), we detected an increase in intron 3 retention in *35S::FPA:YFP*
519 compared to Col-0 that was associated with proximal polyadenylation events in intron 3 or
520 exon 4 (Figure 5 – supplement 6, Supplementary file 3). At *AT1G12210*, encoding a CNL-
521 class NLR, we detected short isoforms ending within the first intron of the 5′ UTR in Col-0
522 (Figure 5 – supplement 7, Supplementary file 3). Loss of FPA function caused transcriptional
523 readthrough at this poly(A) site, resulting in production of full-length *AT1G12210* mRNAs
524 comprising the full protein-coding sequence. FPA also regulated intronic polyadenylation
525 in Rx N-terminal domain-only genes, such as the RPW8-like domain gene *RPW8/HR*
526 (*AT3G50480*) (Figure 5 – supplement 8, Supplementary file 3).
527

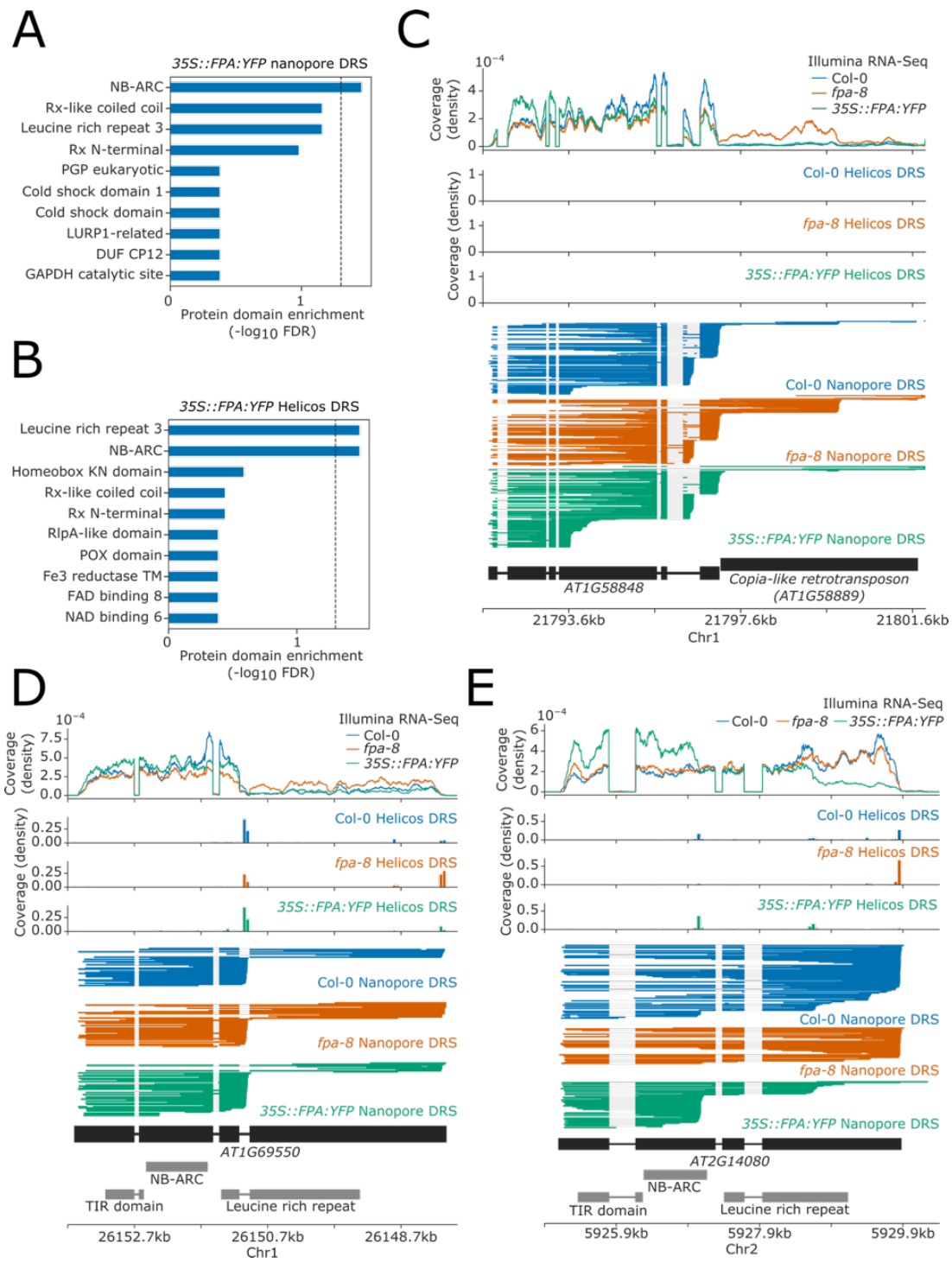
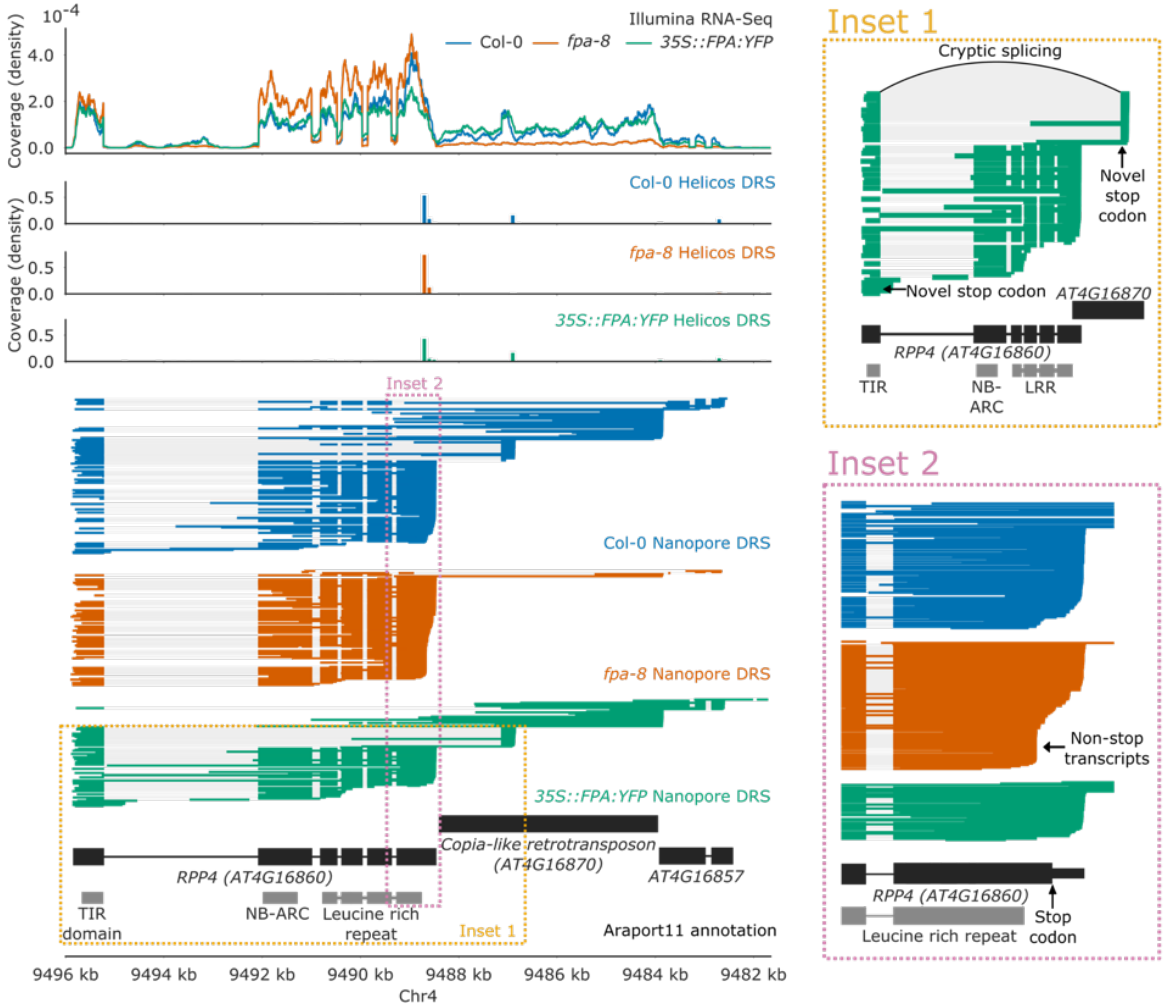


Figure 5: FPA-dependent alternative polyadenylation of NLR transcripts

(A-B) Protein domain enrichment analysis for loci with increased proximal poly(A) site selection in *35S::FPA:YFP* line, detected using **(A)** nanopore DRS or **(B)** Helicos DRS. **(C-E)** FPA regulates **(C)** 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); **(D)** intronic polyadenylation at *AT1G69550;* and **(E)** exonic polyadenylation at *AT2G14080*. Top panels, Illumina RNA-Seq coverage; middle panels, Helicos DRS coverage; bottom panels, nanopore DRS read alignments. Col-0 is shown in blue; *fpa-8*, in orange; *35S::FPA:YFP*, in green; Araport11 annotation, in black; and genomic aligned Interpro domains, in grey.

540 The most common form of NLR regulation was FPA-dependent premature
541 termination within protein-coding exons, which was detected at 46 NLR genes
542 (Supplementary file 3). Many of these proximal exonic poly(A) sites can be identified at low
543 levels in Col-0. For example, at *AT2G14080*, encoding a TNL-class NLR, we detected
544 multiple exonic poly(A) sites located within the second and fourth exons encoding the NB-
545 ARC and LRR domains respectively, selection of which increased in *35S::FPA:YFP* and
546 decreased in *fpa-8* (Figure 5E, Supplementary file 3). At *LOV1* (*AT1G10920*), encoding a
547 CNL-type NLR protein, overexpression of FPA caused more frequent selection of an exonic
548 poly(A) site located within the fourth exon encoding the NB-ARC domain region, whilst loss
549 of FPA function resulted in increased production of full-length transcripts (Figure 5 –
550 supplement 9, Supplementary file 3). FPA was even found to promote premature
551 termination in the protein coding sequence of single exon, intronless NLR genes. For
552 example, at *RPP13* (*AT3G46530),* encoding a CNL-class NLR protein, FPA overexpression
553 causes selection of proximal poly(A) sites located within the region encoding the LRR
554 domain (Figure 5 – supplement 10, Supplementary file 3).
555 An example of complex FPA-dependent regulation is found at *RPP4* (*AT4G16860*).
556 This gene encodes a well-studied TNL-class NLR protein known to mediate Arabidopsis
557 resistance to *Hyaloperonospora arabidopsidis* isolate Emoy2 (*Hpa*-Emoy2) (van der Biezen
558 *et al*, 2002). *RPP4* is part of the *RPP5* cluster comprising seven TNL-class NLRs interspersed
559 with three related sequences and two non-NLR genes (van der Biezen *et al.*, 2002). In
560 agreement with a previous study (Wang & Warren, 2010), we found that in wild-type Col-0,
561 *RPP4* can be transcribed as a chimeric RNA with the downstream *AtCOPIA4* TE
562 (*AT4G16870*), as a result of the selection of two distal poly(A) sites located within the TE
563 (Figure 6) (Wang & Warren, 2010). We also identified several chimeric RNAs containing
564 *RPP4*, the TE and the downstream gene *AT4G16857*. Use of the more proximal poly(A) site
565 within the TE is associated with an approximately 8-kb cryptic splicing event between the 5'
566 splice site of the first exon of *RPP4* and a 3' splice site located within the TE, which skips all
567 *RPP4* exons downstream of exon 1 (Figure 6, Inset 1). This splicing event, detected by
568 nanopore DRS, and which is also supported by Illumina RNA-Seq data (Figure 6), therefore
569 removes the majority of the *RPP4* coding sequence and introduces a novel stop codon
570 (Figure 6, Inset 1). The resulting transcript is predicted to encode a truncated protein,
571 containing only the TIR domain of RPP4. In Col-0, 27% of nanopore DRS reads at the *RPP4*
572 locus terminate at one of the two TE-located poly(A) sites or within the downstream gene
573 (Figure 6, Figure 6 – supplement 1A-B). Loss of FPA function decreases chimeric RNA
574 formation by shifting poly(A) site selection towards a proximal poly(A) site located within
575 the protein-coding region of the final exon (Figure 6, Figure 6 – supplement 1A-B,
576 Supplementary file 3). This results in production of *RPP4* transcripts lacking in-frame stop
577 codons (Figure 6, Inset 2; Figure 6 – supplement 1A-B, Supplementary file 3). Furthermore,
578 in the *35S::FPA:YFP* line, we observed increased selection of a proximal poly(A) site located
579 within the first intron of *RPP4*, which would also encode a truncated RPP4 protein (Figure 6,

580 Inset 1; Figure 6 – supplement 1A-B, Supplementary file 3). We conclude that FPA regulated
581 alternative polyadenylation at *RPP4* produces transcripts with unusually long 3'UTRs,
582 alternative protein isoforms and transcripts that cannot be efficiently translated.
583



584
585 **Figure 6: Complex FPA-dependent patterns of alternative polyadenylation at *RPP4***
586 Top panels, Illumina RNA-Seq coverage; middle panels, Helicos DRS coverage; bottom panels,
587 nanopore DRS read alignments. Col-0 is shown in blue; *fpa-8*, in orange; *35S::FPA:YFP*, in green;
588 Araport11 annotation, in black; and genomic aligned Interpro domains, in grey. **Inset 1** (light orange
589 area) shows a magnified view of TIR-domain only *RPP4* transcripts detected in *35S::FPA:YFP*, caused
590 by proximal polyadenylation in intron 1, and distal polyadenylation within the TE associated with
591 cryptic splicing; **Inset 2** (pink area) shows a magnified view of the stop-codonless transcripts
592 produced within the protein-coding *RPP4* region (Non-stop transcripts) in *fpa-8*.
593

594 **The impact of FPA on NLR gene regulation is independent of its role in controlling**
595 ***IBM1* expression**

596 We have previously shown that FPA controls the expression of the histone demethylase
597 IBM1 (INCREASE IN BONSAI METHYLATION 1; *AT3G07610*) by promoting premature
598 cleavage and polyadenylation within *IBM1* intron 7 (Duc *et al.*, 2013). This regulation is also
599 detectable in nanopore DRS data, with reciprocal changes in poly(A) site choice in *fpa-8*

and *35S::FPA:YFP* (Figure 7 – supplement 1). IBM1 functions to restrict histone H3K9 methylation to genic regions and *ibm1* mutants show developmental defects and ectopic accumulation of H3K9me and DNA methylation at non-CG sites in thousands of transcribed genes (Miura *et al*, 2009; Saze *et al*, 2008). We therefore asked whether the impact of FPA overexpression on global poly(A) site choice could be explained by altered chromatin state, resulting from *IBM1* control. We analysed two independent ChIP-Seq datasets of H3K9me^2 in *ibm1-4* mutants (Inagaki *et al*, 2017; Lai *et al*, 2020). However, we find that only 10.6% of loci with altered poly(A) site choice in the FPA overexpressing line have altered H3K9me^2 in *ibm1* mutants compared to 14.2% of all loci tested (hypergeometric p = 0.97) (Figure 7). We conclude that FPA-dependent regulation of NLR gene expression is not an indirect consequence of FPA controlling *IBM1*.

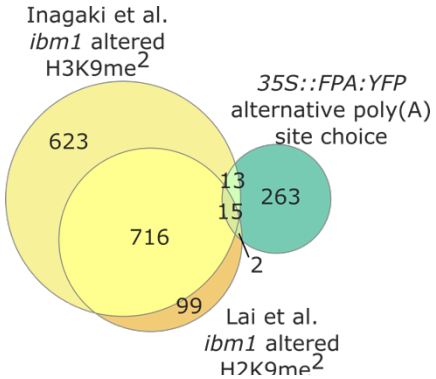


Figure 7: FPA-dependent regulation of NLR expression is independent of IBM1
Venn diagram showing genes with altered H3K9me^2 levels in *ibm1* mutants, in yellow (Inagaki *et al.*, 2017) and orange (Lai *et al.*, 2020); and genes with altered poly(A) site choice in *35S::FPA:YFP,* in green.

FPA promotes premature termination within *RPP7* protein coding exon 6

To understand the impact of FPA on the regulation of NLRs more closely, we focused on the well-studied gene *RPP7* (*AT1G58602). RPP7* encodes a CNL-class NLR protein crucial for strain-specific resistance to *Hpa* isolate Hiks1 (*Hpa*-Hiks1) (Tsuchiya & Eulgem, 2013). Research into *RPP7* has focussed not only on its role in immunity, but also on the regulation of its expression by intragenic heterochromatin (Rigal *et al*, 2012; Saze *et al*, 2013). *RPP7* is approximately 20 kb long and contains nine exons, three comprising the 5'UTR, followed by three protein-coding exons and three exons comprising the 3'UTR (Figure 8A). A *Ty-1* COPIA-type retrotransposon (COPIA-R7) that resides within the relatively long (7 kb) first intron of *RPP7* bears hallmarks of heterochromatin, including m^{5}C DNA methylation and histone H3 dimethylation (H3K9me^2) (Tsuchiya & Eulgem, 2013). The *COPIA-R7* has long terminal repeats (LTRs) at both ends and the 5' LTR region can serve as an alternative poly(A) site. Use of this site is undetectable in wild-type plants in control conditions but is preferentially selected in mutants with disrupted heterochromatic marks. This results in the production of a truncated, non-functional transcript, termed the Exon 1-Containing LTR-

634 terminated transcript (ECL), accompanied by a reduction in the abundance of full-length
635 *RPP7* transcripts and compromised immunity against *Hpa*-Hiks1 infection (Tsuchiya &
636 Eulgem, 2013).
637 Analysis of our combined set of Illumina RNA-Seq, nanopore and Helicos DRS data
638 reveals that the regulation of termination at *RPP7* is more complex than previously reported.
639 We detect ten distinct poly(A) sites at five positions distinguishable in Helicos and nanopore
640 DRS data (Figure 8A). We found at least two poly(A) sites located within the *COPIA-R7*
641 element, including one characteristic of the previously described ECL transcript. Selection
642 of both poly(A) sites was more frequent in *fpa-8* (Figure 8A, Figure 8 – supplement 1A-B,
643 Supplementary file 3). A third poly(A) site is located immediately downstream of the *COPIA-*
644 *R7* element. We also identified at least two poly(A) sites within the second intron of *RPP7*,
645 selection of which is reciprocally regulated by FPA, with a moderate decrease in *fpa-8* and
646 an increase in *35S::FPA:YFP* (Figure 8A, Figure 8 – supplement 1A-B, Supplementary file 3).
647 All of these intronic proximal poly(A) sites occur before the start of the *RPP7* open reading
648 frame, and their selection would generate truncated transcripts that do not encode RPP7
649 protein. At the 3′ end of *RPP7*, we found three alternative poly(A) sites located in the terminal
650 intron (within the 3′ UTR sequence), in addition to the previously identified poly(A) site in
651 the terminal exon (Figure 8A, Inset 1, Supplementary file 3) (Tsuchiya & Eulgem, 2013).
652 Selection of two of these poly(A) sites is associated with cryptic splicing events that lead to
653 the generation of four possible 3′UTR sequences associated with distally processed *RPP7*
654 transcripts. Selection of these 3′UTR intronic poly(A) sites is suppressed by FPA, with
655 increased and decreased usage in *fpa-8* and the FPA overexpressing line respectively
656 (Figure 8A, Inset 1; Figure 8 – supplement 1A-B, Supplementary file 3). Based on these data,
657 we conclude that FPA influences *RPP7* intronic polyadenylation at a larger number of
658 poly(A) sites than previously considered.
659 The major effect of FPA on *RPP7* is within the protein-coding exon 6 of *RPP7*. We
660 identified three poly(A) sites located within the protein-coding sixth exon of *RPP7* (Figure
661 8A). The first two of these are located at the end of the region encoding the NB-ARC domain,
662 and the third is positioned downstream, within the region encoding the LRR repeats.
663 Cleavage and polyadenylation at these three sites results in transcripts without in-frame stop
664 codons, disrupting the coding potential and translatability of *RPP7* mRNA (Figure 8A, Inset
665 2). These poly(A) sites were identified in Helicos and nanopore DRS data, indicating that
666 they are unlikely to be caused by alignment errors. Approximately 25% of all Col-0
667 nanopore DRS reads terminate at one of the three poly(A) sites within the *RRP7* sixth exon.
668 The relative selection of these poly(A) sites depends on FPA expression (Figure 8A, Inset 2,
669 Figure 8 – supplement 1A-B, Supplementary file 3). When FPA is overexpressed,
670 approximately 63% of all *RPP7* transcripts identified by nanopore DRS terminate at proximal
671 poly(A) sites within exon 6, and only 25% of transcripts detected are expected to encode
672 the full *RPP7* open reading frame (Figure 8A, Figure 8 – supplement 1A, Supplementary file
673 3). A clear relative drop in coverage caused by the selection of these exonic proximal poly(A)

674 sites is also observed in *35S::FPA:YFP* Illumina RNA-Seq data (Figure 8A, Supplementary file
675 3). Increased usage of proximal exonic poly(A) sites can be observed, with accompanying
676 reduction in distal poly(A) sites selection, in Helicos DRS data derived from *35S::FPA:YFP*
677 (Figure 8A, Figure 8 – supplement 1B, Supplementary file 3). In contrast, only 4% of *RPP7*
678 nanopore DRS reads identified in *fpa-8* terminate in the sixth exon, and 79% are predicted
679 to encode full-length protein (Figure 8 – supplement 1A, Supplementary file 3). We used
680 RNA gel blot analysis to visualise *RPP7* mRNAs in Col-0, *fpa-8* and *35S::FPA:YFP*
681 backgrounds and detected a clear decrease in signal corresponding to full-length *RPP7*
682 transcripts in *35S::FPA:YFP* (Figure 8B). This orthogonal approach validates the conclusion
683 derived from sequencing analyses that FPA limits the expression of full-length *RPP7* mRNA.
684 These data support previous evidence of FPA-dependent control of *RPP7* (Deremetz *et al.*,
685 2019) but reveal how this processing occurs.
686

