**Materials and Methods**

**Plants**

*Plant material and growth conditions*

The wild-type *Arabidopsis thaliana* accession Col-0 was obtained from Nottingham Arabidopsis Stock Centre. The *vir-1* and VIR complemented (*VIR::GFP-VIR*) lines were provided by Kamil Ruzicka, Brno (*1*). The *hen2-2* (Gabi\_774HO7) mutant was provided by Dominique Gagliardi, Strasbourg. The seeds were sown on MS10 media plates, stratified at 4°C for two days, germinated in a controlled environment at 22°C under 16h light/8h dark conditions, and harvested 14 days after transfer to 22°C.

*Clock phenotype analysis*

Clock phenotype experiments were performed as described in (*42*). Briefly, plants were grown in 12‐h light/12‐h dark cycles at 22oC and 80 μmol m−2 sec−1 light for 9 days. Next, delayed fluorescence measurements were recorded every hour for 6 days at constant temperature (22oC) and under constant light (20 uM red and 20uM blue mix). FFT non-linear Least Squares fitting to estimate period length was conducted using Biodare (*43*).

**RNA**

*RNA isolation*

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

*mGFP in vitro transcription*

mGFP coding sequence was amplified using CloneAmp HiFi PCR Premix (Clontech) and a forward primer containing T7 promoter sequence (Merck) (Table S5). PCR product was purified using GeneJET Gel Extraction (Thermo Fisher Scientific) and DNA Cleanup Micro Kit (Thermo Fisher Scientific) as per manufacturer’s instructions. mGFP was *in vitro* transcribed using mMESSAGE mMACHINE™ T7 ULTRA Transcription Kit (Thermo Fisher Scientific) with and without the addition of a cap analog (Anti-Reverse Cap Analog (ARCA)) according to manufacturer’s instructions. mGFP transcripts were treated with TURBOTM DNase, polyA-tailed using *Escherichia coli* Poly(A) Polymerase (E-PAP) and ATP (Thermo Fisher Scientific) and recovered using MEGAclear™ Kit (Thermo Fisher Scientific) as per manufacturer’s instructions. mGFP mRNAs quantity was assessed using a Qubit 1.0 Fluorometer (as described above), while the quality and integrity was checked using the NanoDrop™ 2000 spectrophotometer and denaturing agarose-gel electrophoresis. ARCA-capped and non-capped *in vitro* mGFP mRNAs were used in the library preparation for DRS using nanopores.

*RT-PCR and RT-qPCR*

Total RNA was reverse transcribed using SuperScript™ III polymerase or SuperScript™ IV VILO™ Master Mix (Thermo Fisher Scientific) according to the manufacturer’s protocol. For RT-PCR, the reactions were performed using the Advantage 2 Polymerase Mix (Clontech) using primers (Merck) listed in Table S5. Next, the PCR products were purified from the gel using GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Fisher Scientific), cloned into the pGEM T-Easy vector (Promega) (according to the manufacturer’s instruction) and sequenced. For RT-qPCR, the reactions were carried out using SYBR Green I (Quiagen) mix with primers (Merck) listed in Table S5 following manufacturer’s instructions.

**Illumina RNA sequencing**

*Preparation of libraries for Illumina RNA sequencing*

Illumina RNA sequencing libraries from purified mRNA were prepared and sequenced by the Centre for Genomic Research at University of Liverpool using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs® Inc.). Paired-end sequencing with read length of 150bp was carried out on Illumina HiSeq 4000. Illumina RNA libraries from ribosome-depleted RNA were prepared using the Illumina TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina®). Paired-end sequencing with read length of 100bp was carried out on Illumina HiSeq2000 at the Genomic Sequencing Unit of the University of Dundee. ERCC RNA Spike-In mixes (Thermo Fisher Scientific) (*2, 3*) were included in each of the libraries using concentrations advised by the manufacturer.

*Mapping of Illumina RNA sequencing data*

Reads were mapped to the TAIR10 reference genome with Araport11 reference annotation (*17*), using STAR version 2.6.1 (*4*), a maximum multimapping rate of 5, a minimum splice junction overhang of 8 nt (3 nt for junctions in the Araport11 reference), a maximum of 5 mismatches per read, and intron length boundaries of 60 - 10,000 nt.

*Differential gene expression analysis using Illumina RNA sequencing data*

Transcript level counts for Illumina RNA sequencing reads were estimated by pseudoalignment with salmon version 0.11.2 (*5*). Counts were aggregated to gene level using tximport (*6*) and differential expression gene expression analyses for *vir-1* mutantvs wild-type and *vir-1* mutant vs VIR complemented line were conducted in R version 3.5 using edgeR version 3.24.3 (*7*).

*Differentially expressed region analysis using Illumina RNA sequencing data*

Mapped read pairs originating from forward and reverse strands were separated and coverage tracks were generated using samtools version 1.9 (*8*). Coverage tracks were then used as input for DERfinder version 1.16.1 (ref). Expressed regions were identified using a minimum coverage of 10 reads, and differential expression between *vir-1* and VIR complemented was conducted using the analyseChr method using 50 permutations.

*Differential exon usage analysis using Illumina RNA sequencing data*

Annotated gene models from Araport11 were divided into transcript chunks (i.e. contiguous regions within which each base is present in the same set of transcript models). Read counts for each chunk were generated using bedtools version 2.27.1 (*9*) intersect in count mode. Chunk counts were then processed using DEXseq version XX (*10*) to identify differentially expressed chunks between *vir-1* and VIR complemented lines. Chunks were annotated as five prime variation if they included a start site of any transcript, and three prime variation if they contained a termination site. Chunks representing overhangs from alternative donor or acceptor sites were also separately classified. Internal exons were subclassified into cassette exons if they could be wholly contained within any intron.

**Nanopore direct RNA sequencing**

*Preparation of libraries for Direct RNA sequencing using nanopores*

mRNA was isolated from approximately 75 μg 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). Nanopore libraries were prepared from 1 μg of poly(A)+ RNA combined with 1 μl of undiluted ERCC RNA Spike-In mix (Thermo Fisher Scientific) using the nanopore DRS Kit (SQK-RNA001 Ltd.) according to manufacturer’s instructions. Quickly poly(T) adapter was ligated to the mRNA using T4 DNA ligase (New England Biolabs® Inc.) in the Quick Ligase reaction buffer (New England Biolabs® Inc.) for 15 min at room temperature. The first strand cDNA was synthesized by SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) using the oligo(dT) adapter. The RNA-cDNA hybrid was purified using Agencourt RNAClean XP magnetic beads (Beckman Coulter). The sequencing adapter was ligated to the mRNA using T4 DNA ligase (New England Biolabs® Inc.) in the Quick Ligase reaction buffer (New England Biolabs® Inc.) for 15 min at room temperature followed by a second purification using Agencourt beads as described above. Libraries were loaded on R9.4 SpotON Flow Cells (Oxford Nanopore Technologies Ltd.) and sequenced on a 48-hour runtime.

To incorporate cap-dependent ligation of a biotinylated 5’ adapter RNA, the following modifications were introduced into the library preparation protocol. 4 µg of mRNA was de-phosphorylated by Calf Intestinal Alkaline Phosphatase (Thermo Fisher Scientific) and 5’ cap was removed by Cap-Clip™ Acid Pyrophosphatase (Cambio) according to the manufacturer’s instructions. Next, 5’ RNA oligo biotinylated at the 5’ end (Integrated DNA Technologies) was ligated to dephosphorylated, de-capped mRNA using T4 RNA ligase I (New England Biolabs® Inc.) and mRNA was purified using Dynabeads™ MyOne™ Streptavidin C1 beads (Thermo Fisher Scientific) according to the manufacturer’s instructions. mRNA quality and quantity were assessed using the NanoDropTM 2000 spectrophotometer and used for nanopore DRS library preparation as described above.

*Processing of nanopore DRS data*

Reads were basecalled with guppy version 2.3.1 (Oxford Nanopore Technologies) using default RNA parameters and converted from RNA to DNA fastq using seqkit version 0.10.0 (*11*). Reads were aligned to the TAIR10 *Arabidopsis thaliana* genome (*12*) and ERCC RNA spike-in sequences (*2, 3*) using minimap2 version 2.8 (*13*) in spliced mapping mode using a kmer size of 14 and a max intron size of 10,000. SAM and BAM file manipulations were performed using samtools version 1.9 (*8*).

proovread version 2.14.1 (*14*)was used to correct errors in the nanopore DRS read. Each nanopore DRS replicate was split into 200 chunks for parallel processing. Each chunk was corrected using 4 samples of Illumina poly(A) RNAseq data, selected randomly from the 36 Illumina files (6 biological replicates sequenced across 6 lanes). Illumina reads 1 and 2 were merged into fragments using FLASh version 1.2.11 (*15*). Unjoined pairs were discarded. Error correction with proovread was conducted in sampling free mode using a minimum nanopore read length of 50 nt. Corrected reads were then mapped to the reference using minimap2 as described above.

*Error profile analysis using Nanopore DRS data*

Error rate analysis of aligned reads was conducted on ERCC RNA Spike-in mix controls using pysam version 0.15.2 (*16*) for BAM file parsing. Matches, mismatches, insertions and deletions to the reference were extracted from the cs tag (a more informative version of CIGAR string, output by minimap2) and normalised by aligned length of the read. Reference bases and mismatch bases per position were also recorded and used to assess the frequency of each substitution and indel type by reference base.

*Over-splitting analysis of Nanopore DRS data*

To identify read pairs which result from over-splitting of signal originating from a single RNA molecule, the sequencing summary files produced by guppy were parsed for sequencing time and channel ID. These were used to identify pairs of consecutively sequenced reads. Genomic locations of reads were parsed from minimap2 mappings and consecutively sequenced reads which aligned adjacently with a genomic distance of between –10 nt and 1000 nt between them were identified. Samples sequenced before or during May 2018 had very low levels of over-splitting (between 0.01 and 0.05% of reads) compared to those sequenced in September 2018 onwards (between 0.8 and 1.5% of reads).

*Analysis of the potential for internal priming in Nanopore DRS data*

To determine if internal priming caused by the RT step can occur in Nanopore data, the location of oligo(A) hexamers within Arabidopsis CDS regions was determined and reads which terminated within a 20 nt window of each hexamer were counted. Of 10,116 CDS oligo(A) runs, 160 (1.58%) had at least one supporting read in one Col-0 nanopore dataset. Of these, 137 were only supported by one replicate, and only 4 were supported by all four biological replicates. 66 (41%) occurred in terminal exons, suggesting they may be genuine APA sites.

