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. Sherwood¹, Daniel Tome², Xuhong Yu³, Pascal Martin³, Jim Beynon², Scott Michaels³, Geoffrey J. Barton¹ and Gordon G. Simpson^{1,4^}

¹School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK.

²School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.

³Department of Biology, Indiana University, Bloomington, IN 47405, USA.

⁴The James Hutton Institute, Invergowrie DD2 5DA, UK.

*Contributed equally

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

Keywords: NLR, nanopore, direct RNA sequencing, spen, FPA, m⁶A, 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

expression quite broadly (Shivaprasad *et al*, 2012; Zhai *et al*, 2011). The first example of alternative splicing-dependent control having a consequence for NLR function came from the study of the *N* gene, which promotes resistance to tobacco mosaic virus (Dinesh-Kumar & Baker, 2000). Likewise, the function of RECOGNITION OF PSEUDOMONAS SYRINGAE 4 (*RPS4*, *AT5G45250*) in resistance to *Pseudomonas syringae* DC3000 depends on alternative splicing of *RPS4* pre-mRNA and the co-ordinated expression of more than one *RPS4* mRNA isoform (Zhang & Gassmann, 2007). The best characterised example of alternative polyadenylation controlling NLR gene expression is at the Arabidopsis RECOGNITION OF PERONOSPORA PARASITICA 7 locus (*RPP7*, *AT1G58602*). Intragenic heterochromatin associated with a retrotransposon located within the relatively long intron 1 of *RPP7* influences the transcription of full-length mRNAs, with a functional consequence for immunity (Tsuchiya & Eulgem, 2013). The analysis of Arabidopsis mutants that disrupt RNA surveillance has uncovered other features of NLR control. For example, null mutants defective in non-sense mediated RNA decay (NMD) are lethal in Arabidopsis because they trigger autoimmunity dependent on the NLR gene *RPS6* (*AT5G46470*) (Gloggnitzer *et al*, 2014). Most recently, mutations in the RNA exosome, which degrades RNAs in a 3' to 5' direction, were reported to suppress autoimmunity by disrupting the function of *RPS6* (Takagi *et al*, 2020). The signalling and activity of NLR proteins is controlled by intra and intermolecular protein interactions (Adachi *et al*, 2019). Specialization and diversification of NLR function has led to the evolution of NLRs that either sense pathogen effectors directly or help to signal effectors detected by other NLRs. Some NLRs control the activity of other NLRs by direct protein interaction. As a result, networks of NLRs can control immune responses. Despite this progress, our overall understanding of how NLRs evolve and are controlled is limited and fragmented (Adachi *et al*, 2019).

The RNA-binding protein FPA was first identified as a factor required for the control of Arabidopsis flowering time (Koornneef *et al*, 1991). Loss of function *fpa* mutants flower late due to elevated levels of the floral repressor, FLC (Schomburg *et al*, 2001). However, this is not the exclusive function of FPA because it is much more widely conserved than FLC. FPA is a member of the spen family of proteins, defined by three N-terminal RNA recognition motifs and a C-terminal protein interaction SPOC domain (Ariyoshi & Schwabe, 2003). We previously showed that FPA could control the site of cleavage and polyadenylation in some mRNAs, including autoregulation of expression at *FPA* pre-mRNA (Duc *et al*, 2013; Hornyik *et al*, 2010). These findings have been extended to show that FPA can affect poly(A) site choice at genes with intronic heterochromatin, including *RPP7* (Deremetz *et al*, 2019). The mechanism by which FPA affects poly(A) site choice remains unclear. FPA might mediate these effects directly by recruiting the RNA 3' end processing machinery to sensitive sites, or indirectly, by influencing splicing, chromatin modifications, or the rate of transcription by RNA Polymerase II (Pol II), for example. We previously used Helicos direct RNA sequencing (DRS) to map the 3' ends of Arabidopsis polyadenylated 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 m⁶A writer complex, which deposits m⁶A 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) m⁶A 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 CFIm 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,

we detected SF3a (SF3A1) and SF3b (SF3B1, B3, B4 and PHF5A) proteins, together with U2AF65A, which make specific interactions with CPSF during terminal exon definition (Kyburz *et al*, 2006). Cleavage and polyadenylation is also intimately linked to transcription termination. Consistent with this, we identified the Pol II subunits NRPB1 and NRPB2, as well as proteins that modify the phosphorylation status of the heptad repeat of the Pol II C-terminal domain (CTD) (Figure 1B, Supplementary file 1). Characteristically, Serine⁵ of the CTD is phosphorylated when Pol II is at the 5' end of genes, and Ser² is phosphorylated when Pol II is at 3' ends (Komarnitsky *et al*, 2000). We found the kinase CDKC2, which phosphorylates Ser² (Wang *et al*, 2014) and the phosphatase CPL1 (homolog of yeast Fcp1), which dephosphorylates Ser⁵ (Koiwa *et al*, 2004), co-purified with FPA (Figure 1B, Supplementary file 1). Finally, we detected the homolog of human XRN2 5' to 3' exonuclease (known as XRN3 in Arabidopsis), which functions in transcription termination (Figure 1B, Supplementary file 1) (Krzyszton *et al*, 2018). These data suggest that we have captured FPA at the 3' end of actively transcribed genes in close proximity to factors that mediate multiple features of RNA 3' end processing and Pol II transcription termination.

FPA is a component of the genetically defined autonomous pathway of flowering time control (Andres & Coupland, 2012; Simpson, 2004). Consistent with this function, a second major class of proteins co-purifying with FPA are components of this flowering pathway (Figure 1C, Supplementary file 1). Autonomous pathway components limit the expression of the floral repressor *FLC*. Regulation of *FLC* expression is complex, and not fully understood, but it is associated with long non-coding antisense RNAs transcribed at this locus that are alternatively polyadenylated (Hornik *et al*, 2010; Liu *et al*, 2007). Components of the autonomous pathway, identified through mutant screens, include conserved CPF proteins such as FY (WDR33) (Simpson *et al*, 2003), CSTF64 and CSTF77 (Liu *et al*, 2010) that co-purified here with FPA. In addition, the PCF11 homolog PCFS4, which also co-purifies with FPA, was identified as a component of the autonomous pathway by reverse genetics (Xing *et al*, 2008). Consequently, there is an overlap in classification of these proteins because of their biochemical role as CPF proteins and the biological impact that they have on flowering time control at *FLC*. We found that other autonomous pathway RNA binding proteins, which are not core CPF proteins but do affect RNA 3' end formation, also co-purify with FPA (Figure 1C, Supplementary file 1). These include HLP1, a heterogenous ribonucleoprotein A (hnRNP A)-related protein (Zhang *et al*, 2015). Additional RNA binding proteins of the autonomous pathway, such as FLK (Mockler *et al*, 2004), KHZ1/KHZ2 (Yan *et al*, 2019), FLY (Dai *et al*, 2020) and GRP7/8 (Streitner *et al*, 2008) were also detected. Not all members of the autonomous pathway are RNA binding proteins, and we detected FVE co-purifying with FPA. FVE homologs function in diverse chromatin remodelling complexes in other organisms (Ausin *et al*, 2004). Notably, we also detected PDP1 and EMB1579/RSA1, which have recently been shown to physically interact with FVE (Zhang *et al*, 2020; Zhou *et al*, 2018). EMB1579/RSA1 controls nuclear condensate formation and global patterns of transcription and splicing, and *emb1579/rsa1* mutants

flower late because of elevated levels of *FLC* (Zhang *et al.*, 2020). In addition, we detected TAF15b which represses *FLC* transcription (Eom *et al.*, 2018). The autonomous pathway components FLD, LD and SDG26, which were recently found to co-purify together (Fang *et al.*, 2020), were not detected here. We identified NSRA, which affects alternative polyadenylation of antisense RNAs at *FLC*, but functions antagonistically to the autonomous pathway by promoting *FLC* expression (Bazin *et al.*, 2018). FLK has been found to associate with PEP, HUA1 and HEN4 (Rodriguez-Cazorla *et al.*, 2015), and here, we found all four of these proteins co-purifying with FPA (Figure 1C, Supplementary file 1). In addition to regulating *FLC* expression, these four proteins have been shown to interact to control alternative polyadenylation within pre-mRNA encoding the floral homeotic transcription factor AGAMOUS (Rodriguez-Cazorla *et al.*, 2015). Their co-purification here suggests that this role may be more global and involve direct interactions at RNA 3' ends.

