Supplemental Materials and Methods

Worm culture

C. elegans worms were cultured on Nematode Growth Medium (NGM) agar seeded with Escherichia coli OP50 according to the standard methods described previously (Brenner 1974).

Plasmid construction

Cloning was performed by PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies) using specific primers (Supplemental Table S4). PCR-amplified or synthesized (Integrated DNA Technologies) DNA fragments were inserted to vectors by Gateway Technology (Life Technologies) or Gibson Assembly (Gibson et al. 2009).

Single-copy transgene insertion

Transgenes were inserted to the *ttTi4348*, *ttTi5605* or *cxTi10882* locus on chromosome I, II or IV, respectively by Mos1-mediated Single-Copy transgene Insertion (MosSCI) (Frokjaer-Jensen et al. 2012).

RNAi clones

RNAi clones were obtained from the Ahringer or the Vidal library (Kamath and Ahringer 2003; Rual et al. 2004).

RNA extraction

Worms were harvested, washed three times with M9 medium, resuspended in 1 ml of TRI Reagent (Molecular Research Center, Cincinnati, OH, USA) and frozen in liquid nitrogen. Worms were broken open by five repeats of freeze and thaw using liquid nitrogen and a 42°C heating block, before RNA was extracted and purified according to the supplier's protocol. The purified RNA was treated with DNA-free Kit (Thermo Fischer Scientific, Waltham, MA, USA) to remove DNA.

RT-qPCR

cDNA was generated from total RNA by ImProm-II Reverse Transcription System (Promega) using the Oligo$(dT)_{15}$ Primer (for mRNA) or Random Primers (for intergenic transcripts) according to the supplier's protocol. qPCR was performed with specific primers (Supplemental Table S4), the SYBR Green PCR Master Mix (Applied Biosystems), and the StepOnePlus Real-time PCR System. Values are shown in Supplemental Table S5.

ChIP

The ChIP protocol was adapted from (Zhong et al. 2010) and modified. For Pol II ChIP, about 30,000 wild-type worms were treated with mock or *xrn-2* RNAi from L1 to L4 stage at 20°C. This took 39 hours for the former and 48 hours for the more slowly developing latter animals. For XRN2 ChIP, about 30,000 worms expressing XRN2-GFP fusion protein (the HW1023 strain in Supplemental Table S2) were cultured from L1 to L4 stage for 39 hours at 20°C. Vulval morphology was observed to confirm mid-L4 stage. Worms were harvested, washed three times with 12 ml of M9 buffer and incubated in 15 ml of M9 buffer with 2% formaldehyde for 30 minutes at room temperature with gentle agitation for protein-DNA cross-linking. The reaction was stopped by incubation with 125 mM glycine for 5 minutes. The worm pellets were washed twice with 12 ml of M9 buffer and frozen in liquid nitrogen. The pellets were resuspended in 800 µl of FA buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1% Triton,

0.1% Na-deoxycholate, 150 mM NaCl) with 0.1% SDS and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche) and disrupted with beads using the FastPrep-24 Instrument (MP Biomedicals) at 4°C. The lysates were transferred to new tubes and sonicated with the Bioruptor Plus sonication device (Diagenode) (setting: 30 sec-on, 30 sec-off for 31 cycles) at 4°C. After centrifugation at 12,000 x g at 4°C for 10 min, the supernatants were transferred to new tubes. The lysates were measured by NanoDrop 1000 (Thermo Fisher Scientific), and 100 µg (at 280 nm) was diluted in 800 µl of FA buffer with 0.1% SDS and 1x protease inhibitor cocktail. The lysate was incubated with 5 µg of mouse anti-Pol II CTD antibody (8WG16, Abcam) (for Pol II-ChIP) or 12 µg of rabbit anti-GFP antibody (A-11122, Thermo Fisher Scientific) at 4°C for 1 hour with gentle agitation. It was then incubated with 45µl of Dynabeads Protein G (Thermo Fisher Scientific) at 4°C for 1 hour with gentle agitation. The beads were washed three times with 1 ml of FA buffer, once with 1 ml of FA-1M buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1% Triton, 0.1% Na-deoxycholate, 1 M NaCl), once with 1 ml of FA-500 mM buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1% Triton, 0.1% Na-deoxycholate, 500 mM NaCl), once with 1 ml of TEL buffer (10 mM Tris-HCl, pH8.0, 250 mM LiCl, 1% NP-40, 1% Na-deoxycholate, 1 mM EDTA) and three times with 1 ml of TE buffer (10 mM Tris-HCl, pH8.0, 1 mM EDTA). The beads were resuspended in 150 µl of Elution buffer (10 mM Tris-HCl, pH8.0, 1 mM EDTA, 1% SDS, 250 mM NaCl) and incubated at 65°C for 15 min with agitation. The elute was transferred to a new tube and incubated with 1 µl /mg RNase A at 37°C for 1 hour. Finally, the elute was incubated with 1 µl /mg proteinase K at 55°C for 1 hour followed by overnight incubation at 65°C. DNA was purified using the MinElute PCR Purification Kit (Qiagen) in 30 µl of DNase-free water. As an input, DNA was purified from 10 µg of the lysate.

