**Direct RNA Sequencing with Nanopores Reveals Arabidopsis Transcriptome Complexity and Maps the m6A Epitranscriptome**

**Understanding genome organization and gene regulation in the model plant Arabidopsis requires insight on RNA transcription, processing and modification. We used nanopores to directly sequence RNA from a wild-type accession and a mutant defective in mRNA methylation (m6A). Here we show that m6A can be mapped in full-length mRNAs transcriptome-wide using nanopore direct RNA sequencing. The same datasets reveal previously unannotated combinations of cap-associated transcription start sites, splicing events, poly(A) site choice, poly(A) tail length and antisense RNAs. Loss of m6A from terminal exons is associated with decreased transcript abundance and increased chimeric RNA formation. A functional consequence of disrupted m6A is a lengthening of the circadian period. We conclude that nanopore direct RNA sequencing can reveal combinations of mRNA processing and modification in full-length single molecule reads. The application of this approach to less well-studied species has the potential to transform our understanding of what their genomes encode.**

Patterns of pre-mRNA processing and base modifications determine eukaryotic mRNA coding potential and fate. Alternative transcripts produced from the same gene can differ in the position of the start site, the site of cleavage and polyadenylation and the combination of exons spliced into the mature mRNA. Collectively termed the epitranscriptome, RNA modifications play crucial context-specific roles in gene expression (*1, 2*). The most abundant internal modification of mRNA is methylation of adenosine at the N6 position (m6A). Revealing RNA modifications and processing combinations is essential to understand gene expression and what genomes really encode.

RNA sequencing (RNAseq) is used to dissect transcriptome complexity. RNAseq involves copying RNA into complementary DNA (cDNA) with reverse transcriptases (RT) and it is the subsequent DNA copies that are sequenced. RNAseq has transformed our understanding of diverse features of transcriptomes, but limitations can include misidentification of 3’ ends through internal priming (*3*), spurious antisense and splicing events produced by RT template switching (*4*) and the inability to detect some base modifications in the copying process (*5*). The fragmentation of RNA prior to short-read sequencing makes the interpretation of the combination of authentic RNA processing events difficult, and it remains an unsolved problem (*6*).

We asked if long read direct RNA sequencing (DRS) with nanopores (*7*) could reveal the complexity of *Arabidopsis thaliana* RNA processing and modifications. In nanopore sequencing, intact RNA is fed through a protein pore (nanopore) by a motor protein. The nanopore sits in a membrane through which an electrical current is passed. As RNA moves through the nanopores, the current is disrupted in a manner by which the raw signal can be interpreted as RNA sequence (*7*). Arabidopsis is a pathfinder model in plant biology, and its genome annotation strongly influences the annotation and understanding of what other plant genomes encode. Viable Arabidopsis mutants defective in RNA processing and modification have helped to establish roles for RNA regulation in controlling fundamental features of plant biology (such as flower development and stress responses) and also facilitate the dissection of features of gene expression that are effectively hidden, for example, if unstable RNAs do not accumulate (*8, 9*).

Here we apply nanopore DRS and Illumina RNAseq to wild-type Arabidopsis (Col-0) and mutants defective in m6A (*10*) and exosome mediated RNA decay (*11*). We reveal mRNA m6A and pre-mRNA processing combinations in full-length Arabidopsis mRNAs transcriptome-wide.

***Nanopore DRS detects long complex mRNAs and short, structured non-coding RNAs***

We purified poly (A)+ RNA from 4 biological replicates of 14-day old *Arabidopsis thaliana* Col-0 seedlings. We incorporated synthetic ERCC RNA Spike-In mixes into all replicates (*12, 13*) and carried out nanopore DRS. Parallel Illumina RNAseq was performed on related material. Using the base-calling and alignment software Guppy (Oxford Nanopore Technologies), we identified around 1 million reads per sample (Table S1). The longest read alignments were 12.7kb for mRNA encoding *AT1G48090* (Figure 1A), spanning 63 exons and 12.8kb for mRNA encoding *At1G67120*, spanning 58 exons (Figure S1A). These represent some of the longest contiguous mRNAs sequenced from Arabidopsis. Among the shortest read-alignments were those spanning genes encoding highly structured non-coding RNAs such as UsnRNAs (Figure 1B). Plant UsnRNAs are transcribed from individual genes with well characterised promoters (*14*). In most cases, UsnRNA read alignments were consistent with the corresponding Araport11 annotation (*15*). However, nanopore sequencing also revealed that U12 snRNA is processed from a longer, spliced precursor RNA. Consistent with this, U12 precursor RNA accumulates in *hen2-2*, a mutant defective in the activity of the RNA exosome which processes non-coding RNAs (*11*) (Figure 1B).

***Base-calling errors in nanopore DRS are non-random***

We used the synthetic ERCC RNA Spike-Ins as internal controls to monitor properties of the sequencing reads. The spike-ins were detected in a quantitative manner (Figure S1B), consistent with the idea that nanopore sequencing is quantitative (*16*). For the portion of the reads that align to the reference, the sequence identity is 92% as measured against the ERCC RNA spike-ins (Figure S1C). The errors showed evidence of base specificity (Figure S1D, E). For example, guanidine was under-represented and uracil over represented relative to the reference nucleotide (nt) distribution. In some situations, this bias could impact the utility of interpreting nanopore sequence errors. In order to address base calling errors, we used the software tool proovread (*17*) and parallel Illumina RNAseq data to error-correct nanopore reads.

***Artefactual splitting of raw signal affects transcript interpretation***

In addition to base-calling errors, we detected artefacts caused by over-splitting of raw signal. As a result of this phenomenon, two distinct transcripts, rather than one, appear to align to the same gene (Figure S1F); the first comprising an apparently novel 3’ end and the second an apparently novel 5’ end. Over-splitting of reads was recently reported in nanopore DNA sequencing (*18*) and can be detected when two reads sequenced consecutively through the same pore map adjacently to the genome (*18*). Over-splitting generally occurs at low frequency (less than 2% reads) and appears to be associated with MinKNOW software. However, RNAs originating from specific gene loci, such as *RH3* (*AT5G26742*), appear to be highly susceptible to this effect with over-splitting affecting up to 20% of reads (Figure S1F).

***Spurious antisense reads are rare or absent in nanopore DRS***

Since only 2 of 9,445 (0.02%) reads mapped antisense to the ERCC RNA Spike-In collection (*12*) and 0 of 19,665 reads mapped anti-sense to the highly expressed gene *RCA* (*AT2G39730*), we concluded that spurious antisense is rare or absent from nanopore data. This simplifies the interpretation of authentic antisense RNAs. This is important in Arabidopsis because the distinction between RT-dependent template switching and authentic antisense RNAs produced by RNA dependent RNA polymerases is not straightforward (*19*). For example, by nanopore DRS, we could identify Arabidopsis long non-coding antisense RNAs, such as those at the auxin efflux carriers *PIN4*, *5* and *7* (Figure 1C). The existence of these previously unannotated antisense RNAs was supported by Illumina RNAseq of wild type Col-0 and the exosome mutant *hen2-2* (Figure 1C), where the abundance of these non-coding antisense RNAs increased 13-fold. Consequently, the low level steady-state accumulation of some antisense RNAs may explain why they are currently unannotated.