The third major group of proteins co-purifying with FPA are conserved members of an mRNA m⁶A writer complex; MTA, MTB, VIR and FIP37 (Ruzicka *et al.*, 2017) (Figure 1D, Supplementary file 1). The m⁶A modification dependent on this writer complex is predominately targeted to the 3' UTR of Arabidopsis protein coding mRNAs (Parker *et al.*, 2020b). The co-purification of FPA with the m⁶A writer complex components could therefore be explained by a direct role for FPA in m⁶A modification, or simply because both CPF and m⁶A writer proteins are found at RNA 3' ends.

The overall picture that emerges from this analysis is that FPA associates in close proximity to diverse proteins that function to promote cleavage, polyadenylation, transcription termination and RNA modification at the 3' end of Pol II-transcribed genes.

FPA co-localises with Pol II phosphorylated at Ser² of the CTD heptad repeat at the 3' end of Arabidopsis genes

We next asked whether FPA was recruited specifically to the 3' ends of genes or was associated with Pol II throughout transcription. To distinguish between these possibilities, we used antibodies against FPA and Pol II phosphorylated at either Ser⁵ or Ser² of the CTD heptad repeat to perform ChIP-Seq analysis (Yu *et al.*, 2019). Our metagene analysis reveals that FPA is enriched at the 3' end of genes and co-localises with Pol II phosphorylated at Ser² of the CTD (Figure 2A, Figure 2 - supplement 1). These findings are consistent with the idea that FPA is specifically recruited to the 3' ends of genes and is not associated with Pol II during elongation.

Next, we investigated the relationship between FPA and the different phosphorylated forms of Pol II at individual genes. We have previously identified specific loci at which loss of FPA function causes defective 3' end formation and readthrough into downstream genes resulting in the formation of chimeric RNAs (Duc *et al.*, 2013). For example, at *PIF5* (*AT3G59060*), chimeric RNAs are formed with the downstream gene *PAO3* (*AT3G59050*) in loss-of-function *fpa-7* and *fpa-8* mutant backgrounds (Duc *et al.*, 2013). Here we found that, in wild type Col-0 plants, FPA and Ser² phosphorylated Pol II co-localise

at the 3' end of *PIF5*, with low levels of Ser^{5/2} phosphorylated Pol II detected at *PAO3* (Figure 2B). In *fpa-7* mutants, however, Pol II phosphorylated at Ser² of the CTD is lost from the 3' end of *PIF5*, with increased levels of Ser⁵ and Ser² phosphorylated Pol II detected at *PAO3* downstream. These data are consistent with FPA influencing 3' end formation and the termination of transcription by Pol II.

Figure 2: FPA co-localises with Pol II Ser² at the 3' end of genes and promotes transcription termination

(A) ChIP-Seq metagene profile showing the normalised occupancy of FPA (green) and Pol II phosphorylated at Ser⁵ (pink) and Ser² (brown) of the CTD (Yu *et al.*, 2019) relative to the major 3' position of each gene as measured using Helicos DRS. Only long genes (> 2.5 kb) are included (n=10,215). **(B)** ChIP-Seq occupancy in counts per million (CPM) of FPA (input shown in gray; FPA, in green) and Pol II phosphorylated at Ser⁵ or Ser² (Col-0 is shown in blue; *fpa-7*, in orange) at the *PIF5* (AT3G59060) and *PAO3* (AT3G59050) loci (top panels); Helicos DRS (middle panels, normalised coverage density) and nanopore DRS (bottom panels, read alignments) - Col-0 is shown in blue; *fpa-8*, in orange; and Araport11 annotation, in black.

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

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

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

We first assessed the utility of the three sequencing technologies to map changes in RNA 3' end processing by focusing on the well characterised *FPA* locus. *FPA* autoregulates its expression by promoting premature cleavage and polyadenylation within intron 1 of *FPA* pre-mRNA (Duc *et al.*, 2013; Hornyik *et al.*, 2010). Consistent with this, a proximal poly(A) site in the first intron of *FPA* and distal sites in the terminal intron and exon can be mapped precisely in Col-0 using nanopore and Helicos DRS (Figure 3). Using all three data types we detected a quantitative shift towards selection of distal poly(A) sites in the loss-of-function *fpa-8* mutant and a strong shift to proximal poly(A) site selection when *FPA* is overexpressed (*35S::FPA:YFP*). Nanopore DRS provides the clearest picture of 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 $p=3.9 \times 10^{-7}$) (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 $p=4.8 \times 10^{-30}$), 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.

In order to analyse the Illumina RNA-Seq data, we developed annotation-agnostic software for detecting alternative RNA 3' end processing events, using a similar approach to the existing tools DERfinder (Collado-Torres *et al*, 2017), RNAprof (Tran Vdu *et al*, 2016) and DEXSeq (Anders *et al*, 2012). We segmented the Illumina RNA-Seq data using coverage thresholding and relative changes in expression in *fpa-8* or *35S::FPA:YFP* compared to Col-0. Segmented regions were grouped into transcriptional loci using the annotations generated from Col-0, *fpa-8*, and *35S::FPA:YFP* nanopore DRS datasets. Differential usage of regions within each locus was then tested using DEXSeq (Figure 4E). Using this approach, we identified 2535 loci with differential RNA processing events in *fpa-8*: 1792 upregulated, 390 downregulated, and 353 with both upregulated and downregulated regions (FDR < 0.05, absolute logFC > 1) (Figure 4E, left panel). 1747 loci with differential RNA processing events were identified in *35S::FPA:YFP*: 997 upregulated, 532 downregulated, and 218 with both upregulation and downregulation (Figure 4E, right panel). In the *fpa-8* mutant, we found that the median effect size for differentially processed regions was greater for upregulated regions than downregulated regions (Figure 4E, left panel). This is consistent with the major effect of loss of FPA function resulting in transcriptional readthrough events and increased expression of intergenic regions, downstream genes and/or transposable elements (TE). In contrast, the median effect size for differentially processed regions was similar for up- and downregulated regions in *35S::FPA:YFP* (Figure 4E, right panel). This is consistent with an increase in the relative expression of proximal exonic and intronic regions, and a loss of expression of distal exonic regions caused by preferential selection of proximal poly(A) sites.

To test whether changes in FPA expression result in alternative splicing, we quantified the number of reads supporting each annotated splice junction, grouped by transcriptional locus, and performed differential junction usage analysis with DEXSeq (Figure 4F). We found that the clearest alterations in splice site usage occurred in *fpa-8*, where 868 loci have splice junctions that are strongly upregulated (Figure 4F, left panel). The effect sizes for these loci tended to be greater than those of differentially spliced loci detected in *35S::FPA:YFP* (Figure 4F, right panel). Closer inspection of the junctions affected in *fpa-8* indicated that they are the result of increased cryptic splicing events associated with altered poly(A) site choice and the increased splicing of downstream genes in readthrough and chimeric RNAs. These changes in processing can be illustrated by examining the chimeric RNAs formed at the *PIF5* locus. Here, readthrough in *fpa-8* mutants is associated with cryptic splicing events between the annotated terminal exon of *PIF5* and the first exon of *PAO3* (Figure 4 - supplement 2). In addition, because the detectable expression of the downstream gene *PAO3* is increased, there is a corresponding increase in reads spanning annotated *PAO3* splice junctions. These changes in splicing can be explained by the effect of FPA on 3' end formation at *PIF5*. We conclude that changes in splicing are indirect effects of the altered 3' end processing which occurs in *fpa-8* mutants, rather than direct effects of FPA on splice site choice.

Our *IVI-MS* analysis revealed that conserved members of the Arabidopsis m⁶A writer complex co-purify with FPA (Figure 1D, Supplementary file 1). The most related proteins to FPA in humans are RBM15/B, which co-purify with the human m⁶A writer complex and are required for m⁶A deposition (Patil *et al.*, 2016). We therefore asked if FPA influenced the level of m⁶A in Arabidopsis mRNA. We used LC-MS/MS to analyse the ratio of m⁶A/A in mRNA purified from Col-0, *fpa-8*, *35S::FPA:YFP* and a mutant defective in the m⁶A writer complex component VIR (*vir-1*). Consistent with previous reports, the level of mRNA m⁶A in the hypomorphic *vir-1* allele was reduced to approximately 10% of wild-type levels (Parker *et al.*, 2020b; Ruzicka *et al.*, 2017) (Figure 4 - supplement 3). However, we did not detect any differences in the m⁶A level between genotypes with altered FPA activity. We conclude that FPA does not influence global levels of mRNA m⁶A methylation.