*Poly (A) length estimation using Nanopore DRS data*

Poly (A) tail length estimations were produced using Nanopolish version 0.11.0 (*17*) and added as tags to bam files using pysam version 0.15.2 (*16*). Per gene length distributions were then produced using the Araport11 annotation, and genes with significant changes in length distribution in the *vir-1* mutant compared to wild-type were identified using a Kolmogorov-Smirnoff test*.* *p*-values were adjusted for multiple testing using Benjamini-Hochberg correction.

*3’ end analyses of nanopore and Helicos DRS reads*

Helicos DRS data were prepared as described in (*22*) and (*23*). Positions with three or more supporting reads were considered to be peaks of Nanopore or Helicos 3’ ends. The distance between each nanopore peak and the nearest Helicos peak was then determined. 37% of nanopore peaks occurred at the same position as a Helicos peak, and the standard deviation in distance was 12.5 nt.

To determine the percentage of nanopore DRS 3’ ends mapping within annotated genic features, transcripts were first flattened into a single record per gene. Exonic annotation was given priority over intronic or intergenic annotation and CDS annotation over UTR. Reads were assigned to genes if they overlapped them by >20% of their aligned length and the annotated feature type of the 3’ end position was determined. Counts were generated both for all reads and for unique positions per gene.

*Isoform collapsing of nanopore DRS data*

Error corrected full length alignments were collapsed into clusters of reads with identical sets of introns. These clusters were then subdivided by 3’ end location by using a Gaussian kernel with sigma of 100 to find local minima between read ends, which were used as cut points to separate clusters. The read with the longest aligned length in each cluster was used as the representative in the figure.

*Splicing analysis of nanopore DRS and Illumina RNAseq data*

Splice junction locations, their flanking sequences, and the read counts supporting them were extracted from Illumina RNA sequencing, Nanopore DRS, and Nanopore error-corrected DRS reads using pysam version 0.14 (*16*), and from Araport11 (*18*) and AtRTD2 (*19*) reference annotations. Splice junctions at the same position, but on opposite strands, were counted independently. Junctions were classified by their most likely snRNP machinery using biopython version 1.71 (*20*), with position weight matrices calculated by (*21*). Position weight matrices were scored against the sequence –3 nt to +10 nt of the donor site, and –14 nt to +3 nt of the acceptor site. Junctions were further categorised as canonical or non-canonical based on the presence or absence of the GT/AG intron motif. For isoform analysis, linked splices from the same read were extracted from full length Nanopore error corrected reads and counted to create unique sets of splice junctions. Intronless reads were not counted. UpSet plots were generated in Python 3.6 using the package upsetplot.

*Validation of novel splice sites*

To validate novel splice junctions detected in nanopore DRS, 5 splice sites out of 20 most highly expressed splice sites were selected for further validation and 3 out of 5 selected splice sites were amplified in RT-PCR followed by DNA sequencing (described above).

*5’ adapter detection analyses using nanopore DRS data*

To produce positive and negative examples of 5’ adapter containing sequences, 5’ soft-clipped regions were extracted from aligned reads for the Col-0 replicate 1 datasets (with and without adapter ligation), using pysam (*16*). These soft-clipped sequences were then searched for the presence of the GeneRacer™ adapter sequence using blastn version 2.7.1 (*22*). Two rules were initially applied to filter blastn results: a match of 10 nt or more to the 44 nt adapter, and an E value of less than 100. Reads from the adapter containing dataset that failed one or both of these criteria were used as negative training examples. A final rule requiring the match to the adapter sequence to occur directly adjacent to the aligned read was also applied. Reads from the adapter containing dataset that passed all three rules were used as the positive training set. When comparing the ratio of positive to negative examples between datasets containing the adapter and those generated from the same tissue without, we found that these three rules gave a signal to noise ratio of >5000 (Table S2).

72,083 positive and 123,739 negative training examples from Col-0 tissue replicate 1 were collected to train the neural network. We then estimated the amount of raw signal from the 5’ end of the squiggle which was required on average to capture the 5’ adapter. To do this, we used Nanopolish eventalign version 0.11.0 (*23*) to identify the interval in the raw read which corresponded to the mRNA alignment to the reference in the adapter positive examples. Since the adapter can be identified immediately adjacent the alignment in sequence space for these reads, the signal after the event alignment must correspond to signal originating from the adapter. The median length of these signals was 1441 points, and 96% of the signals were less than 3000. Therefore, we used a window size of 3,000 to make predictions.

The model was trained in python 3.6 using Keras version 2.2.4 with Tensorflow version 1.10.0 backend (*24, 25*). A ResNet style architecture was used (*26*), composed of 8 residual blocks containing two convolutional layers of kernel size 5 and a shortcut convolution with kernel size 1. Down-sampling using maximum pooling layers with a stride of 2 was used between each residual block. A penultimate densely connected layer of size 16 was used, with training dropout of 0.5. Input signals were standardized by median absolute deviation scaling across the whole read before the final 3,000 points were taken, and the negative samples were augmented by addition of random internal signals from reads and pure gaussian, multi-gaussian, and perlin noise signals. The whole dataset was also augmented on the fly during training by the addition of gaussian noise with standard deviation of 0.1. Models were trained for a maximum of 100 epochs (batch size of 128, 100 batches per epoch, positive and negative examples sampled in a 1:1 ratio) using the RMSprop optimiser with an initial learning rate of 0.001, which was reduced by a factor of 10 after three epochs with no reduction in validation loss. Early stopping was used after five epochs with no reduction in validation loss. Model performance was evaluated using five-fold cross validation and by testing on independently generated datasets from Col-0 replicate 2, produced with and without the adapter ligation protocol (Fig S3B,C).

For evaluation of the reduction in 3’ bias of adapter ligated datasets, we used the Araport11 exon annotations to produce per base coverage for each gene in the Col-0 replicate 2 dataset. Coverage was generated separately for reads predicted to contain adapters and those that did not. Leading zeros at the very 5’ and 3’ of genes were assumed to be caused by mis-annotation of UTRs and so were trimmed. The quartile coefficient of variation (interquartile range / median) was then used as a robust measure of variation in coverage across each gene. To orthogonally validate the 5’ ends of adapter ligated reads, full length cDNA clone sequences were downloaded from RIKEN RAFL. These were mapped with minimap2 (*13*) in spliced mode and distance from each nanopore alignment 5’ end to the nearest RIKEN RAFL alignment (*27*). 5’ end was calculated using bedtools (*9*). The amount of 5’ end sequence which is rescued when 5’ adapters are used was estimated by identifying the largest peak in 5’ end locations per gene in the absence of adapter, and then measured how this peak shifted using reads predicted to contain adapters.

*Differential error site analysis using nanopore DRS data*

To detect sites of Virilizer dependent m6A RNA modifications, we developed scripts to test changes in per base error profiles of aligned reads. Pileup columns for each position with >10 reads coverage were generated using pysam (*16*) and reads in each column were categorised as either A, C, G, T or indel. The relative proportions of each category were counted. Counts from replicates of the same condition were aggregated and a 2x5 contingency table was produced for each base comparing ­*vir-1* and VIR complemented lines. A G-test was performed to identify bases with significantly altered error profiles. For bases with *p-*value less than 0.05, G-tests for homogeneity between replicates of the same condition were then performed. Bases where the sum of the G statistic for homogeneity tests was greater than the G statistic for the *vir-1* and complemented line comparison were filtered. Multiple testing correction was carried out using the Benjamini-Hochberg method, and an FDR threshold of 0.05 was used. The log2 fold change in mismatch to match ratio (compared to the reference base) between VIR complemented and *vir-1* was calculated using the Haldane correction for zero counts. Bases which had a log fold change > 1 were considered to have a reduced error rate in the *vir-1* mutant.

To identify motifs enriched at sites with reduced error rate, reduced error rate sites were increased in size by 5nt in each direction and overlapping sites were merged using bedtools version 2.27.1 (*9*). Sequences corresponding to these sites were extracted from the TAIR10 reference and overrepresented motifs were detected in the sequences using MEME version 5.0.2 (*28*), run in zero or one occurrence mode with a motif size range of 5-7 and a minimum requirement of 100 sequence matches. The presence of these motifs at error sites was then detected using FIMO version 5.0.2 (*29*). A relaxed FDR threshold of 0.1 was used with FIMO to capture more degenerate motifs matching the m6A consensus.

*Differential gene expression analysis using nanopore DRS data*

Gene level counts were produced for each nanopore DRS replicate using featureCounts version 1.6.3 (*30*) in long-read mode with strand specific counting. Differential expression analysis between VIR complemented and *vir-1* lines was then performed in R version 3.5 using edgeR version 3.24.3 (*7*).

*Identification of alternative 3’ end positions and chimeric RNA using nanopore DRS data*

Genes with differential 3’ end usage were identified by producing 3’ profiles of reads which overlapped with each annotated gene locus by >20%. These profiles were then compared between VIR complemented and *vir-1* using a Kolmogorov Smirmov test to identify changes. Multiple testing correction was performed using the Benjamini Hochberg method. To approximately identify the direction and distance of the change, the normalised single base level histograms of the Col-0 profile was subtracted from that of the mutant profile, and the minimum and maximum points in the difference profile were found. These represent the site of most reduced and increased relative usage, respectively. Results were filtered for FDR < 0.05 and absolute change of site > 13 nt (the measured error range of nanopore DRS 3’ end alignment).

To identify genes with significant increases in chimeras in the *vir-1* mutant, we used the Araport11 annotation (*18*) to identify reads which overlapped with multiple adjacent gene loci (chimeric reads) and those that originated from a single locus (non-chimeric reads). Chimeric reads were considered to originate from the most upstream gene with which they overlapped. We pooled reads from replicates of a condition and used 50 bootstrapped 75% samples without replacement to estimate the ratio of chimeric to non-chimeric reads at each gene in that condition. The Haldane correction for zero counts was applied. The distributions of complemented line and *vir-1* chimeric to non-chimeric ratios were tested using a Kolmogorov Smirnoff test to detect loci with altered chimera production. All possible pairwise combinations of VIRc and *vir-1* bootstraps were then compared produce a distribution of estimates of the change in chimeric to non-chimeric ratio in the *vir-1* mutant. Loci which had more than one chimeric read in *vir-1* and demonstrated at least a two-fold increase in chimeric read ratio in more than 50% of bootstrap comparisons, plus were significantly changed at FDR < 0.05, were considered to be sites of increased chimeric RNA formation in *vir-1* mutant.