Sequencing data processing

Alignments were performed using the qAlign function from the QuasR package in R (version 1.14) (Gaidatzis et al. 2015) with the reference genome package downloaded from Bioconductor (https://www.bioconductor.org/). ChIP-seq reads were aligned using the default parameters, while RNA-seq reads were aligned using the parameter "splicedAlignment=TRUE", resulting in splice-aware alignments. The resulting alignments were converted to BAM format, sorted and indexed using Samtools (version 1.2) (Li et al. 2009).

Read-through analysis

In order to identify downstream regions, gene bodies were defined as the distance between the start of the 5'-most annotated exon and the end of the 3'-most annotated exon, then extended by 20 bp in both directions. Downstream regions were then defined as the distance between the 3' end of each gene and the 5' end of the next gene on the same strand, or up to 10kb downstream of the 3' end if no downstream gene occurred within that distance. Reads overlapping with all resulting downstream regions were quantified in a strand-specific manner. Downstream regions with $\geq$ 16 counts in at least 3 samples were then used as input for differential expression analysis using edgeR (Robinson et al. 2010). A thresholded test was performed using the glmTreat function to detect downstream regions with increased expression under *xrn-2* RNAi of at least 2-fold ($\log_2$ fold-change cutoff of 1). Genes whose downstream region passed this threshold with an adjusted p-value < 0.05 were considered to show read-through. To determine a high-confidence set of genes that are dependent on XRN2 for termination, a similar analysis was performed for upstream regions, defined as the 2kb upstream of the 5' end of each gene. Reads overlapping with upstream regions of genes showing downstream read-through were quantified in a strand-specific manner. Only

upstream regions with ≥ 8 counts in at least 3 samples were used as input for differential expression analysis using edgeR. A likelihood ratio test was performed to detect upstream regions with significant changes in expression under *xrn-2* RNAi, using a cutoff of an adjusted p-value < 0.05. Genes whose upstream regions showed a significant increase in expression were excluded from the list of XDT genes. This list was further filtered to remove any genes in operons. See Supplemental Table S1. For the *xrn-2(xe31)* and *xrn-2cd; xrn-2(xe31)* experiments without replicates, a simple threshold of $\log_2$ fold-change ≥ 2 in downstream regions and $\log_2$ fold-change < 1 in upstream regions was applied to determine XDT genes.

ChIP-seq data analysis

A control set of 1,687 genes was randomly chosen from the list of genes not in operons that showed no downstream read-though, such that the expression levels of the control genes as measured by RNA-seq had the same distribution as the expression levels of the XDT genes. Metaplots of Pol II ChIP and XRN2 ChIP signal around both XDT and non-XDT genes were generated using deepTools (Ramirez et al. 2016). Gene bodies were scaled to the same length between the 5'-most annotated transcript start site and the 3'-most annotated transcript end site. ChIP signal was quantified in 50-bp bins, and the mean of two biological replicates was calculated using the bigwigCompare function. To directly compare Pol II vs XRN2 ChIP signal on a per-gene level, reads were quantified in a 1-kb window around both the TSS and TES, normalized to library size, and $\log_2$-transformed after a pseudocount of 8 was added. A linear regression was performed on $\log_2$ XRN2 signal against $\log_2$ Pol II signal for the TSS and TES separately.

For analysis of Ser2p Pol II signal, data was downloaded as fastq files from SRA (accession number SRP072749; Garrido-Lecca et al. 2016). Read were aligned against the ce10 genome as for all other ChIP-seq data. Metaplots were created using deepTools as described above. Since the two replicates of Ser2p Pol II ChIP-seq showed variable degrees of enrichment, they were analyzed separately rather than being averaged.

Sequence motif analysis

De novo motif enrichment analysis in promoters was performed using the HOMER findMotifsGenome.pl command (v.4.8.2) (Heinz et al. 2010). The set of XDT promoters, comprising 2 kb upstream of the TSS, was compared against the set of non-XDT promoters. The OProf tool from the Signal Search Analysis Server (http://ccg.vital-it.ch/ssa/oprof.php) was used to generate motif occurrence profiles for the TATA-box and Inr motifs, using the Promoter Motifs set from the built-in motif library and the ce10 genome (Ambrosini et al. 2003). Default options were used for all other parameters. Nucleotide frequencies around the 5' ends of genes were calculated using alphabetByCycle from the ShortRead R package (Morgan et al. 2009).

EMS mutagenesis screen for *xrn-2* temperature-sensitive alleles

About 6,000 L4-stage larvae of the HW1604 strain with *dpy-13* transcription read-through reporters (Supplemental Table S2) were harvested, washed and incubated with 50 mM EMS in 6 ml of M9 buffer for 4 hours at room temperature. The worms were washed three times with M9 buffer and cultured at 20°C. The L3- or L4-stage larvae of the F2 generation were screened for GFP-positive at 25°C. Four GFP-positive mutant lines were isolated and backcrossed five times with the parental HW1604 strain to remove unrelated mutations. Mutations were identified by *xrn-2* cDNA sequencing (for *xe31[R182W]* and *xe34[E699K]* alleles) or whole-genome sequencing (for

xe32[R304W] and *xe33[P347L]* alleles). Genomic DNA was extracted and purified using Gentra Puregene Tissue Kit (Qiagen, Venlo, Netherlands). DNA libraries were created from 50 ng of genomic DNA using Nextera DNA Library Prep Kit (Illumina, San Diego, CA, USA). The sequencing data were generated using a HiSeq 2500 sequencer (Illumina). Mutations were mapped by single nucleotide variant (SNV) analysis using MiModD (Version 0.1.2) according to the guideline (https://sourceforge.net/projects/mimodd/). For RNA-seq, *xrn-2(xe31)*, *xrn-2cd; xrn-2(xe31)* and control wild-type worms were cultured from L1 to early L4 stage for 59 hours at 15°C before the temperature was shifted to 26°C, the restrictive temperature. The worms were further cultured and harvested every hour for 10 hours until late L4 stage.