Overall, these analyses reveal that the primary function of FPA is to control poly(A) site choice. FPA predominantly promotes poly(A) site selection, hence *fpa* loss-of-function backgrounds exhibit readthrough at sites used in wild-type, while FPA overexpression results in increased selection of proximal poly(A) sites. Changes in splicing patterns in backgrounds with different FPA activities can be explained as consequences of alternative polyadenylation. Moreover, FPA is not required for the global levels of mRNA m⁶A, despite co-purifying with components of the m⁶A writer complex.

FPA controls the processing of NLR transcripts

We next asked what groups of genes were sensitive to FPA-dependent alternative polyadenylation. We found that genes encoding proteins involved in cleavage, polyadenylation and transcription termination, or which co-purified with FPA, were affected by FPA-mediated alternative polyadenylation. These included *CFIS2*, *CPSF160*, *CPSF73-1*, *EMB1579/RSA1* and *ATSPEN2* (Figure 4 - supplement 4-8). FPA was also found to promote exonic proximal polyadenylation within the coding regions of transcripts encoding PCF11-related proteins *PCFS1* and *PCFS5*, despite both genes being encoded by a single exon (Figure 4 - supplement 9-10). Moreover, we found that the *BORDER1* and *BORDER2* genes, which encode proteins that protect expression of adjacent Arabidopsis genes by 3' pausing of Pol II (Yu *et al.*, 2019), were controlled by FPA-dependent alternative polyadenylation (Figure 4 - supplement 11-12).

We next used Interpro annotations (Mitchell *et al.*, 2019) to perform protein family domain enrichment analysis of the genes affected by FPA as revealed by the nanopore and Helicos DRS analyses. This approach is similar to gene ontology analysis but relies on the automatic annotation of domains using profile hidden Markov models, rather than on a mixture of curated automatic and manual annotation. Consequently, annotations of poorly studied genes should be similarly detailed to well-studied ones. We found that genes encoding nucleotide binding NB-ARC, Rx-like coiled coil (CC), and Leucine Rich Repeat (LRR) domains were most enriched amongst those with increased proximal polyadenylation

in 35S::*FPA:YFP* (Figure 5A-B). This combination of protein domains is associated with the family of disease resistance factors known as NLR proteins.

According to the recently published pan-NLRome of Arabidopsis, the Col-0 accession contains at least 206 genes with some combination of TIR (Toll-Interleukin-1 Receptor), CC, RPW8-like CC, NB-ARC, or LRR domains, which might be classified as NLRs or partial NLRs (Van de Weyer *et al.*, 2019). In general, these can be grouped into classes by their N-terminal domain as either TIR (TNLs), CC (CNLs), RPW8-like (RNLs), or NB-ARC and LRR only (NLs) NLR genes. There are also many genes that comprise only TIR, CC, or CCR domains and NLRs that incorporate other domains that function as integrated decoys (Le Roux *et al.*, 2015). We manually examined these genes to identify those with alternative polyadenylation dependent on *FPA*. Nanopore DRS reveals that pre-mRNA processing at many NLR loci is more complicated than current annotation would indicate. As a result, we revised the annotation of some NLR genes in this analysis. For example, we find that the TNL gene *AT5G46490* is incorrectly split into two loci, *AT5G46490* and *AT5G46500* (Figure 5 - supplement 1, Supplementary file 3). This is in fact a single locus with a previously unrecognised 2.7 kb intron containing a proximal poly(A) site, use of which is regulated by *FPA*. This finding was supported by nanoPARE data (Schon *et al.*, 2018), which shows that there are no capped 5' ends originating from the annotated downstream gene. Use of the distal poly(A) site introduces an additional ~400 amino acids to the C-terminus of the protein with predicted homology to RPP5 and lengthens the LRR domain. At the divergent TNL pair *AT1G72840* and *AT1G72850*, an annotated natural antisense RNA opposite *AT1G72850* (*AT1G72852*) was found to be transcribed as a single RNA with *AT1G72840* (Figure 5 - supplement 2, Supplementary file 3). *FPA*-regulated proximal polyadenylation can occur at the annotated *AT1G72852* 3' end, but we saw no evidence in nanoPARE data that downstream *AT1G72840* can be transcribed on its own.

Clearly, correct annotation is a necessary part of understanding how NLRs are expressed and controlled. Notably, we could also reannotate the chromosomal region around *RPS6* (*AT5G46470*). The extreme autoimmunity phenotypes of NMD mutants and mitogen activated kinase pathway mutants are caused by the dysregulation of *RPS6* but the mechanisms involved are not understood (Gloggnitzer *et al.*, 2014; Takagi *et al.*, 2020). Our nanopore DRS analysis reveals a long and complex *RPS6* 3'UTR, with multiple exons and alternative poly(A) sites (Figure 5 - supplement 3). We also detect evidence of rare transcripts in this region that do not appear to be contiguous with *RPS6* 3'UTR reads. Instead, these reads correspond to an independent unannotated downstream gene that overlaps with the *RPS6* 3'UTR. Consistent with this idea, capped 5' ends identified by nanoPARE map to this region (Schon *et al.*, 2018) and our nanopore DRS analysis of the RNA exosome mutant *hen2-2* (Parker *et al.*, 2020a) reveals relatively high levels of independently transcribed RNAs in the corresponding position (Figure 5 - supplement 3). The downstream gene encodes a TIR domain similar to the *RPS6* TIR domain (Figure 5 - supplement 4). Use of the distal *RPS6* poly(A) site therefore constitutes readthrough into the downstream TIR-

domain-only NLR (Figure 5 – supplement 3). We conclude that long read nanopore DRS data has the potential to resolve the complexity of NLR gene annotation, that long read DNA sequence data and short read Illumina RNA-Seq cannot.

Of the 206 NLR genes examined, 124 were sufficiently expressed to identify alternative polyadenylation in the nanopore DRS data, and of these, 62 (50%) were found to have FPA-dependent alternative polyadenylation (Supplementary file 3). These effects could be classified into three modes of regulation: (i) FPA control of readthrough and chimeric RNAs, usually by repressing their formation; (ii) FPA regulation of intronic poly(A) site selection; and (iii) FPA control of poly(A) site selection within protein-coding exons. FPA can affect poly(A) site choice using a combination of all three modes of regulation at a number of complex loci.

We detected 16 NLR genes where a change in FPA levels altered formation of readthrough or chimeric RNAs containing one or more NLR loci (Supplementary file 3). *AT1G58848*, which is part of the *RPP7* cluster containing five CNL-class NLRs, displayed increased readthrough into the downstream TE *AT1G58889* in the *fpa-8* mutant (Figure 5C, Supplementary file 3). Loss of FPA function can also lead to the generation of clusters containing two or more NLR genes transcribed together as one transcriptional unit. For example, the TNL-class gene *AT1G63730* forms chimeric RNA with the downstream TNL-class gene *AT1G63740* in *fpa-8* (Figure 5 – supplement 5, Supplementary file 3).

We identified 18 NLR genes with intronic polyadenylation regulated by FPA (Supplementary file 3). At *AT1G69550*, encoding a TNL-type NLR, we identified a proximal poly(A) site within the third intron (Figure 5D, Supplementary file 3). Use of this poly(A) site results in mRNAs with a premature stop codon, which encodes a form of the *AT1G69550* protein lacking the majority of the predicted LRR domain. In *fpa-8* mutants, readthrough of this poly(A) site increased, resulting in higher levels of full-length transcripts. At the TNL-type NLR *RPS4*, which has previously been shown to be regulated by alternative splicing (Zhang & Gassmann, 2007), we detected an increase in intron 3 retention in *35S::FPA:YFP* compared to Col-0 that was associated with proximal polyadenylation events in intron 3 or exon 4 (Figure 5 – supplement 6, Supplementary file 3). At *AT1G12210*, encoding a CNL-class NLR, we detected short isoforms ending within the first intron of the 5' UTR in Col-0 (Figure 5 – supplement 7, Supplementary file 3). Loss of FPA function caused transcriptional readthrough at this poly(A) site, resulting in production of full-length *AT1G12210* mRNAs comprising the full protein-coding sequence. FPA also regulated intronic polyadenylation in Rx N-terminal domain-only genes, such as the RPW8-like domain gene *RPW8/HR* (*AT3G50480*) (Figure 5 – supplement 8, Supplementary file 3).

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 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.