**miCLIP**

*Preparation of miCLIP libraries*

Total RNA for miCLIP was isolated from 7.5 mg of 14-day old Arabidopsis Col-0 seedlings as previously described (*31*). mRNA was isolated from ~1 mg of total RNA using oligo(dT) and streptavidin paramagnetic beads (PolyATtract® mRNA Isolation Systems, Promega) as per manufacturer’s instructions. miCLIP was carried out using 15 µg of mRNA as described in (*32*) using antibody against N6-methyladenosine (#202 003 Synaptic Systems) with minor modifications. No-antibody controls were processed throughout the experiment. 4-12% Bis-Tris gel electrophoresis was run at 180 V for 50 minutes, while transfer of the RNA-antibody complexes was carried out at 30 V for 60 minutes. Following the transfer, membranes were exposed to the Medical X-Ray Film Blue (Agfa) at -80°C overnight. Reverse transcription was carried out using barcoded RT primers - RT41, RT48, RT49 and RT50 (Integrated DNA Technologies) (Table S5). After reverse transcription, one cDNA fraction corresponding to 70-200 nt was purified from the 6% TBE-urea gel (Thermo Fisher Scientific). After final PCR, all libraries were pooled together and purified using Agencourt Ampure XP magnetic beads (Beckman Coulter) and eluted in nuclease-free water. Paired-end sequencing with read length of 100 bp was carried out on Illumina MiSeq v2 at Edinburgh Genomics of the University of Edinburgh. Input sample libraries were prepared using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs® Inc.) and sequenced on Illumina HiSeq2000 at the Tayside Centre for Genomics analysis of the University of Dundee with pair-end read length of 75bp.

*Processing for miCLIP sequencing data*

miCLIP data were assessed for quality using FastQC version 0.11.8 (*33*) and MultiQC version 1.7 (*34*). Only the forward read was used for analysis as the miCLIP site is located at the 5’ position of the forward read. 3’ adapter and poly(A) sequences were trimmed using cutadapt version 1.18 (*35*) and Unique Molecular Identifiers were extracted from the 5’ of the reads using UMI-tools version 0.5.5 (*36*). Immunoprecipitation and no antibody controls were demultiplexed and multiplexing barcodes were trimmed using seqkit version 0.10.0 (*11*). Reads were mapped to the TAIR10 reference genome with Araport11 reference annotation (*18*), using STAR version 2.6.1 (*4*), a maximum multimapping rate of 5, a minimum splice junction overhang of 8 nt (3 nt for junctions in the Araport11 reference), a maximum of 5 mismatches per read, and intron length boundaries of 60 - 10,000 nt. SAM and BAM file manipulations were performed using samtools version 1.9 (*8*). Removal of PCR duplicates was then performed using UMI-tools in directional model (*36*). miCLIP 5’ coverage and matched input 5’ coverage tracks were generated using bedtools version 2.27.1 (*9*) and these were used to call miCLIP peaks at single nucleotide resolution with Piranha version 1.2.1 (*37*) with relaxed *p*-value thresholds of 0.5. Reproducible peaks across pairwise combinations of the three replicates were identified by Irreproducible Discovery Rate (IDR) analysis using the python package idr version 2.0.3 using an IDR threshold of 0.05 (*38*). The final set of peaks was identified by pooling the three replicates and re-analysing using Piranha, then ranking the peaks by FDR and selecting the top N peaks, where N was the smallest number of reproducible peaks discovered by pairwise comparisons of the three replicates. This yielded 141,198 unique nucleotide level miCLIP peaks.**m6A LC-MS analysis**

The m6A content analysis using Liquid Chromatography - Mass Spectroscopy (LC-MS) was performed as described in (*39*). Chromatography was carried out by the FingerPrints Proteomics facility at the University of Dundee.

**Data Availability**

Sequencing datasets described in this study have been deposited at the European Nucleotide Archive (ENA): Study, PRJEB3993; accession no, ERP003245.

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

2. H. Lee, P. S. Pine, J. McDaniel, M. Salit, B. Oliver, External RNA controls consortium beta version update. *Journal of genomics* **4**, 19-22 (2016).

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

4. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21 (2013).

5. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417 (2017).

6. C. Soneson, M. Love, M. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. *F1000Research* **4**, (2016).

7. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140 (2010).

8. A. Wysoker *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

9. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841-842 (2010).

10. S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data. *Genome Research* **22**, 2008-2017 (2012).

11. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS ONE* **11**, e0163962 (2016).

12. T. A. G. Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).

13. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

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

15. T. Magoč, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* **27**, 2957-2963 (2011).

16. A. Heger, Belgrad, T. G., Goodson, M., & Jacobs, K., pysam: Python interface for the SAM/BAM sequence alignment and mapping format., (2014).

17. R. E. Workman *et al.*, Nanopore native RNA sequencing of a human poly(A) transcriptome. *bioRxiv*, 459529 (2018).

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

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

20. A. Dalke *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

21. N. Sheth *et al.*, Comprehensive splice-site analysis using comparative genomics. *Nucleic acids research* **34**, 3955-3967 (2006).

22. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC bioinformatics* **10**, 421-421 (2009).

23. N. J. Loman, J. Quick, J. T. Simpson, A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature methods* **12**, 733 (2015).

24. F. Chollet. (2018).

25. M. Abadi *et al.*, TensorFlow: large-scale machine learning on heterogeneous distributed systems. (2016).

26. X. Z. Kaiming He, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition. (2015).

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

28. T. L. Bailey *et al.*, MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202-W208 (2009).

29. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)* **27**, 1017-1018 (2011).

30. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41**, e108-e108 (2013).

31. V. Quesada, R. Macknight, C. Dean, G. G. Simpson, Autoregulation of *FCA* pre-mRNA processing controls Arabidopsis flowering time. *The EMBO Journal* **22**, 3142-3152 (2003).

32. A. V. Grozhik, B. Linder, A. O. Olarerin-George, S. R. Jaffrey, Mapping m6A at individual-nucleotide resolution using crosslinking and immunoprecipitation (miCLIP). *Methods in molecular biology (Clifton, N.J.)* **1562**, 55-78 (2017).

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

34. P. Ewels, M. Magnusson, M. Käller, S. Lundin, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).

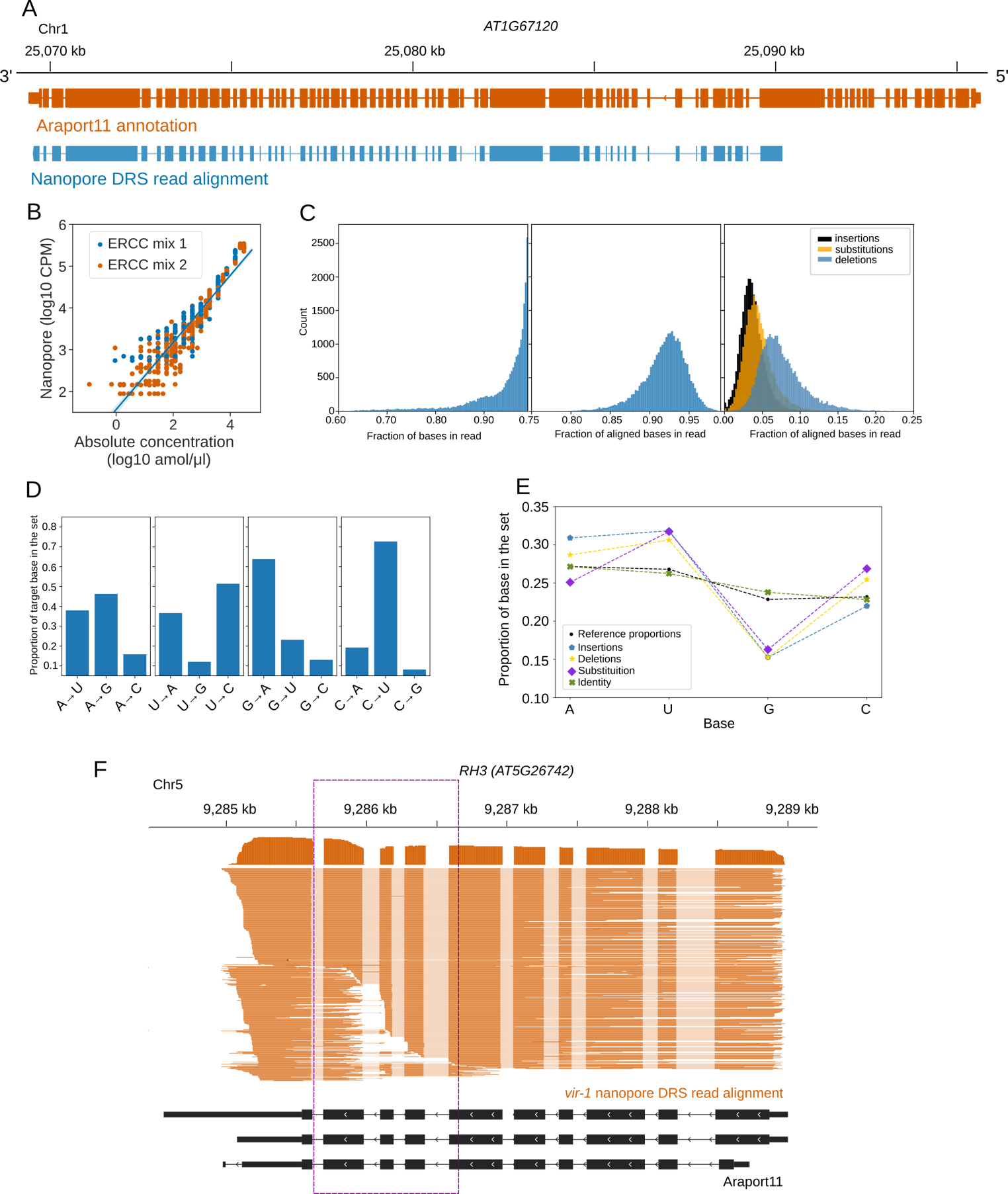
35. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).