***Nanopore DRS confirms sites of RNA 3’ end formation and estimates poly(A) tail length***

The ligation of the motor-protein adapter to RNA 3’ ends results in nanopores sequencing mRNA poly(A) tails first. We used the software tool nanopolish-polyA to estimate poly(A) tail lengths for individual transcripts (*20*). This approach indicated an average length of 76nt for Arabidopsis mRNA poly(A) tails, but with a wide range in estimated lengths for individual mRNAs (95% in the range 13-197nt) (Figure 2A). The generally shorter poly(A) tails of chloroplast and mitochondria-encoded transcripts, that are a feature of RNA decay in these organelles, were also detectable. We found that poly(A) tail length is negatively correlated with gene expression in Arabidopsis (Spearman rho = -0.3, *p*=9.8×10-134, 95% CIs [-0.32, -0.28]) (Figure S2A), consistent with other species previously analysed by short read TAILseq (*21*).

We next identified the sites of mRNA cleavage and polyadenylation. We had previously mapped Arabidopsis mRNA 3’ ends transcriptome-wide using Helicos Bio sequencing (*20*). We compared the position of 3’ ends of nanopore DRS read alignments and Helicos Bio data genome-wide. The median genomic distance between nanopore DRS and Helicos Bio 3’ ends was nt (1 standard deviation) demonstrating close agreement between these orthogonal technologies (Figure 2B). Likewise, the overall distribution of the 3’ end of aligned nanopore DRS reads resembles the pattern we previously reported with Helicos Bio data (*20*). For example, 97% of nanopore DRS 3’ ends (4,152,800 reads at 639,178 unique sites, 93% of all unique sites) mapped to annotated 3’ untranslated regions (UTRs) or downstream of the current annotation. 3’ ends that mapped to coding sequence or 5’UTRs were rare (2.8%, 119,524 reads at 39,610 unique sites, 5.8% of all unique sites). 3’ ends that mapped to introns were even rarer (0.29%, 12,554 reads at 7,791 unique sites, 1.1% of unique sites), but included examples of alternative polyadenylation with well-established regulatory roles, such as at mRNA encoding the RNA binding protein FPA which controls flowering time (*9*) (Figure 2C), and at mRNA encoding the histone H3K9 demethylase IBM1, which controls levels of genic DNA methylation (*21*) (Figure S2B).

Since RT-dependent internal priming can result in the misinterpretation of authentic cleavage and polyadenylation sites (*3*), we asked whether nanopore sequencing was compromised in this way. To address this issue, we examined whether nanopore DRS read 3’ ends mapped to potential internal priming substrates comprised of 6 consecutive adenosines, within transcribed coding sequence (according to Araport11 annotation). Of 10,116 such oligo (A)6 sequences, only 4 have read-alignments terminating within 13nt in all four datasets. Of these, 2 were not detectable after error correction with proovread (suggesting that they resulted from alignment errors) and 2 mapped to the terminal exon of coding sequence annotation, indicating that they may be authentic 3’ ends. Hence, internal priming is rare or absent in nanopore DRS data. Overall, we conclude that nanopore DRS can identify multiple authentic features of RNA 3’ end processing.

***Cap-dependent adapter ligation enables 5’ end detection by nanopore DRS***

In contrast to the clarity of 3’ end detection, nanopore reads are frequently truncated prior to annotated transcription start sites resulting in a 3’ bias of aligned nanopore reads (Figure 3A). Consequently, it is not possible to conclude which, if any, aligned reads correspond to full-length transcripts. To address this issue, we used cap-dependent ligation of a biotinylated 5’ adapter RNA to purify capped mRNAs. We re-sequenced two biological replicates of Arabidopsis Col-0 incorporating 5’ adapter ligation (Table S1). We filtered nanopore DRS reads using the sequence alignment tool BLASTN and specific criteria (Table S2) to detect 5’ adapter RNA sequences. We then used high confidence examples of sequences that passed or failed these filters to train a convolutional neural network to detect the 5’ adapter RNA in raw signal (Figure S3A-C). In this way, we improved the identification of 5’ adapter ligated RNAs without requiring genome alignment. As a result, we demonstrated enrichment of full-length, cap-dependent mRNA sequences, filtered for the presence of the 5’ adapter RNA (Figure 3A, B). This procedure reduced the median 3’ bias of nanopore read alignments per gene (as measured by quartile coefficient of variation of per base coverage) from 0.55 to 0.08 (Figure 3B). In order to determine whether the 5’ ends we detected reflected full-length mRNAs, we compared them against annotated transcription start sites in datasets derived from full-length Arabidopsis cDNA clones (*22*). We found that 41% of adapter ligated nanopore DRS reads mapped within 5nt of transcription start sites and 60% mapped to within 13nt (Figure 3C). We also detected recently defined examples of alternative 5’ transcription start sites (*23*) at specific Arabidopsis genes (Figure S3D). We therefore conclude that this approach is effective in detecting authentic mRNA 5’ ends.

Reads with adapters had, on average, 11nt more mapped sequence at their 5’ ends compared to reads lacking the 5’ adapter RNA (Figure 3D). This distinction may be explained by loss of processive control by the motor protein when the end of an RNA molecule enters the pore. As a result, the 5’ end of RNA is not correctly sequenced. Consistent with our Arabidopsis transcriptome-wide data, reads mapping to the synthetic ERCC RNA Spike-Ins and *in vitro* transcribed RNAs also lacked ~11nt of authentic 5’ sequence (Figures S3E, F). However, the precise length of 5’ sequence missing from all these RNAs varied, suggesting that there are sequence or context-specific effects on sequence accuracy associated with the passage of 5’ RNA through the pore (Figure 3D, S3E, F). Despite the close agreement between nanopore DRS, Illumina RNAseq and full-length cDNA start site data at *RCA*, the annotated start site in Araport11 and AtRTD2 (*24*) is quite different (Figure 3E). This apparent overestimation of 5’UTR length is widespread in Araport11 annotation (Figure S3G). Consequently, with appropriate modification to the protocol, nanopore DRS data can be used to revise Arabidopsis transcription start site annotation.

***Detection of novel splice sites and unannotated combinations of established splicing events***

We next investigated the utility of nanopore DRS to reveal patterns of splicing. In single reads, nanopore sequencing revealed some of the most complex splicing combinations known in the Arabidopsis transcriptome. For example, the splicing pattern of a 12.7kb read alignment, comprised of 63 exons, agreed exactly with the *AT1G48090.4* isoform annotated in Araport11 (Figure 1A). Mutually exclusive alternative splicing of *FLM* *(AT1G77080*)exons that mediate the thermosensitive response controlling flowering time (*25*) was also detected (Figure 4A). However, considering the non-corrected dataset as a whole, a combination of sequencing and alignment errors probably resulted in the mis-calling of splicing events because 58% (170,702) of the unique splice junctions detected in the combined set of replicate data were absent from Araport11 and AtRTD2 annotations, and unsupported by Illumina RNAseq (Figure 4B, Table S3). We applied proovread (*17*) error correction with the parallel Illumina RNAseq data and re-analysed the corrected and uncorrected nanopore DRS data. The four Col-0 biological replicate nanopore DRS datasets capture 75% (102,486) and 69% 104,686) of the Araport11 and AtRTD2 splice site annotations, respectively. Most canonical GU/AG splicing events (100,450; 81%) detected in the error corrected nanopore data were found in both annotations and were supported by Illumina RNAseq (Figure 4B, Table S3). 3,234 unique canonical splicing events in the error corrected nanopore DRS data were supported by Illumina RNAseq, but were absent from both Araport11 and AtRTD2 annotations, highlighting potential gaps in the complexity of Arabidopsis splicing annotation (Figure 4B, Table 3). Consistent with this, we validated 3 of these splicing events by using RT-PCR followed by cloning and sequencing (Figure S4A). In order to examine the features of these unannotated splices, we applied previously determined splice site position weight matrices of the flanking sequences to categorize U2 or 12 class splice sites (*28*) (Figure S4B, C). Of the 3,234 novel events found in error corrected data, which were also supported by Illumina alignments, 73% were classified as canonical U2 or U12 splice sites, reinforcing the idea that they could be authentic.