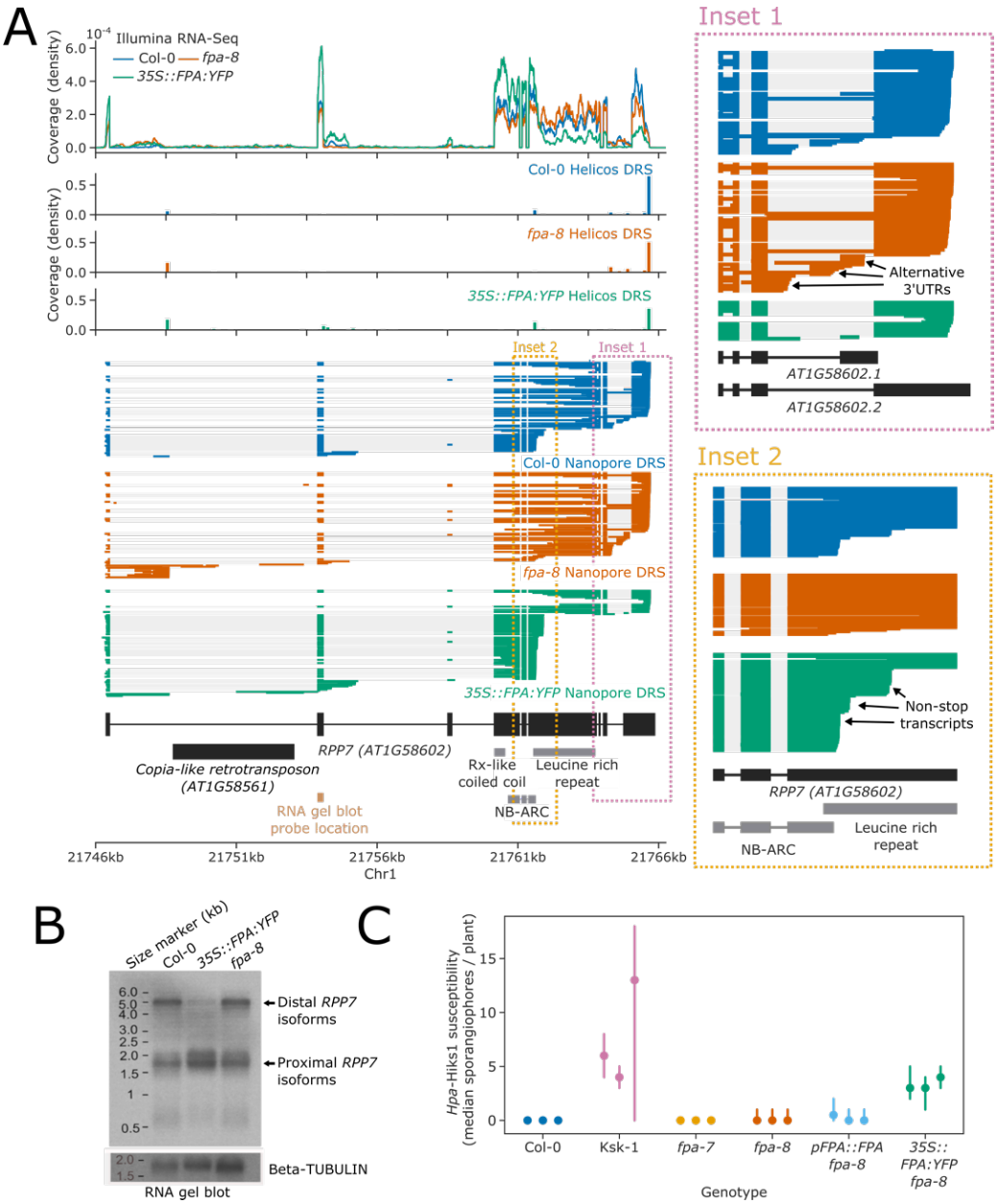


Figure 8: FPA promotes premature cleavage and polyadenylation within *RPP7* protein-coding exon 6 that compromises plant immunity against *Hyaloperonospora arabidopsidis* isolate Hiks1(*Hpa*-Hiks1)

(A) FPA-dependent RNA 3′ end formation changes at the *RPP7 (AT1G58602)* locus. Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8*, in orange; *35S::FPA:YFP*, in green; Araport11 annotation, in black; and genomic aligned Interpro domains, in grey. **Inset 1** (pink area) shows a magnified view of the *RPP7* 3′ UTR region with alternative RNA 3′ ends; **Inset 2** (light orange area) shows a magnified view of the stop-codonless transcripts produced in the 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 four days after *Hpa*-Hiks1 inoculation of Col-0 in dark blue; Ksk-1 in pink; *fpa-7* in light orange; *fpa-8* in dark orange, *pFPA::FPA* in light blue; and *35S::FPA:YFP* in green. Error bars

703 are 95% confidence intervals. Each experimental replicate was generated from 7-45 plants per
704 genotype.
705
706 **FPA modulates *RPP7*-dependent, race-specific susceptibility to the pathogen**
707 ***Hyaloperonospora arabidopsidis***
708 Since the major impact of FPA on *RPP7* is to limit the production of full-length protein coding
709 transcripts, we reasoned that immunity might be compromised when FPA is overexpressed.
710 To test this hypothesis, we carried out pathogenesis assays using the oomycete *Hpa*-Hiks1,
711 against which RPP7 provides race-specific immunity. We inoculated Arabidopsis seedlings
712 with *Hpa*-Hiks1 spores in three independent experiments and checked susceptibility by
713 counting the number of sporangiophores four days after inoculation. We compared the
714 susceptibility of six Arabidopsis genotypes: Col-0, Ksk-1, FPA loss-of-function mutant alleles
715 *fpa-7* and *fpa-8,* a complementing line *pFPA::FPA* expressing FPA gene under the control
716 of its native promoter (Zhang *et al*, 2016) and the FPA overexpressing line *35S::FPA:YFP,*
717 both of which were introduced into the *fpa-8* mutant allele. *fpa-7* is a Transfer-DNA insertion
718 mutant allele (Alonso *et al*, 2003), whilst *fpa-8* is derived from an ethyl methanesulfonate
719 (EMS)-induced mutant screen (Baurle *et al*, 2007). With the exception of the Ksk-1 accession,
720 all these mutant alleles and transgenic lines are in a Col-0 background. Ksk-1 lacks *RPP7*
721 and is therefore an *Hpa*-Hiks1 susceptible ecotype used here as a positive control (median
722 5 sporangiophores per plant) (Figure 8C). Col-0 was used as a resistant control (median 0
723 sporangiophores per plant). Like Col-0, the *fpa-7* mutant displayed inhibited sporulation of
724 *Hpa*-Hiks1 (median 0 sporangiophores per plant), suggesting that loss of FPA function does
725 not reduce plant resistance. Greater variability in pathogen susceptibility was observed in
726 the *fpa-8* mutant and was not restored by complementation with *pFPA::FPA,* possibly
727 indicating background EMS mutations affecting susceptibility. Both lines show low median
728 susceptibility, however (median 0 sporangiophores per plant). In contrast, *35S::FPA:YFP*
729 plants exhibited a similar level of sporulation to the pathogen-sensitive *Ksk-1* accession
730 (median 3 sporangiophores per plant). This suggests that the premature exonic termination
731 of *RPP7* caused by FPA has a functional consequence for Arabidopsis immunity against *Hpa*-
732 Hiks1.
733
734
735 **Discussion**
736 This study reveals that the predominant role of the RNA-binding protein FPA is to promote
737 poly(A) site choice and transcription termination, and links this process to widespread
738 control of Arabidopsis NLR immune-response genes. Using *IVI-MS* proteomics, we showed
739 that FPA is closely associated with proteins involved in RNA 3′ end processing, m^6A
740 modification and transcription termination. Metagene analysis of ChIP-Seq data revealed
741 that FPA co-localised on chromatin with Ser^2 phosphorylated Pol II at the 3′ ends of
742 Arabidopsis genes. The analysis of three genetic backgrounds with different FPA activity

743 levels, using three orthogonal RNA sequencing technologies, revealed that FPA
744 predominantly promotes poly(A) site choice and does not affect other aspects of RNA
745 processing or modification. Almost half of Arabidopsis NLR loci with detectable levels of
746 expression were sensitive to FPA, and in most cases, FPA targeted poly(A) sites within
747 protein-coding exons, limiting the expression of full-length NLR proteins. For example, FPA
748 promoted premature cleavage and polyadenylation within protein-coding exon 6 of *RPP7*
749 transcripts, functionally compromising *Hpa*-Hiks1 strain-specific immunity.
750
751 **Uncovering protein assemblies that mediate 3′ end processing in living plant cells**
752 We used an *in vivo* FA cross-linking approach to identify proteins localised in the
753 neighbourhood of FPA inside living plant cells. We discovered that FPA co-purifies with
754 factors involved in processing and modifying mRNA 3′ ends and promoting transcription
755 termination. These include almost every conserved component of the cleavage and
756 polyadenylation machinery and RNA-binding proteins such as HLP1 (Zhang *et al.*, 2015)
757 that, like FPA, functions in 3′ end formation and controls the expression of the flowering
758 time transcription factor FLC. We found factors involved in modulating the phosphorylation
759 state of Pol II at the 3′ end of genes and components of the m^6A writer complex which target
760 mRNA 3′ UTRs (Parker *et al.*, 2020b). Overall, these data not only reveal proteins closely
761 associated with FPA, but also provide the most in depth view of proteins involved in
762 Arabidopsis RNA 3′ end processing. As a result, this dataset provides clues to the function
763 of some of the uncharacterised proteins detected here.
764 Components of the m^6A writer complex co-purify with FPA. However, we found that
765 FPA did not affect global levels of this modification in Arabidopsis mRNAs. In contrast, FPA-
766 related spen family proteins in humans and Drosophila are required for m^6A deposition
767 (Knuckles *et al*, 2018; Patil *et al.*, 2016). Clearly, distinct spen protein functions have evolved
768 in different species.
769 The small size of FA enables extensive cross-linking but may not necessarily ensure
770 comprehensive recovery of protein assemblies. Likewise, depth of MS/MS coverage will
771 influence the sensitivity of protein detection. Consequently, technical reasons may explain
772 why we did not detect particular CPF components (e.g. CPSF30 and FIP1), or FLL2, which
773 was recently reported to mediate liquid phase transitions of the Arabidopsis 3′ end
774 processing machinery (Fang *et al*, 2019). In addition, we did not detect any members of the
775 PAF1 (RNAPII-associated factor 1) complex, which were first identified in Arabidopsis as
776 genetic suppressors of late flowering autonomous pathway mutants (Oh *et al*, 2004). An
777 alternative explanation for the absence of some proteins detected here might be that FPA
778 functions to exclude them as part of the mechanism by which it mediates poly(A) site choice.
779 Two Arabidopsis PCF11 paralogs with Pol II CTD-interacting domains (CIDs), PCFS2
780 and PCFS4, were found to co-purify with FPA, but two paralogs which lack CIDs, PCFS1 and
781 PCFS5, were not. PCF11 has been shown to have functionally separable roles in
782 transcription termination and cleavage and polyadenylation (Sadowski *et al*, 2003). The N-

783 terminal PCF11 CID is required for transcription termination, whilst the C-terminal domains
784 are required for cleavage and polyadenylation. The specific interaction of FPA with CID-
785 containing PCF11 paralogs may suggest that FPA regulates alternative polyadenylation by
786 altering Pol II speed and termination. Recent biochemical studies have demonstrated that
787 the human SPOC domain protein PHF3 is able to bind to two adjacent Ser^2 phosphorylated
788 heptads of the CTD of Pol II, using two electropositive patches on the surface of the SPOC
789 domain (Appel, Franke et al. 2020). One of these patches, and the key amino acid residues
790 within it, is conserved in the structure of the FPA SPOC domain (Zhang *et al.*, 2016).
791 Consequently, FPA might also be able to interact with the CTD, possibly in conjunction with
792 CID domains of PCFS2 and PCFS4. Such interactions could explain the specific co-
793 association of FPA and Pol II phosphorylated at Ser^2 of the CTD at the 3′ end of genes.
794

795 **FPA-dependent control of transcription termination and the regulatory and**
796 **evolutionary dynamics of NLRs**

797 An unanticipated finding of this study was that Arabidopsis NLR loci were the most enriched
798 group of FPA targets. NLRs function in the immune response of plants and animals, where
799 they sense the presence of pathogens or pathogen activity, for example by detecting
800 pathogen effectors (Jones *et al.*, 2016; Tamborski & Krasileva, 2020; Wu *et al.*, 2017).
801 Consistent with this crucial role in defence, NLR genes are under powerful and dynamic
802 selective pressure, and are subject to rapid birth and death. Defining the inventory of
803 Arabidopsis NLR genes recently depended on long range DNA sequencing of a diverse set
804 of accessions (Van de Weyer *et al.*, 2019). Here we show that long read nanopore DRS
805 enables not only new insight into the authentic complexity of NLR mRNA processing, but
806 also the accurate annotation of NLR genes. For example, our reannotation of the *RPS6* locus
807 provides a new framework to understand the recurring role of RPS6 in autoimmunity. The
808 lethality of null Arabidopsis NMD mutants and phenotypes of mutants defective in the
809 mitogen kinase activated pathway are RPS6-dependent, but the mechanisms involved are
810 unclear (Gloggnitzer *et al.*, 2014; Takagi *et al.*, 2020). We found that *RPS6* is transcribed
811 through a previously unrecognised downstream gene that encodes an RPS6-like TIR
812 domain. We show that the expression of the downstream gene is dependent on the RNA
813 exosome component HEN2. Mutations in *HEN2* were recently identified as suppressors of
814 RPS6-dependent autoimmune phenotypes (Takagi *et al.*, 2020). Consequently, accurate
815 annotation of complex NLR loci facilitates the interpretation of basic features of NLR
816 function.

817 Among the 124 NLRs with detectable expression in nanopore DRS data, we found
818 62 that were sensitive to FPA activity. FPA controls 3′ end formation of NLR genes in three
819 different transcript locations (Figure 9): (i) in 3′UTRs, where it can prevent readthrough and
820 chimeric RNA formation; (ii) in introns, where it promotes proximal polyadenylation; and
821 (iii) in protein-coding exons, where it promotes stop-codonless transcript formation. The
822 consequences of such complex control of RNA 3′ end formation are wide-ranging and likely

823 to be context-dependent (Mayr, 2019). For example, where FPA controls readthrough and
824 chimeric RNA formation, it affects 3'UTR length, sequence composition, and cryptic splice
825 site usage. Long or intron-containing 3'UTRs are targeted by the NMD machinery, leading
826 to RNA decay, suppressed translatability, and/or altered sub-cellular localisation. Long,
827 unstructured 3'UTRs influence intermolecular RNA interactions and phase separation, and
828 the close proximity of mRNAs in the resulting granules may enable co-translational protein
829 complex formation (Ma *et al*, 2020). Readthrough transcription may also disrupt expression
830 of downstream genes by transcription interference (Proudfoot, 1986). Premature
831 transcription termination at intronic poly(A) sites can introduce novel stop-codons, resulting
832 in transcripts that either limit the expression of full-length protein or which code for C-
833 terminally truncated NLR proteins with altered functions. TIR domain-only proteins, for
834 example, are known to be active in NLR regulation, resulting in constitutive signalling
835 activity (Zhang *et al*, 2004) or acting as competitive inhibitors by titrating full-length NLR
836 protein (Williams *et al*, 2014). The TE-containing 3'UTR of *RPP4* appears to be required for
837 resistance to the pathogen *Hpa*-Emoy2, although the mechanism involved is unclear (Wang
838 & Warren, 2010). We discovered that cryptic splicing of exon 1 of *RPP4* to a novel 3' splice
839 site within the TE can produce a unique transcript encoding only the RPP4 TIR domain. It
840 will be interesting to examine whether the TIR-only RPP4 isoform is required for full
841 pathogen resistance. We found that intronic and exonic proximal poly(A) site selection at
842 *RPS4* resulted in the formation of transcripts with the same protein coding potential as those
843 with retained introns that have previously been shown to be essential for RPS4-dependent
844 resistance against *P. syringae* DC3000 (Zhang & Gassmann, 2007). Therefore, a
845 combination of alternative polyadenylation and splicing underpins *RPS4* control. Finally,
846 premature transcription termination in introns and exons can result in the production of
847 stop-codonless transcripts that are unable to be translated into protein. These truncated
848 transcripts may be subject to decay by RNA surveillance mechanisms such as the NMD or
849 non-stop decay pathways, or act as non-coding RNA decoys titrating the levels of regulatory
850 miRNAs (Shivaprasad *et al.*, 2012). Increased rates of NLR transcription in plants under
851 pathogen attack could promote elongation through such "regulatory" poly(A) sites. In this
852 way, the expression of NLR proteins might be limited to low levels during pathogen
853 surveillance but kept poised for rapid activation during pathogen infection.

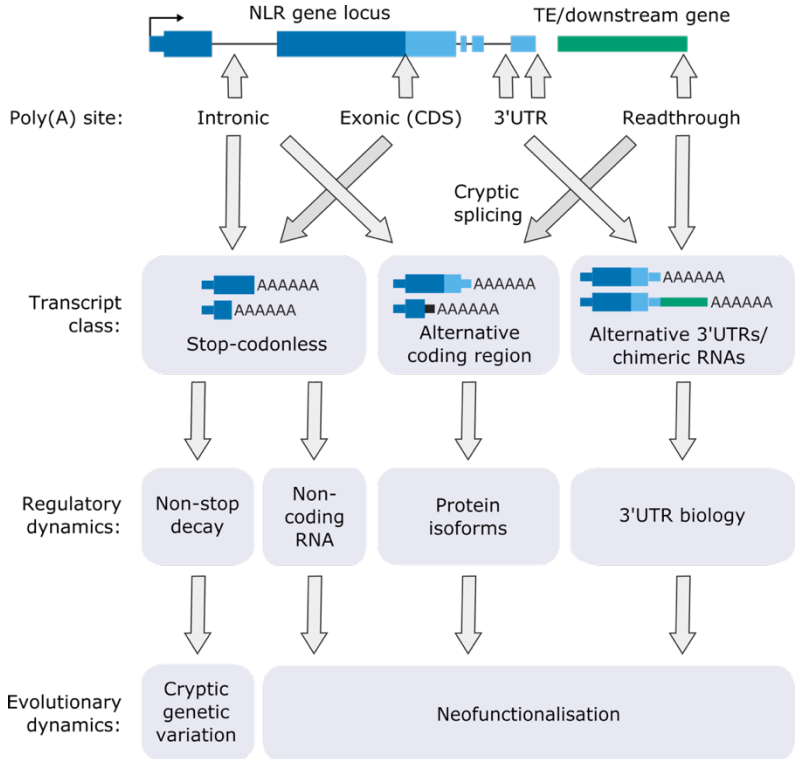


Figure 9: Functional consequences of FPA-dependent alternative polyadenylation at NLR loci
Model diagram showing how FPA-regulated alternative polyadenylation at NLR loci might affect the regulatory and evolutionary dynamics of plant disease resistance.

It is remarkable that FPA mostly targeted the protein-coding exons of NLR genes, and even controlled premature transcription termination within the open reading frame of single-exon NLR genes, like *RPP13*. Since the evolution of *cis*-regulatory elements controlling poly(A) site choice within introns or 3'UTRs is free from the constraints of protein-coding functionality, why are protein-coding exons targeted so frequently? One possibility is that this allows expression of newly created NLR genes to be kept under tight control, facilitating rapid evolution whilst reducing the chances of autoimmunity (Figure 9). Consequently, alternative polyadenylation might function to hide NLR genes from negative selection and contribute to cryptic genetic variation in a similar way to that proposed for NMD and miRNA-mediated NLR control (Raxwal & Riha, 2016; Shivaprasad *et al.*, 2012). The distinct consequences of alternative polyadenylation may facilitate the selective "trialling" of different NLR isoforms, such as TIR-domain-only encoding transcripts, or chimeric RNAs that incorporate sequence encoding integrated decoy domains (Le Roux *et al.*, 2015). Cryptically spliced chimeric RNAs, with subsequent retrotransposition, can be a source of new genes (Akiva *et al*, 2006). Therefore, the control of transcription termination may directly facilitate neofunctionalization of NLRs. The frequent occurrence of NLR genes in clusters may enable variation through unequal crossovers during recombination (Barragan & Weigel, 2020). In addition, as we show here, this feature of NLR genome organization is also sensitive to the formation of cryptically processed chimeric NLR RNAs.

879 We speculate that the diversity of NLR transcript processing might also facilitate pathogen
880 activity detection in a conceptually related way to integrated protein domain decoys but
881 occurring at the RNA level. For example, the formation of chimeric RNAs that comprise NLR
882 transcripts with extremely long 3'UTRs may facilitate the detection of pathogens targeting
883 NMD. Overall, it will be interesting to compare NLR patterns of cleavage, polyadenylation
884 and transcription termination across related Arabidopsis accessions and species to resolve
885 the contribution of controlled transcription termination to NLR evolutionary dynamics.
886 We focused on *RPP7* to understand complex NLR regulation. Although *RPP7*
887 expression has been well studied, our combinatorial sequencing analyses clarified the
888 complexity of processing events at the *RPP7* locus. Our data supports the established
889 pattern of alternative polyadenylation at intragenic heterochromatin within *RPP7* intron 1
890 (Tsuchiya & Eulgem, 2013). We revealed additional complexity that included FPA-
891 dependent promotion of cleavage and polyadenylation within protein coding exon 6. The
892 impact of FPA at exon 6 reduces the expression of full-length *RPP7* transcripts, with
893 functional consequences for strain-specific resistance to the oomycete pathogen *Hpa-*
894 Hiks1. Consequently, two distinct patterns of alternative polyadenylation mediate *RPP7*
895 regulation, one involving FPA, and another involving intragenic heterochromatin.
896 RPP7 is a recurring partner in examples of hybrid weakness, or necrosis (Bomblies
897 & Weigel, 2007). In these cases, autoimmunity is explained by allele-specific physical
898 interactions between RPP7 protein and the atypical disease resistance protein RPW8/HR (Li
899 *et al.*, 2020). We found that not only is *RPP7* pre-mRNA targeted by FPA-dependent
900 premature transcript termination, but so too is *RPW8/HR*. This raises the possibility that FPA
901 could "rescue" hybrid weakness by limiting the expression of these proteins. Consequently,
902 controlling NLR transcription termination may modulate the balance of costs and benefits
903 of NLR gene function.
904 The impact of FPA overexpression revealed here derives from artificial transgene
905 expression. However, pathogens could similarly attenuate NLR activity by evolving effectors
906 that modulate the expression or activity of factors controlling NLR poly(A) site choice.
907 Consistent with this idea, the HopU1 effector of *P. syringae* targets the RNA-binding
908 protein, AtGRP7 (Fu *et al*, 2007), which co-purified with FPA. In addition, the Pi4089 effector
909 of the oomycete pathogen *Phytopthora infestan*s targets the KH-domain RNA-binding
910 protein StKRBP1 in potato. As a result, the abundance of StKRBP1 increases, resulting in
911 enhanced infection by *P. infestans* (Wang *et al*, 2015). This precedent reveals that effector-
912 mediated increases in RNA-binding protein abundance can transform host RNA-binding
913 proteins into susceptibility factors. Phylogenetic analysis of StKRBP1 suggest that a direct
914 homolog is absent in Brassicacae but that the most related Arabidopsis proteins are FLK
915 and PEP, both of which co-purify with FPA and have been shown to regulate
916 polyadenylation (Rodriguez-Cazorla *et al.*, 2015). FPA, GRP7, FLK and PEP, along with other
917 RNA-binding proteins, act in concert to fine tune the timing of flowering through the
918 regulation of *FLC*. Similar RNA-binding protein-dependent modulation of NLR expression

may also explain how quantitative disease resistance can occur (Corwin & Kliebenstein, 2017).