36. T. Smith, A. Heger, I. Sudbery, UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome research* **27**, 491-499 (2017).

37. P. J. Uren *et al.*, Site identification in high-throughput RNA-protein interaction data. *Bioinformatics (Oxford, England)* **28**, 3013-3020 (2012).

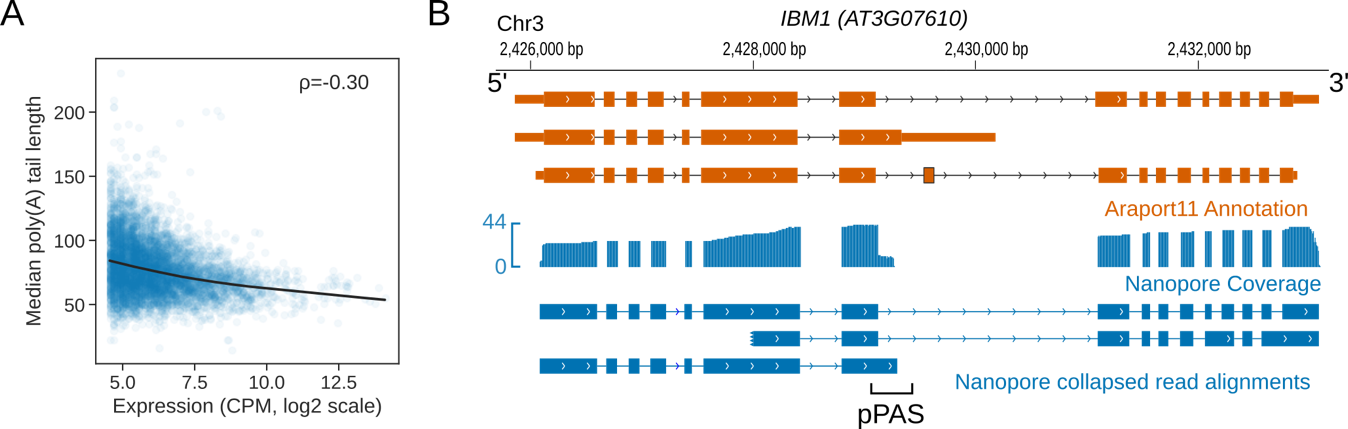
38. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752-1779 (2011).

39. H. Huang *et al.*, Recognition of RNA *N*6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature Cell Biology* **20**, 285-295 (2018).

****

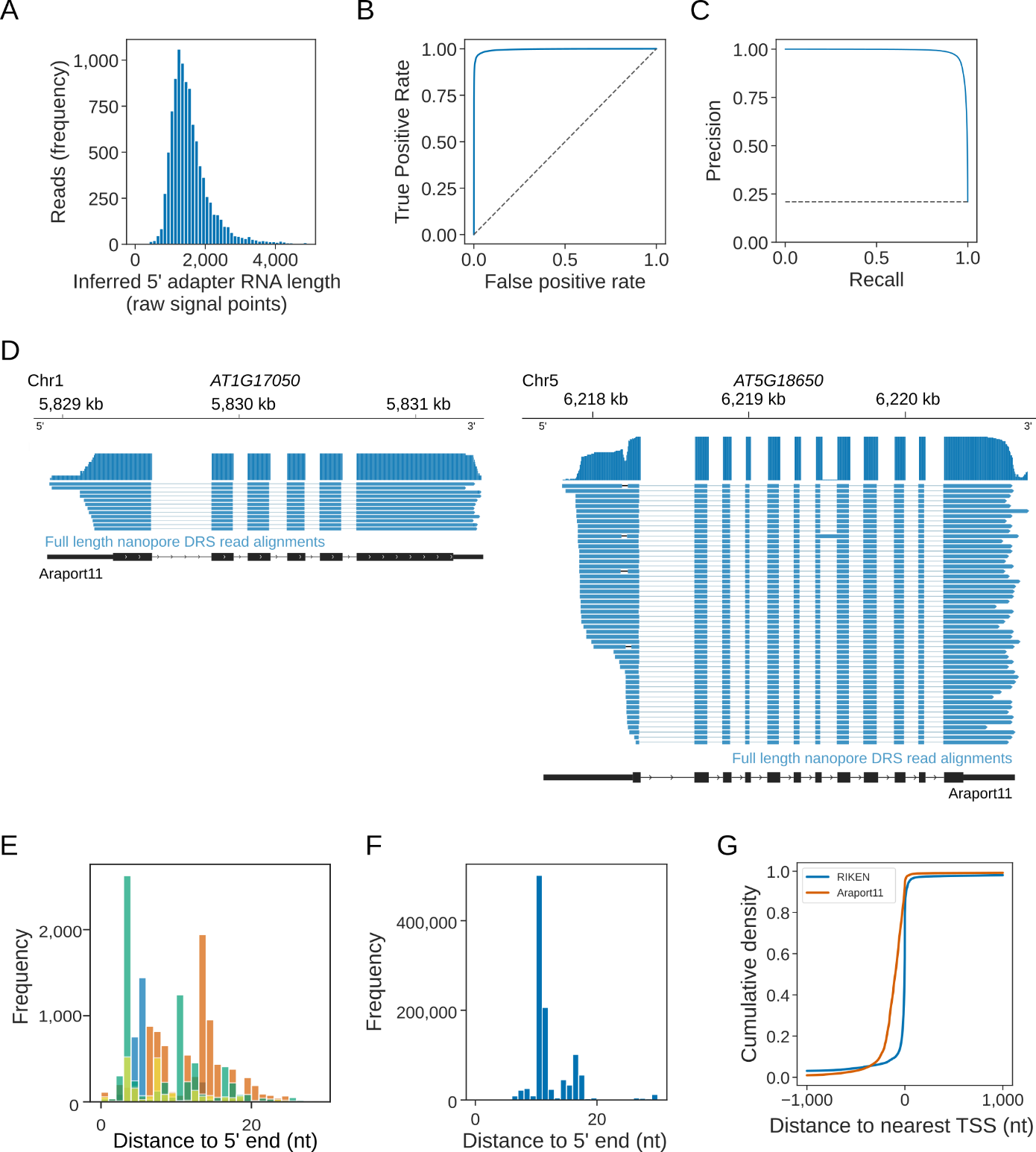
**Figure S1. Properties of the Nanopore DRS sequencing data.**

1. Nanopore DRS identified 12.8kb transcript generated from *AT1G67120* gene that includes 58 exons. An RNA isoform present in the *Araport11* annotation is shown in orange and an RNA isoform identified using nanopore DRS – in blue.
2. Synthetic ERCC RNA Spike-In mixes are detected in a quantitative manner. Absolute concentrations of spike-ins are plotted against counts per million reads in log10 scale. ERCC RNA Spike-In mix 1 is shown in blue and mix 2 – in orange.
3. Overview of the sequencing and alignment characteristics of nanopore DRS data for ERCC RNA Spike-Ins. Distribution of the length fraction of each sequenced read that aligns to the ERCC RNA Spike-In reference is shown on the left; distribution of fraction of identity that matches between the sequence of the read and the ERCC RNA Spike-In reference for the aligned portion of each read – in the centre; distributions of the occurrence of insertions (black), substitutions (orange) and deletions (blue) as a proportion of the number of aligned bases in each read – on the right.
4. Substitution preference for each nucleotide (A, U, G, C, left-to-right). When substituted, guanidine is replaced with adenosine in more than 63% of its substitutions, while cytosine is replaced by uridine 73% of the time. Conversely uridine is rarely replaced with guanidine (12%) and adenine is rarely substituted with cytosine (16%).
5. Nucleotide representation within the ERCC RNA Spike-In reference sequences (black dots) compared with the nucleotide representation within four categories from the nanopore DRS reads. Identity matches between the sequence of the read and the ERCC RNA Spike-In reference (green crosses), insertions (blue pentagons), deletions (yellow stars) and substitutions (purple diamonds). Guanidine (G) is under-represented, and uridine (U) is over-represented, for all three categories of error (insertion, deletion and substitution (relative to the reference nucleotide distribution). Cytosine (C) is over-represented in the set of deletions and substitutions. Adenine (A) is over-represented for insertions and deletions and, under-represented in the set of substitutions.
6. Signals originating from the *RH3* transcripts are susceptible to systematic over-splitting around exons 7-9 (highlighted using a purple dashed box), resulting in reads with apparently novel 5’ or 3’ positions. This appears to only occur at high frequency in datasets collected after May 2018 (Table S1) and may be the result of an update to the MinKNOW software.