References

Brenner S. (1974) The genetics of Caenorhabditis elegans. *Genetics*. **77**: 71-94

Frokjaer-Jensen C, Davis MW, Ailion M, Jorgensen EM. (2012) Improved Mos1-mediated transgenesis in C. elegans. *Nat Methods*. **9**: 117-118

Gaidatzis D, Lerch A, Hahne F, Stadler MB. (2015) QuasR: quantification and annotation of short reads in R. *Bioinformatics*. **31**: 1130-1132

Garrido-Lecca A, Saldi T, Blumenthal T. (2016) Localization of RNAPII and 3' end formation factor CstF subunits on C. elegans genes and operons. *Transcription.* **7** (3): 96–110.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. **6**: 343-345

Kamath RS, Ahringer J. (2003) Genome-wide RNAi screening in Caenorhabditis elegans. *Methods*. **30**: 313-321

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**: 2078-2079

Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics.* **25**: 2607-2608

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. **44**: W160-165

Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**: 139-140

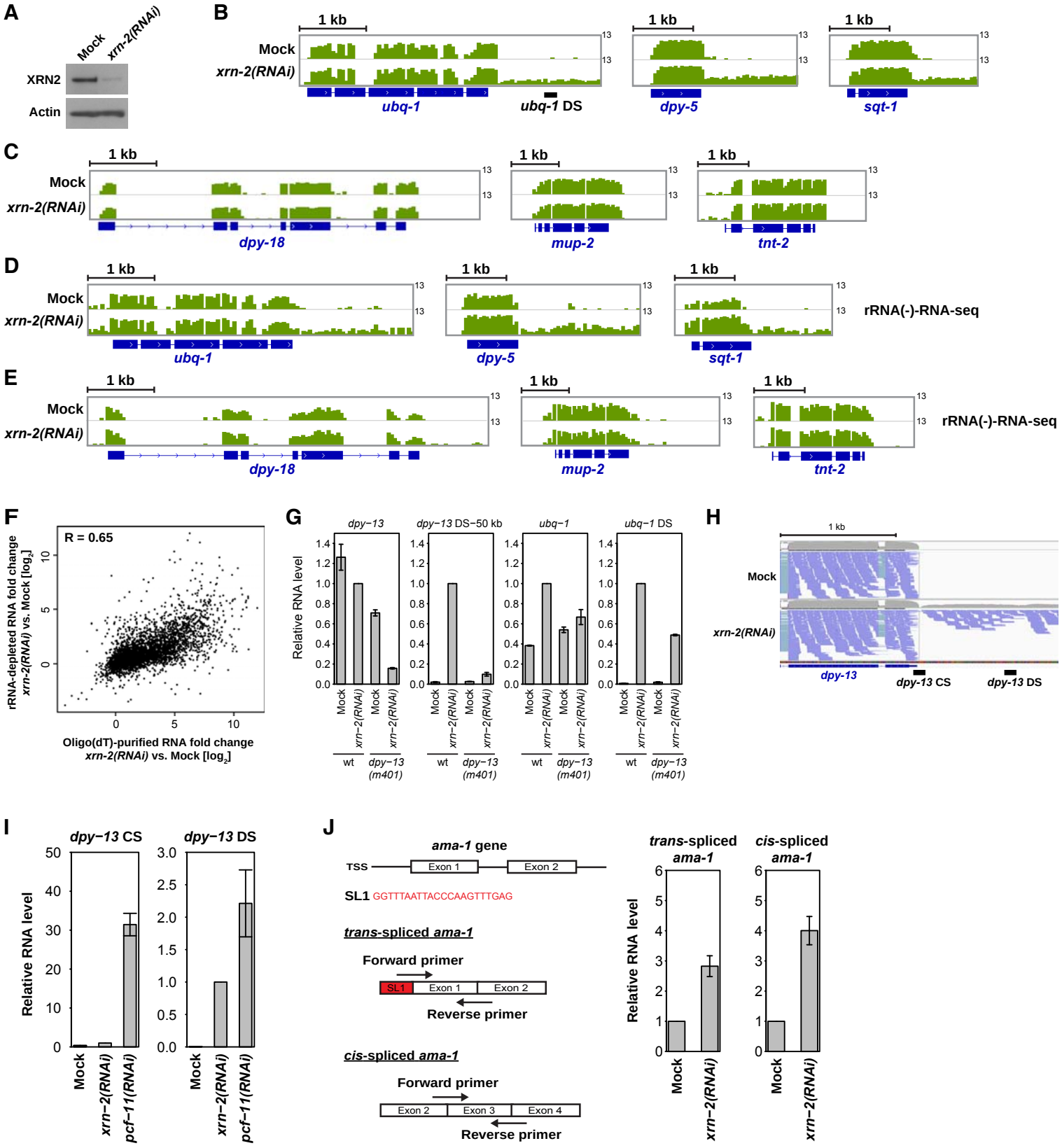
Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den