New ways to analyse RNA processing

An essential feature of our study was the introduction of new approaches to study RNA processing and 3′ end formation. The use of long read nanopore DRS was transformative for our understanding of the complexity of NLR gene expression, providing insight that neither short read Illumina RNA-Seq of fragmented RNA nor Helicos DRS was capable of. We have recently shown that nanopore DRS maps RNA 3′ ends in close agreement with short-read Helicos DRS, and that nanopore DRS is not compromised by internal priming artefacts (Parker *et al.*, 2020b). Consequently, we used nanopore DRS to quantify alternative patterns of cleavage and polyadenylation. Here, we introduce a new approach to analyse alternative polyadenylation by applying the EMD metric. This approach improves the interpretation of changes in RNA processing by incorporating information on the scale of the effect and the sequence distance separating alternative poly(A) sites. This is valuable because larger distance changes between poly(A) sites are more likely to impact mRNA coding potential or trigger mRNA surveillance than subtle changes in the 3′UTR length.

The major advantage of nanopore DRS is that it can reveal multiple authentic features of individual transcripts in the context of the full-length RNA. Illumina RNA-Seq currently provides greater relative coverage and hence more statistical power. A limitation of some short-read RNA-Seq-based analyses of RNA processing is their dependence upon reference transcript annotations that may be incomplete. For example, in disease or mutant conditions, RNA processing often occurs at novel sites, that are not encountered in reference transcriptomes (as was the case here for NLR genes). In this study, we used the nanopore DRS data to generate bespoke reference transcriptomes for the genotypes under analysis, refining the value of the Illumina RNA-Seq data. This combined approach is likely to be generally useful in interpreting transcriptomes.

Concluding Remarks

It is difficult to identify alternative polyadenylation from conventional short read RNA-Seq data. As a result, the impact of alternative polyadenylation is probably under-reported. Here we show that premature transcription termination of NLR genes is widespread. This finding was made possible by recent progress in annotating this highly diverse gene family in Arabidopsis (Van de Weyer *et al.*, 2019), combined with our long-read sequencing of NLR transcripts in different genetic backgrounds with nanopore DRS. As a result, we could improve the accuracy of NLR annotation and reveal a layer of NLR gene regulation that may also influence the dynamics of NLR evolution. The continued development of approaches that reveal full-length native RNA molecules is likely to provide new insight into other important but previously intractable questions of biology.

959

960 # Materials and methods

961

962 **Plants**

963 *Plant material and growth conditions*

964 The wild type Col-0 accession and *fpa-7* were obtained from Nottingham Arabidopsis Stock
965 Centre. The *fpa-8* mutant (Col-0 background) and *35S::FPA:YFP* in *fpa-8* (Baurle *et al.*, 2007)
966 were provided by C. Dean (John Innes Centre). The surface-sterilised seeds were sown on
967 MS10 media plates supplemented with 2% agar, stratified at 4°C for two days, germinated
968 in a controlled environment at 20°C under 16h light/8h dark conditions, and harvested 14
969 days after transfer to 20°C.

970

971 ***IVI-MS***

972 *Preparation of IVI-MS libraries*

973 14-day-old seedlings were harvested and cross-linked with 1 % (v/v) formaldehyde for 15
974 min under vacuum. The cross-linking reaction was stopped by the addition of glycine to a
975 final concentration of 0.125 M under vacuum for a further 5 min. Nuclei were isolated from
976 frozen ground plant tissue using Honda buffer (20 mM Hepes-KOH pH 7.4, 10 mM $MgCl_2$,
977 440 mM sucrose, 1.25 % (w/v) Ficoll, 2.5 % (w/v) Dextran T40, 0.5 % (v/v) Triton X-100, 5
978 mM DTT, 1 mM PMSF, 1% (v/v) Protease Inhibitor Cocktail [Sigma]) and collected by
979 centrifugation at 2,000*g* for 17 min at 4°C. Nuclei were washed twice with Honda buffer
980 (centrifugation at 1,500*g* for 15 min at 4°C between washes) and lysed in nuclei lysis buffer
981 (50 mM Tris-HCl pH 8, 10 mM EDTA, 1 % (w/v) SDS, 1 mM PMSF, 1 % (v/v) Protease Inhibitor
982 Cocktail) by sonication for 4 cycles of 30 s pulses with low power and 60 s cooling between
983 pulses using the Bioruptor UCD-200 (Diagenode). Following centrifugation (16,100*g* for 10
984 min at 4°C), the supernatant was diluted ten-fold with sample dilution buffer (16.7 mM Tris-
985 HCl pH 8, 167 mM NaCl, 1.1 % (v/v) Triton X-100, 1 % (v/v) Protease Inhibitor Cocktail).
986 Cross-linked protein complexes were isolated with GFP-trap agarose beads (Chromotek),
987 incubated at 4°C with constant rotation for 5 h, followed by centrifugation (141*g* for 3 min
988 at 4°C). Beads were washed three times with washing buffer (150 mM NaCl, 20 mM Tris-
989 HCl pH 8, 2 mM EDTA pH 8, 1 % (v/v) Triton X-100, 0.1 % (w/v) SDS, 1 mM PMSF) by
990 centrifugations between washes (400*g* for 3 min at 4°C). Samples were incubated at 90°C
991 for 30 min to reverse cross-linking prior to SDS-PAGE. Each biological replicate was
992 separated into 5 fractions following SDS-PAGE, then subjected to in-gel digestion with
993 trypsin and submitted for LC-MS/MS analysis (LTQ Orbitrap Velos Pro mass spectrometer
994 (ThermoFisher Scientific)). Three biological replicates were performed for each genotype.

995

996 *IVI-MS data analysis*

997 Raw peptide data files from mass spectrometry were analyzed by MaxQuant software
998 (version 1.6.10.43) (Cox & Mann, 2008). Peptide tables were then loaded using Proteus
999 (version 0.2.14) (Gierlinski *et al*, 2018) and summarised to protein level counts using the hi-
1000 flyer method (mean of the top 3 most abundant peptides). Because wild-type plants lacking
1001 GFP were used as controls, a large number of proteins that are enriched by
1002 immunoprecipitation are below the detection threshold in the control. This group of
1003 proteins can be classified as "missing not at random" (MNAR). In all proteomics
1004 experiments, there will also be a number of proteins which are not detected purely by
1005 chance, and these are referred to as "missing at random" (MAR). We treated proteins that
1006 were missing from all replicates of a condition as MNAR, and proteins that were missing
1007 moved only from a subset of replicates as MAR. Using the imputeLCMD package (version
1008 2.0) (Lazar, 2015), a K nearest neighbours' strategy was used to impute MAR examples, and
1009 a quantile regression imputation of left centred data (QRILC) approach was used to impute
1010 MNAR examples. Differential expression analysis was performed on imputed data using
1011 limma (version 3.40.0) (Ritchie *et al*, 2015). Because imputation is not deterministic (i.e. will
1012 lead to different outcomes every time), we improved the robustness of the analysis by
1013 performing 999 bootstraps of the imputation and differential expression, and summarising
1014 the results using the median log fold change and harmonic mean p value.
1015
1016 **ChIP-Seq**
1017 *Preparation of libraries for ChIP-Seq*
1018 ChIP against FPA and Pol II phosphorylated at either Ser^5 or Ser^2 of the CTD heptad repeat
1019 was performed as previously described (Yu *et al.*, 2019). Polyclonal antibodies against FPA
1020 amino acids 536-901 were rasied in rabbit by Covance.
1021
1022 *ChIP-Seq data processing*
1023 FPA and PolII ChIP-Seq data is available at ENA accession PRJNA449914. $H3K9me^2$ ChIP-
1024 seq data was downloaded from ENA accessions PRJDB5192 (Inagaki *et al.*, 2017) and
1025 PRJNA427432 (Lai *et al.*, 2020). Reads were aligned to the TAIR10 reference genome using
1026 Bowtie2 (version 2.3.5.1) (Langmead & Salzberg, 2012), with the parameters --mm --very-
1027 sensitive --maxins 800 --no-mixed --no-discordant. Counts per million normalised coverage
1028 profiles were generated using deeptools (version 3.4.3) (Ramirez *et al*, 2014). For 3' end
1029 centred metagene profiles, we determined the major 3' position per gene using the
1030 Araport11 annotation and existing Col-0 Helicos DRS data (Sherstnev *et al.*, 2012).
1031 Metagenes centred on these positions were then generated in Python 3.6 using pyBigWig
1032 (version 0.3.17) (Ramirez *et al.*, 2014), numpy (version 1.18.1) (Harris *et al*, 2020) and
1033 matplotlib (version 3.1.3) (Hunter, 2007). For differential $H3K9me^2$ analysis, read counts per
1034 gene (including intronic regions) were generated using pysam (version 0.16.0), and
1035 differential expression analysis was performed using edgeR (version 3.22.5) (Robinson *et*
1036 *al*, 2010).

1037
1038 **RNA**
1039 *Total RNA isolation*
1040 Following manufacturers' instructions, total RNA was isolated using RNeasy Plant Mini kit
1041 (QIAGEN) and treated with TURBO DNase (ThermoFisher Scientific). The total RNA
1042 concentration was measured using a Qubit 1.0 Fluorometer and Qubit RNA BR Assay Kit
1043 (ThermoFisher Scientific), while RNA quality and integrity was assessed using the
1044 NanoDrop™ 2000 spectrophotometer (ThermoFisher Scientific) and Agilent 2200
1045 TapeStation System (Agilent).
1046
1047 **Nanopore DRS**
1048 *Preparation of libraries for direct RNA sequencing using nanopores*
1049 Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as detailed above.
1050 mRNA isolation and preparation of nanopore DRS libraries (using the SQK-RNA001
1051 nanopore DRS Kit (Oxford Nanopore Technologies) were performed as previously
1052 described (Parker *et al.*, 2020b). Libraries were loaded onto R9.4 SpotON Flow Cells
1053 (Oxford Nanopore Technologies) and sequenced using a 48-h runtime. Four biological
1054 replicates were performed for each genotype.
1055
1056 *Nanopore DRS data processing*
1057 Nanopore DRS reads were basecalled using Guppy (version 3.6.0) high accuracy model.
1058 Reads were mapped to the Arabidopsis *TAIR10* genome (The Arabidopsis Genome, 2000)
1059 using minimap2 (version 2.17) with the parameters -a -L –cs=short -x splice -G20000 –end-
1060 seed-pen 12 –junc-bonus 12 -uf. Spliced alignment was guided using junctions from the
1061 Araport11 annotation (Cheng *et al*, 2017). Nanopore DRS reads can suffer from
1062 "oversplitting" – where signal originating from a single RNA molecule is incorrectly
1063 interpreted as two or more reads (Parker *et al.*, 2020b). These errors can be systematic and
1064 can result in false positive 3' ends. To filter these errors, we identified reads which were
1065 sequenced consecutively through the same pore and also mapped contiguously on the
1066 genome (within 1 kb of each other). We therefore filtered all but the most 3' reads, which
1067 should contain the genuine RNA 3' end. Pipelines for processing of nanopore DRS data
1068 were built using Snakemake (Koster & Rahmann, 2012).
1069
1070 **Helicos DRS**
1071 *Preparation of samples for Helicos DRS*
1072 Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as specified
1073 above. Next, samples were processed by Helicos BioSciences as previously described
1074 (Ozsolak *et al.*, 2009; Sherstnev *et al.*, 2012). Three biological replicates were performed
1075 for each genotype.
1076

1077 *Helicos DRS data processing*
1078 Helicos DRS reads were mapped to the *Arabidopsis* TAIR10 genome using Heliosphere
1079 (version 1.1.498.63) as previously described (Sherstnev *et al.*, 2012). Reads were filtered to
1080 remove those with >4 nt insertions or deletions, and to mask regions with low complexity
1081 as determined using dustmasker (Camacho *et al*, 2009) (from BLAST+ suite version 2.10.1)
1082 set at DUST level 15 (Sherstnev *et al.*, 2012)
1083
1084 **Differential 3′ end analysis of Nanopore and Helicos DRS datasets**
1085 Transcriptional loci were first discovered in Col-0, *fpa*-8 and *35S::FPA:YFP* nanopore DRS
1086 reads using the long-read transcript assembly tool StringTie2 version 2.1.1 (Pertea *et al.*,
1087 2015). Novel transcriptional loci were merged with annotated loci from the Araport11
1088 reference (Cheng *et al.*, 2017). To detect sites with altered 3′ end distributions in *fpa-8* and
1089 *35S::FPA:YFP* mutants, we pooled the replicates of either nanopore or Helicos DRS data
1090 and identified reads overlapping each transcriptional locus. These reads were used to build
1091 distributions of 3′ end locations. The difference in 3′ end distributions between the
1092 treatment and control (Col-0) was measured using the earth mover distance (EMD, also
1093 known as the Wasserstein distance). To identify loci with statistically significant changes in
1094 3′ distributions, we performed an EMD permutation test using 999 bootstraps – reads for
1095 each locus were randomly shuffled between the control and treatment samples to create
1096 null distributions, and the EMD recalculated. The histogram of null EMDs was fitted using a
1097 gamma distribution, and the *p*-value was calculated from the distribution (the probability
1098 of achieving the observed EMD or greater by chance). *p*-values were corrected for multiple
1099 testing using the Benjamini-Hochberg method. Genes with an EMD ≥ 25 and an FDR < 0.05
1100 were considered differentially alternatively polyadenylated, and directionality of change
1101 was identified using the difference in mean 3′ position. Software developed to perform
1102 differential 3′ analysis is available on GitHub at https://github.com/bartongroup/d3pendr
1103 and 10.5281/zenodo.4319113, and can be used with nanopore DRS, Helicos DRS, or
1104 Illumina 3′ tag-based datasets.
1105
1106 **Illumina RNA sequencing**
1107 *Preparation of libraries for Illumina RNA sequencing*
1108 Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as described
1109 above. mRNA isolation and preparation of the sequencing libraries were performed using
1110 the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) by
1111 the Centre for Genomic Research (University of Liverpool). 150-bp paired-end sequencing
1112 was carried out on Illumina HiSeq 4000. Six biological replicates were performed for each
1113 genotype.
1114
1115 *Illumina RNA sequencing data processing*

1116 Illumina RNA-Seq data was assessed for quality using FastQC (version 0.11.9) and MultiQC
1117 (version 1.8) (Andrews, 2010; Ewels *et al*, 2016). Reads were mapped to the TAIR10
1118 genome using STAR (version 2.7.3a) (Dobin *et al*, 2013) with a splice junction database
1119 generated from the Araport11 reference annotation (Cheng *et al.*, 2017). Counts per million
1120 normalised coverage tracks were created using samtools (version 1.10) and deeptools
1121 (version 3.4.3) (Ramirez *et al.*, 2014). To identify expressed regions in each locus, coverage
1122 profiles for each replicate of control and treatment were first extracted using pyBigWig
1123 (version 0.3.17) (Ramirez *et al.*, 2014). These were normalised so that the area under each
1124 profile was equal to the mean area under the profiles. A normalised coverage threshold of
1125 1 was used to identify expressed regions of the loci. These regions were further segmented
1126 where there were changes in expression between control and treatment conditions of at
1127 least two-fold in a 25 nt window. Regions which were smaller than 50 nt were removed.
1128 Expression of the regions was then calculated using featureCounts (version 2.0.0) (Liao *et*
1129 *al*, 2013) and differential usage within transcriptional loci was assessed using DEXseq
1130 (version 1.32.0) (Reyes *et al*, 2013). Loci were considered differentially processed if they
1131 had a locus-level FDR < 0.05 and at least one region with an absolute logFC > 1 and FDR <
1132 0.05. To conduct differential splice junction usage analysis, counts of splice junctions
1133 annotated in the bespoke nanopore DRS-derived annotation, plus Araport11 annotation,
1134 were generated for each locus using featureCounts (version 2.0.0) (Liao *et al.*, 2013).
1135 Differential splice junction usage was assessed using DEXseq (version 1.32.0) (Reyes *et al.*,
1136 2013). Loci were considered differentially spliced if they had a locus-level FDR < 0.05 and
1137 at least one junction with an absolute logFC > 1 and FDR < 0.05.
1138
1139 **Gene tracks**
1140 Gene track figures were generated in Python 3.6 using matplotlib (version 3.1.3) (Hunter,
1141 2007). For gene tracks where any condition had >200 nanopore DRS read alignments, 200
1142 representative alignments were selected by random sampling without replacement (with
1143 the exception of the *FPA* figure, where 500 read alignments were sampled). nanoPARE data
1144 (Schon *et al.*, 2018) was processed as described in (Parker *et al.*, 2020b).
1145
1146 **Protein domain family enrichment analysis**
1147 To conduct protein domain enrichment analysis, Interpro domain annotations of
1148 Arabidopsis proteins were downloaded from bioMart (Smedley *et al*, 2009) and converted
1149 to genomic co-ordinates using the Araport11 annotation (Cheng *et al.*, 2017). Domain
1150 families overlapping each locus tested for alternative polyadenylation using either
1151 nanopore or Helicos DRS were identified using pybedtools (version 0.8.1) (Dale *et al*, 2011).
1152 To identify enriched domain families, domains were randomly shuffled between tested loci
1153 in 19,999 bootstraps, and the number of times each domain class overlapped significantly
1154 alternatively polyadenylated loci by chance was recorded. This was compared to the

1155 observed overlap of each domain family with alternatively polyadenylated loci to calculate
1156 *p* values, which were corrected for multiple testing using the Benjamini Hochberg method.
1157
1158 **Manual annotation of alternatively polyadenylated NLR genes**
1159 To identify which of the 206 previously annotated NLR genes present in Col-0 were
1160 alternatively polyadenylated in *fpa-8* and *35S::FPA*, we devised a standard operating
1161 procedure for visual inspection. Genes which had nanopore DRS read coverage in at least
1162 two conditions were considered expressed. Genes were considered alternatively
1163 polyadenylated if they had multiple 3′ end locations each supported by at least 4 nanopore
1164 DRS reads, and if there was a clear change in nanopore DRS coverage in the treatment
1165 condition compared to Col-0.
1166
1167 **RNA gel blot analysis of *RPP7* mRNAs**
1168 RNA gel blot analysis was carried out as previously described (Quesada *et al*, 2003) with
1169 minor modifications. *RPP7* mRNA was detected using a probe annealing to the second
1170 exon of *RPP7* (*AT1G58602*) gene (200 bp PCR product amplified with the following primers:
1171 Forward: 5′-TCGGGGACTACTACTACTCAAGA-3′ and Reverse: 5′-
1172 TCTTGATGGTGTGAAAGAATCTAGT-3′). As a loading control, *Beta-TUBULIN* mRNA was
1173 used, which was visualised by a probe annealing to the third exon of the *Beta-TUBULIN*
1174 *(AT1G20010) gene* (550 bp PCR product amplified with the following primers: Forward: 5′-
1175 CTGACCTCAGGAAACTCGCG-3′ and Reverse: 5′- CATCAGCAGTAGCATCTTGG-3′). The
1176 probes were 5′ labelled using [γ-^{32}P]-ATP (Perkin Elmer) and DECAprime™ II DNA labelling
1177 kit (ThermoFisher Scientific) and purified on Illustra G-50 columns (GE Healthcare Life
1178 Sciences). mRNA isoforms were visualized and quantified using Amersham Typhoon Gel
1179 and Blot Imaging System (GE Healthcare Bio-Sciences AB). The RiboRuler High Range RNA
1180 Ladder (ThermoFisher Scientific), used to identify approximate size of RNA bands, was first
1181 dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher
1182 Scientific) and then labelled with [γ-^{32}P]-ATP (Perkin Elmer) using T4 Polynucleotide Kinase
1183 (ThermoFisher Scientific) prior being loaded on a gel.
1184
1185 **m^{6}A liquid chromatography–mass spectroscopy analysis (m^{6}A LC-MS/MS)**
1186 Total RNA was isolated and checked as described above. mRNA was extracted twice from
1187 approximately 75 µg of total RNA using the Dynabeads mRNA purification kit (Thermo
1188 Fisher Scientific) following the manufacturer's instructions. The quality and quantity of
1189 mRNA was assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher
1190 Scientific) and Agilent 2200 TapeStation System (Agilent). Samples for the m^{6}A LC-MS/MS
1191 were prepared as previously described (Huang *et al*, 2018) with several modifications. 100
1192 ng of mRNA were diluted in a total volume of 14 ml of nuclease-free water (ThermoFisher
1193 Scientific) and digested by nuclease P1 (1U, Merck) in 25 µl of buffer containing 20 mM of
1194 NH_4OAc (pH 5.3) at 42°C for 2 h. Next, 3 µl of freshly made 1M NH_4HCO_3 and Calf Intestinal

1195 Alkaline Phosphatase (1U, New England Biolabs) were added, and samples were incubated
1196 at 37°C for 2 h. Then, the samples were diluted to 50 µl with nuclease-free water and filtered
1197 (0.22 µm pore size, 4 mm diameter, Millipore). LC-MS/MS was carried out by the
1198 FingerPrints Proteomics facility at the University of Dundee. m^6A to A ratio quantification
1199 was performed in comparison with the curves obtained from pure A and m^6A nucleoside
1200 standards. Statistical analysis was performed using a 2-way t-test.
1201
1202 **Pathogenesis assays**
1203 The pathogenesis assays were carried out as previously described (Tome *et al*, 2014). The
1204 *Hpa*-Hiks1 isolate was maintained by weekly sub-culturing on the Ksk-1 plants. A solution
1205 containing *Hpa*-Hiks1 spores was used to inoculate the Col-0, Ksk-1, *fpa-7, fpa-8, pFPA::FPA*
1206 and *35S::FPA:YFP* 14-day-old seedlings. Sporangiophores were counted four days after
1207 inoculation. Three biological replicates were performed for each genotype using up to 45
1208 plants per line per each repeat. Statistical analysis was performed using a 2-way t-test.
1209
1210 **Code availability**
1211 All pipelines, scripts and notebooks used to generate figures are available from GitHub at
1212 https://github.com/bartongroup/Simpson_Barton_FPA_NLRs and Zenodo at
1213 10.5281/zenodo.4319109. The software tool developed for detecting changes in poly(A)
1214 site choice in Nanopore and Helicos DRS data is available from GitHub at
1215 https://github.com/bartongroup/d3pendr and Zenodo at 10.5281/zenodo.4319113.
1216
1217 **Data availability**
1218 *IVI-MS* data is available from the ProteomeXchange Consortium via the PRIDE (Perez-
1219 Riverol *et al*, 2019) partner repository with the dataset identifier PXD022684. FPA and Pol II
1220 ChIP-Seq data is available from ENA accession PRJNA449914 (Yu *et al.*, 2019). Col-0
1221 nanopore DRS data is available from ENA accession PRJEB32782 (Parker *et al.*, 2020b). *fpa-*
1222 *8* and *35S::FPA:YFP* nanopore DRS data is available from ENA accession PRJEB41451.
1223 *hen2-2* nanopore DRS data is available from ENA accession PRJEB41381. Col-0, *fpa-8* and
1224 *35S::FPA:YFP* Helicos DRS data is available from Zenodo DOI 10.5281/zenodo.4309752
1225 ahead of submission to ENA. Col-0, *fpa-8* and *35S::FPA:YFP* Illumina RNA-Seq data is
1226 available from ENA accession PRJEB41455.
1227
1228 **Acknowledgments**
1229 We thank Paul Birch and Ingo Hein for comments on the manuscript, and David Baulcombe,
1230 Ian Henderson and Wenbo Ma for helpful NLR discussions. We thank Dr Abdelmadjid Atrih
1231 (Centre for Advanced Scientific Technologies, School of Life Sciences) for m^6A LC-MS/MS
1232 analysis.
1233
1234 **Competing Interests**

The authors declare no competing interests.