****

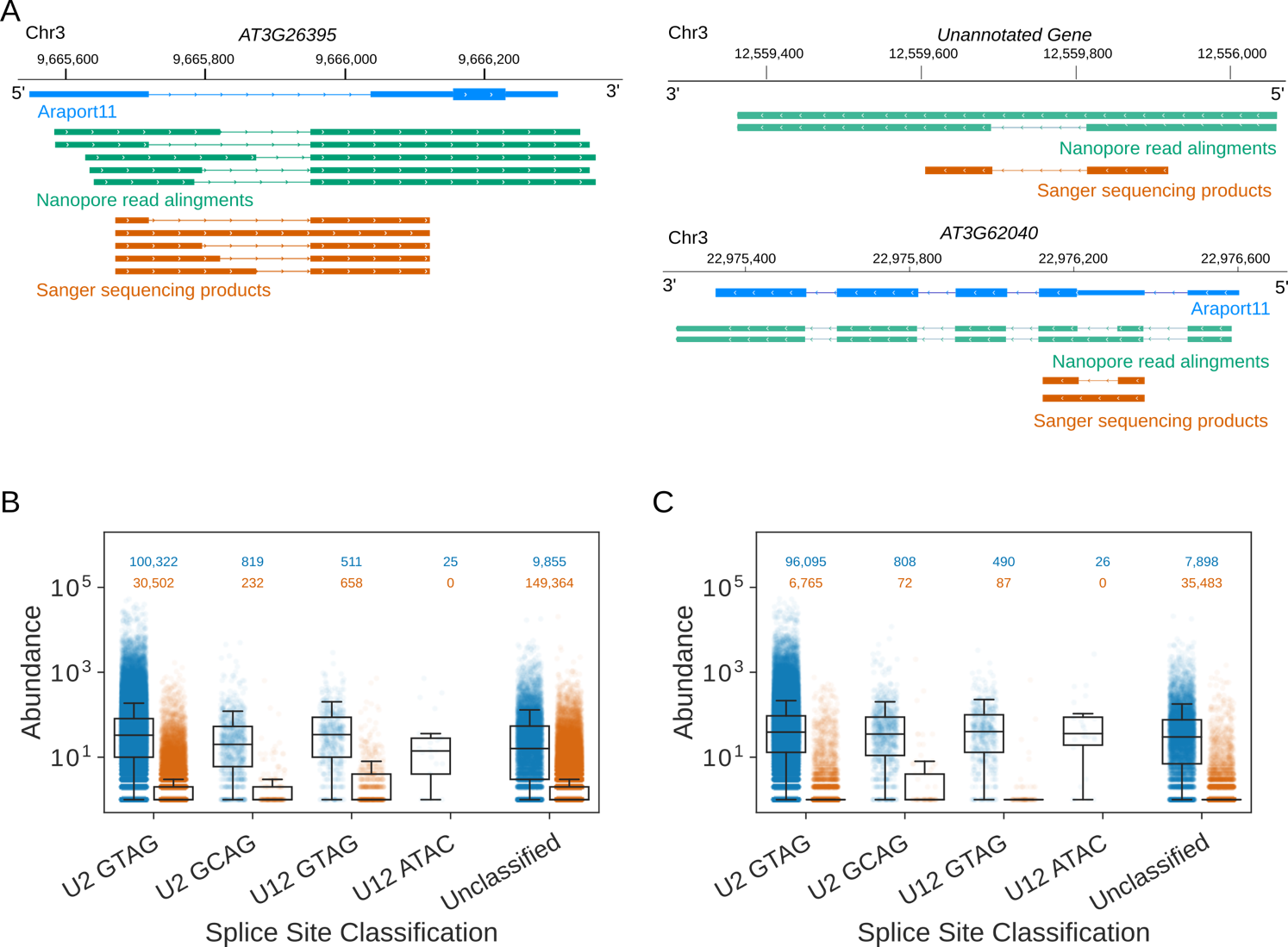
**Figure S2. 3’ end processing information provided by nanopore DRS.**

1. Poly(A) tail length of RNA negatively correlates with the expression of the gene. Expression in log scale of counts per million (CPM) obtained from nanopore DRS data is plotted against median poly(A) tail length. ρ - Spearman rho value.
2. Nanopore DRS identified new 3’ polyadenylation sites in RNAs transcribed from *IBM1 (AT3G07610)* gene. Blue track shows coverage of nanopore DRS reads. Isoforms found in the Araport11 annotation are shown in orange and those detected by nanopore DRS are in blue. pPAS is proximal polyadenylation site.



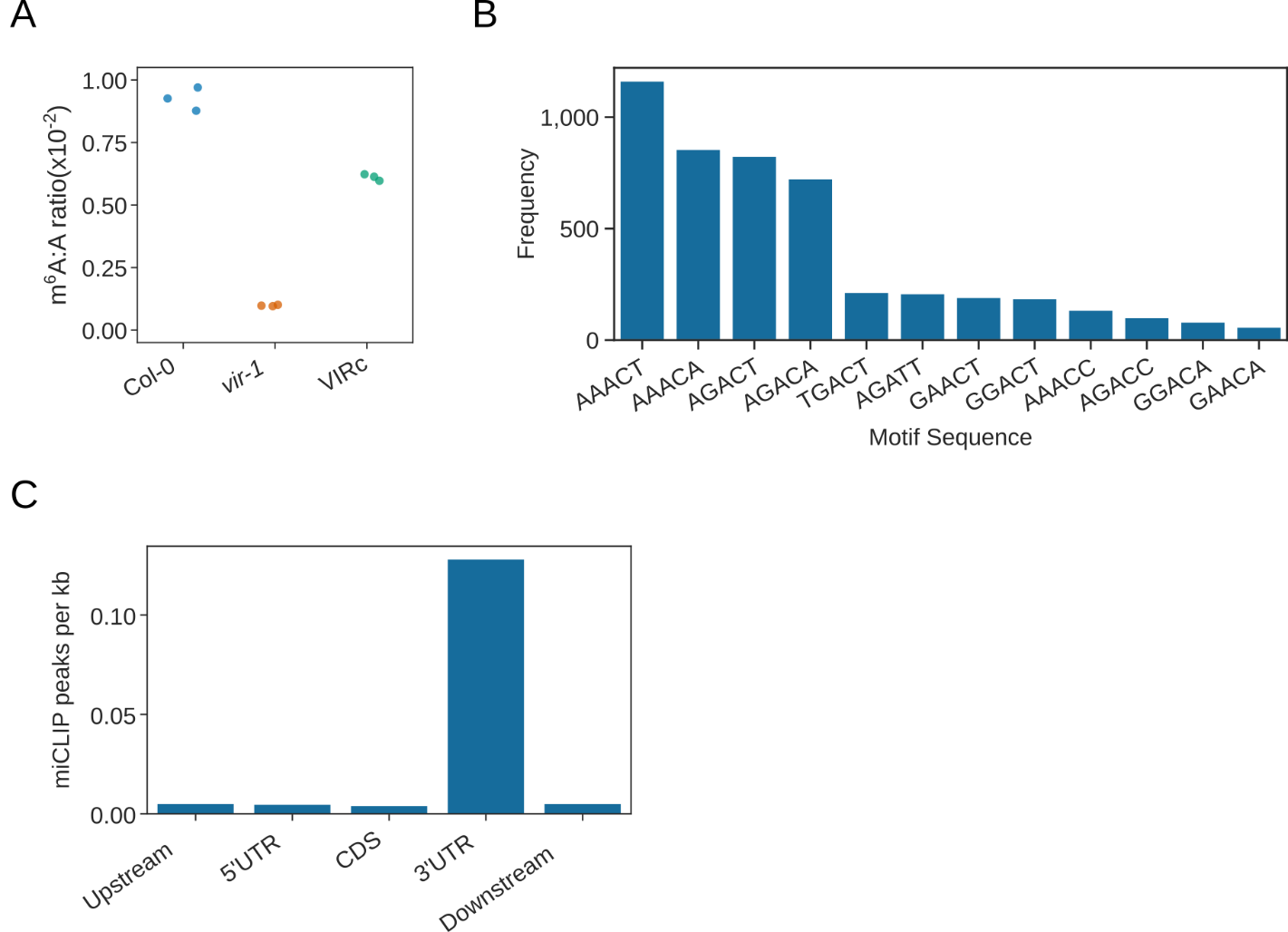
**Figure S3. 5’ end processing information provided by nanopore DRS with cap-dependent ligation of 5’ adapter RNA.**

1. Histogram showing the distribution of 5’ adapter RNA length in nanopore raw current signal, inferred from alignment of mRNA sequence to signal using nanopolish eventalign. The median signal length was 1441 points, and 96% of adapter signals were 3000 points or less.
2. Out-of-bag Receiver Operator Characteristic curve showing performance of trained convolutional neural network at detecting 5’ adapter RNA using 3000 points of signal. Curve was generated using 5-fold cross validation.
3. Out-of-bag precision recall curve showing performance of trained neural network, generated using 5-fold cross validation.
4. Alternative transcription start sites were identified using nanopore DRS with cap-dependent ligation of 5’ end adapter at *AT1G17050* and *AT5G18650* genes. Blue track shows coverage of nanopore DRS reads. Isoforms found in Araport11 annotation are shown in black and those detected by nanopore DRS with cap-dependent ligation of 5’ adapter RNA are in blue.
5. Reads mapping to ERCC RNA Spike-Ins lack approximately 11nt of sequence at the 5’ end. Histogram shows distance to the 5’ end for ERCC RNA Spike-In reads (each shown in different colours, only spike ins with more than 1000 supporting reads shown).
6. Reads mapping to *in vitro* transcribed mGFP lack approximately 11 nt of sequence at the 5’ end. Histogram shows distance to the 5’ end for *in vitro* transcribed mGFP.
7. Araport11 annotation overestimates the length of 5’ UTRs. Cumulative distribution function shows the distance to the nearest TSS identified from full length transcripts cloned as part of the RIKEN RAFL project (blue) and Araport11 annotation (orange).

****

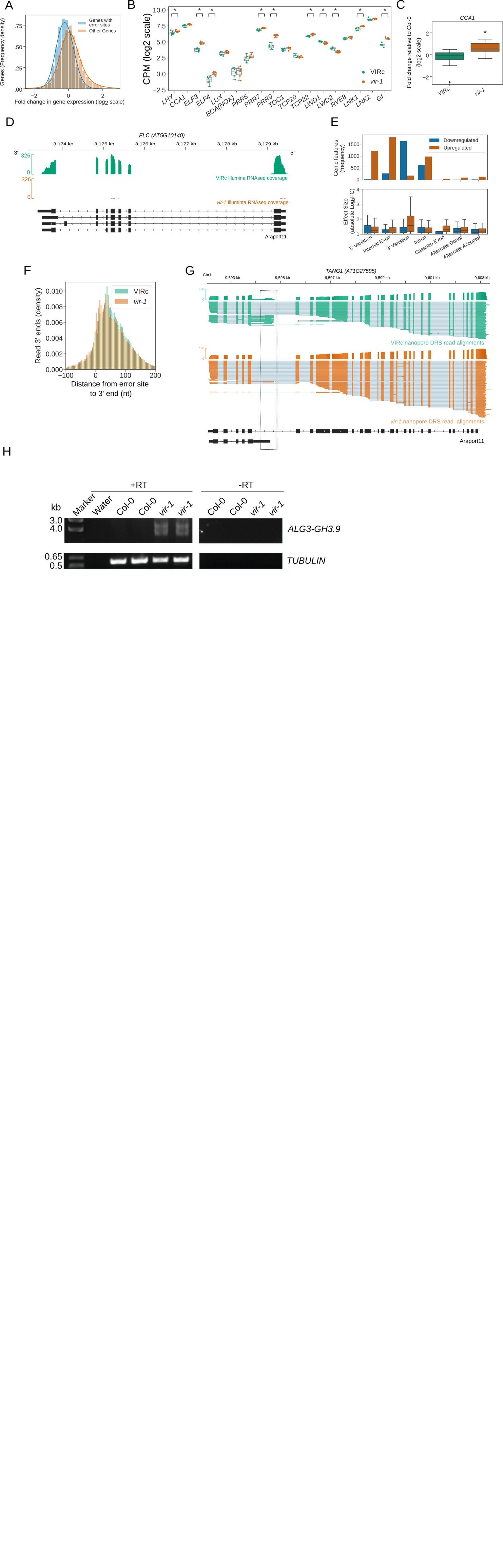
**Figure S4. New details on splicing identified using nanopore DRS.**

1. Nanopore DRS can be validated using RT-PCR. From the top 20 most highly expressed RNAs with novel splice sites, three were selected and validated by RT-PCR followed by sequencing of the DNA products. RNA isoforms present in Araport11 annotation are shown in blue; RNA isoforms found using nanopore DRS – in green; Sanger sequencing products obtained using RT-PCR – in orange.
2. Classification of annotated (blue) and novel (orange) splice sites found non-error corrected nanopore DRS data. Counts are plotted in log10 scale and the exact numbers are indicated above each category in blue for annotated events and in orange for novel events.
3. Classification of annotated (blue) and unannotated (orange) splice sites found error corrected nanopore DRS data. Counts are plotted in log10 scale and the exact numbers are indicated above each category in blue for annotated events and in orange for novel events. 50% of unannotated splice junctions identified before error correction could not be classified by the position weight matrix approach. In addition, these sites were not detected after error correction, suggesting that they result from alignment errors.

