**Molecular Cell, Volume 50**
**Supplemental Information**

# ZFP-1(AF10)/DOT-1 Complex Opposes H2B Ubiquitination to Reduce Pol II Transcription

**Germano Cecere, Sebastian Hoersch, Morten B. Jensen, Shiv Dixit, and Alla Grishok**

## Inventory of Supplemental Information

**Figure S1, Figure S2, Figure S3, and Figure S4, Related to Figure 1.**

**Figure S5, Related to Figure 2**

**Figure S6, Related to Figure 4**

**Figure S7, Related to Figure 5**

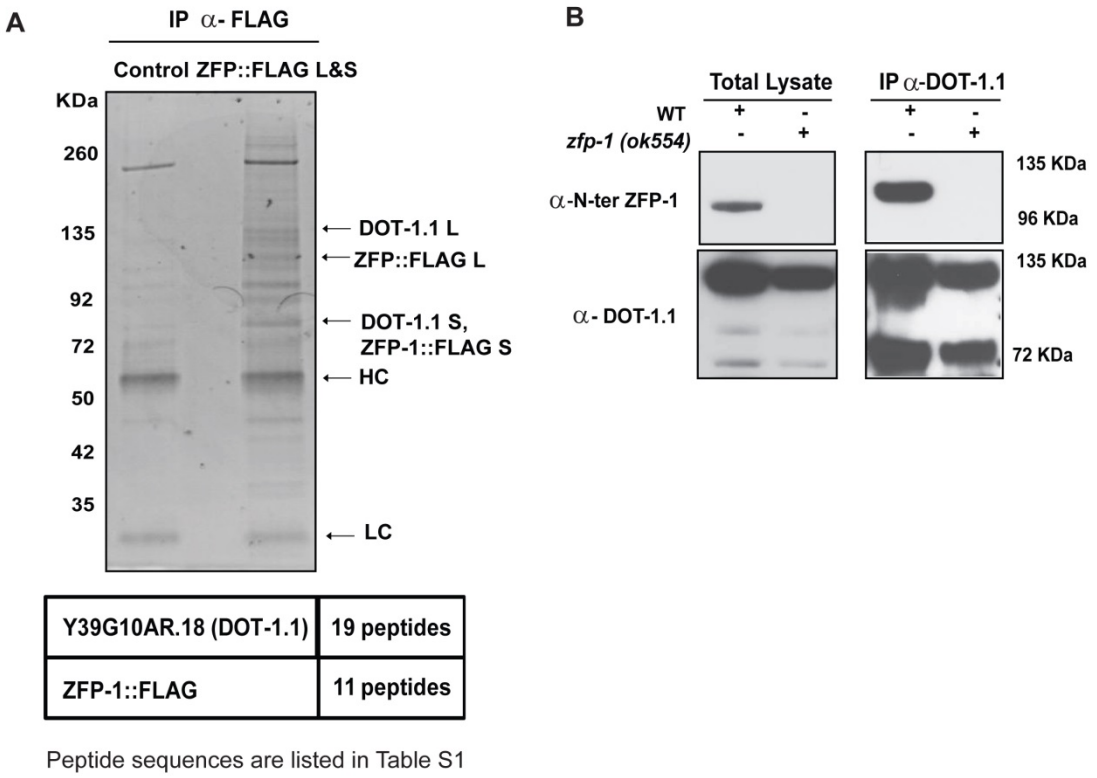
**Table S1. Mass Spectrometry Analysis Results of ZFP-1::FLAG Immunoprecipitation, Related to Figures 1 and S1**

**Table S2. Gene Lists Used for Statistical Analyses, Related to Figures 1–5 and S4–S6**

**Table S3. Functional Enrichment Analysis of Data Sets, Related to Figure 3**

**Supplemental Experimental Procedures**

**Supplemental References**




| | |
|---|---|
| Y39G10AR.18 (DOT-1.1) | 19 peptides |
| ZFP-1::FLAG | 11 peptides |

Peptide sequences are listed in Table S1

**Figure S1. ZFP-1 Interacts with Y39G10AR.18 (DOT-1.1), Related to Figure 1**

(A, top) Immunoprecipitation of ZFP-1 isoforms tagged with FLAG at the C-terminus with the anti-FLAG antibody followed by mass spectrometry of co-immunoprecipitating proteins. A gel stained with SYPRO is shown: left lane - control immunoprecipitation using mixed embryo extracts of non-transgenic strain, right lane – immunoprecipitation using mixed embryo extracts of integrated ZFP-1::FLAG strain (AGK 128, armIs5 Is[ZFP-1::FLAG,unc-119+] (Mansisidor et al., 2011)). Bands corresponding to the two DOT-1.1 (Y39G10AR.18) isoforms (L –long and S- short) as well as ZFP-1::FLAG L (long) and ZFP-1::FLAG S (short) isofroms are shown; HC and LC indicate heavy and light antibody chains. (A, bottom) Summary of mass spectrometry results; a complete list of peptide sequences is available in Table S1.

(B) The ZFP-1 protein co-immunoprecipitates with DOT-1.1. Top panels: western blot

using an antibody specific for the N-terminus of ZFP-1 and recognizing only the long ZFP-1 isoform, left – total protein lysates, right – proteins immunoprecipitated with anti-DOT-1.1 antibody in wild type and *zfp-1(ok554)* mutant. Bottom panels: western with the anti-DOT-1.1 antibody.