Funding

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

References

Adachi H, Derevnina L, Kamoun S (2019) NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr Opin Plant Biol* 50: 121-131

Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, Novik A, Sorek R (2006) Transcription-mediated gene fusion in the human genome. *Genome Res* 16: 30-6

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* 301: 653-7

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22: 2008-17

Andres F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet* 13: 627-39

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Ariyoshi M, Schwabe JW (2003) A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev* 17: 1909-20

Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat Genet* 36: 162-166

Bach-Pages M, Homma F, Kourelis J, Kaschani F, Mohammed S, Kaiser M, van der Hoorn RAL, Castello A, Preston GM (2020) Discovering the RNA-Binding Proteome of Plant Leaves with an Improved RNA Interactome Capture Method. *Biomolecules* 10

Barragan AC, Weigel D (2020) Plant NLR Diversity: The Known Unknowns of Pan-NLRomes. *The Plant Cell*

Baurle I, Smith L, Baulcombe DC, Dean C (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science* 318: 109-12

Bazin J, Romero N, Rigo R, Charon C, Blein T, Ariel F, Crespi M (2018) Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses. *Front Plant Sci* 9: 1209

1278 Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier
1279 in plant species. *Nat Rev Genet* 8: 382-93
1280 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009)
1281 BLAST+: architecture and applications. *BMC bioinformatics* 10: 421-421
1282 Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017)
1283 Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The*
1284 *Plant Journal* 89: 789-804
1285 Collado-Torres L, Nellore A, Frazee AC, Wilks C, Love MI, Langmead B, Irizarry RA, Leek JT,
1286 Jaffe AE (2017) Flexible expressed region analysis for RNA-seq with derfinder. *Nucleic Acids*
1287 *Res* 45: e9
1288 Corwin JA, Kliebenstein DJ (2017) Quantitative Resistance: More Than Just Perception of a
1289 Pathogen. *Plant Cell* 29: 655-665
1290 Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized
1291 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:
1292 1367-72
1293 Dai G-Y, Chen D-K, Sun Y-P, Liang W-Y, Liu Y, Huang L-Q, Li Y-K, He J-F, Yao N (2020) The
1294 Arabidopsis KH-domain protein FLOWERING LOCUS Y delays flowering by upregulating
1295 FLOWERING LOCUS C family members. *Plant Cell Reports* 39: 1705-1717
1296 Dale RK, Pedersen BS, Quinlan AR (2011) Pybedtools: a flexible Python library for
1297 manipulating genomic datasets and annotations. *Bioinformatics* 27: 3423-4
1298 Deremetz A, Le Roux C, Idir Y, Brousse C, Agorio A, Gy I, Parker JE, Bouche N (2019)
1299 Antagonistic Actions of FPA and IBM2 Regulate Transcript Processing from Genes Containing
1300 Heterochromatin. *Plant Physiol* 180: 392-403
1301 Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their
1302 possible role in tobacco mosaic virus resistance. *Proc Natl Acad Sci U S A* 97: 1908-13
1303 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
1304 TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21
1305 Duc C, Sherstnev A, Cole C, Barton GJ, Simpson GG (2013) Transcription termination and
1306 chimeric RNA formation controlled by *Arabidopsis thaliana* FPA. *PLoS Genet* 9: e1003867
1307 Eom H, Park SJ, Kim MK, Kim H, Kang H, Lee I (2018) TAF15b, involved in the autonomous
1308 pathway for flowering, represses transcription of FLOWERING LOCUS C. *Plant J* 93: 79-91
1309 Ewels P, Magnusson M, Käller M, Lundin S (2016) MultiQC: summarize analysis results for
1310 multiple tools and samples in a single report. *Bioinformatics* 32: 3047-3048
1311 Fang X, Wang L, Ishikawa R, Li Y, Fiedler M, Liu F, Calder G, Rowan B, Weigel D, Li P, Dean C
1312 (2019) Arabidopsis FLL2 promotes liquid-liquid phase separation of polyadenylation
1313 complexes. *Nature* 569: 265-269
1314 Fang X, Wu Z, Raitskin O, Webb K, Voigt P, Lu T, Howard M, Dean C (2020) The 3' processing
1315 of antisense RNAs physically links to chromatin-based transcriptional control. *Proc Natl Acad*
1316 *Sci U S A* 117: 15316-15321
1317 Fu ZQ, Guo M, Jeong BR, Tian F, Elthon TE, Cerny RL, Staiger D, Alfano JR (2007) A type III
1318 effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature* 447: 284-8
1319 Gierlinski M, Gastaldello F, Cole C, Barton GJ (2018)
1320 Gloggnitzer J, Akimcheva S, Srinivasan A, Kusenda B, Riehs N, Stampfl H, Bautor J, Dekrout B,
1321 Jonak C, Jimenez-Gomez JM, Parker JE, Riha K (2014) Nonsense-mediated mRNA decay
1322 modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host*
1323 *Microbe* 16: 376-90

1324 Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E,
1325 Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A,
1326 Del Rio JF, Wiebe M, Peterson P, Gerard-Marchant P et al. (2020) Array programming with
1327 NumPy. *Nature* 585: 357-362
1328 Hornyik C, Terzi LC, Simpson GG (2010) The spen family protein FPA controls alternative
1329 cleavage and polyadenylation of RNA. *Dev Cell* 18: 203-13
1330 Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, Zhao BS, Mesquita A, Liu C, Yuan CL, Hu Y-C,
1331 Hüttelmaier S, Skibbe JR, Su R, Deng X, Dong L, Sun M, Li C, Nachtergaele S, Wang Y et al.
1332 (2018) Recognition of RNA N^6-methyladenosine by IGF2BP proteins enhances mRNA stability
1333 and translation. *Nature Cell Biology* 20: 285-295
1334 Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science &*
1335 *Engineering* 9: 90-95
1336 Inagaki S, Takahashi M, Hosaka A, Ito T, Toyoda A, Fujiyama A, Tarutani Y, Kakutani T (2017)
1337 Gene-body chromatin modification dynamics mediate epigenome differentiation in
1338 Arabidopsis. *EMBO J* 36: 970-980
1339 Jones JD, Vance RE, Dangl JL (2016) Intracellular innate immune surveillance devices in
1340 plants and animals. *Science* 354
1341 Knuckles P, Lence T, Haussmann IU, Jacob D, Kreim N, Carl SH, Masiello I, Hares T, Villasenor
1342 R, Hess D, Andrade-Navarro MA, Biggiogera M, Helm M, Soller M, Buhler M, Roignant JY
1343 (2018) Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding
1344 factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Gene Dev* 32: 415-
1345 429
1346 Koiwa H, Hausmann S, Bang WY, Ueda A, Kondo N, Hiraguri A, Fukuhara T, Bahk JD, Yun DJ,
1347 Bressan RA, Hasegawa PM, Shuman S (2004) Arabidopsis C-terminal domain phosphatase-
1348 like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases. *Proc Natl Acad Sci*
1349 *U S A* 101: 14539-44
1350 Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA
1351 polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14:
1352 2452-60
1353 Koornneef M, Hanhart CJ, van der Veen JH (1991) A genetic and physiological analysis of late
1354 flowering mutants in Arabidopsis thaliana. *Mol Gen Genet* 229: 57-66
1355 Koster J, Rahmann S (2012) Snakemake--a scalable bioinformatics workflow engine.
1356 *Bioinformatics* 28: 2520-2522
1357 Krzyszton M, Zakrzewska-Placzek M, Kwasnik A, Dojer N, Karlowski W, Kufel J (2018)
1358 Defective XRN3-mediated transcription termination in Arabidopsis affects the expression of
1359 protein-coding genes. *Plant J* 93: 1017-1031
1360 Kyburz A, Friedlein A, Langen H, Keller W (2006) Direct interactions between subunits of
1361 CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and
1362 splicing. *Mol Cell* 23: 195-205
1363 Lai Y, Eulgem T (2018) Transcript-level expression control of plant NLR genes. *Mol Plant*
1364 *Pathol* 19: 1267-1281
1365 Lai Y, Lu XM, Daron J, Pan S, Wang J, Wang W, Tsuchiya T, Holub E, McDowell JM, Slotkin RK,
1366 Le Roch KG, Eulgem T (2020) The Arabidopsis PHD-finger protein EDM2 has multiple roles in
1367 balancing NLR immune receptor gene expression. *PLoS Genet* 16: e1008993
1368 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:
1369 357-9

Lazar C (2015) imputeLCMD: a collection of methods for left-censored missing data imputation. *R package, version* 2

Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, Zhou B, Levaillant M, Adachi H, Yoshioka H, Raffaele S, Berthome R, Coute Y, Parker JE, Deslandes L (2015) A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161: 1074-1088

Li L, Habring A, Wang K, Weigel D (2020) Atypical Resistance Protein RPW8/HR Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity. *Cell Host Microbe*

Liao Y, Smyth GK, Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41: e108-e108

Liu F, Quesada V, Crevillen P, Baurle I, Swiezewski S, Dean C (2007) The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Mol Cell* 28: 398-407

Liu FQ, Marquardt S, Lister C, Swiezewski S, Dean C (2010) Targeted 3 ' Processing of Antisense Transcripts Triggers Arabidopsis FLC Chromatin Silencing. *Science* 327: 94-97

Ma W, Zhen G, Xie W, Mayr C (2020)

Mayr C (2019) What Are 3' UTRs Doing? *Cold Spring Harb Perspect Biol* 11

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47: D351-D360

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* 28: 1078-86

Mockler TC, Yu X, Shalitin D, Parikh D, Michael TP, Liou J, Huang J, Smith Z, Alonso JM, Ecker JR, Chory J, Lin C (2004) Regulation of flowering time in Arabidopsis by K homology domain proteins. *Proc Natl Acad Sci U S A* 101: 12759-64

Oh S, Zhang H, Ludwig P, van Nocker S (2004) A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the Arabidopsis FLC/MAF MADS box gene family. *Plant Cell* 16: 2940-53

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM (2009) Direct RNA sequencing. *Nature* 461: 814-8

Parker MT, Barton GJ, Simpson GG (2020a)

Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, Hall AJ, Barton GJ, Simpson GG (2020b) Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m(6)A modification. *Elife* 9

Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537: 369-373

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

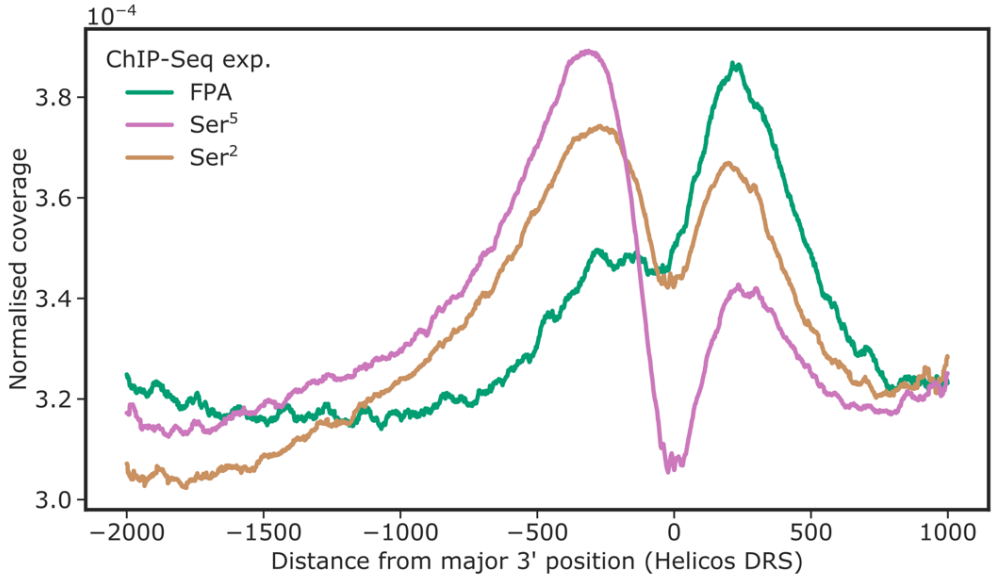
Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33: 290-5

1416 Proudfoot NJ (1986) Transcriptional interference and termination between duplicated
1417 alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature* 322:
1418 562-5
1419 Quesada V, Macknight R, Dean C, Simpson GG (2003) Autoregulation of *FCA* pre-mRNA
1420 processing controls *Arabidopsis* flowering time. *The EMBO Journal* 22: 3142-3152
1421 Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T (2014) deepTools: a flexible platform for
1422 exploring deep-sequencing data. *Nucleic Acids Res* 42: W187-91
1423 Raxwal VK, Riha K (2016) Nonsense mediated RNA decay and evolutionary capacitance.
1424 *Biochim Biophys Acta* 1859: 1538-1543
1425 Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume AM, Horos R, Hentze MW, Preiss T,
1426 Millar AA (2016) *In planta* determination of the mRNA-binding proteome of Arabidopsis
1427 etiolated seedlings. *Plant Cell* 28: 2435-2452
1428 Reyes A, Anders S, Weatheritt RJ, Gibson TJ, Steinmetz LM, Huber W (2013) Drift and
1429 conservation of differential exon usage across tissues in primate species. *Proc Natl Acad Sci*
1430 *U S A* 110: 15377-82
1431 Rigal M, Kevei Z, Pelissier T, Mathieu O (2012) DNA methylation in an intron of the *IBM1*
1432 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* 31: 2981-93
1433 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers
1434 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*
1435 *Res* 43: e47
1436 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential
1437 expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140
1438 Rodriguez E, El Ghoul H, Mundy J, Petersen M (2016) Making sense of plant autoimmunity
1439 and 'negative regulators'. *FEBS J* 283: 1385-91
1440 Rodriguez-Cazorla E, Ripoll JJ, Andujar A, Bailey LJ, Martinez-Laborda A, Yanofsky MF, Vera A
1441 (2015) K-homology nuclear ribonucleoproteins regulate floral organ identity and
1442 determinacy in arabidopsis. *PLoS Genet* 11: e1004983
1443 Ruzicka K, Zhang M, Campilho A, Bodi Z, Kashif M, Saleh M, Eeckhout D, El-Showk S, Li H,
1444 Zhong S, De Jaeger G, Mongan NP, Hejatko J, Helariutta Y, Fray RG (2017) Identification of
1445 factors required for m^6A mRNA methylation in Arabidopsis reveals a role for the conserved
1446 E3 ubiquitin ligase HAKAI. *New Phytol* 215: 157-172
1447 Sadowski M, Dichtl B, Hubner W, Keller W (2003) Independent functions of yeast Pcf11p in
1448 pre-mRNA 3' end processing and in transcription termination. *EMBO J* 22: 2167-77
1449 Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, Ito T, Kakutani T (2013) Mechanism
1450 for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin.
1451 *Nature Communications* 4
1452 Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC
1453 domain-containing protein in Arabidopsis thaliana. *Science* 319: 462-5
1454 Schomburg FM, Patton DA, Meinke DW, Amasino RM (2001) FPA, a gene involved in floral
1455 induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. *Plant Cell*
1456 13: 1427-36
1457 Schon MA, Kellner MJ, Plotnikova A, Hofmann F, Nodine MD (2018) NanoPARE: parallel
1458 analysis of RNA 5' ends from low-input RNA. *Genome Res* 28: 1931-1942
1459 Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton GJ, Simpson
1460 GG (2012) Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and
1461 polyadenylation. *Nat Struct Mol Biol* 19: 845-52

1462 Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC (2012) A microRNA
1463 superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant*
1464 *Cell* 24: 859-74
1465 Simpson GG (2004) The autonomous pathway: epigenetic and post-transcriptional gene
1466 regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol* 7: 570-574
1467 Simpson GG, Dijkwel PP, Quesada V, Henderson I, Dean C (2003) FY is an RNA 3' end-
1468 processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell*
1469 113: 777-87
1470 Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A (2009)
1471 BioMart--biological queries made easy. *BMC Genomics* 10: 22
1472 Streitner C, Danisman S, Wehrle F, Schoning JC, Alfano JR, Staiger D (2008) The small
1473 glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana.
1474 *Plant Journal* 56: 239-250
1475 Takagi M, Iwamoto N, Kubo Y, Morimoto T, Takagi H, Takahashi F, Nishiuchi T, Tanaka K, Taji
1476 T, Kaminaka H, Shinozaki K, Akimitsu K, Terauchi R, Shirasu K, Ichimura K (2020) Arabidopsis
1477 SMN2/HEN2, Encoding DEAD-Box RNA Helicase, Governs Proper Expression of the
1478 Resistance Gene SMN1/RPS6 and Is Involved in Dwarf, Autoimmune Phenotypes of mekk1
1479 and mpk4 Mutants. *Plant Cell Physiol* 61: 1507-1516
1480 Tamborski J, Krasileva KV (2020) Evolution of Plant NLRs: From Natural History to Precise
1481 Modifications. *Annu Rev Plant Biol* 71: 355-378
1482 The Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant
1483 *Arabidopsis thaliana*. *Nature* 408: 796-815
1484 Tian B, Manley JL (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell*
1485 *Biol* 18: 18-30
1486 Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated
1487 resistance in Arabidopsis thaliana. *Nature* 423: 74-7
1488 Tome DF, Steinbrenner J, Beynon JL (2014) A growth quantification assay for
1489 Hyaloperonospora arabidopsidis isolates in Arabidopsis thaliana. *Methods Mol Biol* 1127:
1490 145-58
1491 Tran Vdu T, Souiai O, Romero-Barrios N, Crespi M, Gautheret D (2016) Detection of generic
1492 differential RNA processing events from RNA-seq data. *RNA Biol* 13: 59-67
1493 Tsuchiya T, Eulgem T (2013) An alternative polyadenylation mechanism coopted to the
1494 Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci*
1495 *U S A* 110: E3535-43
1496 Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl
1497 JL, Weigel D, Bemm F (2019) A Species-Wide Inventory of NLR Genes and Alleles in
1498 Arabidopsis thaliana. *Cell* 178: 1260-1272 e14
1499 van der Biezen EA, Freddie CT, Kahn K, Parker JE, Jones JD (2002) Arabidopsis RPP4 is a
1500 member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew
1501 resistance through multiple signalling components. *Plant J* 29: 439-51
1502 Wang X, Boevink P, McLellan H, Armstrong M, Bukharova T, Qin Z, Birch PR (2015) A Host KH
1503 RNA-Binding Protein Is a Susceptibility Factor Targeted by an RXLR Effector to Promote Late
1504 Blight Disease. *Mol Plant* 8: 1385-95
1505 Wang YH, Warren JT, Jr. (2010) Mutations in retrotransposon AtCOPIA4 compromises
1506 resistance to Hyaloperonospora parasitica in Arabidopsis thaliana. *Genet Mol Biol* 33: 135-
1507 40