In addition to previously unannotated splicing events, we identified unannotated combinations of previously established splice sites. For example, we identified 19 *FLM* splicing patterns which all adhered to known splice junction sites (Figure 4A). However, 11 of these transcript isoforms were not previously annotated. In order to investigate this phenomenon transcriptome-wide, we analysed the 5’ cap-dependent nanopore DRS datasets of full-length mRNAs. Unique sets of co-splicing events were extracted from error-corrected reads (in order to focus on splicing, we did not consider single exon reads or 5’ and 3’ positions). The read coverage of the 5’ adapter RNA-ligated datasets is lower than the conventional nanopore DRS (Table S1), but 13,064 unique splicing patterns were detected that matched annotations in one or both of Araport11 and AtRTD2 (Figure 4C). 8,659 unique splicing patterns were identified in the error corrected nanopore data that were not identified in either annotation (Figure 4C, Table S3). Of these, 50% (4,293) used only splice donor and acceptor pairs that were already annotated in either Araport11 or AtRTD2. Hence, this approach defines splicing patterns produced from alternative combinations of known splice sites.

Overall, we conclude that nanopore DRS can reveal the complexity of splicing in the context of full-length mRNAs. However, accurate splice pattern detection requires error correction with, for example, high-accuracy orthogonal sequence data. Accurate splice detection, even with error-free sequences, can be confounded because equivalent alternative junctions often exist (*29*). Therefore, improved computational tools for not only error correction, but also splicing-aware long-read alignment are required.

***Differential error site analysis reveals the m6A epitranscriptome***

The epitranscriptome has emerged recently as a crucial, but relatively neglected, layer of gene regulation (*1, 2*). m6A has been mapped transcriptome-wide using approaches based on antibodies that recognize this mark (*5, 28*). However, in principle, m6A can be detected by nanopore DRS (*7*). Since m6A is not included in the training data for nanopore base-calling software, we asked whether its incorrect interpretation could be used to identify Arabidopsis m6A transcriptome-wide. We applied nanopore DRS to 4 biological replicates of an Arabidopsis mutant defective in the function of a conserved m6A writer complex component, Virilizer (*vir-1*), and 4 biological replicates of a line expressing VIR fused to GFP that restores the activity of VIR in the *vir-1* mutant background (*10*) (Figure S5A). In addition, we sequenced a parallel set of biological replicates with Illumina RNAseq. We then used a G-test statistical analysis to determine if there was a differential error profile in alignments at each reference base between the mutant (defective m6A) and the complementing line. We identified 17,491 sites with greater than two-fold higher error rate (compared to the reference) in the complementing line with restored m6A (Figure 5A). No VIR-dependent error sites mapped to either chloroplast or mitochondrial-encoded RNAs. 99.8% of differential error sites map within Araport11 annotated protein-coding genes. These sites were enriched in the consensus m6A target sequence DRAYH (D=G or U or A; R=G or A; Y=C or U; H=A or C or U) (Figure 5B) and preferentially found in 3’UTRs (Figure 5C). Since approximately 5nt contribute to the observed current at a given time point in nanopore sequencing (*7*), the presence of a methylated adenosine could affect the accuracy of base-calling of surrounding nucleotides. Consistent with this, we identified 4,749 consensus motifs at error sites (FDR < 0.1, Figure S5 B), with a median of 2 error sites per m6A motif (95% CIs [1, 7]). These results agree with the established and conserved properties of authentic m6A sites, suggesting that differential alignment errors can be used to identify thousands of m6A modifications in nanopore DRS datasets.

In order to examine the validity of m6A sites identified by the differential error site analysis, we used an orthogonal technique to map m6A. Previous maps of Arabidopsis m6A are based on Me-RIP (*8, 29*) and limited by a resolution of around 200nt (*30*). Therefore, to examine Arabidopsis m6A sites with a more accurate method, we used miCLIP (*28*) analysis of 3 biological replicates of Arabidopsis Col-0. We found that, like the differential error sites uncovered in nanopore DRS analysis, the Arabidopsis miCLIP reads were enriched in 3’ UTRs, with no enrichment over stop codons (Figure 5D, S5C). 77% of called nanopore DRS differential error sites fall within 5nt of an miCLIP peak (Figure 5E, F). We therefore conclude that our analysis of nanopore data can detect authentic VIR-dependent m6A sites transcriptome-wide.

***The m6A writer complex is required for circadian rhythm and flowering time gene control***

The combination of transcript processing and modification data obtained using nanopore DRS enabled us to investigate the impact of m6A on Arabidopsis gene expression. We found a global reduction in protein-coding gene expression in *vir-1* (using either nanopore DRS or Illumina RNAseq data) corresponding to transcripts that were methylated in the VIR complemented line (Figure 6A, S6A). These findings are consistent with the recent discovery that m6A predominantly protects Arabidopsis mRNAs from endonucleolytic cleavage (*8*). Therefore, although the m6A writer complex comprises conserved components that target a conserved consensus sequence and distribution of m6A enriched in the last exon, it appears that this modification predominantly promotes the decay of mRNAs in human cells (*31*), but promotes the stability of mRNAs in Arabidopsis (*8*).

The changes in gene expression in *vir-1* were wide ranging (Figure S6B, C, D). For example, we found that the abundance of mRNAs encoding components of the interlocking transcriptional feedback loops that comprise the Arabidopsis circadian oscillator, such as *CCA1*  (*34*), was increased in *vir-1* (Figure S6B, C). This distinction had a biological consequence because the *vir-1* mutant exhibited an increase in clock period (Figure 6B, C). We also found that previously identified targets of the circadian clock, such as *CAB1,* showed increases in poly(A) tail length in *vir-1* compared to Col-0 and the complemented line (Figure 6D)*,* likely identifying transcripts where poly(A) tail length is controlled in a circadian manner (*35*)*.* One output of the circadian clock is the control of flowering time and we found that, not only was the expression of photoperiod pathway components altered in *vir-1*, but other flowering time genes too (12% of established flowering time genes differentially expressed using Illumina RNAseq data; Table S4). Notably, detectable levels of sense and antisense RNA at the *FLC* locus were reduced compared to wild-type (Figure S6D). Consequently, the proper control of circadian rhythms, flowering time and the regulatory module at *FLC* ultimately requires the m6A writer complex component, VIR.