****

**Figure S5. Identification of VIR-dependent modifications transcriptome-wide.**

1. VIR complemented line shows increase in m6A methylation compared to *vir-1* mutant**.** Col-0 ratio of methylated adenosine at the 6th position to unmethylated adenosine obtained using LC-MS analysis is shown in blue; *vir-1* mutant ratio – in orange; VIR complemented line ratio – in green.
2. Frequency of m6A motifs detected at *vir-1* reduced error sites, as detected by FIMO using the motif detected *de novo* by MEME (Fig 5B) and an FDR threshold of 0.1.
3. Bar plot shows the number of miCLIP peaks 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.



**Figure S6. Changes in circadian clock and generation of chimeric RNAs are observed in the mutant with reduced m6A modification.**

1. Histogram shows log2 fold change in gene expression based on Illumina RNAseq data from *vir-1* mutant and VIR complemented line. Genes with differential error rate sites are in blue, and those without – in orange.
2. Expression of core circadian clock components is perturbed in *vir-1* mutant. Boxplots showing normalised gene expression measured using Illumina RNAseq in log2 counts per million, for VIR complemented line – in green; and *vir-1* mutant*–* in orange. Asterisks denote expression changes which are significant using an FDR threshold of 0.05.
3. Expression of *CCA1*, encoding a regulator of circadian rhythm in Arabidopsis, is increased in *vir-1* mutant. Boxplot shows gene expression change from Col-0 by RT-qPCR for VIR complemented line – in green; and *vir-1 –* in orange. Expression change is significant using a p value threshold of 0.05.
4. Expression of *Flowering Locus C* (*FLC*) gene is decreased in the *vir-1* mutant. Gene tracks show gene expression from Illumina RNAseq data for VIR complemented line (green) and *vir-1* mutant (orange).
5. Splicing is moderately disrupted in *vir-1* mutant. Less retained (blue) and more retained (orange) introns are shown in various categories of splice sites.
6. A shift in use of more proximal polyadenylation site is observed in *vir-1* mutant. Histogram shows distance from error site to 3’ end in *vir-1* mutant (orange) and VIR complemented line (green).
7. *ALG3-GH3.9* chimeric RNAs are generated in *vir-1* mutant. RT-PCR gel shows formation of chimeric RNAs in the ­*vir-1* mutant compared to Col-0.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Estimated Over-splitting rate (%)** | 0.047 | 0.045 | 0.111 | 0.035 | 0.047 | 0.040 | 0.019 | 0.019 | 0.069 | 0.809 | 0.959 | 1.600 | 0.878 | 1.467 | 0.887 |
| **Percent Mapped to TAIR10** | 60.89 | 70.25 | 79.04 | 76.94 | 25.32 | 74.89 | 70.12 | 75.98 | 27.30 | 90.52 | 94.62 | 91.85 | 94.74 | 93.78 | 93.53 |
| **Median length of reads mapped to TAIR10** | 865 | 878 | 873 | 878 | 737 | 858 | 840 | 840 | 732 | 871 | 846 | 832 | 884 | 802 | 849 |
| **Longest read mapped to TAIR10** | 12,607 | 12,446 | 11,994 | 11,852 | 5,663 | 11,434 | 12,744 | 11,542 | 5,596 | 12,015 | 15,040 | 10,836 | 15,291 | 11,522 | 11,471 |
| **Median length of reads mapped to ERCC spike-ins** | 481 | 492 | 493 | 492 | 486 | 488 | 486 | 495 | 507 | 497 | 491 | 493 | 480 | 487 | 500 |
| **Longest read mapped to ERCC spike-ins** | 1,905 | 2,304 | 1,108 | 1,099 | 1,928 | 1,094 | 1,772 | 1,292 | 2,308 | 2,236 | 1,880 | 1,947 | 1,134 | 1,128 | 2,127 |
| **Reads mapped to TAIR10** | 1,003,137 | 1,005,664 | 1,101,368 | 743,684 | 345,799 | 930,629 | 757,028 | 765,322 | 557,949 | 1,538,040 | 812,517 | 1,318,857 | 1,746,153 | 1,271,511 | 1,570,040 |
| **Reads mapped to ERCC spike-ins** | 6756 | 1074 | 713 | 556 | 7005 | 701 | 907 | 525 | 11324 | 1824 | 1094 | 2710 | 865 | 1436 | 1548 |
| **Total Reads Base-called** | 1,647,484 | 1,431,457 | 1,393,351 | 966,529 | 1,365,809 | 1,242,616 | 1,079,578 | 1,007,278 | 2,043,751 | 1,699,123 | 858,747 | 1,435,808 | 1,843,192 | 1,355,795 | 1,678,723 |
| **5' Adapter Ligation** | - | *-* | - | - | + | - | - | - | + | - | *-* | - | *-* | *-* | - |
| **Bio Replicate** | 1 | *1* | 1 | 2 | 1 | 2 | 3 | 4 | 2 | 2 | *2* | 4 | *3* | *4* | 3 |
| **Genotype** | Col-0 | *vir-1* | VIRc | Col-0 | Col-0 | Col-0 | Col-0 | Col-0 | Col-0 | VIRc | *vir-1* | VIRc | *vir-1* | *vir-1* | VIRc |
| **Sequencing Date** | 01/02/2018 | 27/02/2018 | 01/03/2018 | 05/04/2018 | 11/04/2018 | 13/04/2018 | 16/04/2018 | 18/04/2018 | 08/05/2018 | 07/09/2018 | 12/09/2018 | 25/09/2018 | 28/09/2018 | 03/10/2018 | 19/10/2018 |

**Table S1. Properties of the nanopore DRS sequencing data.**

Dataset statistics for all nanopore DRS sequencing runs conducted. Datasets are sorted by date of sequencing run. All data was collected using a MinION with R9.4 flow cell and SQK-RNA001 library kit. Increases in mapping and oversplitting rate which occur in samples collected after September 2018 are therefore likely to have resulted from changes in the MinKNOW software.

A)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Replicate 1 | Adapter | | No Adapter | | Signal to Noise |
| Fails | Passes | Fails | Passes |
| Match length > 10 nt | 58,078 | 211,786 | 47,548 | 11,427 | 15.17 |
| Log10 E Value < 2 | 76,926 | 192,938 | 57,874 | 1,101 | 131.8 |
| Adapter matches directly after alignment | 198,296 | 71,568 | 58,936 | 39 | 545.4 |

B)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Replicate 2 | Adapter | | No Adapter | | Signal to Noise |
| Fails | Passes | Fails | Passes |
| Match length > 10 nt | 37,631 | 342,158 | 100,791 | 12,653 | 72.43 |
| Log10 E Value < 2 | 60,605 | 319,184 | 112,781 | 663 | 895.9 |
| Adapter matches directly after alignment | 266,812 | 112,977 | 113,429 | 15 | 3,202 |

**Table S2. Adapter detection using BLASTN rules approach.**

Table shows the number of reads with adapters detected in two biological replicates of Col-0 sequenced (A and B) with and without adapter ligation protocol. Rules are applied cumulatively, *i.e*., row one shows reads that pass first rule row two shows reads that pass first and second rule, etc. Signal to noise ratio shows the number of positive examples detected using rules in adapter ligated dataset divided by the number of false positives from dataset collected without adapters.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **AtRTD2** | **Araport11** | **Illumina** | **Nanopore EC** | **Nanopore** | **Canonical** | **Noncanonical** |
|  |  |  |  |  | 99,966 | 805 |
|  |  |  |  |  | 84,630 | 79,377 |
|  |  |  |  |  | 25,769 | 57,298 |
|  |  |  |  |  | 20,893 | 404 |
|  |  |  |  |  | 10,882 | 21,484 |
|  |  |  |  |  | 7,301 | 545 |
|  |  |  |  |  | 6,941 | 210 |
|  |  |  |  |  | 6,256 | 879 |
|  |  |  |  |  | 4,801 | 1,894 |
|  |  |  |  |  | 3,939 | 160 |
|  |  |  |  |  | 3,665 | 31 |
|  |  |  |  |  | 2,844 | 59 |
|  |  |  |  |  | 2,424 | 28 |
|  |  |  |  |  | 2,006 | 18 |
|  |  |  |  |  | 832 | 453 |
|  |  |  |  |  | 828 | 5 |
|  |  |  |  |  | 585 | 6 |
|  |  |  |  |  | 512 | 6 |
|  |  |  |  |  | 484 | 31 |
|  |  |  |  |  | 390 | 53 |
|  |  |  |  |  | 369 | 33 |
|  |  |  |  |  | 326 | 2 |
|  |  |  |  |  | 193 | 6 |
|  |  |  |  |  | 148 | 8 |
|  |  |  |  |  | 76 | 3 |
|  |  |  |  |  | 47 | 4 |
|  |  |  |  |  | 24 | 0 |
|  |  |  |  |  | 21 | 0 |
|  |  |  |  |  | 20 | 0 |
|  |  |  |  |  | 18 | 0 |
|  |  |  |  |  | 1 | 0 |