Heuvel S et al. (2004) Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res*. **14**: 2162-2168

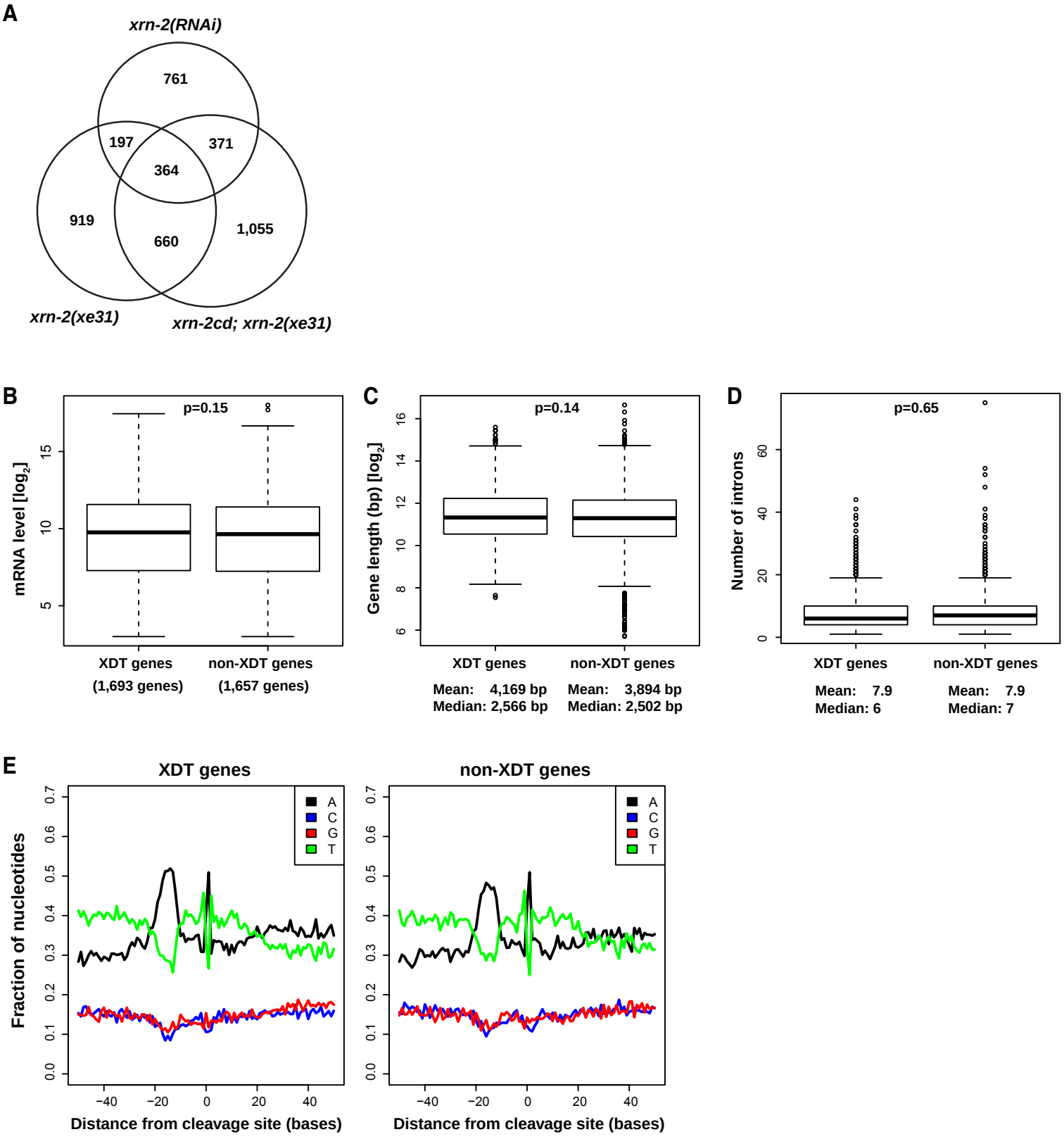
Shi Y, Manley JL. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* **29**: 889-897

von Mende N, Bird DM, Albert PS, Riddle DL. (1988) dpy-13: a nematode collagen gene that affects body shape. *Cell* **55**: 567-576