***Defective m6A writer complex function results in chimeric RNA formation***

In addition to measuring RNA expression, we asked what impact loss of m6A had on pre-mRNA processing. Detectable disruptions to splicing in *vir-1* were modest. For example, using DEX-Seq analysis of annotated splice sites, we found only weak effect size changes to cassette exons, retained introns or alternative donor/acceptor sites compared to the complemented line (Figure S6E). In contrast, a clear defect in RNA 3’ end formation in *vir-1* was apparent. Using a Kolmogorov Smirnov test, we identified 3,579 genes which had a change in nanopore DRS 3’ position profile in the *vir-1* mutant compared to the complemented line (FDR < 0.05, absolute change in position > 13nt). Of these, 3,008 displayed a shift to usage of more proximal poly(A) sites in *vir-1*. 60% of these genes also contain m6A sites detectable by nanopore DRS (70% detectable by miCLIP) and correspond to locations of increased cleavage downstream of m6A sites in the *vir-1* mutant (Figure S6F). 571 genes showed increased transcriptional readthrough beyond the 3’ end in *vir-1*. 73% of these loci also contained nanopore-mapped m6A sites (78% by miCLIP). Impacts of altered 3’ processing are complex, but have the potential to change the relative abundance of certain CDS transcripts. For example, we detected increased readthrough of an intronic cleavage site in the Symplekin-like gene *TANG1* (*AT1G27595*) (Figure S6G); and increased readthrough and cryptic splicing at *ALG3* (*AT2G47760*)that alsoresults in chimeric RNA formation with the downstream gene *GH3.9* (*AT2G4G7750*) (Figure 6E). The existence of the *ALG3-GH3.9* chimeric RNAs was supported by Illumina RNAseq (Figure 6E) and confirmed by RT-PCR, cloning and sequencing (Figure S6H). We detected 523 loci with increased levels of chimeric RNAs in *vir-1* resulting from unterminated transcription proceeding into downstream genes on the same strand. However, only 33% of upstream genes that comprised chimeric RNAs had detectable m6A sites in the complementing line restoring VIR activity. Consequently, these findings might be explained either by an m6A-independent role for VIR or the writer complex in 3’ end formation or an indirect effect of the writer complex on factors required for 3’ processing. m6A independent roles for the human m6A methyltransferases METTL3 (*35*) and METTL16 (*36*) have been found previously, and a role for the writer complex in controlling Arabidopsis RNA processing independent of m6A cannot be overlooked (*35*). In mammals, recognition of the canonical poly(A) signal AAUAAA involves direct binding by CPSF30 (*36, 37*). Notably, an alternatively spliced isoform of Arabidopsis CPSF30 encodes a YTH domain with the potential to bind and read m6A (*39*). Consequently, in plants, m6A may also contribute to the recognition of specific RNA 3’ ends.

***Concluding Remarks***

We show that nanopore DRS has the potential to refine multiple features of Arabidopsis genome annotation, despite the sequence of the genome being available since 2000 (*39*). In addition, we uncover the clearest view of the Arabidopsis epitranscriptome. Modern agriculture is dominated by a handful of intensely researched crops (*41*). However, to diversify global food supply, enhance agricultural productivity and tackle malnutrition, there is a need to focus on crops utilized in rural societies that have received little attention for crop improvement (*42*). Based on our experience with Arabidopsis and water yam (*Dioscorea alata*) (ref), we anticipate that the combination of nanopore DRS and other sequencing approaches will enhance orphan crop annotation.

***Acknowledgements***

This work was funded by BBSRC grants BB/J00247X/1, BB/M004155/1, BB/M010066/1, the University of Dundee Global Challenges Research Fund and the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 799300.

***References***

1. K. D. Meyer, S. R. Jaffrey, Rethinking m6A readers, writers, and erasers. *Annual Review of Cell and Developmental Biology* **33**, 319-342 (2017).

2. I. A. Roundtree, M. E. Evans, T. Pan, C. He, Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187-1200 (2017).

3. C. H. Jan, R. C. Friedman, J. G. Ruby, D. P. Bartel, Formation, regulation and evolution of *Caenorhabditis elegans* 3 ' UTRs. *Nature* **469**, 97-U114 (2011).

4. J. Houseley, D. Tollervey, Apparent non-canonical *trans*-splicing is generated by reverse transcriptase *in vitro*. *Plos One* **5**, (2010).

5. M. Helm, Y. Motorin, Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet* **18**, 275-291 (2017).

6. T. Steijger *et al.*, Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* **10**, 1177-+ (2013).

7. D. R. Garalde *et al.*, Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201-206 (2018).

8. S. J. Anderson *et al.*, *N*-6-methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in Arabidopsis. *Cell Rep* **25**, 1146-+ (2018).

9. C. Hornyik, L. C. Terzi, G. G. Simpson, The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev Cell* **18**, 203-213 (2010).

10. K. Ruzicka *et al.*, Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol* **215**, 157-172 (2017).

11. H. Lange *et al.*, The RNA helicases AtMTR4 and HEN2 target specific subsets of nuclear transcripts for degradation by the nuclear exosome in *Arabidopsis thaliana*. *PLoS Genet* **10**, (2014).

12. L. C. Jiang *et al.*, Synthetic spike-in standards for RNA-seq experiments. *Genome Res* **21**, 1543-1551 (2011).

13. L. H. Reid *et al.*, Proposed methods for testing and selecting the ERCC external RNA controls. *Bmc Genomics* **6**, (2005).

14. F. Waibel, W. Filipowicz, Rna-polymerase specificity of transcription of Arabidopsis U-snRNA genes determined by promoter element spacing. *Nature* **346**, 199-202 (1990).

15. C. Y. Cheng *et al.*, Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant Journal* **89**, 789-804 (2017).

16. D. R. Garalde *et al.*, Highly parallel direct RNA sequencing on an array of nanopores. *Nature methods* **15**, 201 (2018).

17. T. Hackl, R. Hedrich, J. Schultz, F. Forster, proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004-3011 (2014).

18. A. Payne, N. Holmes, V. Rakyan, M. Loose, BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, (2018).

19. S. A. Lima *et al.*, Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* **24**, 1057-+ (2017).

20. A. Sherstnev *et al.*, Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nat Struct Mol Biol* **19**, 845-852 (2012).

21. M. Rigal, Z. Kevei, T. Pelissier, O. Mathieu, DNA methylation in an intron of the *IBM1* histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* **31**, 2981-2993 (2012).

22. M. Seki *et al.*, Functional annotation of a full-length Arabidopsis cDNA collection. *Science* **296**, 141-145 (2002).

23. T. Ushijima *et al.*, Light controls protein localization through phytochrome-mediated alternative promoter selection. *Cell* **171**, 1316-+ (2017).

24. R. X. Zhang *et al.*, A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res* **45**, 5061-5073 (2017).

25. D. Pose *et al.*, Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* **503**, 414-417 (2013).

26. N. Sheth *et al.*, Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* **34**, 3955-3967 (2006).

27. R. Dehghannasiri, L. Szabo, J. Salzman, Ambiguous splice sites distinguish circRNA and linear splicing in the human genome. *Bioinformatics*, (2018).

28. B. Linder *et al.*, Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772 (2015).

29. L. S. Shen *et al.*, *N*-6-methyladenosine RNA modification regulates shoot stem cell fate in Arabidopsis. *Dev Cell* **38**, 186-200 (2016).

30. S. D. Ke *et al.*, A majority of m(6)A residues are in the last exons, allowing the potential for 3 ' UTR regulation. *Gene Dev* **29**, 2037-2053 (2015).

31. X. Wang *et al.*, *N*-6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-+ (2014).

32. C. R. McClung, The Plant Circadian Oscillator. *Biology (Basel)* **8**, (2019).

33. P. Lidder, R. A. Gutierrez, P. A. Salome, C. R. McClung, P. J. Green, Circadian control of messenger RNA stability. Association with a sequence-specific messenger RNA decay pathway. *Plant Physiol* **138**, 2374-2385 (2005).