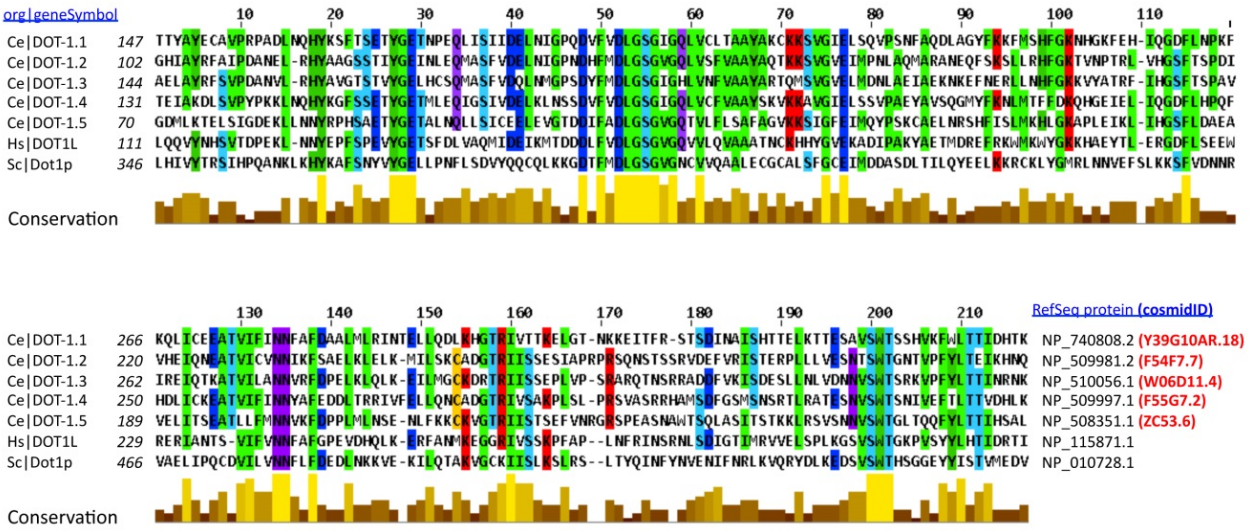



**Figure S2. Clustal-Based Alignment with Conservation Histogram of the Catalytic Domain of Dot1-like Proteins from *C. elegans*, *H. sapiens*, and *S. cerevisisae*, Related to Figure 1**

Official gene symbols are given at the beginning of the alignment, while underlying RefSeq protein accessions and corresponding *C. elegans* cosmid IDs are given at the end. Position numbers (in italics) at the beginning of both alignments blocks refer to positions within the respective RefSeq sequences. The catalytic domain of DOT1L was defined by (Min et al., 2003). *C. elegans* (Ce, 5 proteins, DOT-1.1 – DOT-1.5), *H. sapiens* (Hs, 1 protein, DOT1L), and *S. cerevisisae* (Sc, 1 protein, Dot1p).


Nematoda
Platyhelminthes
Hemichordata
Cephalochordata
Teleostei
Amphibia
Aves
Mammalia
Insecta
Porifera
consensus/100%
consensus/90%
consensus/80%
consensus/70%
consensus/60%
consensus/50%

**Figure S3. PsiBlast Reveals Homologies to DOT1L Genes for the Poorly Characterized C-Terminal Portion of *C. elegans* Y39G10AR.18 Across Metazoans, Including Mammals, Insects, and Sponges, related to Figure 1**

A 221aa subsequence in the C-terminal half of *C. elegans* Y39G10AR.18 (the longest subsequence aligning without gaps with the *C. brenneri* ortholog of Y39G10AR.18) was submitted for PSI-BLAST (Altschul et al., 1997) analysis against a non-redundant protein database (NCBI), matching human DOT1L after 3 iterations. In this display of post-processed PsiBlast output generated using MView (Brown et al., 1998), local alignments are arranged into a quasi-multiple sequence alignment, with E-values / bit scores reported on the left as after the 5th iteration, which marked convergence. Sequences are manually arranged according to taxonomic groupings (color bars). Above the alignment block, MView-generated consensus codes at different percentage cutoffs are given, with capital letters denoting single amino acids and small letters denoting certain amino acid groups based on (Taylor, 1986). Zoom in viewing of the electronic version of this figure is recommended.