The most common form of NLR regulation was FPA-dependent premature termination within protein-coding exons, which was detected at 46 NLR genes (Supplementary file 3). Many of these proximal exonic poly(A) sites can be identified at low levels in Col-0. For example, at *AT2G14080*, encoding a TNL-class NLR, we detected multiple exonic poly(A) sites located within the second and fourth exons encoding the NB-ARC and LRR domains respectively, selection of which increased in *35S::FPA:YFP* and decreased in *fpa-8* (Figure 5E, Supplementary file 3). At *LOV1* (*AT1G10920*), encoding a CNL-type NLR protein, overexpression of FPA caused more frequent selection of an exonic poly(A) site located within the fourth exon encoding the NB-ARC domain region, whilst loss of FPA function resulted in increased production of full-length transcripts (Figure 5 – supplement 9, Supplementary file 3). FPA was even found to promote premature termination in the protein coding sequence of single exon, intronless NLR genes. For example, at *RPP13* (*AT3G46530*), encoding a CNL-class NLR protein, FPA overexpression causes selection of proximal poly(A) sites located within the region encoding the LRR domain (Figure 5 – supplement 10, Supplementary file 3).

An example of complex FPA-dependent regulation is found at *RPP4* (*AT4G16860*). This gene encodes a well-studied TNL-class NLR protein known to mediate Arabidopsis resistance to *Hyaloperonospora arabidopsidis* isolate Emoy2 (*Hpa*-Emoy2) (van der Biezen *et al.*, 2002). *RPP4* is part of the *RPP5* cluster comprising seven TNL-class NLRs interspersed with three related sequences and two non-NLR genes (van der Biezen *et al.*, 2002). In agreement with a previous study (Wang & Warren, 2010), we found that in wild-type Col-0, *RPP4* can be transcribed as a chimeric RNA with the downstream *AtCOPIA4* TE (*AT4G16870*), as a result of the selection of two distal poly(A) sites located within the TE (Figure 6) (Wang & Warren, 2010). We also identified several chimeric RNAs containing *RPP4*, the TE and the downstream gene *AT4G16857*. Use of the more proximal poly(A) site within the TE is associated with an approximately 8-kb cryptic splicing event between the 5' splice site of the first exon of *RPP4* and a 3' splice site located within the TE, which skips all *RPP4* exons downstream of exon 1 (Figure 6, Inset 1). This splicing event, detected by nanopore DRS, and which is also supported by Illumina RNA-Seq data (Figure 6), therefore removes the majority of the *RPP4* coding sequence and introduces a novel stop codon (Figure 6, Inset 1). The resulting transcript is predicted to encode a truncated protein, containing only the TIR domain of *RPP4*. In Col-0, 27% of nanopore DRS reads at the *RPP4* locus terminate at one of the two TE-located poly(A) sites or within the downstream gene (Figure 6, Figure 6 – supplement 1A-B). Loss of FPA function decreases chimeric RNA formation by shifting poly(A) site selection towards a proximal poly(A) site located within the protein-coding region of the final exon (Figure 6, Figure 6 – supplement 1A-B, Supplementary file 3). This results in production of *RPP4* transcripts lacking in-frame stop codons (Figure 6, Inset 2; Figure 6 – supplement 1A-B, Supplementary file 3). Furthermore, in the *35S::FPA:YFP* line, we observed increased selection of a proximal poly(A) site located within the first intron of *RPP4*, which would also encode a truncated *RPP4* protein (Figure 6,

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

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

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. **Inset 1** (light orange area) shows a magnified view of TIR-domain only *RPP4* transcripts detected in 35S::FPA:YFP, caused by proximal polyadenylation in intron 1, and distal polyadenylation within the TE associated with cryptic splicing; **Inset 2** (pink area) shows a magnified view of the stop-codonless transcripts produced within the protein-coding *RPP4* region (Non-stop transcripts) in *fpa-8*.

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

We have previously shown that FPA controls the expression of the histone demethylase *IBM1* (INCREASE IN BONSAI METHYLATION 1; AT3G07610) by promoting premature cleavage and polyadenylation within *IBM1* intron 7 (Duc *et al.*, 2013). This regulation is also 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 H3K9me² 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 H3K9me² 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 H3K9me² 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 m⁵C DNA methylation and histone H3 dimethylation (H3K9me²) (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-

terminated transcript (ECL), accompanied by a reduction in the abundance of full-length *RPP7* transcripts and compromised immunity against *Hpa*-Hiks1 infection (Tsuchiya & Eulgem, 2013).

Analysis of our combined set of Illumina RNA-Seq, nanopore and Helicos DRS data reveals that the regulation of termination at *RPP7* is more complex than previously reported. We detect ten distinct poly(A) sites at five positions distinguishable in Helicos and nanopore DRS data (Figure 8A). We found at least two poly(A) sites located within the *COPIA-R7* element, including one characteristic of the previously described ECL transcript. Selection of both poly(A) sites was more frequent in *fpa-8* (Figure 8A, Figure 8 - supplement 1A-B, Supplementary file 3). A third poly(A) site is located immediately downstream of the *COPIA-R7* element. We also identified at least two poly(A) sites within the second intron of *RPP7*, selection of which is reciprocally regulated by FPA, with a moderate decrease in *fpa-8* and an increase in *35S::FPA:YFP* (Figure 8A, Figure 8 - supplement 1A-B, Supplementary file 3). All of these intronic proximal poly(A) sites occur before the start of the *RPP7* open reading frame, and their selection would generate truncated transcripts that do not encode RPP7 protein. At the 3' end of *RPP7*, we found three alternative poly(A) sites located in the terminal intron (within the 3' UTR sequence), in addition to the previously identified poly(A) site in the terminal exon (Figure 8A, Inset 1, Supplementary file 3) (Tsuchiya & Eulgem, 2013). Selection of two of these poly(A) sites is associated with cryptic splicing events that lead to the generation of four possible 3'UTR sequences associated with distally processed *RPP7* transcripts. Selection of these 3'UTR intronic poly(A) sites is suppressed by FPA, with increased and decreased usage in *fpa-8* and the FPA overexpressing line respectively (Figure 8A, Inset 1; Figure 8 - supplement 1A-B, Supplementary file 3). Based on these data, we conclude that FPA influences *RPP7* intronic polyadenylation at a larger number of poly(A) sites than previously considered.

The major effect of FPA on *RPP7* is within the protein-coding exon 6 of *RPP7*. We identified three poly(A) sites located within the protein-coding sixth exon of *RPP7* (Figure 8A). The first two of these are located at the end of the region encoding the NB-ARC domain, and the third is positioned downstream, within the region encoding the LRR repeats. Cleavage and polyadenylation at these three sites results in transcripts without in-frame stop codons, disrupting the coding potential and translatability of *RPP7* mRNA (Figure 8A, Inset 2). These poly(A) sites were identified in Helicos and nanopore DRS data, indicating that they are unlikely to be caused by alignment errors. Approximately 25% of all Col-0 nanopore DRS reads terminate at one of the three poly(A) sites within the *RPP7* sixth exon. The relative selection of these poly(A) sites depends on FPA expression (Figure 8A, Inset 2, Figure 8 - supplement 1A-B, Supplementary file 3). When FPA is overexpressed, approximately 63% of all *RPP7* transcripts identified by nanopore DRS terminate at proximal poly(A) sites within exon 6, and only 25% of transcripts detected are expected to encode the full *RPP7* open reading frame (Figure 8A, Figure 8 - supplement 1A, Supplementary file 3). A clear relative drop in coverage caused by the selection of these exonic proximal poly(A)

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

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 fpa-8* in light blue; and *35S::FPA:YFP fpa-8* in green. Error bars

are 95% confidence intervals. Each experimental replicate was generated from 7-45 plants per genotype.

FPA modulates *RPP7*-dependent, race-specific susceptibility to the pathogen *Hyaloperonospora arabidopsidis*

Since the major impact of FPA on *RPP7* is to limit the production of full-length protein coding transcripts, we reasoned that immunity might be compromised when FPA is overexpressed. To test this hypothesis, we carried out pathogenesis assays using the oomycete *Hpa*-Hiks1, against which *RPP7* provides race-specific immunity. We inoculated Arabidopsis seedlings with *Hpa*-Hiks1 spores in three independent experiments and checked susceptibility by counting the number of sporangiophores four days after inoculation. We compared the susceptibility of six Arabidopsis genotypes: Col-0, Ksk-1, FPA loss-of-function mutant alleles *fpa-7* and *fpa-8*, a complementing line *pFPA::FPA* expressing FPA gene under the control of its native promoter (Zhang *et al*, 2016) and the FPA overexpressing line *35S::FPA:YFP*, both of which were introduced into the *fpa-8* mutant allele. *fpa-7* is a Transfer-DNA insertion mutant allele (Alonso *et al*, 2003), whilst *fpa-8* is derived from an ethyl methanesulfonate (EMS)-induced mutant screen (Baurle *et al*, 2007). With the exception of the Ksk-1 accession, all these mutant alleles and transgenic lines are in a Col-0 background. Ksk-1 lacks *RPP7* and is therefore an *Hpa*-Hiks1 susceptible ecotype used here as a positive control (median 5 sporangiophores per plant) (Figure 8C). Col-0 was used as a resistant control (median 0 sporangiophores per plant). Like Col-0, the *fpa-7* mutant displayed inhibited sporulation of *Hpa*-Hiks1 (median 0 sporangiophores per plant), suggesting that loss of FPA function does not reduce plant resistance. Greater variability in pathogen susceptibility was observed in the *fpa-8* mutant and was not restored by complementation with *pFPA::FPA*, possibly indicating background EMS mutations affecting susceptibility. Both lines show low median susceptibility, however (median 0 sporangiophores per plant). In contrast, *35S::FPA:YFP* plants exhibited a similar level of sporulation to the pathogen-sensitive *Ksk-1* accession (median 3 sporangiophores per plant). This suggests that the premature exonic termination of *RPP7* caused by FPA has a functional consequence for Arabidopsis immunity against *Hpa*-Hiks1.