34. S. Lin, J. Choe, P. Du, R. Triboulet, R. I. Gregory, The m6A methyltransferase METTL3 promotes translation in human cancer cells. *Mol Cell* **62**, 335-345 (2016).

35. K. E. Pendleton *et al.*, The U6 snRNA m6A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell* **169**, 824-835 e814 (2017).

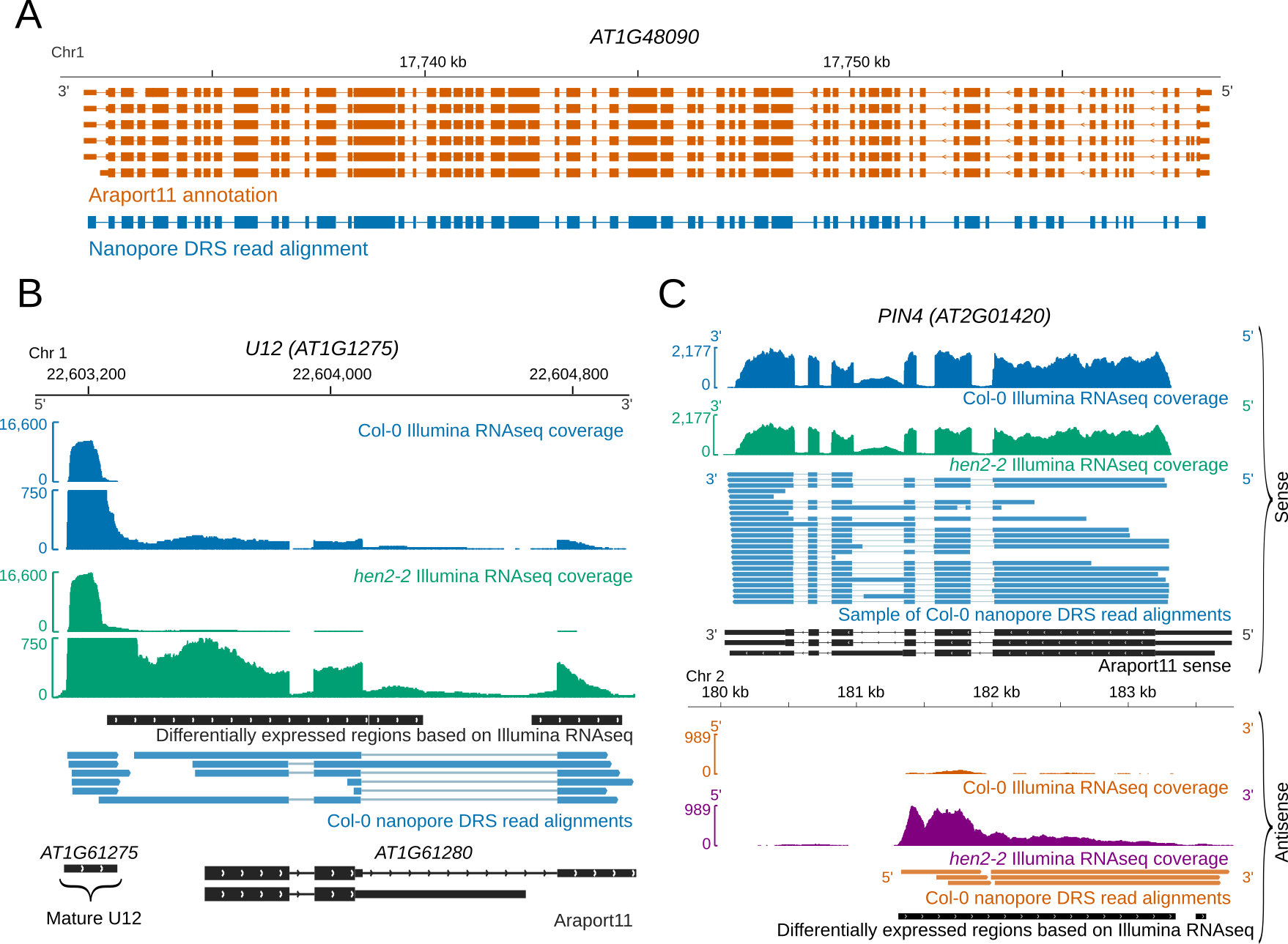
36. S. L. Chan *et al.*, CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3 ' processing. *Gene Dev* **28**, 2370-2380 (2014).

37. L. Schonemann *et al.*, Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Gene Dev* **28**, 2381-2393 (2014).

38. A. T. Stevens, D. K. Howe, A. G. Hunt, Characterization of mRNA polyadenylation in the apicomplexa. *Plos One* **13**, (2018).

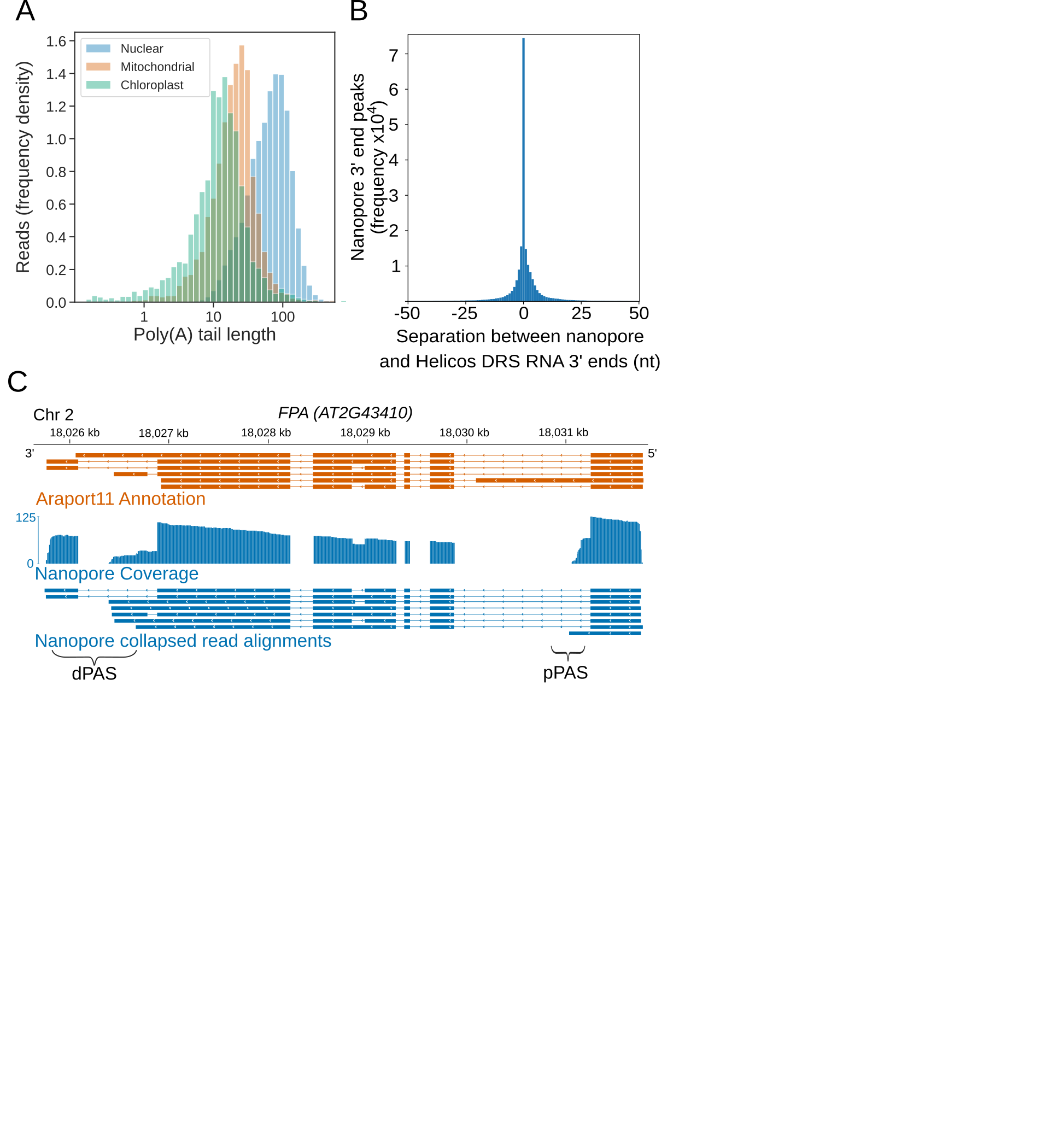
39. S. Kaul *et al.*, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).