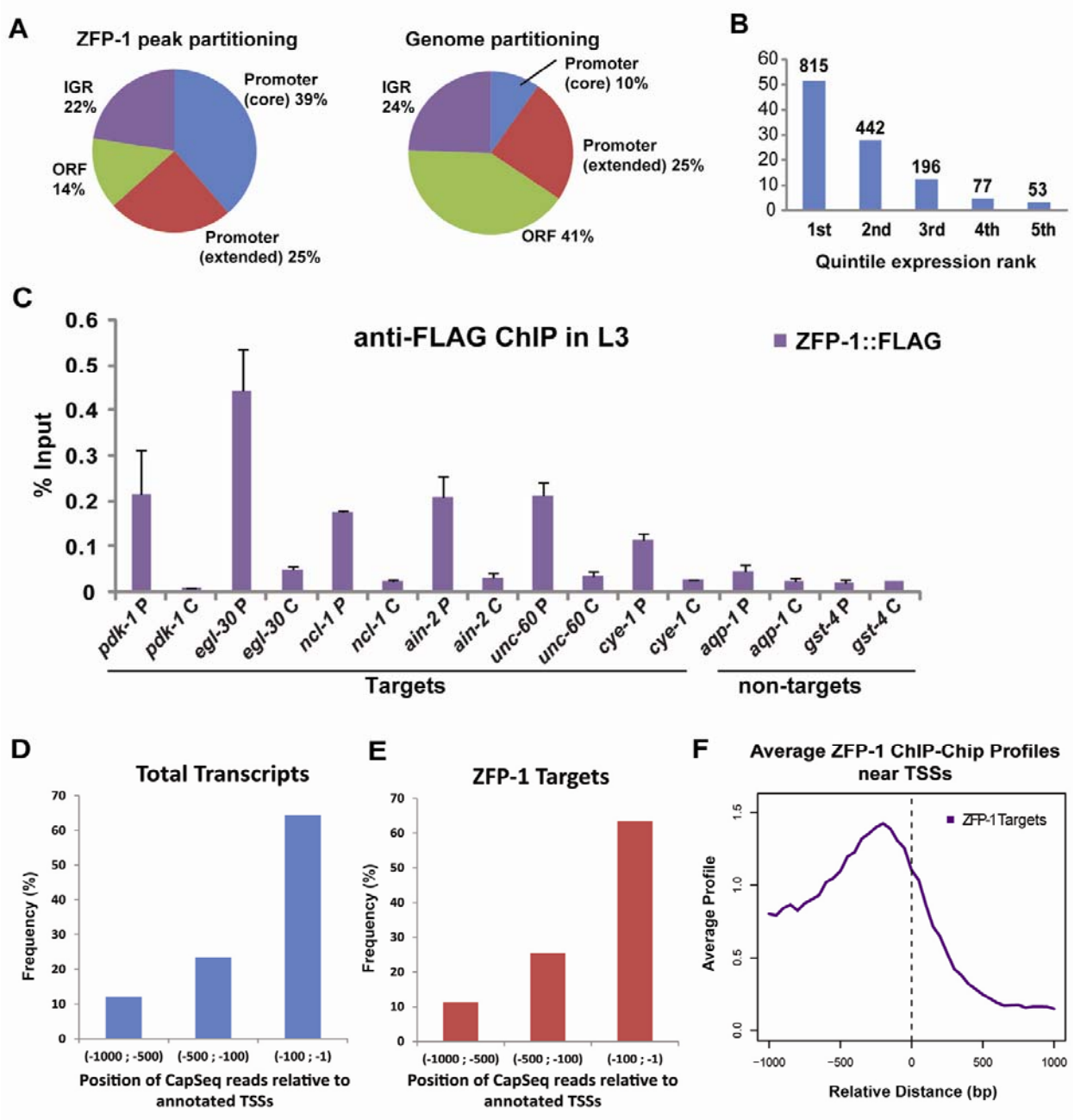



**Figure S4. ZFP-1 Binds to Promoters of Highly Expressed Genes Enriched in H3K79 Methylation, Related to Figures 1 and 2**

(A) ZFP-1 binding sites are enriched at promoters. The distribution of ZFP-1 binding

sites within sectors of the *C. elegans* genome including the 500bp upstream of TSS - Promoter (core), promoter regions spanning 0.5-2kb upstream of TSS – Promoter (extended), open reading frames (ORF) or the remaining of the genome representing intergenic regions (IGR) is shown (left). For comparison, the portion of base pairs within each category genome-wide is shown on the right. Every bp was assigned to only one category within a hierarchy (highest to lowest): Promoter (core), Promoter (extended), ORF, IGR.

(B) Genes bound by ZFP-1 tend to be highly expressed. Genes having a ZFP-1 binding site called within 500 bp upstream of TSS were placed in bins according to their expression rank ($1^{st}$ quintile being the highest and $5^{th}$ the lowest expression). The actual number of ZFP-1-bound genes in each quintile is shown above each bar. Expression data used are from modENCODE.

(C) Validation of ZFP-1 ChIP-chip data using ChIP-qPCR; target and non-target genes were included as indicated. Larval stage L3 animals of the ZFP-1::FLAG strain were used for IP with anti-FLAG antibody; "P" after the gene name indicates promoter, "C" after the gene name indicates coding region, here and in the following figures. Results of three independent experiments are shown; error bars represent standard deviation.

(D) Mapping of the 5' ends of capped RNA reads (from (Gu et al., 2012)) relative to the annotated start sites of Wormbase transcripts. The data were binned into three regions: from -1000nt to -500nt, from -500nt to 100nt, and from -100nt to -1nt relative to the TSSs and presented as % frequency of mapping to these regions.

(E) Same analysis as in (D) considering only ZFP-1 target transcripts.

(F) The average profile of ZFP-1 ChIP-chip data showing the density reads in 10bp

windows from −1000bp to +1000bp relative to the annotated transcriptional start sites (TSSs) of ZFP-1 target genes.