1508 Wang ZW, Wu Z, Raitskin O, Sun Q, Dean C (2014) Antisense-mediated FLC transcriptional
1509 repression requires the P-TEFb transcription elongation factor. *Proc Natl Acad Sci U S A* 111:
1510 7468-73
1511 Wei C, Chen J, Kuang H (2016) Dramatic Number Variation of R Genes in Solanaceae Species
1512 Accounted for by a Few R Gene Subfamilies. *Plos One* 11: e0148708
1513 Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, Ve T, Ma Y, Saucet SB,
1514 Ericsson DJ, Casey LW, Lonhienne T, Winzor DJ, Zhang X, Coerdt A, Parker JE, Dodds PN,
1515 Kobe B, Jones JD (2014) Structural basis for assembly and function of a heterodimeric plant
1516 immune receptor. *Science* 344: 299-303
1517 Wu CH, Abd-El-Haliem A, Bozkurt TO, Belhaj K, Terauchi R, Vossen JH, Kamoun S (2017) NLR
1518 network mediates immunity to diverse plant pathogens. *Proc Natl Acad Sci U S A* 114: 8113-
1519 8118
1520 Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W (2014) Dynamic
1521 analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across
1522 seven tumour types. *Nat Commun* 5: 5274
1523 Xing DH, Zhao HW, Xu RQ, Li QSQ (2008) Arabidopsis PCFS4, a homologue of yeast
1524 polyadenylation factor Pcf11p, regulates FCA alternative processing and promotes flowering
1525 time. *Plant Journal* 54: 899-910
1526 Yan Z, Shi H, Liu Y, Jing M, Han Y (2019) KHZ1 and KHZ2, novel members of the autonomous
1527 pathway, repress the splicing efficiency of FLC pre-mRNA in Arabidopsis. *J Exp Bot*
1528 Yang S, Tang F, Gao M, Krishnan HB, Zhu H (2010) R gene-controlled host specificity in the
1529 legume-rhizobia symbiosis. *Proc Natl Acad Sci U S A* 107: 18735-40
1530 Yu XH, Martin PGP, Michaels SD (2019) BORDER proteins protect expression of neighboring
1531 genes by promoting 3 ' Pol II pausing in plants. *Nature Communications* 10
1532 Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA,
1533 Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC (2011) MicroRNAs as
1534 master regulators of the plant NB-LRR defense gene family via the production of phased,
1535 trans-acting siRNAs. *Genes Dev* 25: 2540-53
1536 Zhang XC, Gassmann W (2007) Alternative splicing and mRNA levels of the disease
1537 resistance gene RPS4 are induced during defense responses. *Plant Physiol* 145: 1577-87
1538 Zhang Y, Dorey S, Swiderski M, Jones JD (2004) Expression of RPS4 in tobacco induces an
1539 AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *Plant J* 40: 213-24
1540 Zhang Y, Gu L, Hou Y, Wang L, Deng X, Hang R, Chen D, Zhang X, Zhang Y, Liu C, Cao X (2015)
1541 Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates
1542 plant flowering by targeting alternative polyadenylation. *Cell Res* 25: 864-76
1543 Zhang Y, Li Z, Chen N, Huang Y, Huang S (2020) Phase separation of Arabidopsis EMB1579
1544 controls transcription, mRNA splicing, and development. *Plos Biol* 18: e3000782
1545 Zhang Y, Rataj K, Simpson GG, Tong L (2016) Crystal Structure of the SPOC Domain of the
1546 Arabidopsis Flowering Regulator FPA. *Plos One* 11: e0160694
1547 Zhou JX, Liu ZW, Li YQ, Li L, Wang B, Chen S, He XJ (2018) Arabidopsis PWWP domain
1548 proteins mediate H3K27 trimethylation on FLC and regulate flowering time. *J Integr Plant*
1549 *Biol* 60: 362-368
1550
1551
1552

1553
1554 ## Supplementary Figures
1555



1556
1557 **Figure 2 - supplement 1: FPA co-localises with Pol II Ser^2 at the 3′ end of genes**
1558 **(A)** ChIP-Seq metagene profile showing the normalised occupancy of FPA (green) and Pol
1559 II phosphorylated at Ser^5 (pink) and Ser^2 (brown) of the CTD relative to the major 3′ position
1560 of each gene as measured using Helicos DRS. Only short genes (<2.5 kb) are included
1561 (n=17,440). The figure was compiled from the same ChIP data used in Figure 2.
1562

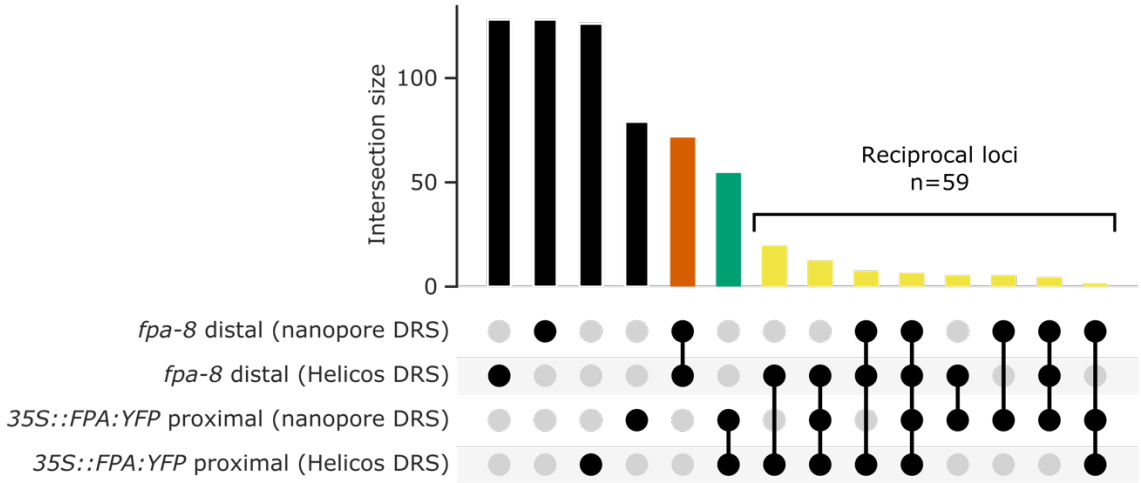


Figure 4 – supplement 1: Nanopore and Helicos DRS reveal FPA-dependent RNA 3' end processing changes

(A) Comparison of RNA 3' ends identified in nanopore and Helicos DRS datasets in *fpa-8* and *35S::FPA:YFP* (compared to Col-0). Bar size represents the number of alternatively polyadenylated loci common to an intersection highlighted using circles below. Bars indicating loci which 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 which are identified as reciprocally regulated by FPA (distal polyadenylation in *fpa-8,* proximal in *35S::FPA:YFP*) by at least one technique, in yellow.

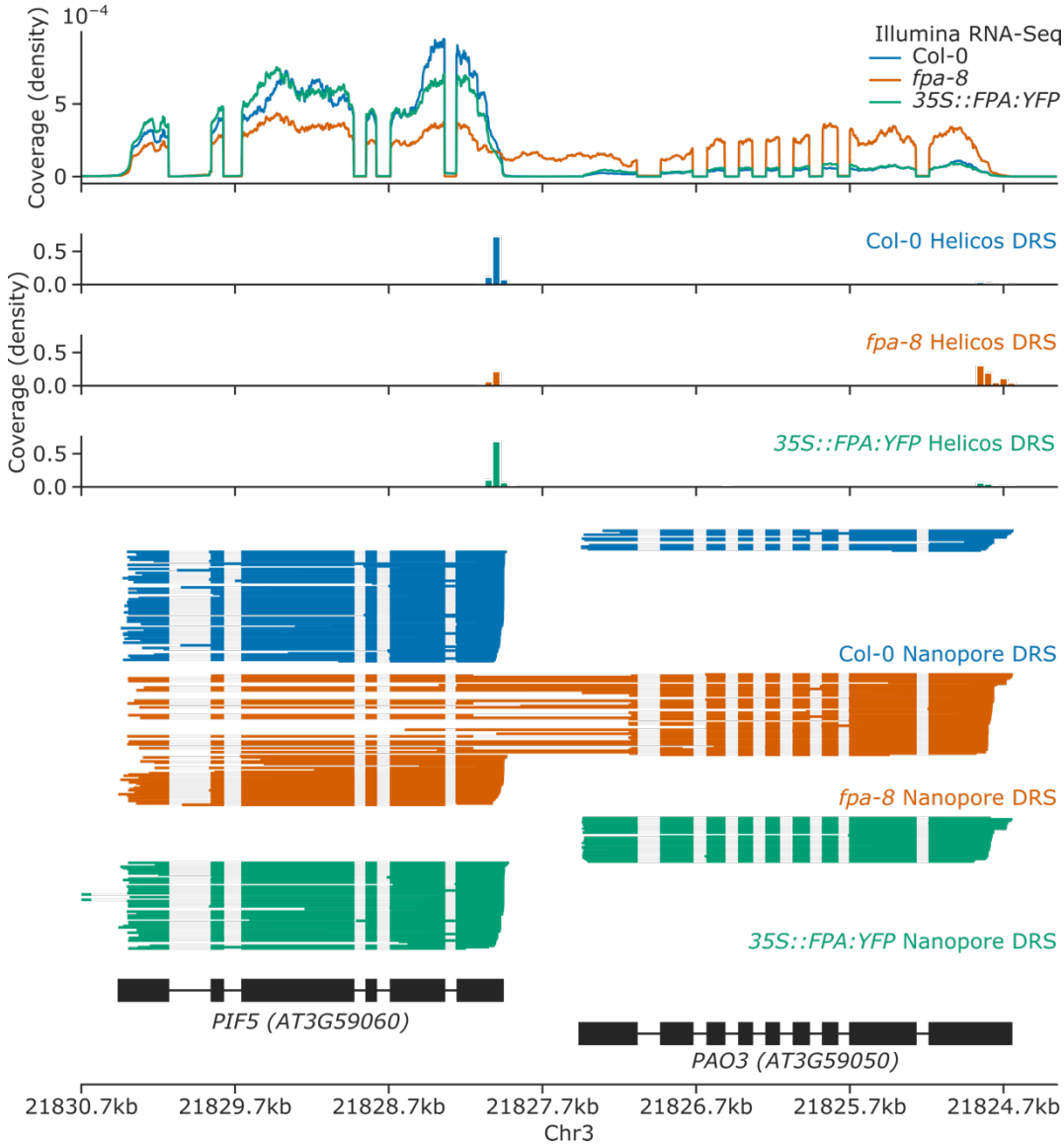
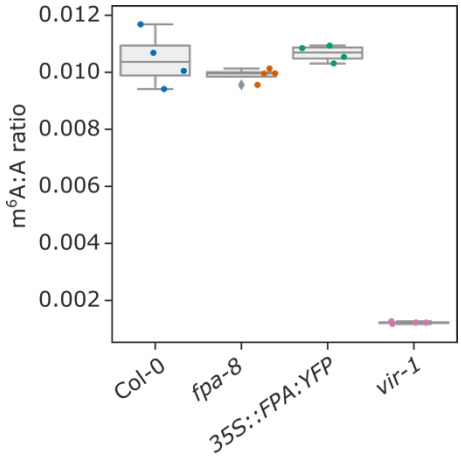


Figure 4 – supplement 2: Splicing alterations in *fpa-8* can be explained by changes in RNA 3' end formation

Chimeric RNA formation at the *PIF5* gene locus. Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.



1585

1586 **Figure 4 – supplement 3: FPA does not affect global mRNA m^{6}A methylation level**

1587 Box plot showing m^{6}A:A ratio analysed using LC-MS/MS. Col-0 is shown in blue; *fpa-8*, in

1588 orange; *35S::FPA:YFP*, in green; low m^{6}A mutant *vir-1* (used as a positive control), in pink.

1589

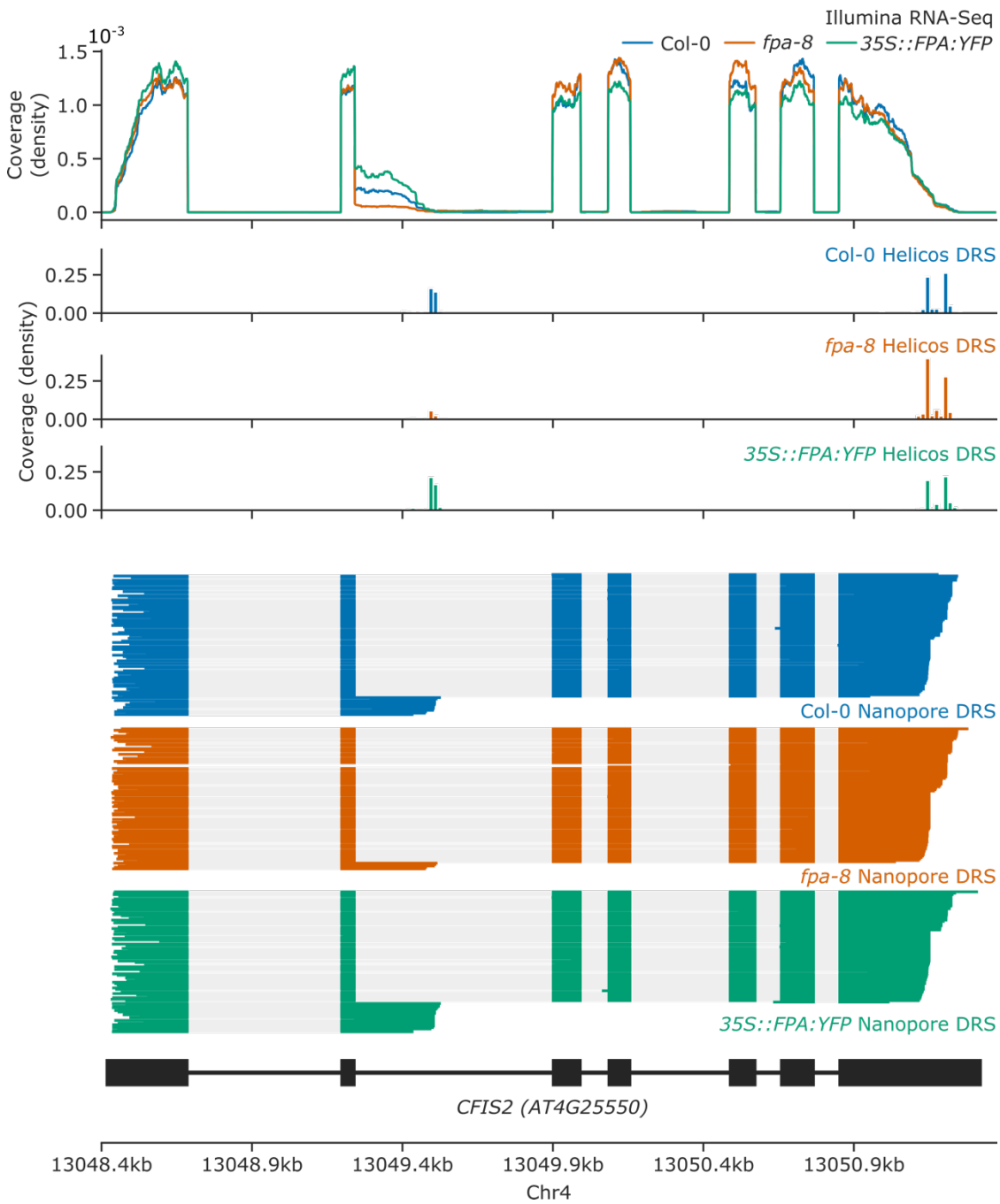


Figure 4 – supplement 4: FPA regulates RNA 3' end processing at *CFIS2* (*AT4G2550)* locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

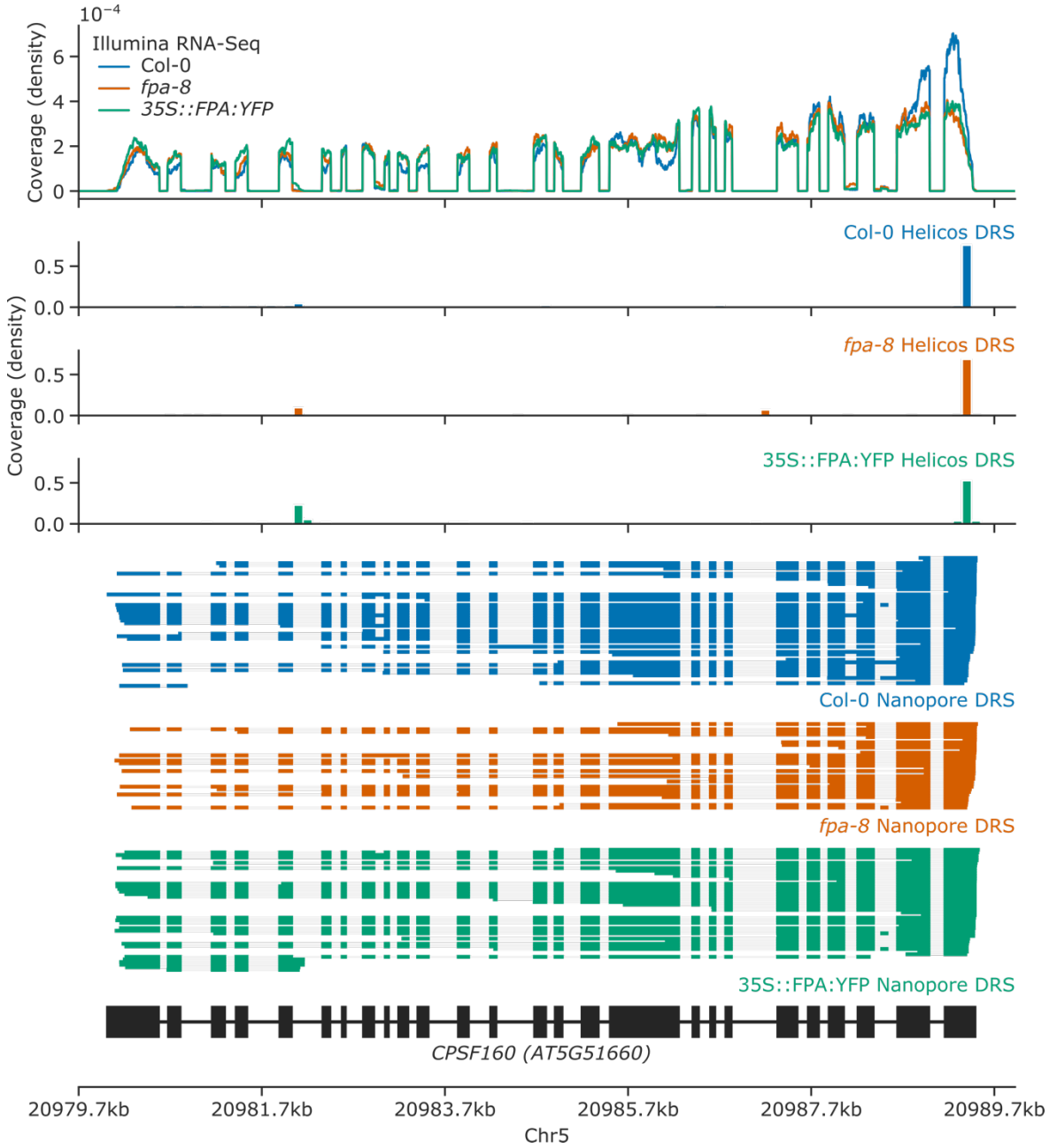


Figure 4 – supplement 5: FPA regulates RNA 3' end processing at *CPSF160* (*AT5G51660*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

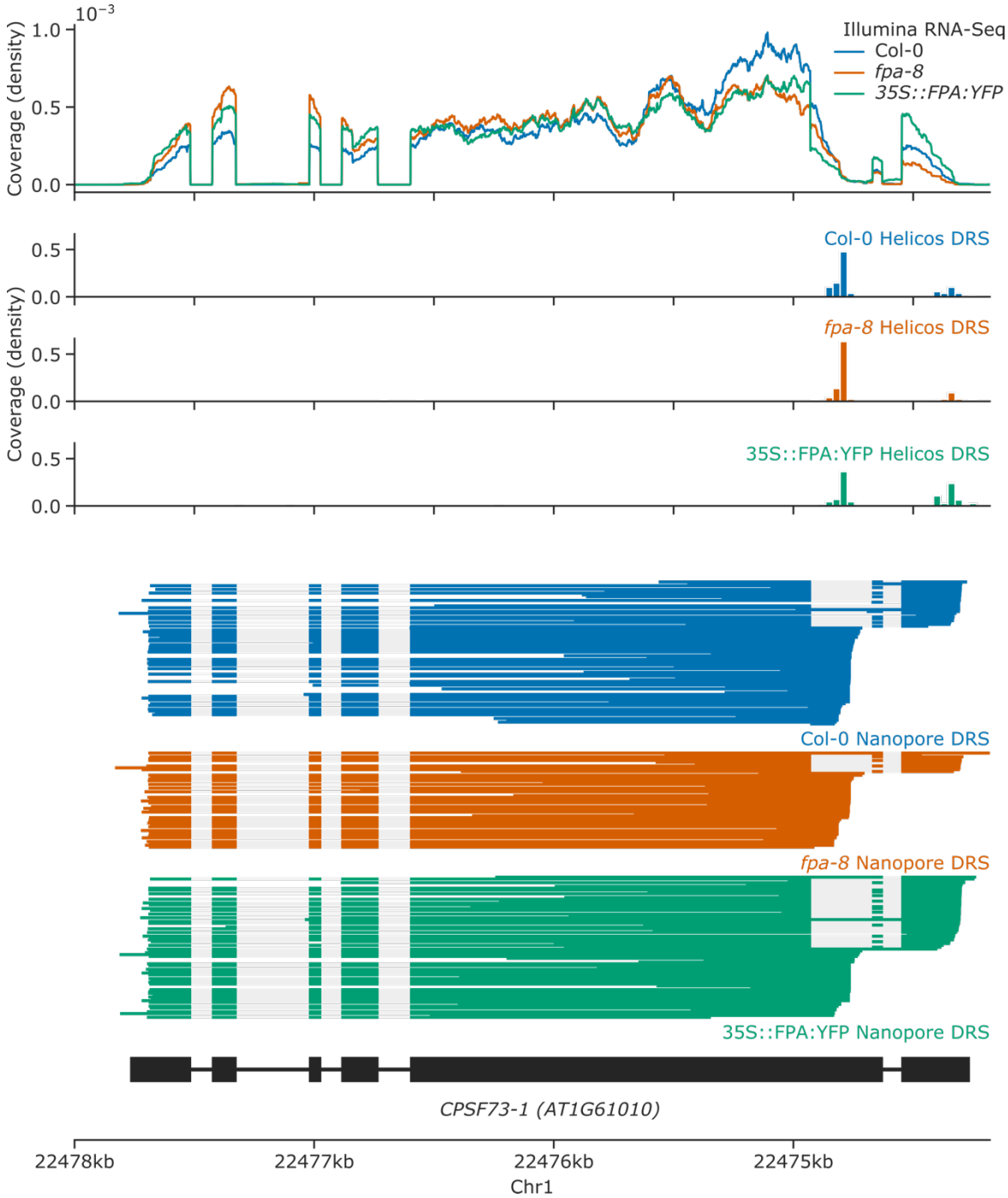


Figure 4 – supplement 6: FPA regulates RNA 3′ end processing at *CPSF73-1* (*AT1G61010*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