40. *World population prospects: the 2017 revision, Key Findings and Advance Tables.* (2017).



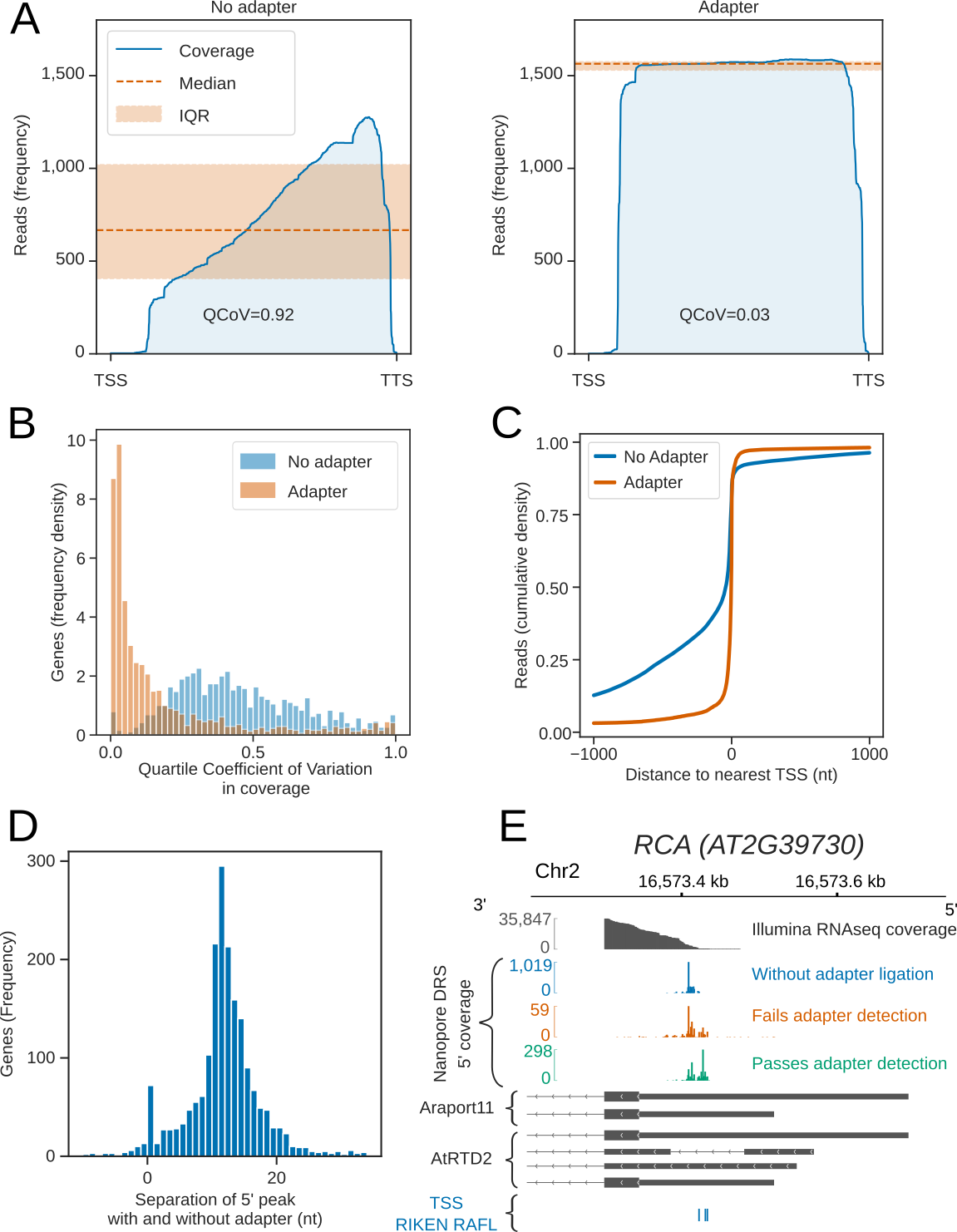
**Figure 1. Diverse Arabidopsis RNAs are detected by Nanopore DRS**

1. Nanopore DRS 12.7kb read alignment at *AT1G48090* comprising 63 exons. Araport11 annotation is shown in orange, nanopore DRS read alignment – in blue.
2. Nanopore DRS in combination with Illumina RNAseq data of the *hen2-2* exosome mutant identified an unannotated precursor RNA of U12 snRNA. Col-0 Illumina RNAseq coverage and Nanopore read alignments are shown in blue; *hen2-2* Illumina RNAseq coverage – in green; Araport11 annotation – in black.
3. *PIN4* long non-coding antisense RNAs detected using nanopore DRS. Col-0 sense Illumina RNAseq coverage and nanopore sense read alignments are shown in blue; Col-0 antisense Illumina RNAseq coverage and Nanopore antisense read alignments – in orange; *hen2-2* mutant sense Illumina RNAseq coverage – in green; *hen2-2* mutant antisense Illumina RNAseq coverage – in purple. Sense RNA isoforms found in Araport11 and antisense differentially expressed regions detected with DERfinder are shown in black.



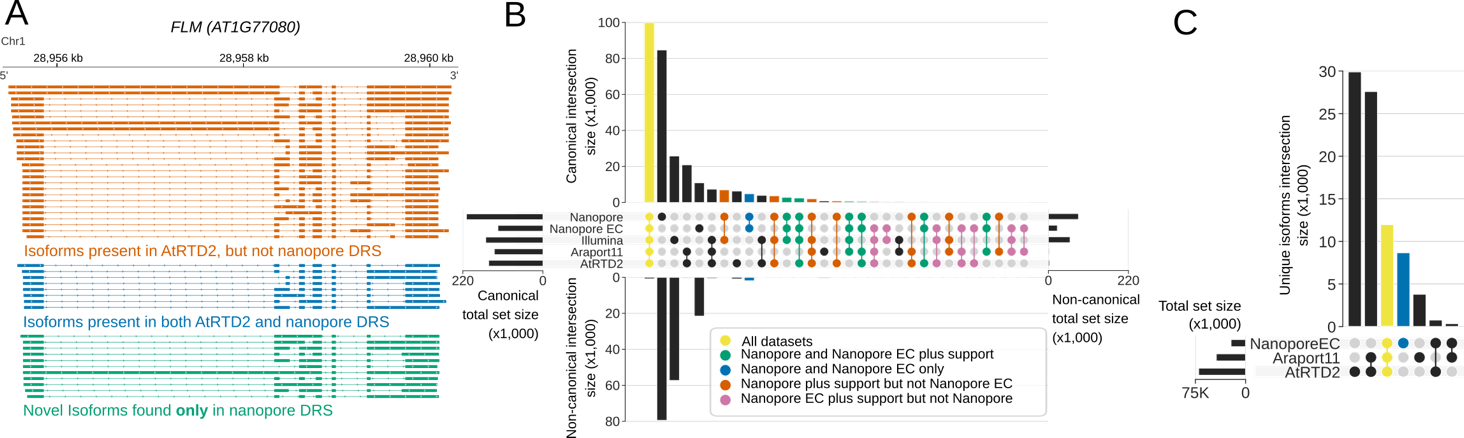
**Figure 2. Nanopore DRS reveals poly(A) tail length and maps 3’ cleavage and polyadenylation sites.**