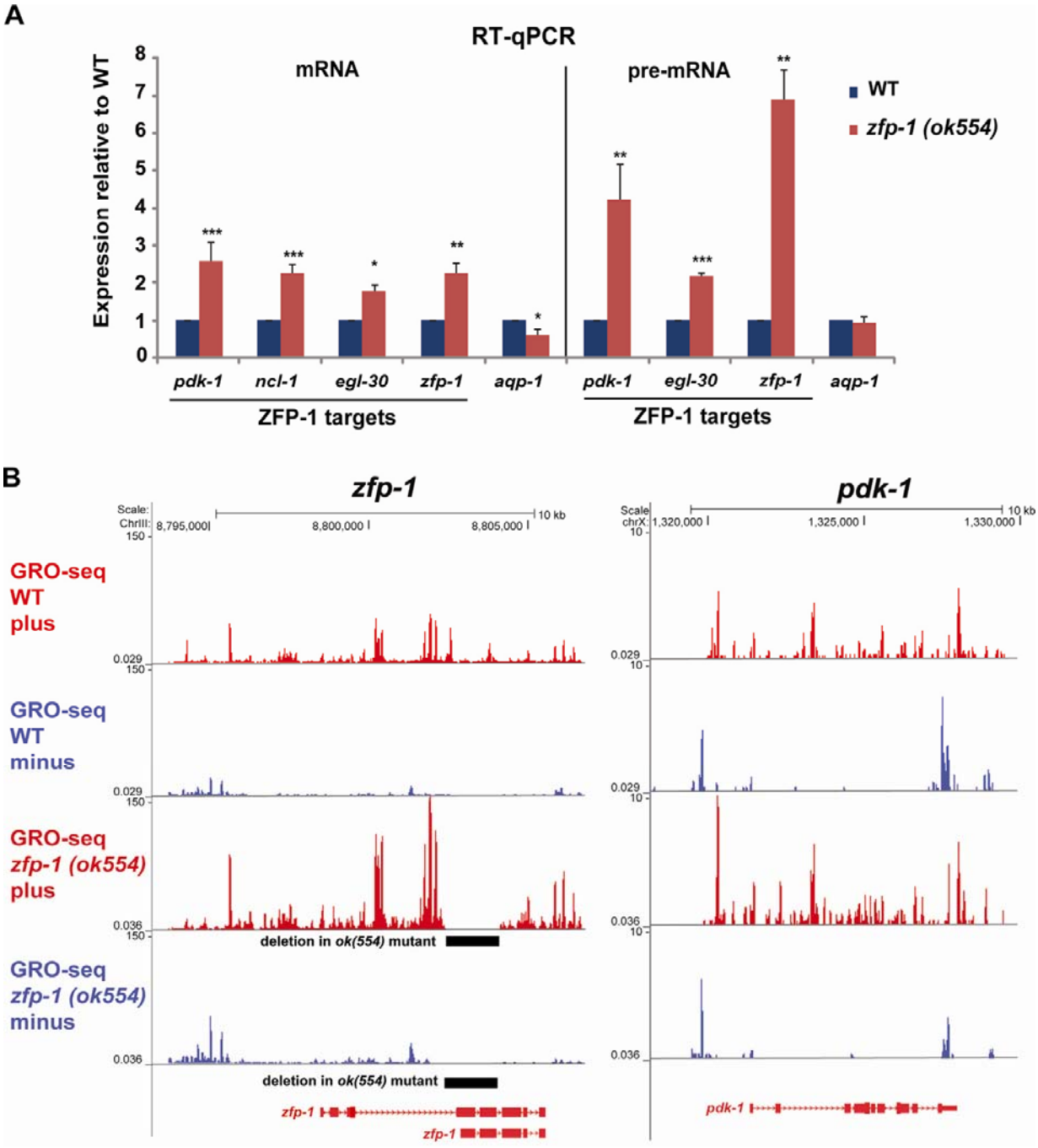



**Figure S5. ZFP-1 Target Genes with Increased Pol II Transcription in *zfp-1(ok554)*, Related to Figure 2**

(A) Expression analysis of ZFP-1 target genes and control non-target gene *aqp-1* in wild type and *zfp-1(ok554)* L3 larvae by RT-qPCR. Relative mRNA expression is shown on the left, pre-mRNA expression is on the right. *act-3* mRNA expression is used as an

internal control. Results of at least three independent experiments are shown, error bars represent standard deviation. * indicates significance of $p < 0.05$, **- $p < 0.01$ and ***- $p < 0.005$ compared to wild type according to Student's t-test.

(B) Examples of ZFP-1 target genes that show an increase in transcriptionally-engaged Pol II in *zfp-1(ok554)*. Track listing from top to bottom: GRO-seq normalized reads aligned to the plus strand (red) in wild type larvae, GRO-seq normalized reads aligned to the minus strand (blue) in wild type larvae, GRO-seq normalized reads aligned to the plus strand (red) in *zfp-1(ok554)* larvae, GRO-seq normalized reads aligned to the minus strand (blue) in *zfp-1(ok554)* larvae. Gene models are based on UCSC Genome Browser (ce06): *zfp-1* (F54F2.2 (left)), *pdk-1* (H42K12.1b (right)). The black bar indicates the deletion present in the *zfp-1(ok554)* mutant, therefore, there are no GRO-seq reads corresponding to this region in the *zfp-1 (ok554)* mutant larvae.