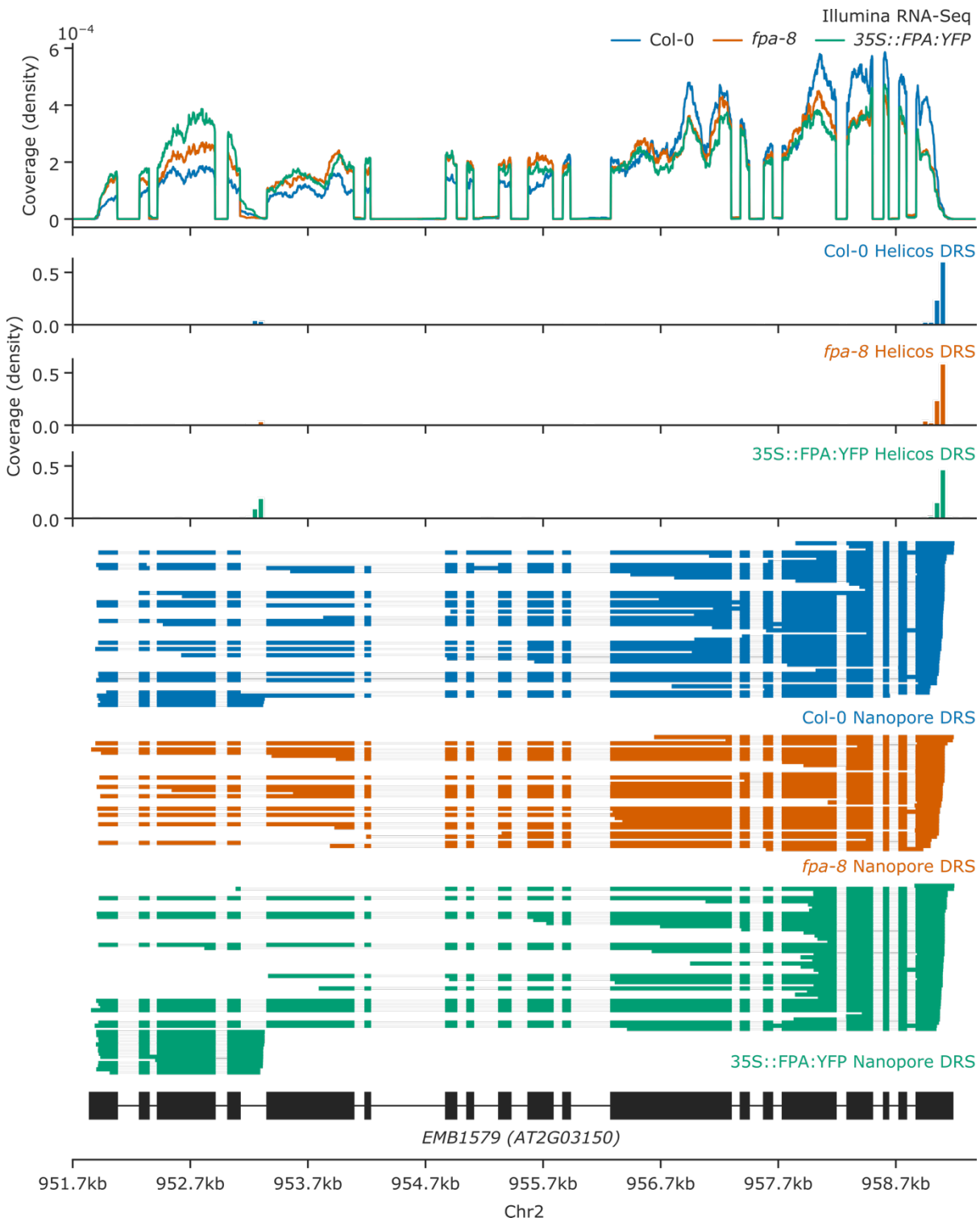


Figure 4 – supplement 7: FPA regulates RNA 3' end processing at *EMB1579/RSA1 (AT2G03150)* locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

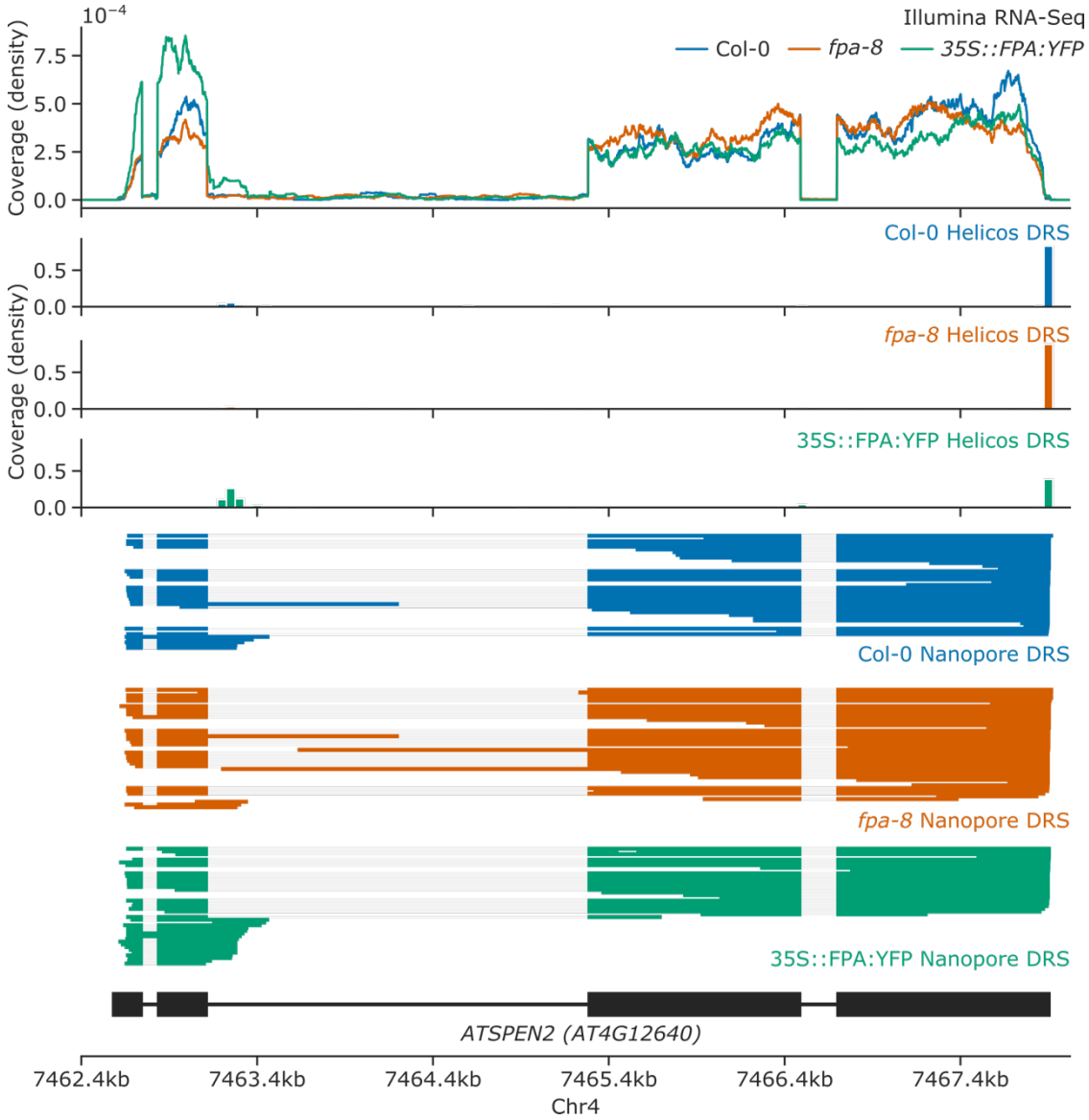


Figure 4 – supplement 8: FPA regulates RNA 3' end processing at *ATSPEN2* (*AT4G12640*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

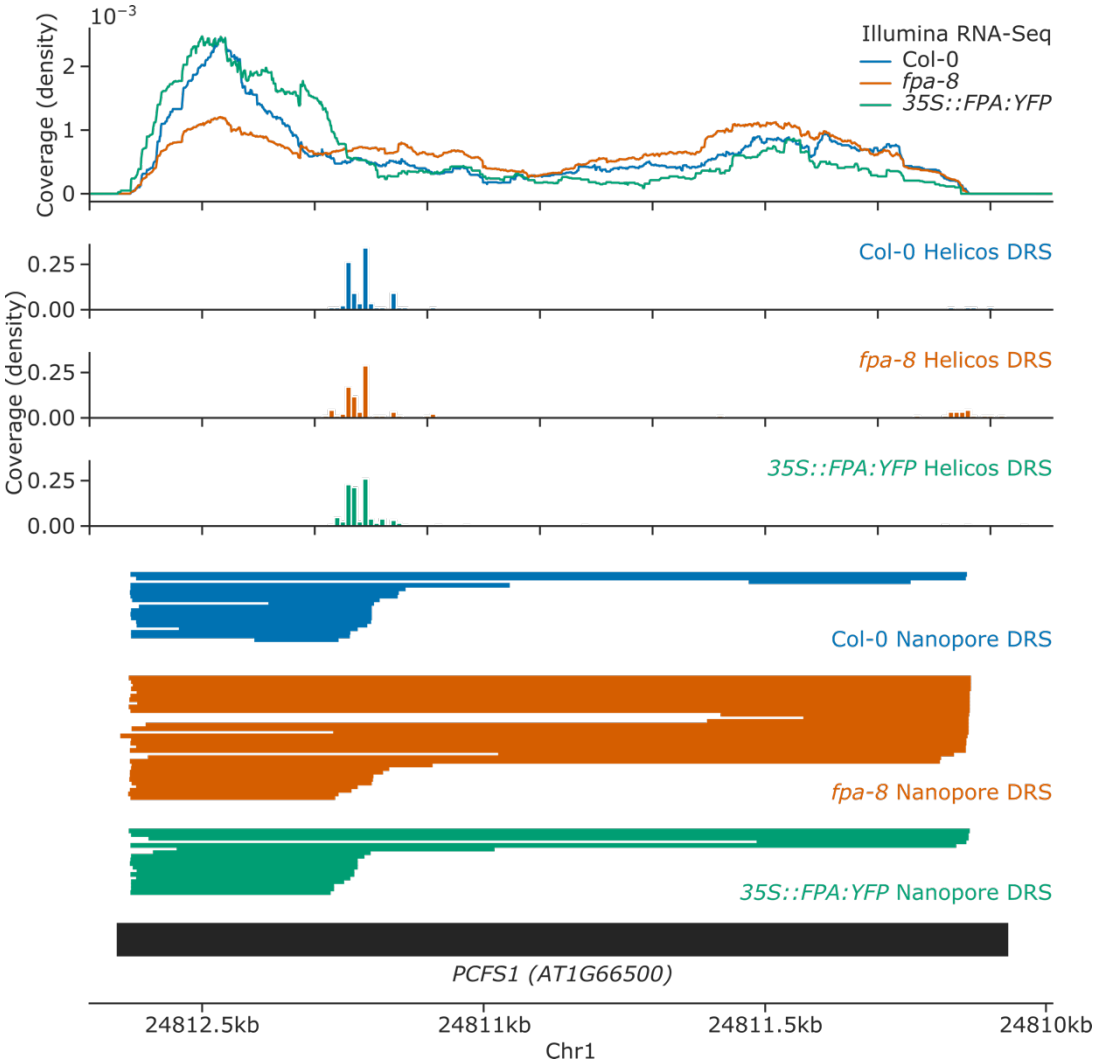


Figure 4 – supplement 9: FPA regulates RNA 3′ end processing at the single exon *PCFS1* (*AT1G66500*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

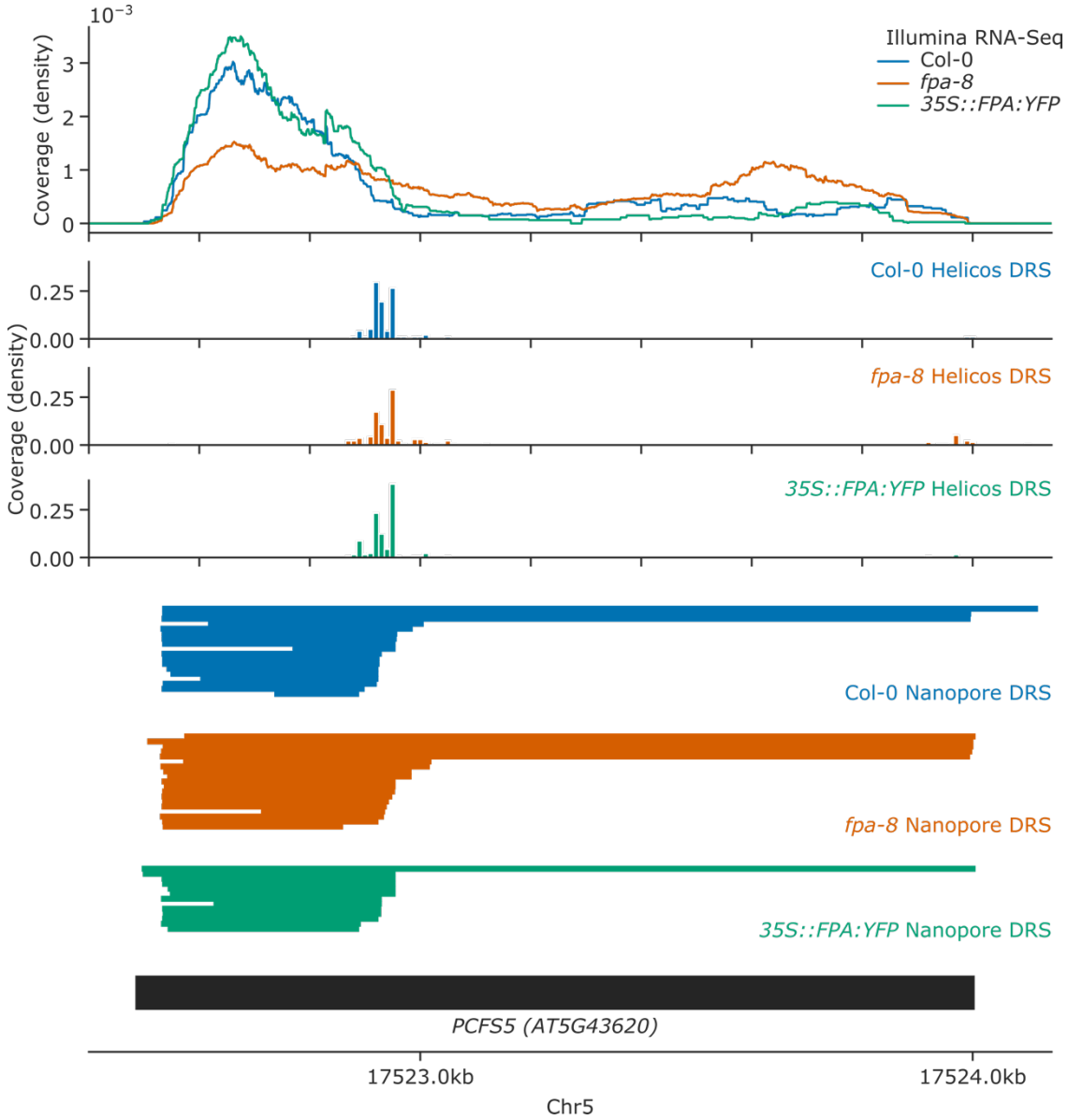


Figure 4 – supplement 10: FPA regulates RNA 3' end processing at the single exon *PCFS5* (*AT5G43620*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

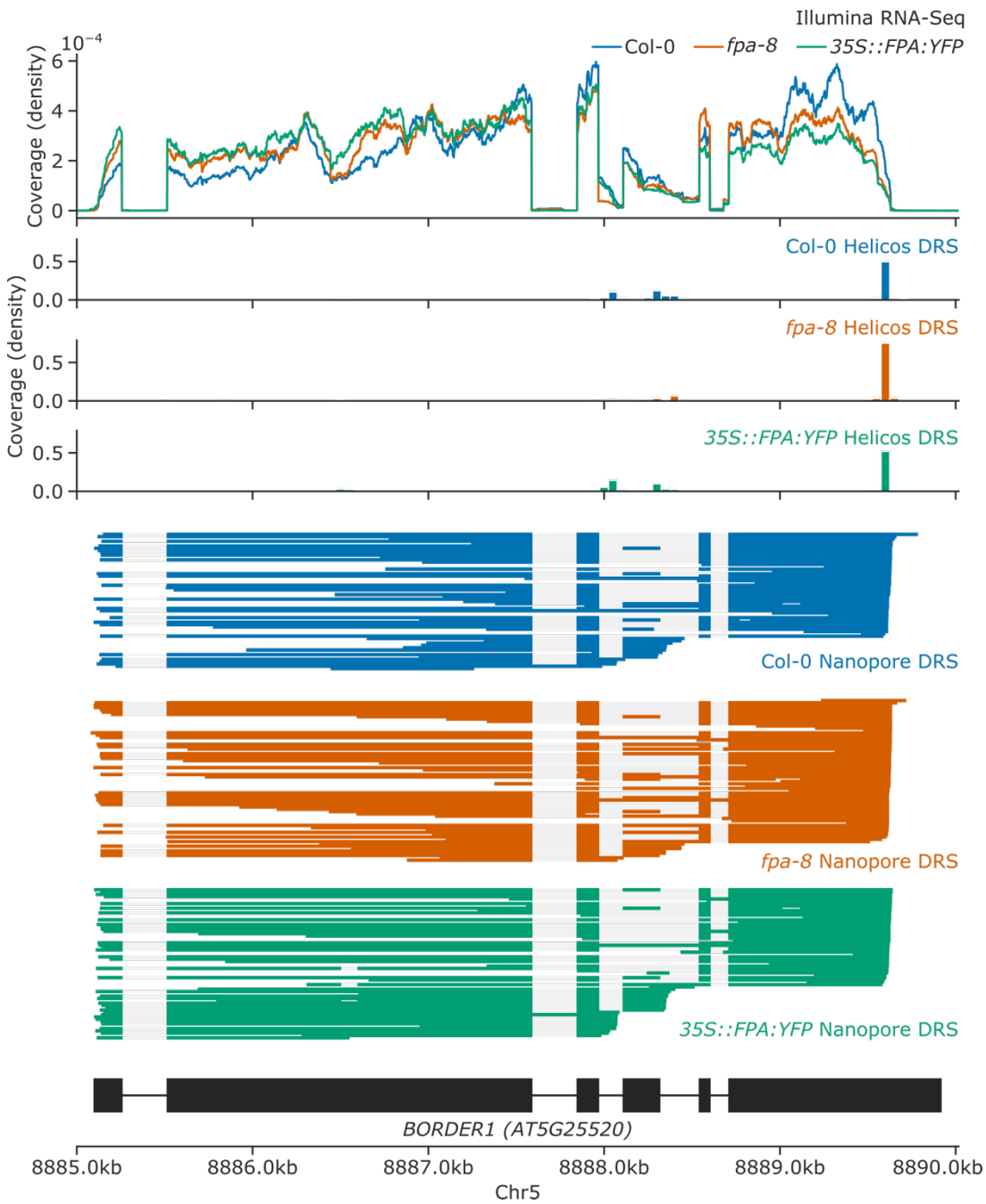


Figure 4 – supplement 11: FPA regulates RNA 3' end processing at *BORDER1* (*AT5G25520*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

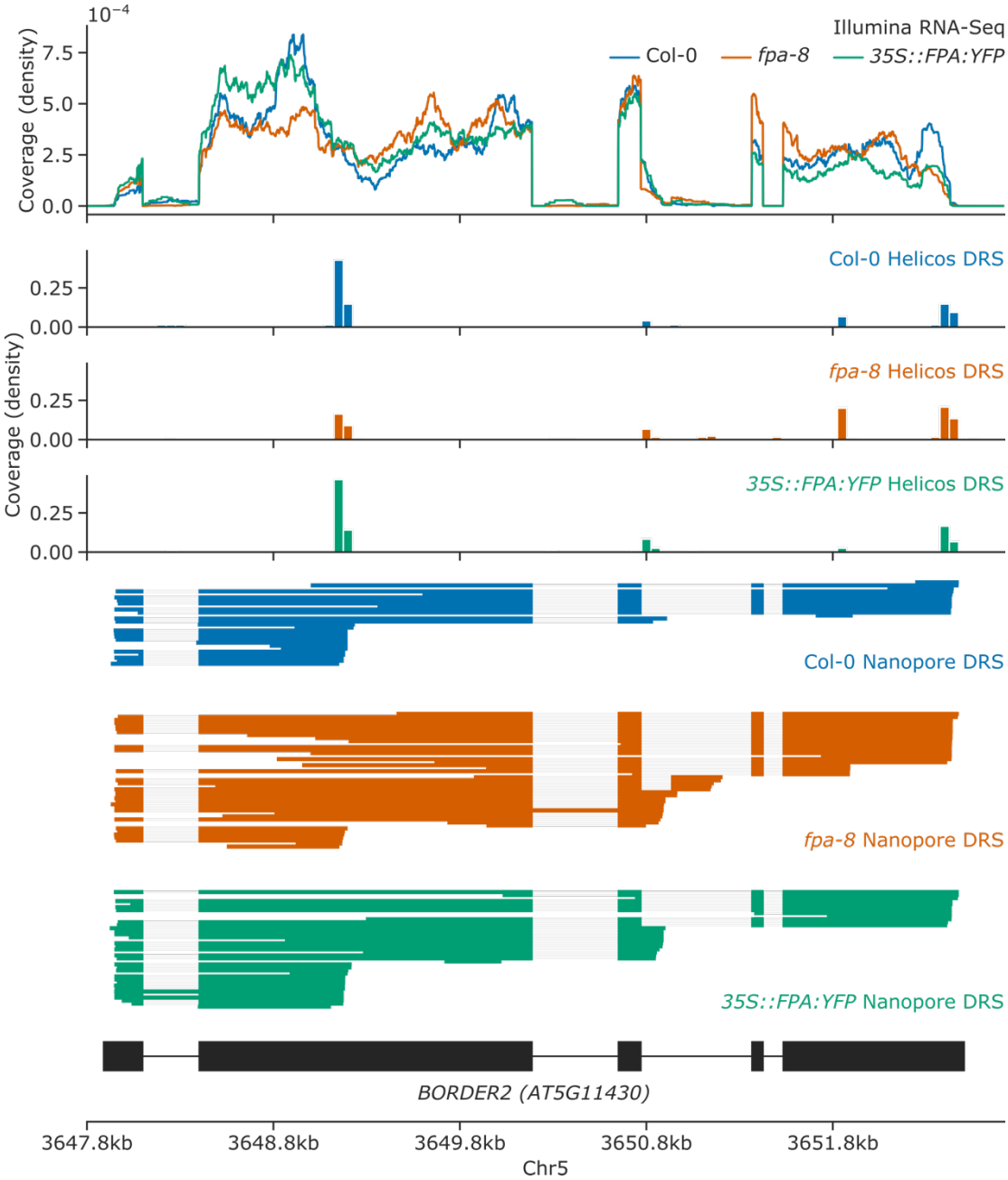


Figure 4 - supplement 12: FPA regulates RNA 3' end processing at *BORDER2* (*AT5G11430*) locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