1. Normalised histogram of poly(A) tail length of RNAs encoded by different genomes. The poly(A) tail lengths of nuclear (blue), mitochondrial (orange) and chloroplast ( in green) genomes are shown.
2. Separation between the RNA 3’ end positions in Nanopore DRS read alignments and the nearest polyadenylation sites identified in Helicos Bio data.
3. Nanopore DRS identified 3’ polyadenylation sites in RNAs transcribed from *FPA (AT2G43410)* . Blue track shows coverage of Nanopore DRS read alignments, and collapsed read alignments representing putative transcript annotations detected by Nanopore DRS. Isoforms found in Araport11 annotation are shown in orange and read alignments from nanopore DRS are in blue.pPAS is the proximal polyadenylation site; dPAS – distal polyadenylation sites.



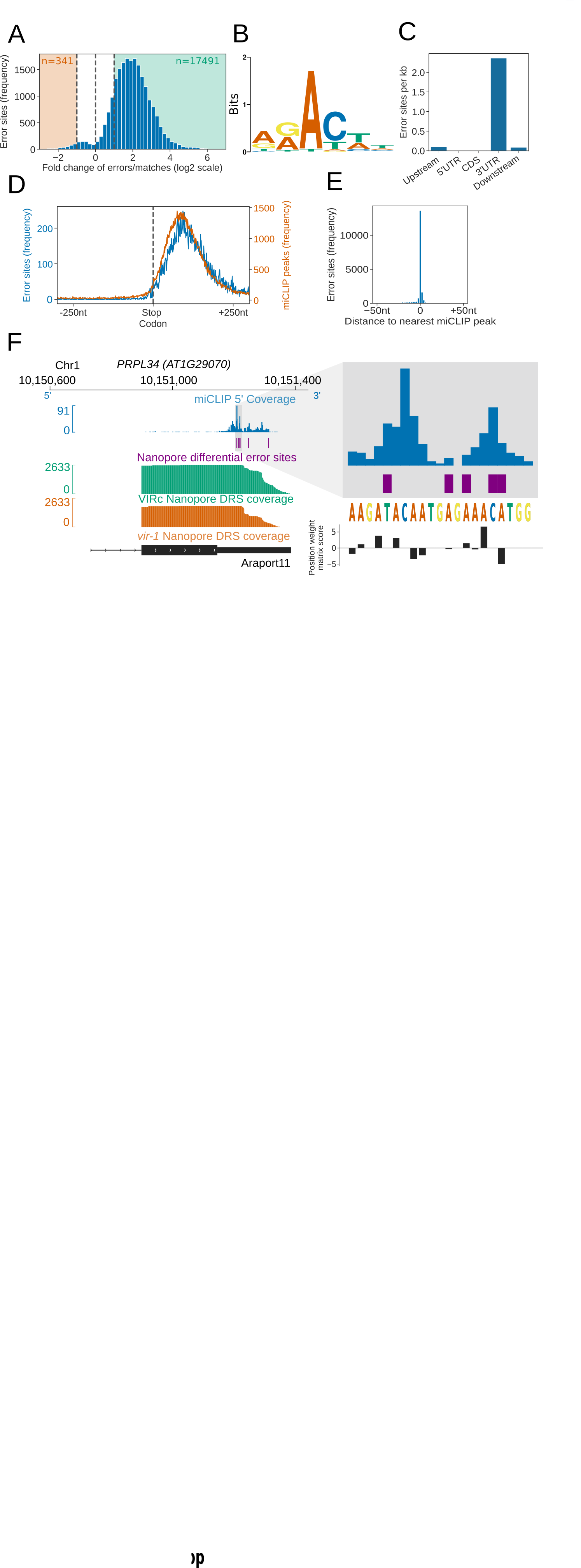
**Figure 3. Cap-dependent ligation of an RNA adapter enables detection of authentic RNA 5’ ends.**

1. 5’ adapter RNA ligation reduces 3’ bias in nanopore DRS data at *RCA (AT2G39730)*. Blue line shows exonic read coverage at *RCA* for reads without adapter – left, and with adapter – right. Orange line shows the median coverage and orange shaded area shows inter-quartile range (IQR). Change in 3’ bias can be measured using the IQR / median (quartile coefficient of variation). 5’ adapter ligation reduces 3’ bias at *RCA* from 0.92 to 0.03.
2. 5’ adapter RNA ligation reduces 3’ bias in nanopore DRS data. Histogram shows quartile coefficient of variation in per base coverage for reads with 5’ adapter RNA – in orange, compared to reads without 5’ adapter RNA – in blue.
3. Cap-dependent adapter ligation allows identification of authentic 5’ ends using nanopore DRS. Cumulative distribution function shows the distance to the nearest TSS identified from full length transcripts cloned as part of the RIKEN RAFL project for reads with 5’ adapter RNA (orange), compared to reads without 5’ adapter RNA (blue).
4. For *RCA (AT2G39730),* the 5’ end identified using cap-dependent 5’ adapter RNA ligation protocol was consistent with Illumina RNAseq and full-length cDNA start site data, but differed from the 5’ ends in the Araport11 and AtRTD2 annotations. Illumina RNAseq coverage is shown in grey and nanopore DRS 5’ end coverage generated without cap-dependent ligation protocol – in blue. Coverage for read alignments generated using the cap-dependent ligation protocol with and without 5’ adapter RNA are shown in green and orange, respectively. RNA isoforms found in Araport11 and AtRTD2 annotations are shown in grey and TSSs identified from full length transcripts cloned as part of the RIKEN RAFL project - in blue.
5. Cap-dependent adapter ligation allowed resolution of an additional 11 nt of sequence at the RNA 5’ end. Histogram shows nt shift in the largest peak of 5’ coverage for each gene in data obtained using protocol with 5’ adapter vs protocol without 5’ adapter.