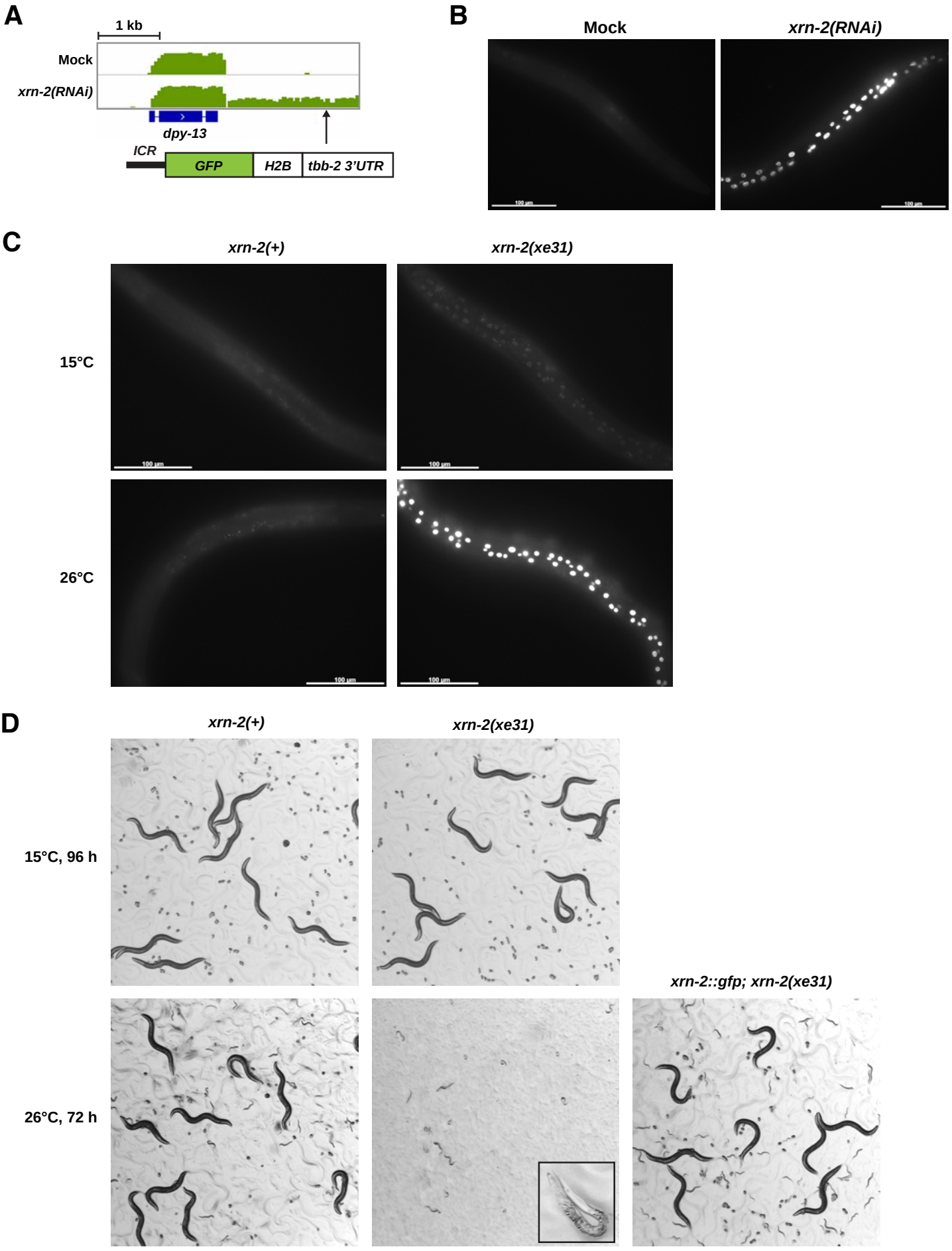
Zhong M, Niu W, Lu ZJ, Sarov M, Murray JI, Janette J, Raha D, Sheaffer KL, Lam HY, Preston E, Slightham C, Hillier LW, Brock T, Agarwal A, Auerbach R, Hyman AA, Gerstein M, Mango SE, Kim SK, Waterston RH, Reinke V, Snyder M. (2010) Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. *PLoS Genet.* **6**: e1000848