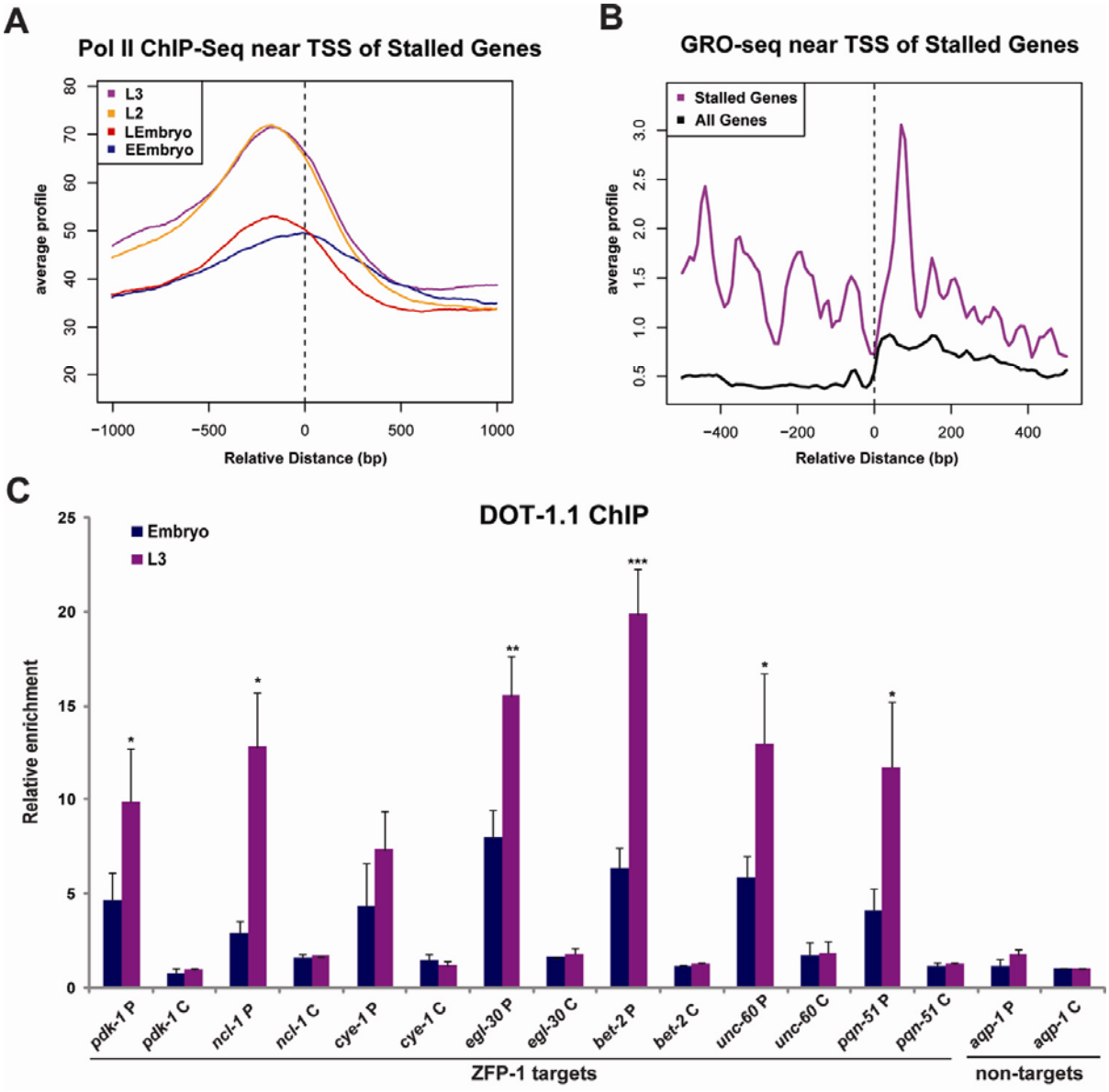



**Figure S6. Progressive Pausing of Pol II during *C. elegans* Development, Related to Figure 4**

(A) The average profile of Pol II ChIP-seq data (from modENCODE) showing the density reads in 10bp windows from −1000bp to +1000bp relative to transcriptional start sites (TSSs) on genes defined stalled (Zhong et al., 2010). (Red) ChIP-seq data from early embryo (EEMB), (blue) ChIP-seq data from late embryo (LEMB), (orange) ChIP-seq data from L2 larvae (L2), (purple) ChIP-seq data from L3 larvae (L3). The average

profile was calculated using Cistrome/Galaxy (Liu et al., 2011).

(B) The average profile of GRO-seq data showing the density reads in 10bp windows from −500bp to +500bp relative to TSSs on stalled genes as defined by (Zhong et al., 2010) (purple) or all genes (black). The profile was calculated using Cistrome (Liu et al., 2011).

(C) An increase in DOT-1.1 occupancy measured by ChIP-qPCR at the promoters of ZFP-1 target genes in L3 larvae compared to the embryo stage; ChIP-qPCR data are presented as an enrichment relative to the coding region of *aqp-1*. Results of three independent experiments are shown, error bars represent standard deviation. * indicates significance of $p < 0.05$ and **- $p < 0.01$.