Figure 5 – supplement 1: Nanopore DRS informs reannotation of the complex NLR locus encompassing *AT546490* and *ATG46500* annotations

Illumina RNA-Seq (top panels, normalised coverage density), nanoPARE (middle panels, normalise coverage density, in pink), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

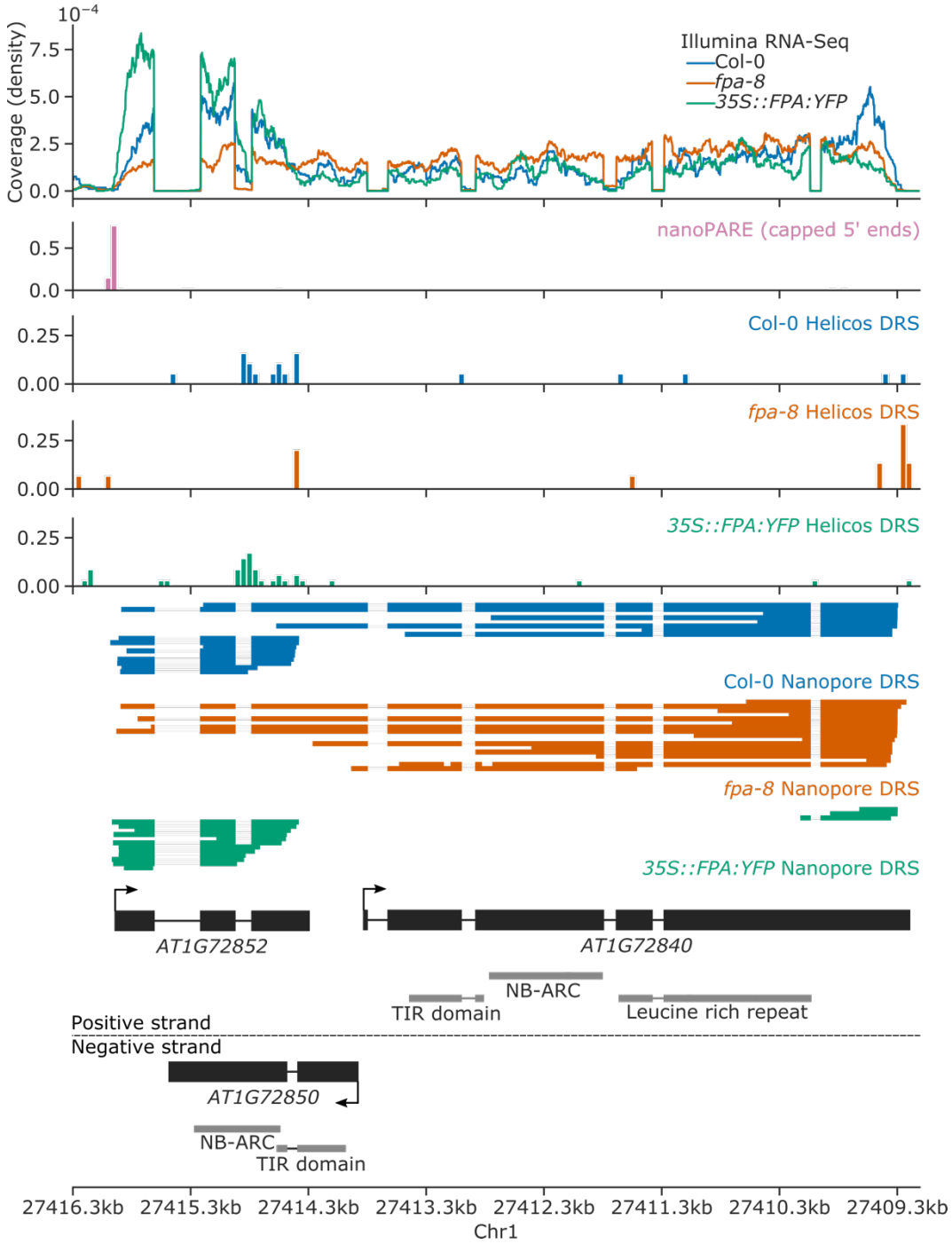


Figure 5 – supplement 2: Nanopore DRS informs reannotation of the complex NLR locus encompassing *AT1G72840*, *AT1G72850* and *AT1G72852* annotations

Illumina RNA-Seq (top panels, normalised coverage density), nanoPARE (middle panels, normalise coverage density, in pink), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8*, in orange; *35S::FPA:YFP*, in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

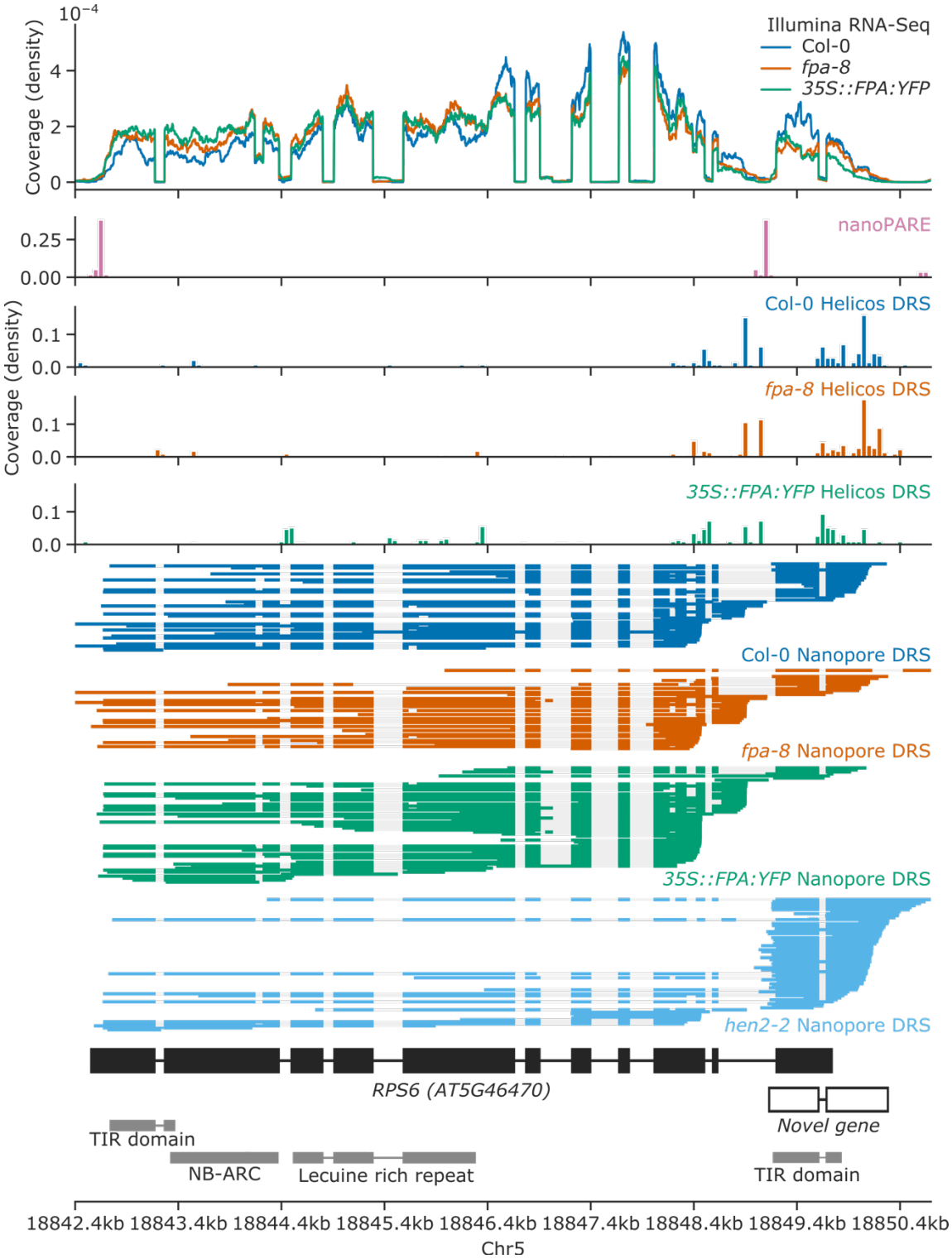


Figure 5 - supplement 3: Nanopore DRS informs reannotation of the complex NLR locus encompassing *RPS6* (*AT5G46470*) annotation

Illumina RNA-Seq (top panels, normalised coverage density), nanoPARE (middle panels, normalise coverage density, in pink), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; *hen2-2*, in light blue; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

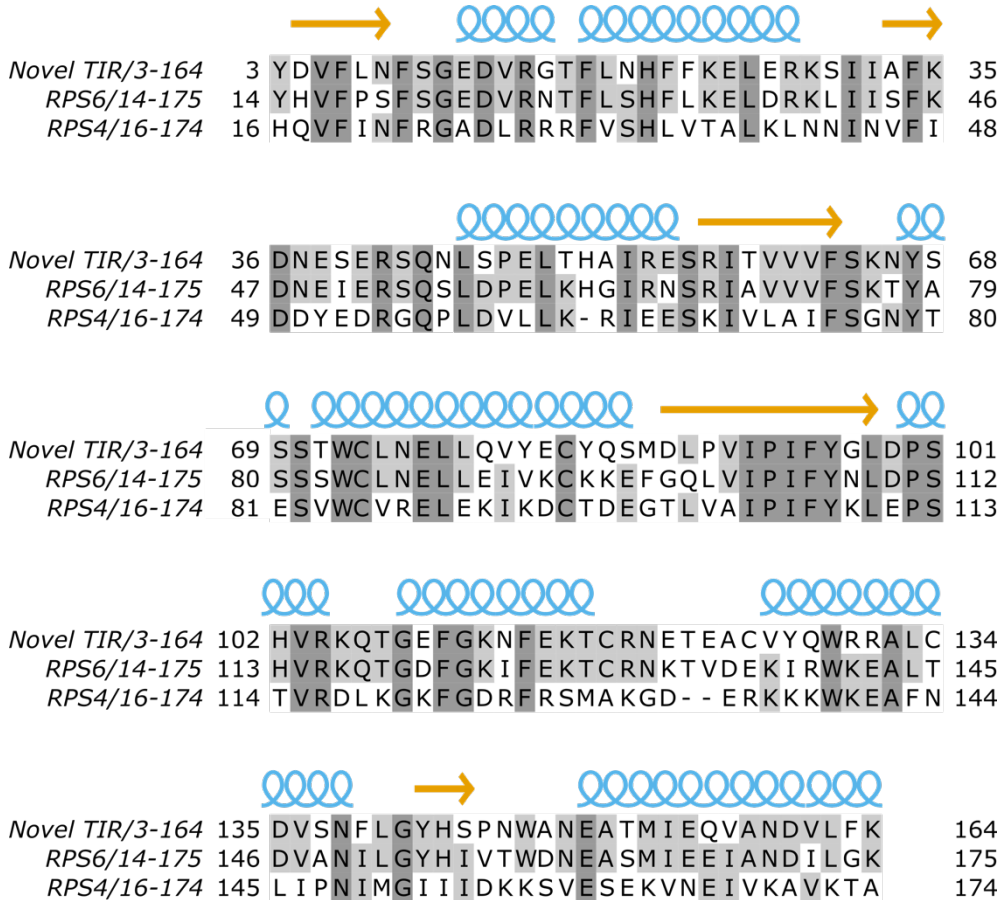


Figure 5 – supplement 4: The novel downstream gene at the *RPS6* locus contains a TIR domain similar to that of *RPS6*.

Protein alignment for the predicted open reading frame of the novel gene downstream of *RPS6* (Novel TIR), with the sequence of the TIR domains from *RPS6* and *RPS4*. Helix and strand secondary structures from UniProt are shown in blue and yellow, respectively. Residues are coloured by conservation (darker grey is more conserved).

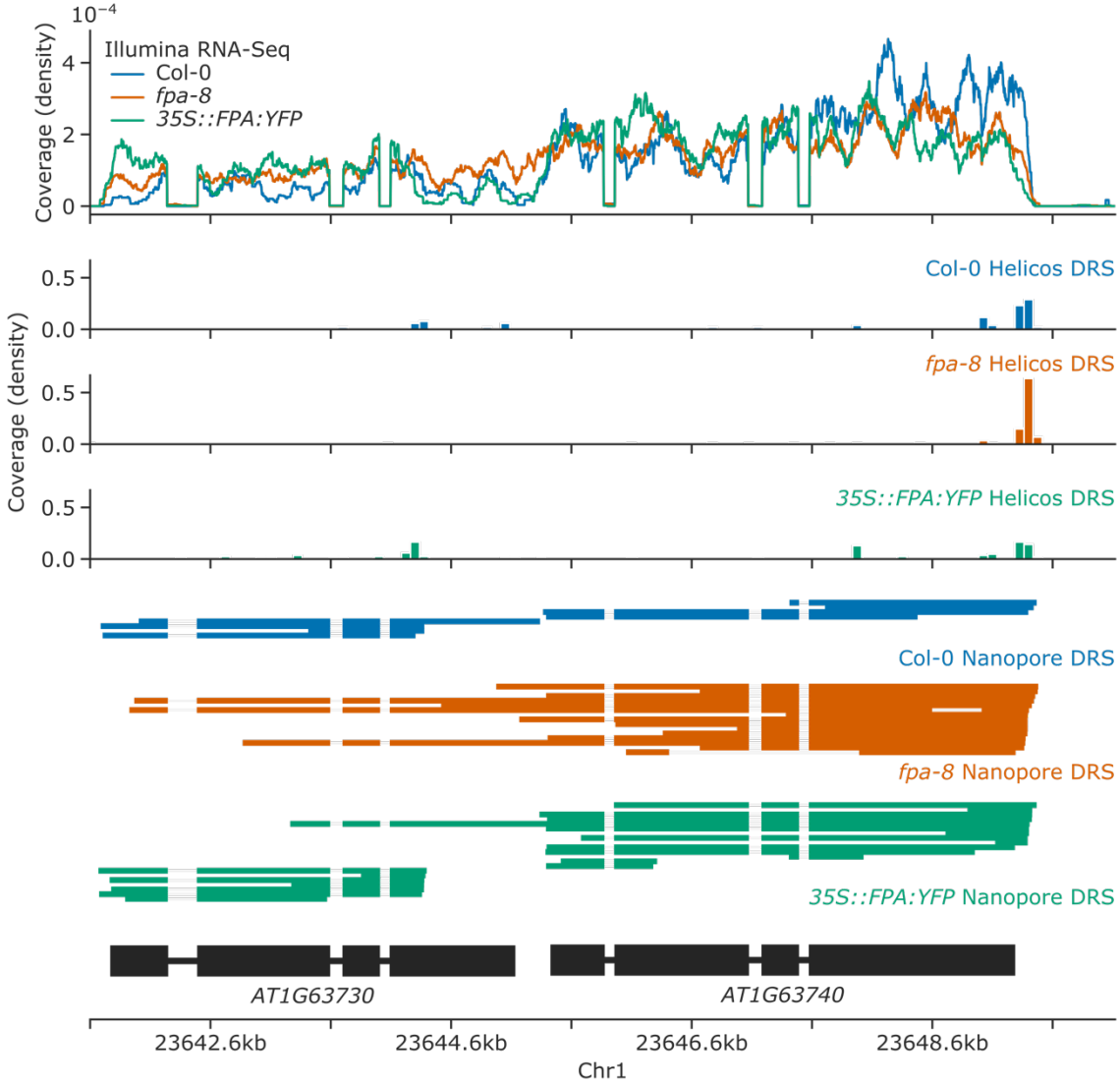


Figure 5 – supplement 5: Loss of FPA function causes chimeric RNA formation at *AT1G63730* and *AT1G63740* NLR loci

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

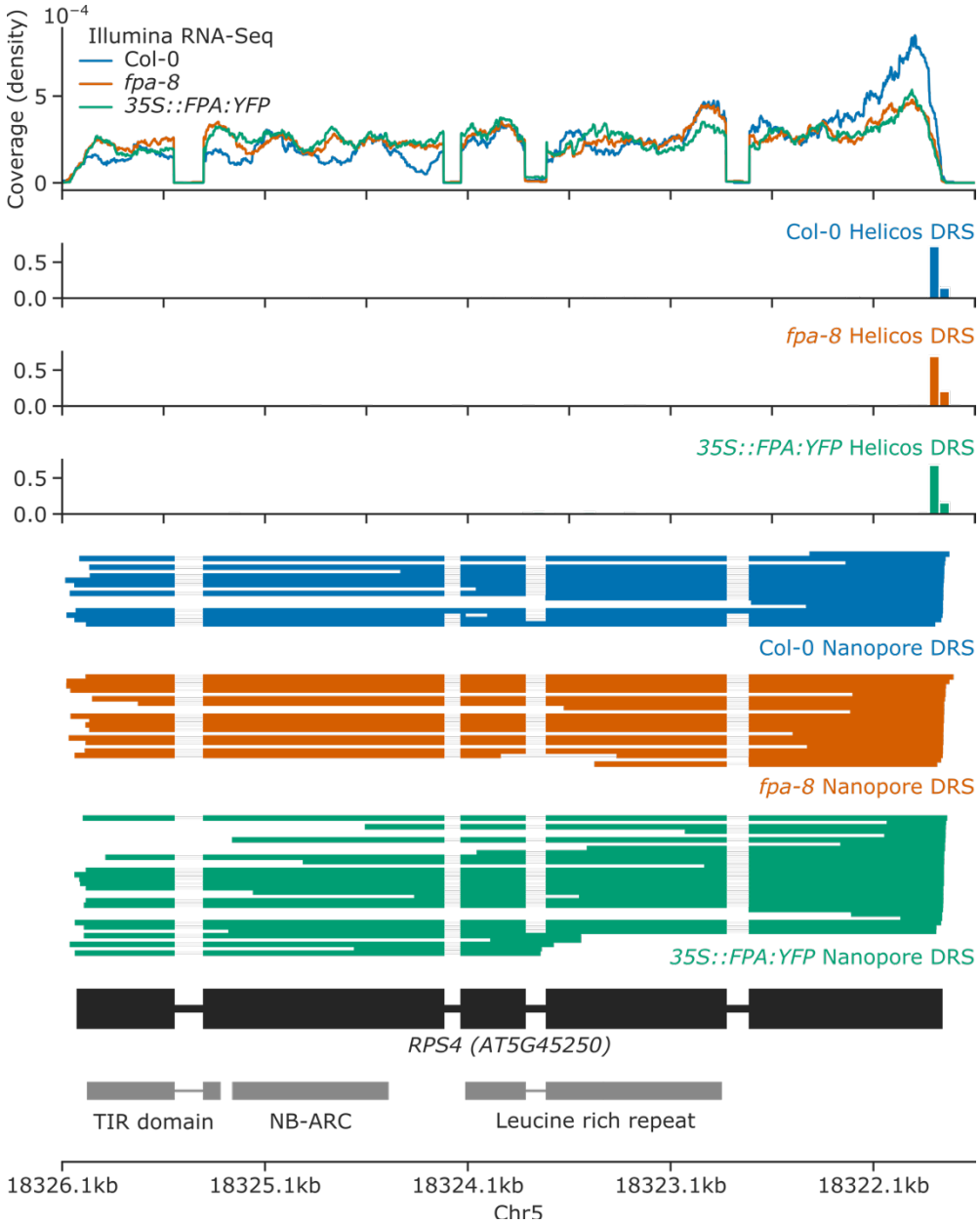


Figure 5 – supplement 6: FPA promotes premature termination events at intron 3 or exon 4 of *RPS4* (*AT5G45250*) NLR locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