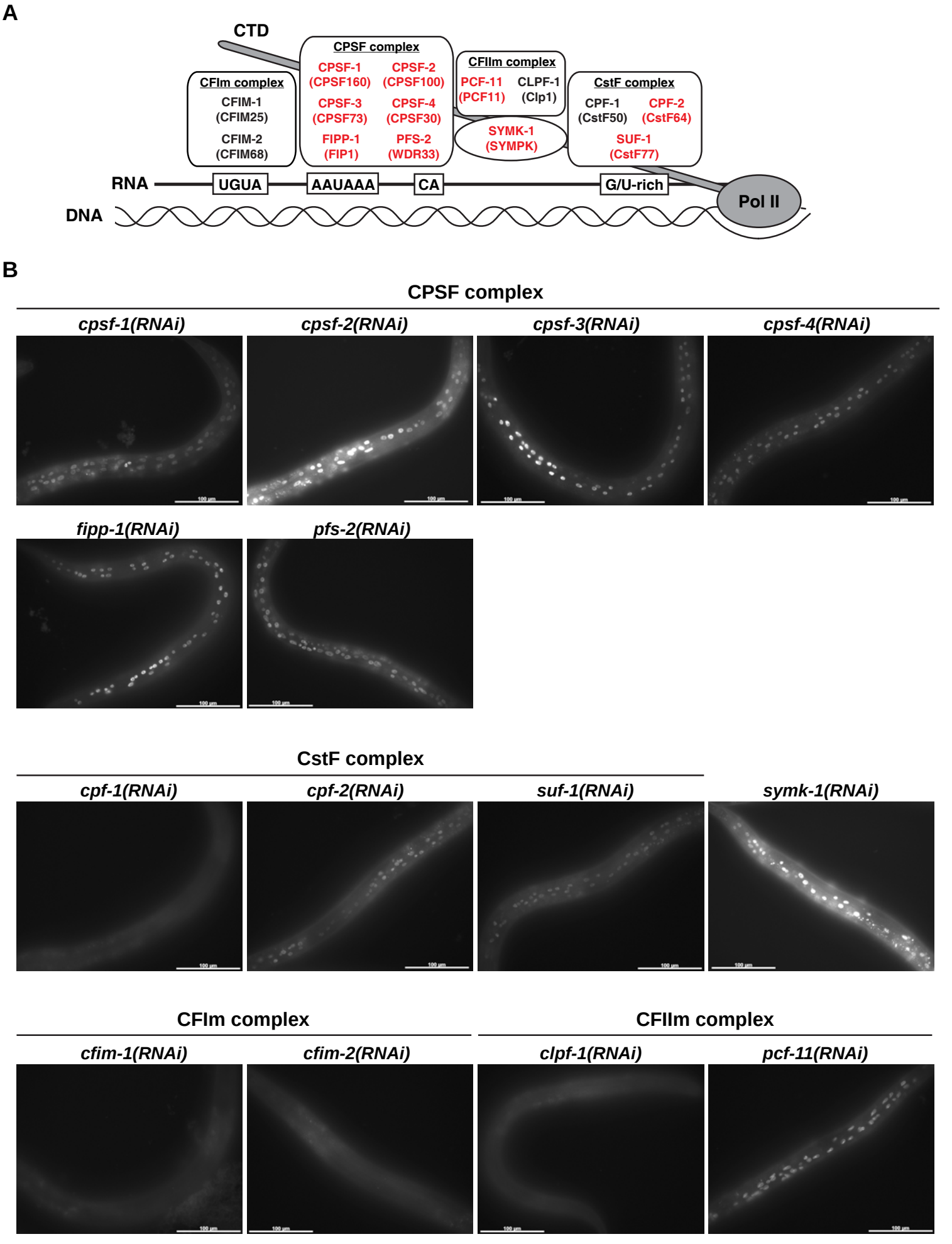
Supplemental Fig. S1. XRN2 inactivation yields read-through transcripts on a specific subset of genes. (A) XRN2 expression was examined by Western blot with Actin as a loading control. (B-E) Snapshots of (B, D) XRN2-dependent and (C, E) XRN2-independent genes from (B, C) poly(A)-RNA-seq (using oligo(dT) purified RNA) or (D, E) rRNA(-)-RNA-seq data. RNA levels on the same strand as the gene of interest are shown in green in a $\log_2$ scale (0-13). RNA-seq libraries of the same type are normalized for total library size. (F) Scatter plot showing $\log_2$ fold-changes of read counts, as determined by edgeR analysis, in downstream regions in *xrn-2* RNAi versus mock for rRNA(-)-RNA-seq versus poly(A)-RNA-seq. (G-J) Wild-type or *dpy-13(m401)* animals were treated with mock, *xrn-2* RNAi or *pcf-11* RNAi from L1 to L4 stage at 20°C. Levels of indicated RNA were quantified by RT-qPCR and normalized to *act-1* (G, J) or *dpy-13* (I) mRNA levels with values in *xrn-2(RNAi)*-treated wild-type animals defined as 1 (n = 3, means ± SEM). Values are shown in Supplemental Table S5. (G) XRN2 knockdown caused >50 kb of read-through from dpy-13.The *dpy-13(m401)* mutation, a transposon insertion in the *dpy-13* promoter (von Mende et al. 1988), was used to impair *dpy-13* transcription. Reduced levels of intergenic transcripts in a region ~50 kb downstream of dpy-13 (dpy-13 DS-50 kb: Fig. 1C) resulted in the presence of this *dpy-13(m401)* mutant relative to the *dpy-13(+)* wild-type (wt) allele when XRN2 was knocked down. Read-through of another XRN2-dependent gene *ubq-1* was examined to confirm knockdown of XRN2 in *dpy-13(m401)* animals. (H) A snapshot of the 3' end of the *dpy-13* gene from the mock vs. *xrn-2(RNAi)* poly(A)-RNA-seq data. Sequenced reads are shown in light blue, coverage in light gray and splice junctions in green. CS: cleavage site; DS: downstream site. (I) *dpy-13* CS: RNA spanning the 3' end cleavage site; *dpy-13* DS: *dpy-13* downstream site. (J) *trans*-spliced and *cis*-spliced (5'-capped) mRNA levels were examined. The design of primers is shown.