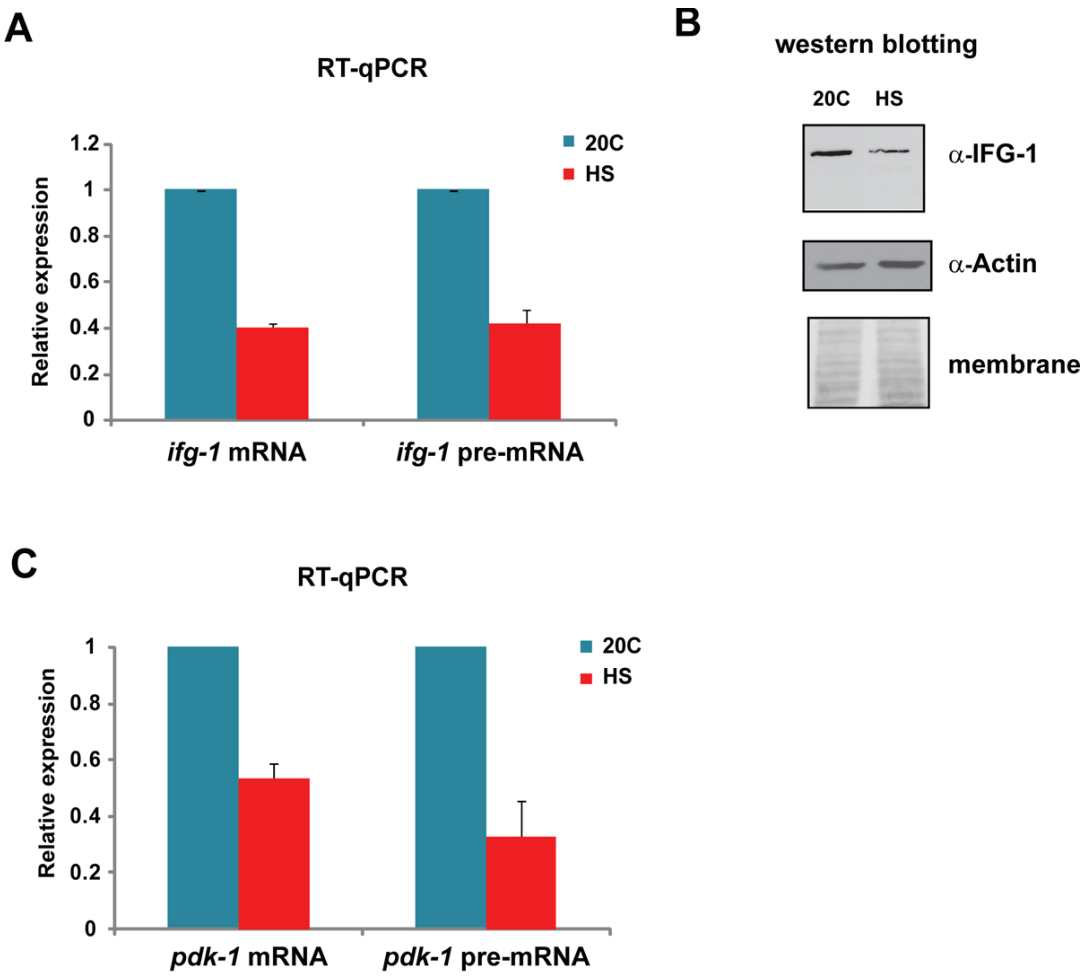



**Figure S7. Expression of *ifg-1* (eIF4G) and *pdk-1* Is Downregulated in Response to Heat Shock, Related to Figure 5**

(A and C) RT-qPCR of *ifg-1* and *pdk-1* gene products using specific primers to detect mRNA or pre-mRNA in normal growth conditions (20°C) and upon heat shock treatment (HS). *act-3* mRNA expression is used as an internal control. Error bars represent standard deviation of the mean from at least two biological replicas.

(B) Western blotting with the anti-IFG-1 antibody shows a reduction in the protein levels of IFG-1 in total protein extract from heat shock-treated worms (HS). Western blot with anti-actin antibody is shown as a loading control.

## Supplemental Experimental Procedures

### *C. elegans* Strains

*zfp-1(ok554)III, armIs5 Is[ZFP-1::FLAG,unc-119(+)]*, *zfp-1(ok554)III; armEx14 Ex[ZFP-1 PHD1-PHD2::FLAG,unc119(+)]*

Procedures for generating transgenic *armIs5,* and *armEx14* lines were described previously (Mansisidor et al., 2011; Avgousti et al., 2013).

### Immunoprecipitation and Mass Spectrometry

*C. elegans* mixed embryos were suspended in the extraction buffer (50mM Tris, 300mM NaCl, 1mM EDTA, 0.1% NP40, protease inhibitor cocktails (Fermentas)) and sonicated with 30 sec pulses 10 times using Branson microtip sonicator at 10% power in the cold room. Protein extracts were span at 12,000rpm at 4°C for 10 minutes. Protein concentration was quantified by the Bradford assay, and 3mg of protein extract was incubated with 5µg of antibody for 1 hour at 4°C; the immune complexes were then incubated with 60 µl IgG Dynabeads (Invitrogen) for 1h at 4°C. After four washes with the extraction buffer the beads were resuspended in 50µl of NuPAGE LDS sample buffer (Invitrogen) and incubated at 95°C for 5 minutes. For FLAG immunoprecipitation the elution of the immune complexes was done using the FLAG peptide (Sigma) for one hour at 4°C. For Mass spectrometry analyses five IP samples were pulled together and resolved on precast NuPAGE Novex 4-12% Bis-Tris gels (Invitrogen). Gels were stained with SilverSNAP (Termo Scientific) and bands were excised and sent to Midwest Bio Services for mass spectrometry analysis. The antibodies used were anti-FLAG M2 (Sigma) and anti-DOT-1.1 (modENCODE, SDQ4129_Y39G10AR18).

**Chromatin Immunoprecipitation**

Procedures were as described in (Mansisidor et al., 2011) with the following modification: ~ 2.5mg of protein chromatin extract was used for each ChIP experiment. We harvested eggs by hypochlorite treatment, and synchronous populations of worms were grown for 36 hours post-hatching at 20ºC on OP-50 *E. coli* at a density of approximately 100,000 animals per 15cm Petri dish. We used 10µg of antibody for each ChIP experiment. The antibodies used are as follows: anti-FLAG M2 Affinity Gel (Sigma), anti-DOT-1.1 (SDQ3964_Y39G10AR.18), anti-H3 (Millipore, 05-928), anti-H3K79me2 (Millipore, 04-835), anti-H3K79me3 (Abcam, ab2621), anti-H2B (Abcam, ab1790), anti-H2Bub1 (Millipore, 05-1312), anti-Pol II 8WG16 (Covance, MMS-126R).