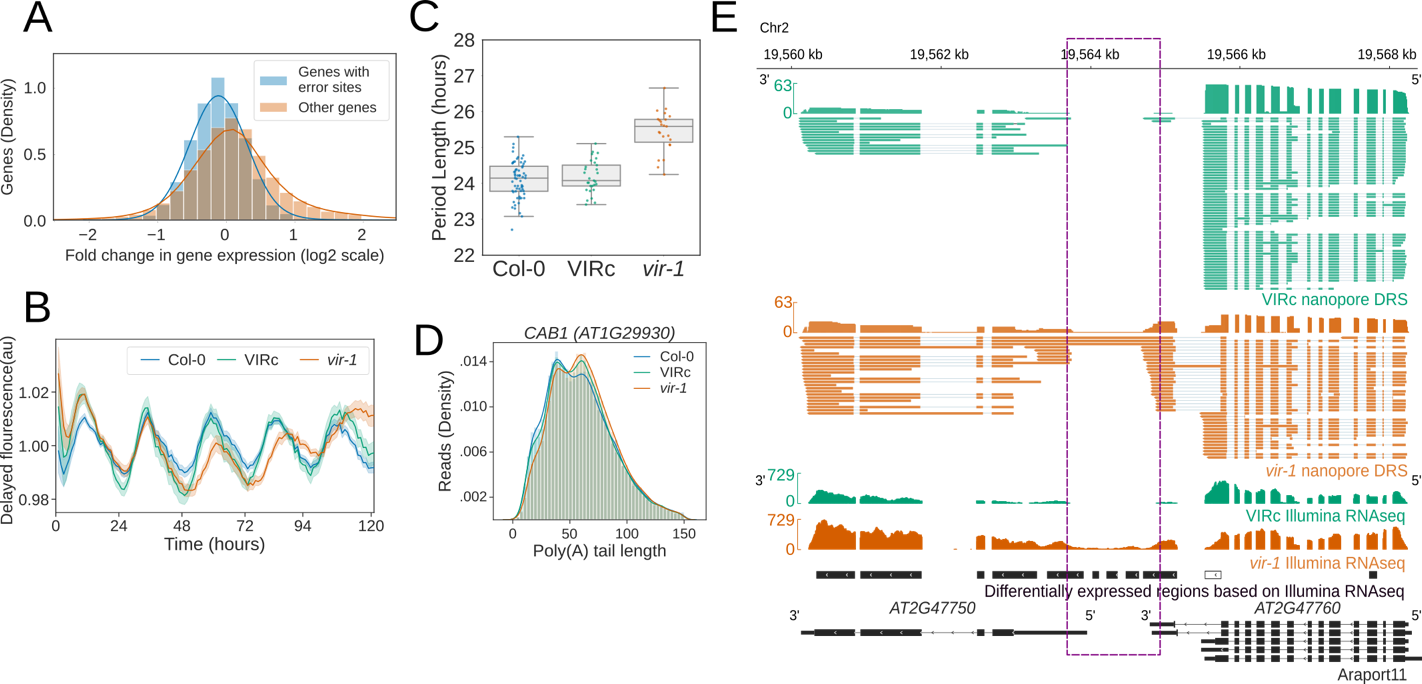
**Figure 4. Nanopore DRS reveals the complexity of alternative splicing.**

1. Nanopore DRS identified unannotated combinations of established splice sites in alternative splicing pattern of *FLOWERING LOCUS M* (*FLM, AT1G77080)*. Isoforms present in AtRTD2 annotation, but not identified using nanopore DRS are shown in orange, isoforms common to both AtRTD2 and nanopore DRS are in blue, novel isoforms identified in nanopore DRS are shown in green.
2. Comparison of splicing events identified in error-corrected and non-error-corrected nanopore DRS, Illumina RNA sequencing, Araport11 and AtRTD2 annotations. Bar size represents number of unique splicing events common to the set intersection highlighted using circles (please refer to Table S3 for the exact values). GT/AG splicing events are shown on the top and non GT/AG on the bottom of the plot. Splicing events common to all five datasets are shown in yellow; events common to both error-corrected and non-error-corrected nanopore DRS with support in orthogonal datasets are in green; events common to both nanopore DRS datasets without orthogonal support are in blue; events found in uncorrected nanopore DRS (but not error corrected) with orthogonal support are in orange; and events found in error-corrected nanopore DRS (but not uncorrected) with orthogonal support are in pink.
3. Comparison of RNA isoforms (defined as sets of co-spliced introns) identified in error-corrected full-length nanopore DRS, Araport11 and AtRTD2 annotations. Bar size represents number of splicing events common to a group highlighted using circles below (please refer to Table S3 for the exact values). Unique splicing patterns nanopore DRS and both reference annotations are shown in yellow. Novel isoforms are shown in blue.

****

**Figure 5. Differential error rate analysis identifies sites of VIR-dependent m6A modifications transcriptome-wide**

1. Loss of VIR function is associated with reduced error rate in nanopore DRS. Histogram shows the log2 fold change in ratio of errors to reference matches at bases with a significant change in error profile in *vir-1* mutant compared to the VIR complemented line. Orange and green shaded regions show sites with increased and reduced errors in *vir-1,* respectively.
2. Motif of the error rate sites matches consensus m6A target sequence. Sequence logo of the motif enriched at sites with reduced error rate in the *vir-1* mutant.
3. Differential error rate sites are primarily found in 3’ UTRs. Bar plot shows the number of differential error rate sites per kb of genic feature types in the Araport11 reference. Upstream and downstream regions are 200 nt regions before and after the annotated transcription termination sites (TTSs), respectively.
4. Differential error rate sites and miCLIP peaks are similarly distributed within the 3’ UTR without accumulation at the stop codon. Metagene plot centred on stop codons shows the frequency of nanopore DRS error sites in blue and miCLIP peaks in orange.
5. The location of differential error rate sites is in good agreement with the locations of miCLIP sites. Histogram shows distribution of distances to the nearest miCLIP peak for each site of reduced error. 77% of error sites are within 5nt of an miCLIP peak.
6. Nanopore DRS differential error sites analysis and miCLIP identify m6A sites in the 3’ UTR of *PRPL34* RNA. miCLIP 5’ end coverage is shown in blue; nanopore DRS differential error sites – in purple; nanopore DRS coverage of VIR complemented line (VIRc) – in green; nanopore DRS coverage of *vir-1* mutant – in orange; RNA isoform from Araport11 annotation – in black. Zoomed region shows miCLIP coverage in blue and error sites in purple labelled with score using m6A consensus position weight matrix (Figure 5B) in black. More positive scores denote higher likelihood of a match to the consensus sequence.

****

**Figure 6. Reduction in m6A RNA modification results in changes to circadian clock and generation of chimeric RNAs.**

1. Genes with differential error rate sites have lower detectable RNA levels. Histogram shows log2 fold change in gene expression based on counts from nanopore DRS reads in *vir-1* mutant and VIR complemented line. Genes with differential error rate sites are in blue, and those without – in orange.
2. Circadian period is lengthened in *vir-1* mutant. Mean delayed fluorescence measurements for Col-0 are in blue, VIR complemented line – in green and *vir-1* mutant – in orange. Shaded areas show bootstrapped 95% confidence intervals for mean.
3. Boxplot shows period lengths of Col-0 (blue), VIR complemented line (green) and *vir-1* mutant (orange) calculated from delayed fluorescence measurements shown in (B).
4. Poly(A) tail length alterations in *vir-1* mutant. Histogram shows poly(A) tail length distribution of *CAB1 (AT1G29930)* in Col-0 (blue), VIR complemented line (green) and *vir-1* mutant (orange).
5. Readthough events and chimeric RNAs are detected in *vir-1*. Nanopore DRS and Illumina RNAseq data for VIR complemented line are shown in green; nanopore DRS and Illumina RNAseq data for *vir-1* mutant – in orange; RNA isoforms found in Araport11 annotation – in black. Differentially expressed regions detected using Illumina RNAseq data with DERfinder are shown in black (for upregulated regions) or white (for downregulated regions). Intergenic readthrough regions are highlighted using a purple dashed rectangle.