Discussion

This study reveals that the predominant role of the RNA-binding protein FPA is to promote poly(A) site choice and transcription termination, and links this process to widespread control of Arabidopsis NLR immune-response genes. Using *IVI-MS* proteomics, we showed that FPA is closely associated with proteins involved in RNA 3' end processing, m⁶A modification and transcription termination. Metagene analysis of ChIP-Seq data revealed that FPA co-localised on chromatin with Ser² phosphorylated Pol II at the 3' ends of Arabidopsis genes. The analysis of three genetic backgrounds with different FPA activity

levels, using three orthogonal RNA sequencing technologies, revealed that FPA predominantly promotes poly(A) site choice and does not affect other aspects of RNA processing or modification. Almost half of Arabidopsis NLR loci with detectable levels of expression were sensitive to FPA, and in most cases, FPA targeted poly(A) sites within protein-coding exons, limiting the expression of full-length NLR proteins. For example, FPA promoted premature cleavage and polyadenylation within protein-coding exon 6 of *RPP7* transcripts, functionally compromising *Hpa*-Hiks1 strain-specific immunity.

Uncovering protein assemblies that mediate 3' end processing in living plant cells

We used an *in vivo* FA cross-linking approach to identify proteins localised in the neighbourhood of FPA inside living plant cells. We discovered that FPA co-purifies with factors involved in processing and modifying mRNA 3' ends and promoting transcription termination. These include almost every conserved component of the cleavage and polyadenylation machinery and RNA-binding proteins such as HLP1 (Zhang *et al.*, 2015) that, like FPA, functions in 3' end formation and controls the expression of the flowering time transcription factor FLC. We found factors involved in modulating the phosphorylation state of Pol II at the 3' end of genes and components of the m⁶A writer complex which target mRNA 3' UTRs (Parker *et al.*, 2020b). Overall, these data not only reveal proteins closely associated with FPA, but also provide the most in depth view of proteins involved in Arabidopsis RNA 3' end processing. As a result, this dataset provides clues to the function of some of the uncharacterised proteins detected here.

Components of the m⁶A writer complex co-purify with FPA. However, we found that FPA did not affect global levels of this modification in Arabidopsis mRNAs. In contrast, FPA-related spen family proteins in humans and *Drosophila* are required for m⁶A deposition (Knuckles *et al.*, 2018; Patil *et al.*, 2016). Clearly, distinct spen protein functions have evolved in different species.

The small size of FA enables extensive cross-linking but may not necessarily ensure comprehensive recovery of protein assemblies. Likewise, depth of MS/MS coverage will influence the sensitivity of protein detection. Consequently, technical reasons may explain why we did not detect particular CPF components (e.g. CPSF30 and FIP1), or FLL2, which was recently reported to mediate liquid phase transitions of the Arabidopsis 3' end processing machinery (Fang *et al.*, 2019). In addition, we did not detect any members of the PAF1 (RNAPII-associated factor 1) complex, which were first identified in Arabidopsis as genetic suppressors of late flowering autonomous pathway mutants (Oh *et al.*, 2004). An alternative explanation for the absence of some proteins detected here might be that FPA functions to exclude them as part of the mechanism by which it mediates poly(A) site choice.

Two Arabidopsis PCF11 paralogs with Pol II CTD-interacting domains (CIDs), PCFS2 and PCFS4, were found to co-purify with FPA, but two paralogs which lack CIDs, PCFS1 and PCFS5, were not. PCF11 has been shown to have functionally separable roles in transcription termination and cleavage and polyadenylation (Sadowski *et al.*, 2003). The N-

terminal PCF11 CID is required for transcription termination, whilst the C-terminal domains are required for cleavage and polyadenylation. The specific interaction of FPA with CID-containing PCF11 paralogs may suggest that FPA regulates alternative polyadenylation by altering Pol II speed and termination. Recent biochemical studies have demonstrated that the human SPOC domain protein PHF3 is able to bind to two adjacent Ser² phosphorylated heptads of the CTD of Pol II, using two electropositive patches on the surface of the SPOC domain (Appel, Franke et al. 2020). One of these patches, and the key amino acid residues within it, is conserved in the structure of the FPA SPOC domain (Zhang *et al.*, 2016). Consequently, FPA might also be able to interact with the CTD, possibly in conjunction with CID domains of PCFS2 and PCFS4. Such interactions could explain the specific co-association of FPA and Pol II phosphorylated at Ser² of the CTD at the 3' end of genes.

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

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

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

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

We speculate that the diversity of NLR transcript processing might also facilitate pathogen activity detection in a conceptually related way to integrated protein domain decoys but occurring at the RNA level. For example, the formation of chimeric RNAs that comprise NLR transcripts with extremely long 3'UTRs may facilitate the detection of pathogens targeting NMD. Overall, it will be interesting to compare NLR patterns of cleavage, polyadenylation and transcription termination across related Arabidopsis accessions and species to resolve the contribution of controlled transcription termination to NLR evolutionary dynamics.

We focused on *RPP7* to understand complex NLR regulation. Although *RPP7* expression has been well studied, our combinatorial sequencing analyses clarified the complexity of processing events at the *RPP7* locus. Our data supports the established pattern of alternative polyadenylation at intragenic heterochromatin within *RPP7* intron 1 (Tsuchiya & Eulgem, 2013). We revealed additional complexity that included FPA-dependent promotion of cleavage and polyadenylation within protein coding exon 6. The impact of FPA at exon 6 reduces the expression of full-length *RPP7* transcripts, with functional consequences for strain-specific resistance to the oomycete pathogen *Hpa-Hiks1*. Consequently, two distinct patterns of alternative polyadenylation mediate *RPP7* regulation, one involving FPA, and another involving intragenic heterochromatin.

RPP7 is a recurring partner in examples of hybrid weakness, or necrosis (Bomblies & Weigel, 2007). In these cases, autoimmunity is explained by allele-specific physical interactions between *RPP7* protein and the atypical disease resistance protein *RPW8/HR* (Li *et al.*, 2020). We found that not only is *RPP7* pre-mRNA targeted by FPA-dependent premature transcript termination, but so too is *RPW8/HR*. This raises the possibility that FPA could "rescue" hybrid weakness by limiting the expression of these proteins. Consequently, controlling NLR transcription termination may modulate the balance of costs and benefits of NLR gene function.

The impact of FPA overexpression revealed here derives from artificial transgene expression. However, pathogens could similarly attenuate NLR activity by evolving effectors that modulate the expression or activity of factors controlling NLR poly(A) site choice. Consistent with this idea, the HopU1 effector of *P. syringae* targets the RNA-binding protein, AtGRP7 (Fu *et al.*, 2007), which co-purified with FPA. In addition, the Pi4089 effector of the oomycete pathogen *Phytophthora infestans* targets the KH-domain RNA-binding protein StKRBP1 in potato. As a result, the abundance of StKRBP1 increases, resulting in enhanced infection by *P. infestans* (Wang *et al.*, 2015). This precedent reveals that effector-mediated increases in RNA-binding protein abundance can transform host RNA-binding proteins into susceptibility factors. Phylogenetic analysis of StKRBP1 suggest that a direct homolog is absent in Brassicaceae but that the most related Arabidopsis proteins are FLK and PEP, both of which co-purify with FPA and have been shown to regulate polyadenylation (Rodriguez-Cazorla *et al.*, 2015). FPA, GRP7, FLK and PEP, along with other RNA-binding proteins, act in concert to fine tune the timing of flowering through the 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.