**Table S3. Splice junctions supported by nanopore DRS and Illumina RNAseq**

Table shows the numbers for the unique splice junction set intersections upset plot (Figure 4B). Shaded cells denote sets included in the intersection for that row, whilst unshaded cells denote sets excluded from the intersection. Rows are sorted by the size of the intersection for canonical splice junctions.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **AGI ID** | **Gene Name** | **Pathway** | **Log2**  **Fold**  **Change** | **Log2**  **counts**  **per**  **million** | **False**  **Discovery**  **Rate** | **Differentially**  **Expressed**  **(abs logFC > 1,**  **FDR < 0.05)** |
| *AT2G19520* | FVE | Autonomous | -0.40 | 5.49 | 3.95E-08 | - |
| *AT2G21660* | ATGRP7 | Autonomous | 0.03 | 6.11 | 8.57E-01 | - |
| *AT2G43410* | FPA | Autonomous | 0.70 | 5.03 | 1.75E-11 | - |
| *AT3G04610* | FLK | Autonomous | 0.29 | 5.18 | 5.90E-04 | - |
| *AT3G10390* | FLD | Autonomous | 0.31 | 3.80 | 6.68E-03 | - |
| *AT4G02560* | LD | Autonomous | 0.68 | 4.36 | 7.49E-09 | - |
| *AT4G16280* | FCA | Autonomous | 0.71 | 5.02 | 2.90E-09 | - |
| *AT5G13480* | FY | Autonomous | 0.49 | 3.80 | 8.78E-04 | - |
| *AT4G22950* | AGL19 | FLC-independent vernalization pathway | 1.10 | 0.81 | 9.17E-04 | + |
| *AT2G45660* | SOC1 | Floral integrator | -1.64 | 6.72 | 4.55E-84 | + |
| *AT5G60910* | FUL | Floral integrator and flower meristem identity | -1.13 | 2.09 | 1.84E-09 | + |
| *AT1G25560* | TEM1 | Floral repressor | -0.81 | 6.50 | 3.83E-02 | - |
| *AT1G68840* | TEM2 | Floral repressor | -0.19 | 7.08 | 6.30E-01 | - |
| *AT2G22540* | SVP | Floral repressor | -0.29 | 6.96 | 2.20E-03 | - |
| *AT3G57390* | AGL18 | Floral repressor | 0.10 | 3.30 | 5.36E-01 | - |
| *AT5G13790* | AGL15 | Floral repressor | 0.52 | 1.02 | 7.22E-02 | - |
| *AT1G24260* | SEP3 | Flower development | 1.02 | 2.82 | 9.91E-09 | + |
| *AT1G02400* | GA2ox6 | Gibberellin | 0.43 | 2.09 | 3.69E-02 | - |
| *AT1G14920* | GAI | Gibberellin | -0.27 | 6.22 | 5.95E-03 | - |
| *AT1G15550* | GA3ox1 | Gibberellin | -0.68 | 1.36 | 1.72E-02 | - |
| *AT1G18075* | miR159b | Gibberellin | 0.72 | 0.13 | 4.98E-02 | - |
| *AT1G30040* | GA2ox2 | Gibberellin | -0.08 | 1.86 | 8.15E-01 | - |
| *AT1G47990* | GA2ox4 | Gibberellin | -0.12 | 0.76 | 7.18E-01 | - |
| *AT1G78440* | GA2ox1 | Gibberellin | -1.43 | -0.19 | 2.57E-04 | + |
| *AT1G79460* | GA2 | Gibberellin | -0.33 | 2.80 | 4.69E-02 | - |
| *AT2G01570* | RGA | Gibberellin | -0.09 | 7.22 | 4.01E-01 | - |
| *AT3G05120* | GID1A | Gibberellin | -0.33 | 4.93 | 1.74E-03 | - |
| *AT3G11440* | MYB65 | Gibberellin | -0.37 | 1.99 | 4.23E-02 | - |
| *AT3G63010* | GID1B | Gibberellin | 0.76 | 3.51 | 4.26E-12 | - |
| *AT4G02780* | CPS1 | Gibberellin | 0.68 | 0.33 | 2.28E-02 | - |
| *AT4G24210* | SLY1 | Gibberellin | -0.39 | 4.02 | 2.31E-03 | - |
| *AT4G25420* | GA20ox1 | Gibberellin | 0.18 | 1.48 | 5.58E-01 | - |
| *AT5G06100* | MYB33 | Gibberellin | 0.17 | 3.76 | 1.20E-01 | - |
| *AT5G27320* | GID1C | Gibberellin | -0.06 | 4.74 | 6.60E-01 | - |
| *AT5G51810* | GA20ox2 | Gibberellin | -1.75 | 1.94 | 7.84E-14 | + |
| *AT5G03840* | TFL1 | Inflorescence architecture | -1.68 | -0.43 | 1.64E-05 | + |
| *AT2G27990* | PNF | Inflorescence meristem development | 0.52 | 1.78 | 1.60E-03 | - |
| *AT5G02030* | PNY | Inflorescence meristem development | 0.08 | 4.08 | 5.24E-01 | - |
| *AT1G04400* | CRY2 | Light perception | -0.28 | 7.25 | 4.69E-03 | - |
| *AT1G09570* | PHYA | Light perception | 0.01 | 7.56 | 9.35E-01 | - |
| *AT2G18790* | PHYB | Light perception | 0.13 | 6.35 | 2.45E-01 | - |
| *AT1G25540* | PFT1 | Light perception | 0.28 | 4.49 | 9.86E-03 | - |
| *AT4G34530* | CIB1 | Light signalling | -0.99 | 4.14 | 2.13E-04 | - |
| *AT5G46210* | CUL4 | Light signalling | 0.14 | 6.54 | 7.76E-02 | - |
| *AT4G24540* | AGL24 | Meristem response | -1.14 | 2.49 | 1.99E-08 | + |
| *AT4G35900* | FD | Meristem response | -0.31 | 1.78 | 2.46E-01 | - |
| *AT1G53090* | SPA4 | Photoperiod | 0.17 | 4.96 | 6.88E-02 | - |
| *AT1G65480* | FT | Photoperiod | -1.75 | -0.32 | 6.10E-08 | + |
| *AT1G69570* | CDF5 | Photoperiod | -0.34 | 4.05 | 7.97E-02 | - |
| *AT2G28550* | TOE1 | Photoperiod | 0.94 | 6.42 | 1.16E-24 | - |
| *AT2G32950* | COP1 | Photoperiod | 0.53 | 5.43 | 5.84E-09 | - |
| *AT2G34720* | NFYA4 | Photoperiod | -0.31 | 5.45 | 7.93E-03 | - |
| *AT2G38880* | NFYB1 | Photoperiod | -0.44 | 3.93 | 8.79E-06 | - |
| *AT2G39250* | SNZ | Photoperiod | 0.65 | 4.99 | 2.46E-09 | - |
| *AT2G46340* | SPA1 | Photoperiod | 0.58 | 7.45 | 4.89E-11 | - |
| *AT2G47700* | RFI2 | Photoperiod | -0.44 | 5.55 | 3.52E-05 | - |
| *AT3G15354* | SPA3 | Photoperiod | 0.84 | 6.06 | 1.01E-13 | - |
| *AT3G47500* | CDF3 | Photoperiod | 0.62 | 6.44 | 8.79E-11 | - |
| *AT3G48590* | NFYC1 | Photoperiod | -0.67 | 4.61 | 2.85E-09 | - |
| *AT3G54990* | SMZ | Photoperiod | 0.54 | 2.37 | 1.39E-02 | - |
| *AT4G11110* | SPA2 | Photoperiod | 0.28 | 5.26 | 2.56E-03 | - |
| *AT4G14540* | NFYB3 | Photoperiod | 0.23 | 5.21 | 1.93E-02 | - |
| *AT5G12840* | NFYA1 | Photoperiod | -0.30 | 5.15 | 2.97E-04 | - |
| *AT5G39660* | CDF2 | Photoperiod | -0.25 | 5.03 | 1.65E-01 | - |
| *AT5G47640* | NFYB2 | Photoperiod | -0.30 | 6.06 | 2.59E-03 | - |
| *AT5G57660* | COL5 | Photoperiod | -0.67 | 8.85 | 7.86E-07 | - |
| *AT5G60120* | TOE2 | photoperiod | 0.38 | 4.71 | 6.29E-05 | - |
| *AT5G62430* | CDF1 | Photoperiod | -0.01 | 5.50 | 9.25E-01 | - |
| *AT3G22380* | TIC | Photoperiod, circadian clock | 0.68 | 8.16 | 5.64E-15 | - |
| *AT5G57360* | ZTL | Photoperiod, circadian clock | 0.18 | 6.13 | 3.36E-02 | - |
| *AT1G01060* | LHY | Photoperiod, circadian clock | 0.32 | 6.54 | 1.85E-02 | - |
| *AT1G09530* | PIF3 | Photoperiod, circadian clock | -0.31 | 4.76 | 4.49E-02 | - |
| *AT1G12910* | LWD1 | Photoperiod, circadian clock | -0.29 | 4.94 | 3.13E-03 | - |
| *AT1G22770* | GI | Photoperiod, circadian clock | 0.93 | 5.13 | 3.98E-16 | - |
| *AT2G18915* | LKP2 | Photoperiod, circadian clock | -0.38 | 4.47 | 2.23E-03 | - |
| *AT2G21070* | FIO1 | Photoperiod, circadian clock | 0.31 | 2.93 | 1.67E-02 | - |
| *AT2G25930* | ELF3 | Photoperiod, circadian clock | 0.99 | 4.41 | 1.04E-14 | - |
| *AT2G40080* | ELF4 | Photoperiod, circadian clock | 1.01 | -0.36 | 2.20E-03 | + |
| *AT2G46790* | PRR9 | Photoperiod, circadian clock | 1.55 | 5.39 | 3.77E-20 | + |
| *AT2G46830* | CCA1 | Photoperiod, circadian clock | 0.04 | 7.64 | 7.18E-01 | - |
| *AT3G04910* | WNK1 | Photoperiod, circadian clock | -0.41 | 7.16 | 2.29E-04 | - |
| *AT3G26640* | LWD2 | Photoperiod, circadian clock | -0.57 | 3.75 | 1.89E-06 | - |
| *AT3G46640* | LUX | Photoperiod, circadian clock | 0.26 | 3.28 | 1.86E-01 | - |
| *AT3G60250* | CKB3 | Photoperiod, circadian clock | -0.53 | 5.49 | 3.48E-06 | - |
| *AT4G08920* | CRY1 | Photoperiod, circadian clock | -0.36 | 8.36 | 5.54E-05 | - |
| *AT4G16250* | PHYD | Photoperiod, circadian clock | 1.14 | 4.13 | 1.97E-19 | + |
| *AT4G18130* | PHYE | Photoperiod, circadian clock | 0.77 | 5.50 | 1.49E-14 | - |
| *AT5G02810* | PRR7 | Photoperiod, circadian clock | 0.23 | 7.01 | 5.61E-03 | - |
| *AT5G08330* | CHE | Photoperiod, circadian clock | -0.86 | 6.21 | 8.99E-11 | - |
| *AT5G24470* | PRR5 | Photoperiod, circadian clock | 0.31 | 2.67 | 2.29E-01 | - |
| *AT5G35840* | PHYC | Photoperiod, circadian clock | 0.46 | 5.53 | 2.50E-06 | - |
| *AT5G37260* | CIR1 | Photoperiod, circadian clock | 0.25 | 3.12 | 1.60E-01 | - |
| *AT5G59560* | SRR1 | Photoperiod, circadian clock | -0.28 | 3.43 | 4.74E-02 | - |
| *AT5G60100* | PRR3 | Photoperiod, circadian clock | 1.69 | 0.92 | 5.93E-11 | + |
| *AT5G61380* | PRR1 | Photoperiod, circadian clock | 0.12 | 3.91 | 3.92E-01 | - |
| *AT5G64813* | LIP1 | Photoperiod, circadian clock | -0.33 | 4.75 | 5.34E-04 | - |
| *AT1G68050* | FKF1 | Photoperiod, circadian clock | 0.68 | -0.32 | 3.91E-02 | - |
| *AT5G17690* | TFL2 | Photoperiod, vernalization and flower development | 0.00 | 4.46 | 9.97E-01 | - |
| *AT2G42200* | SPL9 | Vegetative and reproductive phase change | 0.42 | 3.03 | 8.60E-03 | - |
| *AT3G57920* | SPL15 | Vegetative and reproductive phase change | 0.89 | 0.49 | 1.34E-02 | - |
| *AT1G53160* | SPL4 | Vegetative to reproductive transition | -1.75 | 2.89 | 1.38E-20 | + |
| *AT2G25095* | miR156a | Vegetative to reproductive transition | 0.49 | 0.32 | 9.78E-02 | - |
| *AT2G33810* | SPL3 | Vegetative to reproductive transition | -1.55 | 4.69 | 4.76E-20 | + |
| *AT3G15270* | SPL5 | Vegetative to reproductive transition | -1.94 | 1.48 | 9.23E-15 | + |
| *AT4G31877* | miR156c | Vegetative to reproductive transition | -0.67 | 0.51 | 2.22E-02 | - |
| *AT1G17760* | Cstf77 | Vernalization | 0.23 | 4.42 | 4.52E-02 | - |
| *AT1G71800* | Cstf64 | Vernalization | 0.52 | 3.29 | 4.32E-05 | - |
| *AT3G18990* | VRN1 | Vernalization | -0.12 | 4.78 | 1.54E-01 | - |
| *AT4G00650* | FRI | Vernalization | 0.26 | 2.78 | 1.51E-01 | - |
| *AT5G10140* | FLC | Vernalization | -6.06 | 1.29 | 2.44E-64 | + |
| *AT1G61040* | VIP5 |  | 0.42 | 4.98 | 3.00E-08 | - |
| *AT2G06210* | VIP6 |  | 0.09 | 5.54 | 3.63E-01 | - |
| *AT2G31650* | ATX1 |  | 0.51 | 2.96 | 7.01E-03 | - |
| *AT3G48430* | REF6 |  | 0.26 | 5.21 | 1.89E-02 | - |
| *AT4G20400* | Jmj4 |  | 0.22 | 5.26 | 1.81E-02 | - |
| *AT5G04240* | ELF6 |  | 0.34 | 4.96 | 4.65E-04 | - |
| *AT1G08970* | NFYC3 |  | -0.64 | 5.32 | 1.57E-15 | - |
| *AT1G14400* | UBC1 |  | -0.59 | 6.65 | 1.10E-10 | - |
| *AT1G54830* | NFYC9 |  | -0.57 | 4.25 | 8.50E-08 | - |
| *AT1G55250* | HUB2 |  | -0.12 | 4.83 | 2.25E-01 | - |
| *AT1G56170* | NFYC2 |  | -0.29 | 3.59 | 2.10E-02 | - |
| *AT1G76710* | SDG26 |  | -0.01 | 3.53 | 9.47E-01 | - |
| *AT1G77080* | MAF1 |  | -0.31 | 5.01 | 2.71E-04 | - |
| *AT1G77300* | EFS |  | 0.87 | 5.69 | 2.31E-11 | - |
| *AT1G79730* | ELF7 |  | 0.58 | 5.26 | 2.31E-11 | - |
| *AT2G02760* | UBC2 |  | -0.35 | 6.12 | 2.08E-05 | - |
| *AT2G23380* | CLF |  | 0.18 | 4.07 | 2.68E-01 | - |
| *AT2G33835* | FES |  | 0.18 | 3.32 | 1.16E-01 | - |
| *AT2G44950* | HUB1 |  | 0.11 | 4.92 | 2.45E-01 | - |
| *AT2G45640* | SAP18 |  | -0.86 | 5.11 | 1.98E-15 | - |
| *AT3G12810* | PIE1 |  | 0.93 | 5.33 | 1.07E-17 | - |
| *AT3G24440* | VIL1 |  | 0.41 | 3.99 | 3.18E-04 | - |
| *AT3G33520* | ESD1 |  | -0.16 | 4.09 | 1.60E-01 | - |
| *AT3G49600* | UBP26 |  | 0.41 | 4.92 | 2.39E-05 | - |
| *AT4G15880* | ESD4 |  | 0.00 | 4.68 | 9.60E-01 | - |
| *AT4G16845* | VRN2 |  | -0.03 | 4.17 | 7.95E-01 | - |
| *AT4G29830* | VIP3 |  | -0.22 | 3.96 | 1.23E-02 | - |
| *AT4G32980* | ATH1 |  | 0.94 | 4.91 | 2.84E-15 | - |
| *AT4G39400* | BRI1 |  | 0.22 | 7.58 | 1.34E-01 | - |
| *AT5G11530* | EMF1 |  | 0.43 | 4.70 | 4.55E-04 | - |
| *AT5G16320* | FRL1 |  | 0.09 | 2.57 | 6.41E-01 | - |
| *AT5G23150* | HUA2 |  | 1.38 | 5.55 | 2.97E-26 | + |
| *AT5G37055* | SEF |  | -0.44 | 2.64 | 2.11E-03 | - |
| *AT5G51230* | EMF2 |  | 0.36 | 5.12 | 8.51E-06 | - |
| *AT5G61150* | VIP4 |  | 0.07 | 5.73 | 4.71E-01 | - |
| *AT5G65050* | MAF2 |  | -0.27 | 5.31 | 2.58E-03 | - |
| *AT5G65060* | MAF3 |  | -0.64 | 3.89 | 9.91E-09 | - |
| *AT5G65070* | MAF4 |  | -0.20 | 0.46 | 4.02E-01 | - |