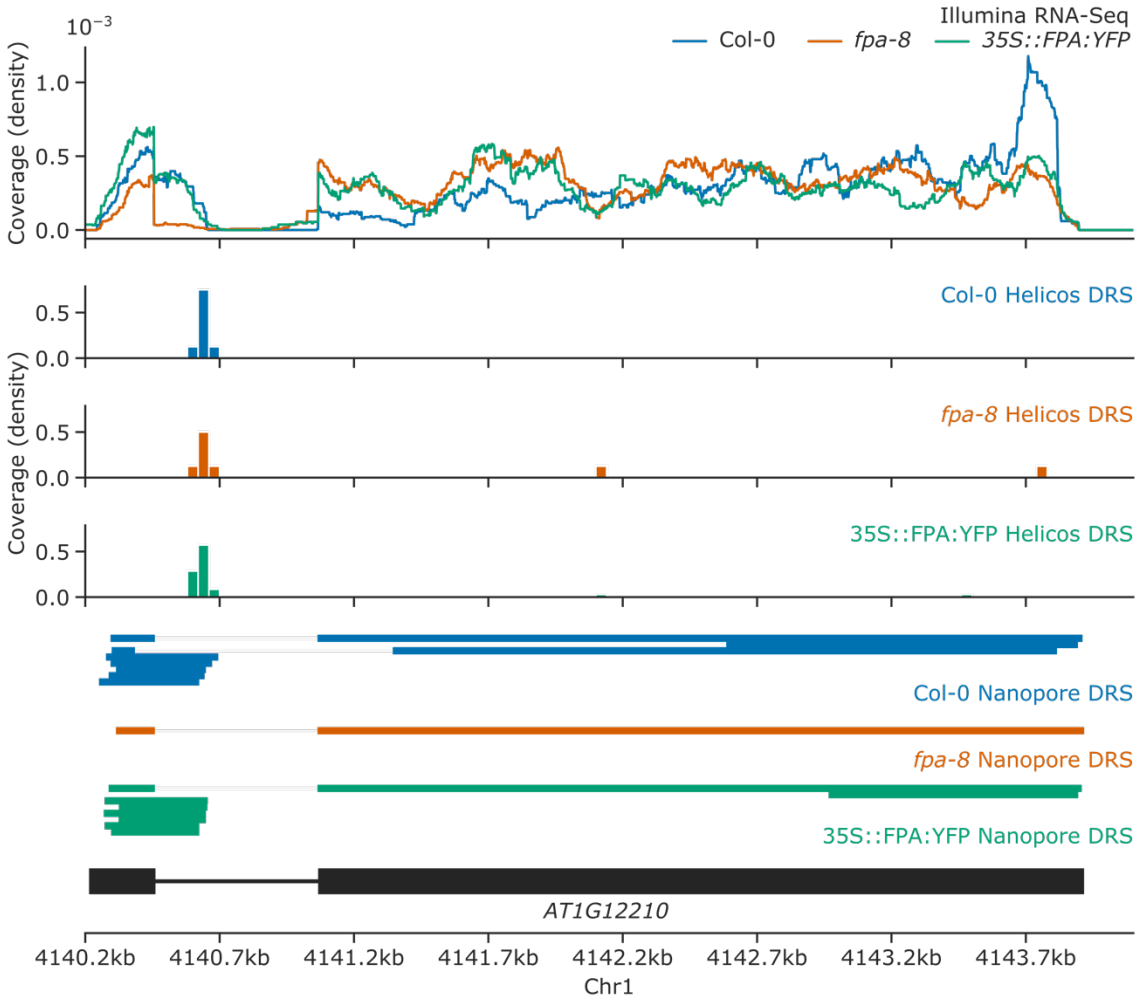


Figure 5 – supplement 7: FPA promotes premature termination at intron 1 of *AT1G12210* NLR locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

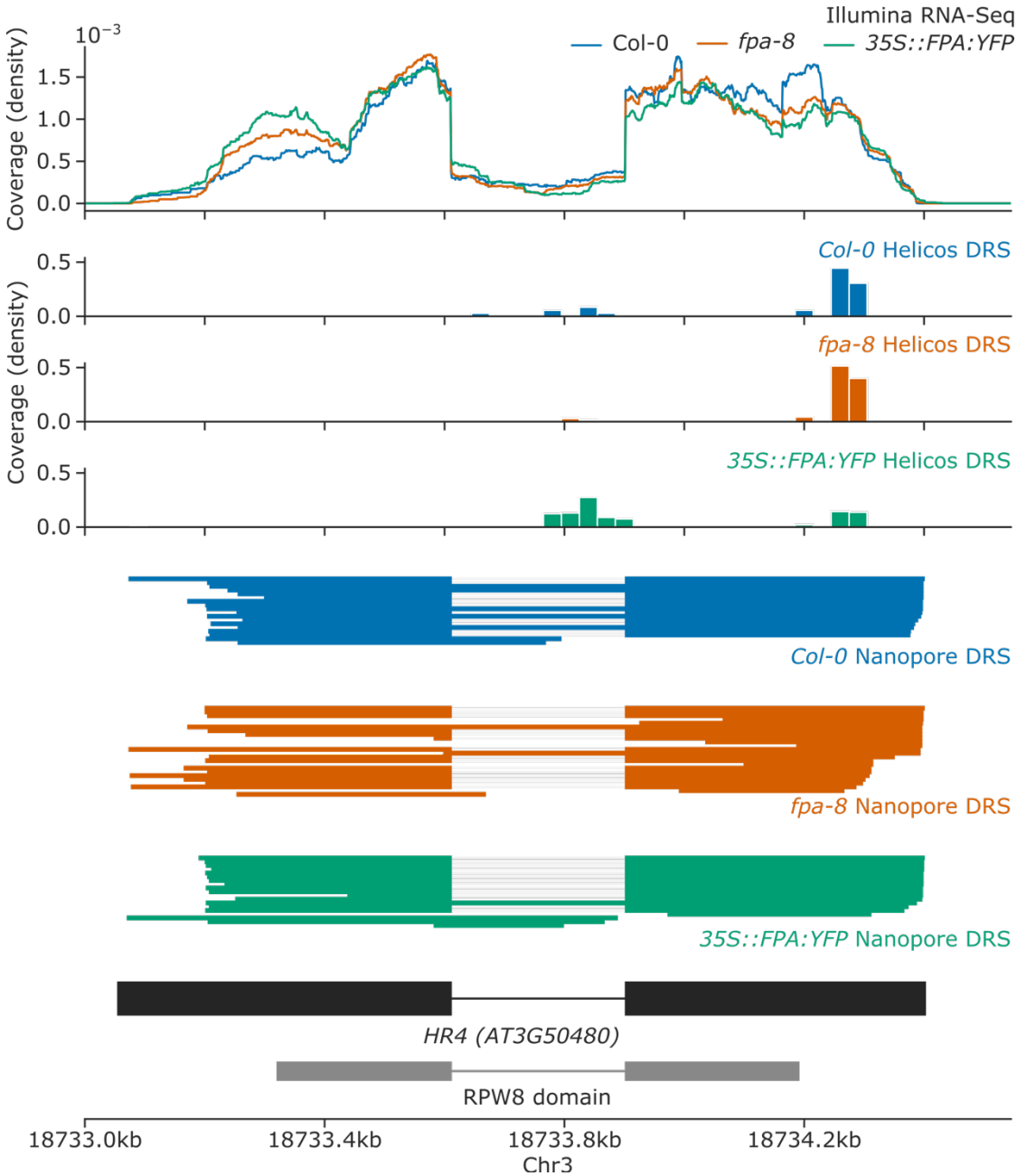


Figure 5 - supplement 8: FPA promotes premature termination at intron 1 of *RPW8/HR4* (*AT3G50480*) non-NLR locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

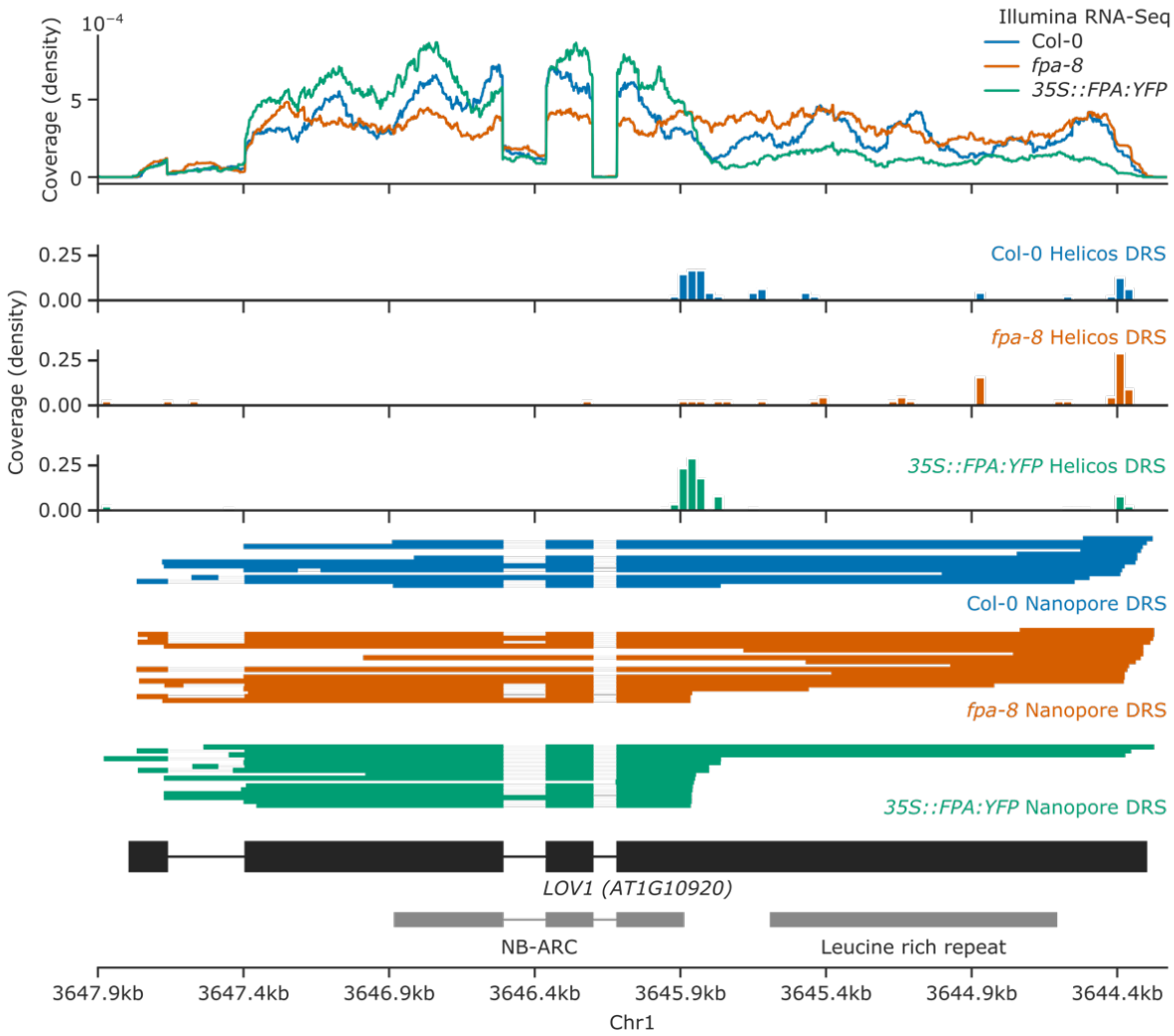


Figure 5 – supplement 9: FPA promotes premature termination within protein-coding exon 4 of *LOV1* (*AT1G10920*) NLR locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

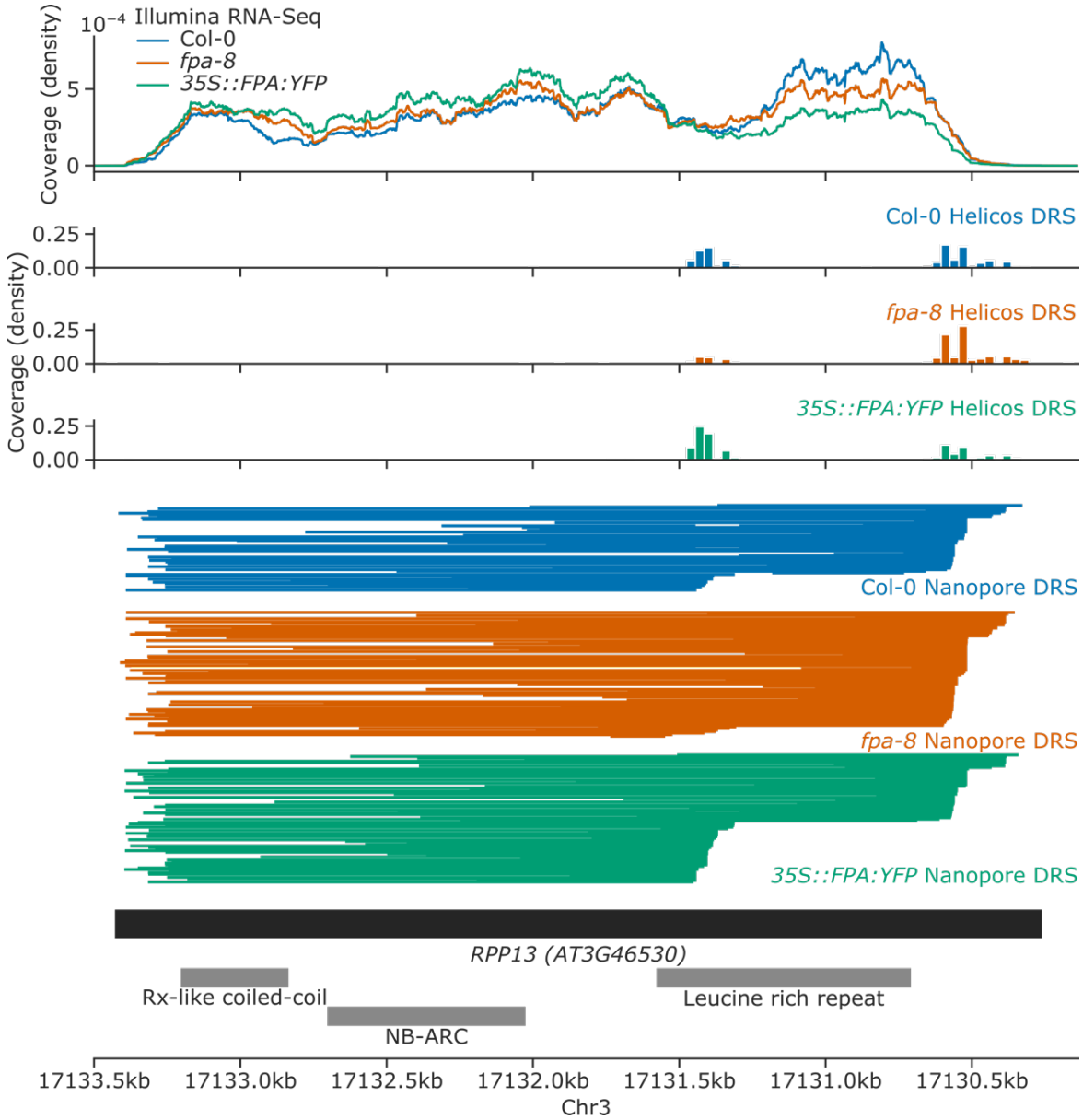


Figure 5 – supplement 10: FPA promotes premature termination within protein coding region of intronless *RPP13* (*AT3G46530*) NLR locus

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black; genomic aligned Interpro domains, in grey.

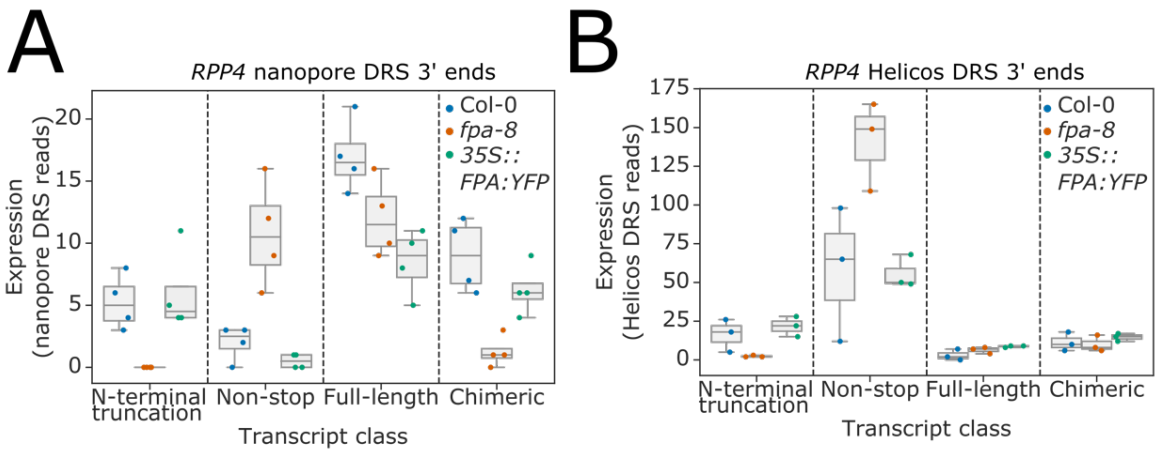


Figure 6 - supplement 1: Complex FPA-dependent patterns of alternative polyadenylation at *RPP4* locus

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*, TE (*AT4G16870*) and/or downstream *AT4G16857*; in Col-0, in blue; *fpa-8, in* orange; and *35S::FPA:YFP,* in green.

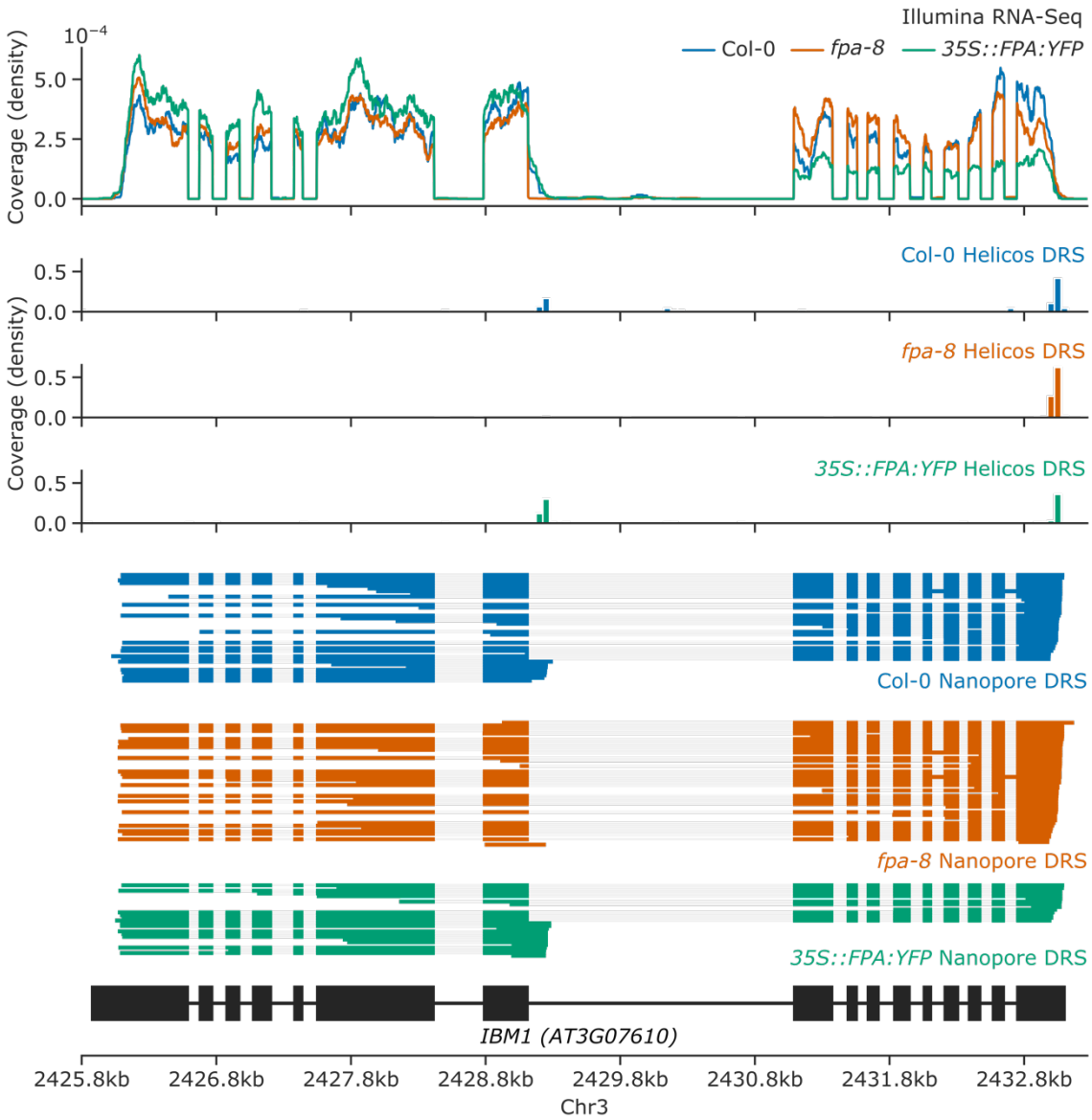


Figure 7 - supplement 1: FPA controls expression at *IBM1* locus by promoting premature termination within *IBM1* intron 7

Illumina RNA-Seq (top panels, normalised coverage density), Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments). Col-0 is shown in blue; *fpa-8,* in orange; *35S::FPA:YFP,* in green; Araport11 annotation, in black.

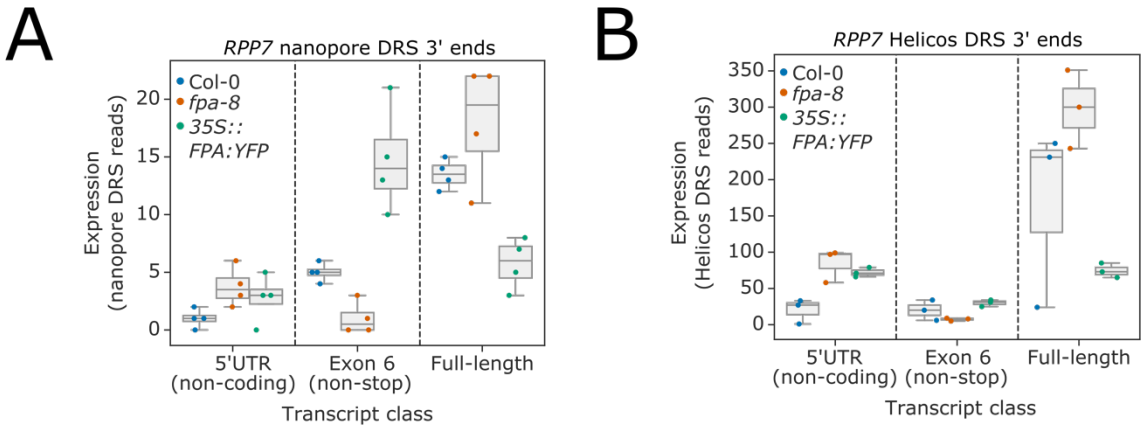


Figure 8 - supplement 1: Complex FPA-dependent patterns of alternative polyadenylation at *RPP7* locus

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; in Col-0, in blue; *fpa-8* in orange; and *35S::FPA:YFP,* in green.

1762 **List of supplementary files:**

1763 Supplementary file 1: Proteins co-purifying with FPA identified by *IVI-MS* **[Linked to Figure**
1764 **1]**.

1765 Supplementary file 2: Properties of the sequencing datasets produced using Nanopore
1766 DRS, Helicos DRS, and Illumina RNA-Seq **[Linked to Figure 4]**.

1767 Supplementary file 3: Manual annotation of 3' processing at the 206 NLR loci present in the
1768 TAIR10 genome **[Linked to Figures 5-7]**.