Materials and methods

Plants

Plant material and growth conditions

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

IVI-MS

Preparation of IVI-MS libraries

14-day-old seedlings were harvested and cross-linked with 1 % (v/v) formaldehyde for 15 min under vacuum. The cross-linking reaction was stopped by the addition of glycine to a final concentration of 0.125 M under vacuum for a further 5 min. Nuclei were isolated from frozen ground plant tissue using Honda buffer (20 mM Hepes-KOH pH 7.4, 10 mM MgCl₂, 440 mM sucrose, 1.25 % (w/v) Ficoll, 2.5 % (w/v) Dextran T40, 0.5 % (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF, 1% (v/v) Protease Inhibitor Cocktail [Sigma]) and collected by centrifugation at 2,000g for 17 min at 4°C. Nuclei were washed twice with Honda buffer (centrifugation at 1,500g for 15 min at 4°C between washes) and lysed in nuclei lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1 % (w/v) SDS, 1 mM PMSF, 1 % (v/v) Protease Inhibitor Cocktail) by sonication for 4 cycles of 30 s pulses with low power and 60 s cooling between pulses using the Bioruptor UCD-200 (Diagenode). Following centrifugation (16,100g for 10 min at 4°C), the supernatant was diluted ten-fold with sample dilution buffer (16.7 mM Tris-HCl pH 8, 167 mM NaCl, 1.1 % (v/v) Triton X-100, 1 % (v/v) Protease Inhibitor Cocktail). Cross-linked protein complexes were isolated with GFP-trap agarose beads (Chromotek), incubated at 4°C with constant rotation for 5 h, followed by centrifugation (141g for 3 min at 4°C). Beads were washed three times with washing buffer (150 mM NaCl, 20 mM Tris-HCl pH 8, 2 mM EDTA pH 8, 1 % (v/v) Triton X-100, 0.1 % (w/v) SDS, 1 mM PMSF) by centrifugations between washes (400g for 3 min at 4°C). Samples were incubated at 90°C for 30 min to reverse cross-linking prior to SDS-PAGE. Each biological replicate was separated into 5 fractions following SDS-PAGE, then subjected to in-gel digestion with trypsin and submitted for LC-MS/MS analysis (LTQ Orbitrap Velos Pro mass spectrometer (ThermoFisher Scientific)). Three biological replicates were performed for each genotype.

IVI-MS data analysis

Raw peptide data files from mass spectrometry were analyzed by MaxQuant software (version 1.6.10.43) (Cox & Mann, 2008). Peptide tables were then loaded using Proteus (version 0.2.14) (Gierlinski *et al.*, 2018) and summarised to protein level counts using the hi-flyer method (mean of the top 3 most abundant peptides). Because wild-type plants lacking GFP were used as controls, a large number of proteins that are enriched by immunoprecipitation are below the detection threshold in the control. This group of proteins can be classified as “missing not at random” (MNAR). In all proteomics experiments, there will also be a number of proteins which are not detected purely by chance, and these are referred to as “missing at random” (MAR). We treated proteins that were missing from all replicates of a condition as MNAR, and proteins that were missing moved only from a subset of replicates as MAR. Using the imputeLCMD package (version 2.0) (Lazar, 2015), a K nearest neighbours’ strategy was used to impute MAR examples, and a quantile regression imputation of left centred data (QRILC) approach was used to impute MNAR examples. Differential expression analysis was performed on imputed data using limma (version 3.40.0) (Ritchie *et al.*, 2015). Because imputation is not deterministic (i.e. will lead to different outcomes every time), we improved the robustness of the analysis by performing 999 bootstraps of the imputation and differential expression, and summarising the results using the median log fold change and harmonic mean p value.

ChIP-Seq

Preparation of libraries for ChIP-Seq

ChIP against FPA and Pol II phosphorylated at either Ser⁵ or Ser² of the CTD heptad repeat was performed as previously described (Yu *et al.*, 2019). Polyclonal antibodies against FPA amino acids 536-901 were raised in rabbit by Covance.

ChIP-Seq data processing

FPA and PolII ChIP-Seq data is available at ENA accession PRJNA449914. H3K9me² ChIP-seq data was downloaded from ENA accessions PRJDB5192 (Inagaki *et al.*, 2017) and PRJNA427432 (Lai *et al.*, 2020). Reads were aligned to the TAIR10 reference genome using Bowtie2 (version 2.3.5.1) (Langmead & Salzberg, 2012), with the parameters --mm --very-sensitive --maxins 800 --no-mixed --no-discordant. Counts per million normalised coverage profiles were generated using deeptools (version 3.4.3) (Ramirez *et al.*, 2014). For 3’ end centred metagene profiles, we determined the major 3’ position per gene using the Araport11 annotation and existing Col-0 Helicos DRS data (Sherstnev *et al.*, 2012). Metagenes centred on these positions were then generated in Python 3.6 using pyBigWig (version 0.3.17) (Ramirez *et al.*, 2014), numpy (version 1.18.1) (Harris *et al.*, 2020) and matplotlib (version 3.1.3) (Hunter, 2007). For differential H3K9me² analysis, read counts per gene (including intronic regions) were generated using pysam (version 0.16.0), and differential expression analysis was performed using edgeR (version 3.22.5) (Robinson *et al.*, 2010).

RNA

Total RNA isolation

Following manufacturers' instructions, total RNA was isolated using RNeasy Plant Mini kit (QIAGEN) and treated with TURBO DNase (ThermoFisher Scientific). The total RNA concentration was measured using a Qubit 1.0 Fluorometer and Qubit RNA BR Assay Kit (ThermoFisher Scientific), while RNA quality and integrity was assessed using the NanoDrop™ 2000 spectrophotometer (ThermoFisher Scientific) and Agilent 2200 TapeStation System (Agilent).

Nanopore DRS

Preparation of libraries for direct RNA sequencing using nanopores

Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as detailed above. mRNA isolation and preparation of nanopore DRS libraries (using the SQK-RNA001 nanopore DRS Kit (Oxford Nanopore Technologies) were performed as previously described (Parker *et al.*, 2020b). Libraries were loaded onto R9.4 SpotON Flow Cells (Oxford Nanopore Technologies) and sequenced using a 48-h runtime. Four biological replicates were performed for each genotype.

Nanopore DRS data processing

Nanopore DRS reads were basecalled using Guppy (version 3.6.0) high accuracy model. Reads were mapped to the Arabidopsis *TAIR10* genome (The Arabidopsis Genome, 2000) using minimap2 (version 2.17) with the parameters -a -L -cs=short -x splice -G20000 -end-seed-pen 12 -junc-bonus 12 -uf. Spliced alignment was guided using junctions from the Araport11 annotation (Cheng *et al.*, 2017). Nanopore DRS reads can suffer from "oversplitting" - where signal originating from a single RNA molecule is incorrectly interpreted as two or more reads (Parker *et al.*, 2020b). These errors can be systematic and can result in false positive 3' ends. To filter these errors, we identified reads which were sequenced consecutively through the same pore and also mapped contiguously on the genome (within 1 kb of each other). We therefore filtered all but the most 3' reads, which should contain the genuine RNA 3' end. Pipelines for processing of nanopore DRS data were built using Snakemake (Koster & Rahmann, 2012).

Helicos DRS

Preparation of samples for Helicos DRS

Total RNA was isolated from the Col-0, *fpa-8* and *35S::FPA:YFP* seedlings as specified above. Next, samples were processed by Helicos BioSciences as previously described (Ozsolak *et al.*, 2009; Sherstnev *et al.*, 2012). Three biological replicates were performed for each genotype.

Helicos DRS data processing

Helicos DRS reads were mapped to the *Arabidopsis* TAIR10 genome using Heliosphere (version 1.1.498.63) as previously described (Sherstnev *et al.*, 2012). Reads were filtered to remove those with >4 nt insertions or deletions, and to mask regions with low complexity as determined using dustmasker (Camacho *et al.*, 2009) (from BLAST+ suite version 2.10.1) set at DUST level 15 (Sherstnev *et al.*, 2012)

Differential 3' end analysis of Nanopore and Helicos DRS datasets

Transcriptional loci were first discovered in Col-0, *fpa-8* and 35S::*FPA:YFP* nanopore DRS reads using the long-read transcript assembly tool StringTie2 version 2.1.1 (Pertea *et al.*, 2015). Novel transcriptional loci were merged with annotated loci from the Araport11 reference (Cheng *et al.*, 2017). To detect sites with altered 3' end distributions in *fpa-8* and 35S::*FPA:YFP* mutants, we pooled the replicates of either nanopore or Helicos DRS data and identified reads overlapping each transcriptional locus. These reads were used to build distributions of 3' end locations. The difference in 3' end distributions between the treatment and control (Col-0) was measured using the earth mover distance (EMD, also known as the Wasserstein distance). To identify loci with statistically significant changes in 3' distributions, we performed an EMD permutation test using 999 bootstraps - reads for each locus were randomly shuffled between the control and treatment samples to create null distributions, and the EMD recalculated. The histogram of null EMDs was fitted using a gamma distribution, and the *p*-value was calculated from the distribution (the probability of achieving the observed EMD or greater by chance). *p*-values were corrected for multiple testing using the Benjamini-Hochberg method. Genes with an EMD ≥ 25 and an FDR < 0.05 were considered differentially alternatively polyadenylated, and directionality of change was identified using the difference in mean 3' position. Software developed to perform differential 3' analysis is available on GitHub at <https://github.com/bartongroup/d3pendr> and [10.5281/zenodo.4319113](https://doi.org/10.5281/zenodo.4319113), and can be used with nanopore DRS, Helicos DRS, or Illumina 3' tag-based datasets.