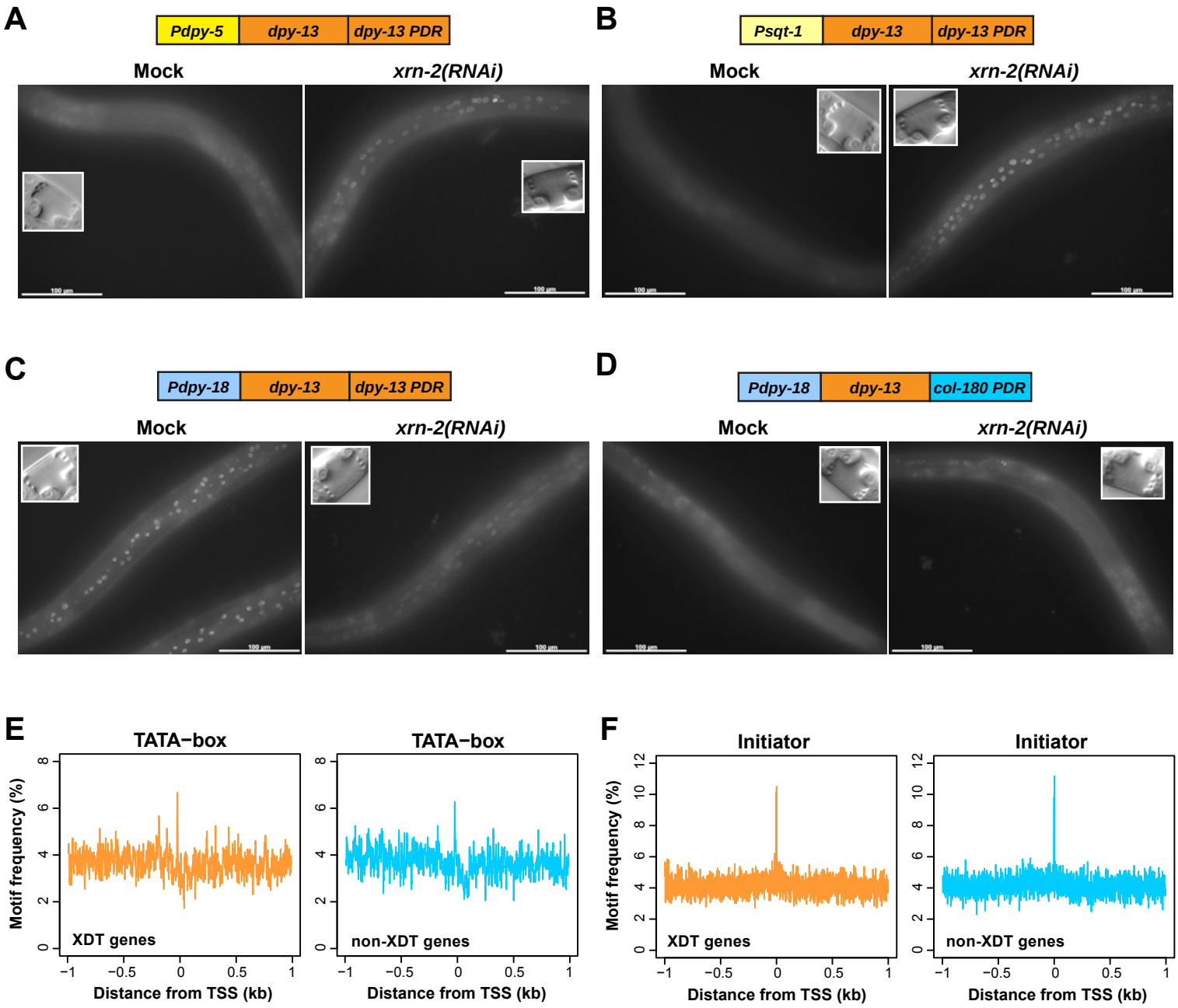
Supplemental Fig. S2. (A) Venn diagram showing overlaps between XDT genes defined in the three experiments. (B-D) Boxplots showing (B) the mRNA level in mock conditions, (C) the gene length, and (D) the number of introns of XDT and non-XDT genes (Supplemental Table S1). (E) Nucleotide distribution from -50 bp to +50 bp around the cleavage sites of XDT and non-XDT genes.