**Identification of genes bound by DOT-1.1**

*C. elegans* genes (RefSeq based on WS170/ce4) were extracted from the UCSC genome browser. A gene was called bound by DOT-1.1 if the center base pair of a ChIP-chip peak overlapped the ORF or the 1,500 bp upstream region. Overlap calls were done using the Galaxy web tool. Of the total 24,901 genes, 2,903 were bound by DOT-1.1 (see Table S2). Genome-wide DOT-1.1 (Y39G10AR.18) localization data are available at modENCODE: http://intermine.modencode.org/.

**Data Analysis**

The overlaps between gene lists were calculated using the online tool at http://www.nemates.org. Representation factors "r" to quantify the enrichment or impoverishment of gene sets were calculated according to r = (n1,2)/[(n1xn2)/N], with n1,2 = number of genes common to set 1 and 2, n1 = number genes in set 1, n2 = number of genes in set 2, N = total number of genes considered. Significance of enrichment and

impoverishment was determined separately by using Fisher's Exact test as implemented in the statistical computing package R (http://www.r-project.org/). The Gene Ontology and functional annotation clustering was done using DAVID bioinformatics resource 6.7 (Huang da et al., 2009).

**Global Run-On Sequencing**

GRO-seq protocol from (Core et al., 2008) was adapted to *C. elegans* as follows.

*Nuclei Extraction for GRO-seq*

Around 300,000 worms were grown on a large plate and harvested at L3 stage (36 hours after hatching at 20°C). Worms were washed off the plate with cold M9 and washed 3-4 times. The worm pellet was resuspended in 2ml of cold Nuclear Run-On (NRO) lysis buffer (10mM Tris pH 7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, 0.5% IGEPAL and 10% glycerol), transferred to a steel dounce (on ice) and stroked 20 times. The worm lysate was centrifuged for 2min at 40g and the supernatant was centrifuged again for 5min at 1000g to pellet the nuclei. Next, the supernatant was discarded and the nuclei pellet was washed two times with 1ml of cold NRO lysis buffer and one time with 1ml of cold Freezing Buffer (50mM Tris-CL pH 8.3, 40% glycerol, 5mM MgCl2, 0.1 mM EDTA). Nuclei were then resuspended in 100µl of Freezing Buffer and used for the NRO reaction.

*NRO Reaction and DNase Digestion*

100µl of nuclei were mixed with an equal volume of reaction buffer (10mM Tris-Cl pH 8.0, 5mM MgCl2, 1mM DTT, 300mM KCL, 20 units of SUPERase Inhibitor, 1% sarkosyl, 500µM ATP, GTP, and Br-UTP, 2µM CTP) and incubated for 5 minutes at 32°C. The reaction was stopped with 750µl of TRIzol reagent (Invitrogen) and the RNA was extracted following the manufacturer's instructions. The RNA pellet was

resuspended in 85µl water and treated with 5µl of turbo DNase (Ambion) in a total volume of 100µl and incubated at 37°C for 30 minutes. The RNA was then extracted by phenol:chloroform and precipitated with one volume of isopropanol with 1/10 volume of Sodium Acetate 3M and 10µg of Glycoblue (Ambion).

*Base Hydrolysis of RNA*

The hydrolysis of RNA was performed using fragmentation reagent (Ambion) for 15min at 70°C and the product was purified by p-30 RNase free spin column (BioRad), following the manufacturer's instruction.

*Immunopurification of Br-U RNA and RNA-End Repair*

The purification of the Br-U RNA and the end repair of the purified Br-U RNA was performed as described in (Core et al., 2008).

*Adapter Ligation and Library Preparation*

For the ligation of 5' and 3' adapters to Br-U RNA we used reagents contained in the ScriptMiner™ Small RNA-sequencing kit (Epicentre) and we immunopurified the ligated Br-U RNA product after each ligation with anti-deoxy-BrU beads (Santa Cruz Biotech (sc-32323-AC). The ligated Br-U RNA was then reverse-transcribed and PCR-amplified with properly-indexed primers for 18 cycles, according the ScriptMiner™ Small RNA-sequencing kit (Epicentre) manufacturer's instructions. The NRO-cDNA libraries were purified by phenol:chloroform, the DNA was precipitated as above and then ran on a non-denaturing 1xTBE, 8% acrylamide gel and stained with SYBR gold (Invitrogen). DNA fragments ranging in size from 120bp to 300bp were excised from the gel and eluted by incubating in TE + 300mM NaCl overnight with rotation. The DNA libraries

were then extracted, precipitated, and sent to sequencing on Illumina Hi-Seq 2000 at the Koch Institute Biopolymers & Proteomics Core Facility, MIT, Cambridge.

*Bioinformatic Analysis of GRO-Seq Data*

Two independent biological replicates were generated for each strain (N2 (WT) and *zfp-1 (ok554)*). Sequencing was performed to a length of 50 nucleotides. Since cloned RNA ranged from 16 to 180 nucleotides in length, the sequencing reads were post-processed to eliminate the 3' linker sequence, if present, so that resulting reads ranged from 16 to 50 nucleotides in length. The reads were then aligned using "bwa 0.6.1" to the *C. elegans* genome (ce6 assembly), and rRNA reads were removed based on genomic locations:

chrI 15060000 15072421 Ce_18S_rDNA 0 +

chrV 17115500 17132700 Ce_5S_rDNA 0 +

chrV 17428800 17429040 Ce_5S_rDNA-singleton 0 +

chrX 16929621 16940441 Ce_repetitiveLocus 0 +

After filtering, we obtained 8,484,939 reads from wild type (5,039,164 reads from the first replicate and 3,445,775 reads from the second replicate) and 7,539,039 reads from *zfp-1(ok554)* (4,765,744 reads from the first replicate and 2,773,315 reads from the second replicate). After ascertaining concordant results with the individual replicate pairs (WT vs. mutant) (data not shown), replicate reads for each WT and mutant were pooled, using downsampling on the replicate with higher read count to ensure equal representation of either replicate, and the pooled reads were processed together to increase statistical power.