Illumina RNA sequencing

Preparation of libraries for Illumina RNA sequencing

Total RNA was isolated from the Col-0, *fpa-8* and 35S::*FPA:YFP* seedlings as described above. mRNA isolation and preparation of the sequencing libraries were performed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) by the Centre for Genomic Research (University of Liverpool). 150-bp paired-end sequencing was carried out on Illumina HiSeq 4000. Six biological replicates were performed for each genotype.

Illumina RNA sequencing data processing

Illumina RNA-Seq data was assessed for quality using FastQC (version 0.11.9) and MultiQC (version 1.8) (Andrews, 2010; Ewels *et al.*, 2016). Reads were mapped to the TAIR10 genome using STAR (version 2.7.3a) (Dobin *et al.*, 2013) with a splice junction database generated from the Araport11 reference annotation (Cheng *et al.*, 2017). Counts per million normalised coverage tracks were created using samtools (version 1.10) and deeptools (version 3.4.3) (Ramirez *et al.*, 2014). To identify expressed regions in each locus, coverage profiles for each replicate of control and treatment were first extracted using pyBigWig (version 0.3.17) (Ramirez *et al.*, 2014). These were normalised so that the area under each profile was equal to the mean area under the profiles. A normalised coverage threshold of 1 was used to identify expressed regions of the loci. These regions were further segmented where there were changes in expression between control and treatment conditions of at least two-fold in a 25 nt window. Regions which were smaller than 50 nt were removed. Expression of the regions was then calculated using featureCounts (version 2.0.0) (Liao *et al.*, 2013) and differential usage within transcriptional loci was assessed using DEXseq (version 1.32.0) (Reyes *et al.*, 2013). Loci were considered differentially processed if they had a locus-level FDR < 0.05 and at least one region with an absolute logFC > 1 and FDR < 0.05. To conduct differential splice junction usage analysis, counts of splice junctions annotated in the bespoke nanopore DRS-derived annotation, plus Araport11 annotation, were generated for each locus using featureCounts (version 2.0.0) (Liao *et al.*, 2013). Differential splice junction usage was assessed using DEXseq (version 1.32.0) (Reyes *et al.*, 2013). Loci were considered differentially spliced if they had a locus-level FDR < 0.05 and at least one junction with an absolute logFC > 1 and FDR < 0.05.

Gene tracks

Gene track figures were generated in Python 3.6 using matplotlib (version 3.1.3) (Hunter, 2007). For gene tracks where any condition had >200 nanopore DRS read alignments, 200 representative alignments were selected by random sampling without replacement (with the exception of the *FPA* figure, where 500 read alignments were sampled). nanoPARE data (Schon *et al.*, 2018) was processed as described in (Parker *et al.*, 2020b).

Protein domain family enrichment analysis

To conduct protein domain enrichment analysis, Interpro domain annotations of Arabidopsis proteins were downloaded from bioMart (Smedley *et al.*, 2009) and converted to genomic co-ordinates using the Araport11 annotation (Cheng *et al.*, 2017). Domain families overlapping each locus tested for alternative polyadenylation using either nanopore or Helicos DRS were identified using pybedtools (version 0.8.1) (Dale *et al.*, 2011). To identify enriched domain families, domains were randomly shuffled between tested loci in 19,999 bootstraps, and the number of times each domain class overlapped significantly alternatively polyadenylated loci by chance was recorded. This was compared to the

observed overlap of each domain family with alternatively polyadenylated loci to calculate p values, which were corrected for multiple testing using the Benjamini Hochberg method.

Manual annotation of alternatively polyadenylated NLR genes

To identify which of the 206 previously annotated NLR genes present in Col-0 were alternatively polyadenylated in *fpa-8* and *35S::FPA*, we devised a standard operating procedure for visual inspection. Genes which had nanopore DRS read coverage in at least two conditions were considered expressed. Genes were considered alternatively polyadenylated if they had multiple 3' end locations each supported by at least 4 nanopore DRS reads, and if there was a clear change in nanopore DRS coverage in the treatment condition compared to Col-0.

RNA gel blot analysis of *RPP7* mRNAs

RNA gel blot analysis was carried out as previously described (Quesada *et al*, 2003) with minor modifications. *RPP7* mRNA was detected using a probe annealing to the second exon of *RPP7* (*AT1G58602*) gene (200 bp PCR product amplified with the following primers: Forward: 5'-TCGCGGACTACTACTCAAGA-3' and Reverse: 5'-TCTTGATGGTGTGAAAGAATCTAGT-3'). As a loading control, *Beta-TUBULIN* mRNA was used, which was visualised by a probe annealing to the third exon of the *Beta-TUBULIN* (*AT1G20010*) gene (550 bp PCR product amplified with the following primers: Forward: 5'-CTGACCTCAGGAACTCGCG-3' and Reverse: 5'-CATCAGCAGTAGCATCTTGG-3'). The probes were 5' labelled using [γ -³²P]-ATP (Perkin Elmer) and DECAprime™ II DNA labelling kit (ThermoFisher Scientific) and purified on Illustra G-50 columns (GE Healthcare Life Sciences). mRNA isoforms were visualized and quantified using Amersham Typhoon Gel and Blot Imaging System (GE Healthcare Bio-Sciences AB). The RiboRuler High Range RNA Ladder (ThermoFisher Scientific), used to identify approximate size of RNA bands, was first dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher Scientific) and then labelled with [γ -³²P]-ATP (Perkin Elmer) using T4 Polynucleotide Kinase (ThermoFisher Scientific) prior being loaded on a gel.

m⁶A liquid chromatography-mass spectroscopy analysis (m⁶A LC-MS/MS)

Total RNA was isolated and checked as described above. mRNA was extracted twice from approximately 75 μ g of total RNA using the Dynabeads mRNA purification kit (Thermo Fisher Scientific) following the manufacturer's instructions. The quality and quantity of mRNA was assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and Agilent 2200 TapeStation System (Agilent). Samples for the m⁶A LC-MS/MS were prepared as previously described (Huang *et al*, 2018) with several modifications. 100 ng of mRNA were diluted in a total volume of 14 μ l of nuclease-free water (ThermoFisher Scientific) and digested by nuclease P1 (1U, Merck) in 25 μ l of buffer containing 20 mM of NH₄OAc (pH 5.3) at 42°C for 2 h. Next, 3 μ l of freshly made 1M NH₄HCO₃ and Calf Intestinal

Alkaline Phosphatase (1U, New England Biolabs) were added, and samples were incubated at 37°C for 2 h. Then, the samples were diluted to 50 µl with nuclease-free water and filtered (0.22 µm pore size, 4 mm diameter, Millipore). LC-MS/MS was carried out by the FingerPrints Proteomics facility at the University of Dundee. m⁶A to A ratio quantification was performed in comparison with the curves obtained from pure A and m⁶A nucleoside standards. Statistical analysis was performed using a 2-way t-test.

Pathogenesis assays

The pathogenesis assays were carried out as previously described (Tome *et al*, 2014). The *Hpa*-Hiks1 isolate was maintained by weekly sub-culturing on the Ksk-1 plants. A solution containing *Hpa*-Hiks1 spores was used to inoculate the Col-0, Ksk-1, *fpa-7*, *fpa-8*, *pFPA::FPA* and *35S::FPA:YFP* 14-day-old seedlings. Sporangiophores were counted four days after inoculation. Three biological replicates were performed for each genotype using up to 45 plants per line per each repeat. Statistical analysis was performed using a 2-way t-test.