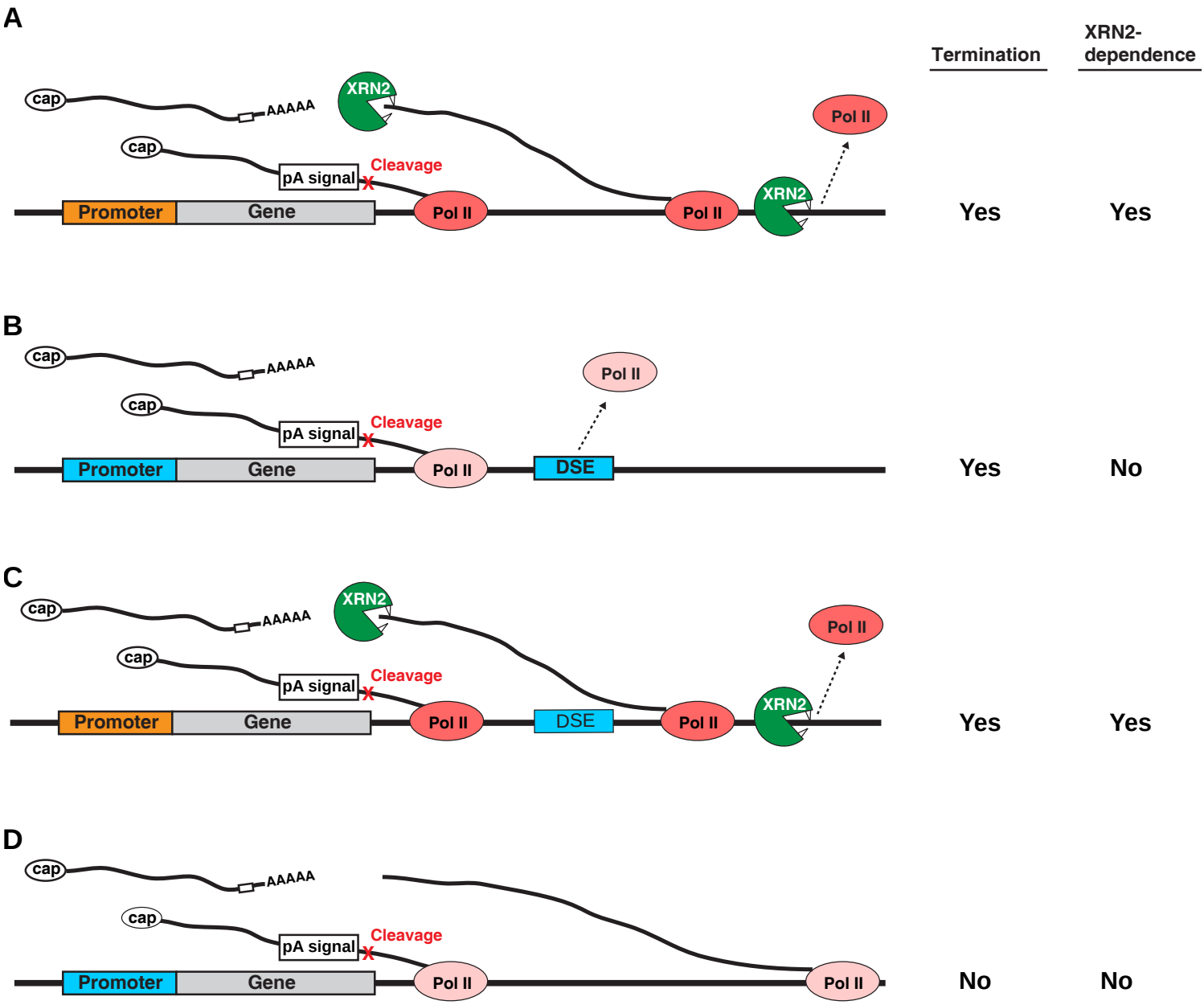
Supplemental Fig. S3. *dpy-13* transcription read-through reporter. (A) The indicated construct was integrated into an intergenic region downstream of endogenous *dpy-13*. (B) The transgenic animals were treated with mock or *xrn-2* RNAi from L1 to L4 stage at 20°C and examined for GFP signal. Scale bar: 100 µm. (C) *xrn-2(+)* wild-type and *xrn-2(xe31)* animals with *dpy-13* transcription read-through reporters were cultured from L1 to L4 stage at 15°C or 26°C and examined for GFP signal. Scale bar: 100 µm. (D) *xrn-2(+)* wild-type, *xrn-2(xe31)* and *xrn-2:gfp; xrn-2(xe31)* animals were cultured from L1 at 15°C or 26°C for indicated durations and observed. The images were acquired by the M205 FA stereo microscope (Leica).



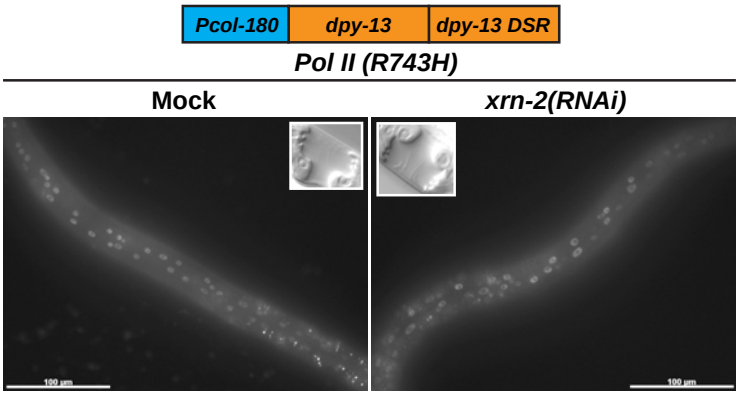
Supplemental Fig. S4. CPFs are required for transcription termination of *dpy-13*. (A) CPFs tested for *dpy-13* transcription termination. Their human homologs are shown in brackets. See (Shi et al. 2015) for their functions. Those whose knockdown caused read-through are shown in red. (B) *dpy-13* transcription read-through reporter animals (Supplemental Fig. S3A, B) were treated with RNAi for indicated factors from L1 to L4 stage at 20°C and examined for GFP signal. Scale bar: 100 μm.



Supplemental Fig. S5. (A-D) Transgenic animals expressing the indicated reporter constructs were treated with mock or *xrn-2* RNAi from L1 to L4 stage at 20°C and examined for GFP signal. Insets show DIC images of mid-L4 stage vulvae. Scale bar: 100 µm. (E, F) Profiles of (E) TATA-box and (F) Initiator motif occurrences from -1kb to +1kb around the TSS of XDT and non-XDT genes.



Supplemental Fig. S6. Schematic summary of the results shown in Fig. 4 and the inferred mechanisms of XDT (XRN2-dependent) and non-XDT (DSE-dependent) transcription termination mechanisms. (A) XRN2-dependent and (B) DSE-dependent transcription termination. (C) Transcription initiated at an XDT gene promoter (orange) is not terminated by a non-XDT gene DSE but relies on XRN2. (D) Transcription initiated at a non-XDT gene promoter (blue) is not terminated by XRN2. Different Pol II colors illustrate presumed differences in TEC properties and/or composition, determined by promoters, which result in differences in termination modes. See main text for details.



Supplemental Fig. S7. The indicated transgenic animals were treated with mock or *xrn-2* RNAi from L1 to L4 stage at 20°C and examined for GFP signal. Insets show DIC images of mid-L4 stage vulvae. Scale bar: 100 μm.