**Table S4 Flowering time gene expression**

Change in gene expression of curated genes involved in flowering time in Arabidopsis, detected using Illumina RNAseq of *vir-1* compared to the VIR complemented line. 12.2% of flowering time genes change mRNA level expression in the *vir-1* mutant, including FLC and FT.

|  |  |  |  |
| --- | --- | --- | --- |
| **Primer name** | **Primer sequence** | **Experiment** | **Gene or gene coordinates** |
| At1g29550 F | ATGGTGGTTACGGATTCTCCA | Chimeric RNA detection RT-PCR | *AT1G29550-AT1G29560* |
| At1g29560 R | TTAAACTCCATCTCCCTCTTT |
| F\_spl2\_A | GCTGTGTTCCTTTTCTCAGCA | Splicing events validation RT-PCR | *AT2G27385* |
| R\_spl2\_A | CTGCCTTGGAGTTCTGCTTC |
| F\_spl5\_A | AGGGATAAGCTCTTCTTTTTCC | Splicing events validation RT-PCR | 3:12559692-12559814 |
| R\_spl5\_A | CGACGGTGACAACGAACC |
| F\_spl7\_AB | GTGGTGTTGAGCCGGAGT | Splicing events validation RT-PCR | *AT3G42806* |
| R\_spl7\_B | CTTTCTCGGTGGCTTTTCTC |
| F\_spl8\_B | ATATGGACGATACTTTATACCCTTTG | Splicing events validation RT-PCR | *AT3G62040* |
| R\_spl8\_B | AGCCATGGTGGTTCCATATT |
| F\_spl9\_AB | AATATTTTCCGGTGAATAGCC | Splicing events validation RT-PCR | *AT3G26395* |
| R\_spl9\_B | AAGCCCAGAATCCGGTGA |
| CCA1\_For | GATGATGTTGAGGCGGATG | RT-qPCR (clock phenotype) | *AT2G46830* |
| CCA1\_Rev | TGGTGTTAACTGAGCTGTGAAG |
| UBC\_For | CTGCGACTCAGGGAATCTTCTAA | RT-qPCR (clock phenotype) | *AT5G25760* |
| UBC\_Rev | TTGTGCCATTGAATTGAACCC |
| F\_GFP\_CDS\_T7prom\_NotI | GCGGCCGC TAATACGACTCACTATAGGGAGA ATGAGTAAAGGAGAAGAACTTTTCACTG | PCR | *mGFP* |
| R\_GFP\_CDS\_AscI | GGCGCGCC TTATTTGTATAGTTCATCCATGCCATG |
| FLC-Fwd | GAGCCAAGAAGACCGAACTC | RT-qPCR (flowering phenotype) | *AT5G10140* |
| FLC-Rev | TTCTGCTCCCACATGATGA |
| RT41clip | NNGTATNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC | miCLIP library preparation |  |
| RT48clip | NNTGTGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC |
| RT49clip | NNTTCTNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC |
| RT50clip | NNTTTCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC |

**Table S5 Primers used in this study**