Code availability

All pipelines, scripts and notebooks used to generate figures are available from GitHub at https://github.com/bartongroup/Simpson_Barton_FPA_NLRs and Zenodo at [10.5281/zenodo.4319109](https://doi.org/10.5281/zenodo.4319109). The software tool developed for detecting changes in poly(A) site choice in Nanopore and Helicos DRS data is available from GitHub at <https://github.com/bartongroup/d3pendr> and Zenodo at [10.5281/zenodo.4319113](https://doi.org/10.5281/zenodo.4319113).

Data availability

IVI-MS data is available from the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al*, 2019) partner repository with the dataset identifier PXD022684. FPA and Pol II ChIP-Seq data is available from ENA accession PRJNA449914 (Yu *et al*, 2019). Col-0 nanopore DRS data is available from ENA accession PRJEB32782 (Parker *et al*, 2020b). *fpa-8* and *35S::FPA:YFP* nanopore DRS data is available from ENA accession PRJEB41451. *hen2-2* nanopore DRS data is available from ENA accession PRJEB41381. Col-0, *fpa-8* and *35S::FPA:YFP* Helicos DRS data is available from Zenodo DOI [10.5281/zenodo.4309752](https://doi.org/10.5281/zenodo.4309752) ahead of submission to ENA. Col-0, *fpa-8* and *35S::FPA:YFP* Illumina RNA-Seq data is available from ENA accession PRJEB41455.

Acknowledgments

We thank Paul Birch and Ingo Hein for comments on the manuscript, and David Baulcombe, Ian Henderson and Wenbo Ma for helpful NLR discussions. We thank Dr Abdelmadjid Atrih (Centre for Advanced Scientific Technologies, School of Life Sciences) for m⁶A LC-MS/MS analysis.

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 Proteomics 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 N⁶-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, Villaseñor
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/FI(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

1370 Lazar C (2015) imputeLCMD: a collection of methods for left-censored missing data
1371 imputation. *R package, version 2*
1372 Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, Zhou B, Levailant M,
1373 Adachi H, Yoshioka H, Raffaele S, Berthome R, Coute Y, Parker JE, Deslandes L (2015) A
1374 receptor pair with an integrated decoy converts pathogen disabling of transcription factors
1375 to immunity. *Cell* 161: 1074-1088
1376 Li L, Habring A, Wang K, Weigel D (2020) Atypical Resistance Protein RPW8/HR Triggers
1377 Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity. *Cell Host Microbe*
1378 Liao Y, Smyth GK, Shi W (2013) The Subread aligner: fast, accurate and scalable read
1379 mapping by seed-and-vote. *Nucleic Acids Res* 41: e108-e108
1380 Liu F, Quesada V, Crevillen P, Baurle I, Swiezewski S, Dean C (2007) The Arabidopsis RNA-
1381 binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC.
1382 *Mol Cell* 28: 398-407
1383 Liu FQ, Marquardt S, Lister C, Swiezewski S, Dean C (2010) Targeted 3' Processing of
1384 Antisense Transcripts Triggers Arabidopsis FLC Chromatin Silencing. *Science* 327: 94-97
1385 Ma W, Zhen G, Xie W, Mayr C (2020)
1386 Mayr C (2019) What Are 3' UTRs Doing? *Cold Spring Harb Perspect Biol* 11
1387 Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-
1388 Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F,
1389 Marchler-Bauer A, Mi H, Natale DA et al. (2019) InterPro in 2019: improving coverage,
1390 classification and access to protein sequence annotations. *Nucleic Acids Res* 47: D351-D360
1391 Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T (2009) An Arabidopsis jmjC
1392 domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J* 28:
1393 1078-86
1394 Mockler TC, Yu X, Shalitin D, Parikh D, Michael TP, Liou J, Huang J, Smith Z, Alonso JM, Ecker
1395 JR, Chory J, Lin C (2004) Regulation of flowering time in Arabidopsis by K homology domain
1396 proteins. *Proc Natl Acad Sci U S A* 101: 12759-64
1397 Oh S, Zhang H, Ludwig P, van Nocker S (2004) A mechanism related to the yeast
1398 transcriptional regulator Paf1c is required for expression of the Arabidopsis FLC/MAF MADS
1399 box gene family. *Plant Cell* 16: 2940-53
1400 Ozsolak F, Platt AR, Jones DR, Reifenger JG, Sass LE, McInerney P, Thompson JF, Bowers
1401 J, Jarosz M, Milos PM (2009) Direct RNA sequencing. *Nature* 461: 814-8
1402 Parker MT, Barton GJ, Simpson GG (2020a)
1403 Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, Hall AJ, Barton GJ,
1404 Simpson GG (2020b) Nanopore direct RNA sequencing maps the complexity of Arabidopsis
1405 mRNA processing and m(6)A modification. *Elife* 9
1406 Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR (2016) m(6)A RNA
1407 methylation promotes XIST-mediated transcriptional repression. *Nature* 537: 369-373
1408 Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A,
1409 Griss J, Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S,
1410 Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF et al. (2019) The PRIDE database and
1411 related tools and resources in 2019: improving support for quantification data. *Nucleic Acids*
1412 *Res* 47: D442-D450
1413 Perteu M, Perteu GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie
1414 enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*
1415 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 m⁶A 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

Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24: 859-74

Simpson GG (2004) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol* 7: 570-574

Simpson GG, Dijkwel PP, Quesada V, Henderson I, Dean C (2003) FY is an RNA 3' end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* 113: 777-87

Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A (2009) BioMart--biological queries made easy. *BMC Genomics* 10: 22

Streitner C, Danisman S, Wehrle F, Schoning JC, Alfano JR, Staiger D (2008) The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in *Arabidopsis thaliana*. *Plant Journal* 56: 239-250

Takagi M, Iwamoto N, Kubo Y, Morimoto T, Takagi H, Takahashi F, Nishiuchi T, Tanaka K, Taji T, Kaminaka H, Shinozaki K, Akimitsu K, Terauchi R, Shirasu K, Ichimura K (2020) Arabidopsis SMN2/HEN2, Encoding DEAD-Box RNA Helicase, Governs Proper Expression of the Resistance Gene SMN1/RPS6 and Is Involved in Dwarf, Autoimmune Phenotypes of mekk1 and mpk4 Mutants. *Plant Cell Physiol* 61: 1507-1516

Tamborski J, Krasileva KV (2020) Evolution of Plant NLRs: From Natural History to Precise Modifications. *Annu Rev Plant Biol* 71: 355-378

The Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815

Tian B, Manley JL (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* 18: 18-30

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423: 74-7

Tome DF, Steinbrenner J, Beynon JL (2014) A growth quantification assay for *Hyaloperonospora arabidopsidis* isolates in *Arabidopsis thaliana*. *Methods Mol Biol* 1127: 145-58

Tran Vdu T, Souiai O, Romero-Barrios N, Crespi M, Gautheret D (2016) Detection of generic differential RNA processing events from RNA-seq data. *RNA Biol* 13: 59-67

Tsuchiya T, Eulgem T (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci U S A* 110: E3535-43

Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl JL, Weigel D, Bemm F (2019) A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis thaliana*. *Cell* 178: 1260-1272 e14

van der Biezen EA, Freddie CT, Kahn K, Parker JE, Jones JD (2002) Arabidopsis RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant J* 29: 439-51

Wang X, Boevink P, McLellan H, Armstrong M, Bukharova T, Qin Z, Birch PR (2015) A Host KH RNA-Binding Protein Is a Susceptibility Factor Targeted by an RXLR Effector to Promote Late Blight Disease. *Mol Plant* 8: 1385-95

Wang YH, Warren JT, Jr. (2010) Mutations in retrotransposon AtCOPIA4 compromises resistance to *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Genet Mol Biol* 33: 135-40

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

Supplementary Figures

Figure 2 - supplement 1: FPA co-localises with Pol II Ser² at the 3' end of genes

(A) ChIP-Seq metagenome profile showing the normalised occupancy of FPA (green) and Pol II phosphorylated at Ser⁵ (pink) and Ser² (brown) of the CTD relative to the major 3' position of each gene as measured using Helicos DRS. Only short genes (<2.5 kb) are included (n=17,440). The figure was compiled from the same ChIP data used in Figure 2.

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.

Figure 4 - supplement 3: FPA does not affect global mRNA m⁶A methylation level

Box plot showing m⁶A:A ratio analysed using LC-MS/MS. Col-0 is shown in blue; *fpa-8*, in orange; 35S::FPA:YFP, in green; low m⁶A mutant *vir-1* (used as a positive control), in pink.

Figure 4 - supplement 4: FPA regulates RNA 3' end processing at *CFIS2* (AT4G25550) 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 AT5G46490 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]**.