For each annotated *C. elegans* transcript, the aligned reads were binned based on the annotated transcriptional start site (TSS) and transcriptional termination site (TTS) into 3

regions as follows: 'promoter' (from 600bp upstream of the annotated TSS to 200bp downstream), 'gene body' (from 200bp downstream of the annotated TSS to 200bp upstream of the annotated TTS), and 'gene end' (from 200bp upstream of the annotated TTS to 400bp downstream). We then calculated region-specific RPKM values (#total reads / (region length (KB) X # mapped reads (M)) for these regions and considered for further analysis only transcripts with RPKM values of > 1 for the 'gene body' (= transcribed genes). The pausing index was calculated as the log2 ratio of 'promoter' RPKM over 'gene body' RPKM for each gene. Genes with a pausing index more than 1 (i.e. at least twice as many promoter reads as gene body reads) were considered paused.

The average Pol II ChIP-seq and GRO-seq profiles in Figure S6A,B and S4F were calculated using Cistrome/Galaxy (Liu et al., 2011) using the CEAS tool (enrichment on chromosome and annotation).

Primers used in this study are listed below.

**Primers used for ChIP:**

| | |
|---|---|
| ***pdk-1* Promoter** | ***pdk-1* Coding** |
| F_pdk-1P | F_pdk-1ex7 |
| AAACAACACATAGACTTGTGCC | CAAATATCAAGCCACCAGTC |
| R_pdk-1P | R_pdk-1ex7 |
| GTACGGTTGTTATCGCTTTCAG | TTCCCAAATTCATCAAACGG |
| ***cye-1* Promoter** | ***cye-1* Coding** |
| F_cye-1P | F_cye-1ex5 |
| GTAAGACCATCAGGGAGGAC | TGGTCTCTGATGGTCAAACGA |
| R_cye-1P | R_cye-1ex5 |
| GGACATTACGCAATCAATCTG | CCAAATGAAATGTTTCTCGGTG |
| ***ncl-1* Promoter** | ***ncl-1* Coding** |
| F_ncl-1P | F_ncl-1ex2 |
| ACTTCAGTCTTCCTCAAACAG | TACCTCCACCTCAACAAGAC |
| R_ncl-1P | R_ncl-1ex2 |
| TCTCTTTCTCACTCTTTCTCAG | GATTCCCTCGATAACCACCT |
| ***ain-2* Promoter** | ***ain-2* Coding** |
| F_ ain-2P | F_ain-2ex5 |
| CTTGTTCATGTCTCTGCGTC | GATATTTCCCTACTCTCACTTCC |

R_ ain-2P
CGTGGAGGAATAAGAAACAAAGAG
***egl-30* Promoter**
F_egl-30P1
GGTTCTTCACCAAATTGCGA
R_egl-30P1
CGTTCTAATGAAACCGATCTCAC
***bet-2* (F57C7.1) Promoter**
Fbet-2 P
GTGAGTGAGTAGATGAGCGT
Rbet-2 P
AGAGACATAGAATGGGAGGG
***unc-60* Promoter**
Func-60P
CCCTCTCCTCTACTTGTCAC
Runc-60P
GCTTAGATAGACGAGGACGAG
***pqn-51* Promoter**
F_ pqn-51P
ATATACTCAGTCGATCGTCGC
R_pqn-51P
CCCTCTGATCACAGAAAGAC
***zfp-1* Promoter**
F_zfp-1P
GTCGTCAATTCTATTTCTCGT
R_zfp-1P
GATAGTAGCCGAAAGGAACAG
***aqp-1* Promoter**
F_aqp-1P
TTTCAGAACTATCATGCCACG
R_ aqp-1P
TCTCTGAGCACACTTTGAGG
***gst-4* Promoter**
F_gst-4P
TTAGATAGAGAATTGGCGAGAG
R_ gst4P
CAAGTAGCAAAGCGATAAACC

R_ain-2ex5
CTCATCCTCATAGTCGTCCA
***egl-30* Coding**
F_egl-30EX3
AGTTTGGTAACGAATCAGAGGA
R_egl-30EX3
CCAGAATCCTCCCATAGCTC
***bet-2* (F57C7.1) Coding**
Fbet-2 C
GAAGTTGTTTCCTCGTCCAG
Rbet-2 C
GCTCGTCCTCAATAAGATCAG
***unc-60* Coding**
Func-60C
CACTCAATAAGGTCATCTTCGT
Runc-60C
TAGATCAAACCGCAATGGGA
***pqn-51* Coding**
F_pqn-51C
CTTGGGATCACTATTGAAAGA
R_pqn-51C
CTTCAACTTCACCATTTCCC
***zfp-1* Coding**
F_zfp-1ex7
CTGAAGTACACAAACAAGAACC
R_zfp-1ex7
TGACAGTGGAACAGTTAGAG
***aqp-1* Coding**
F_aqp-1EX5
TTGCCAGTTATCCATCTCCA
R_ aqp-1EX5
CTCTCATCAATAACAACGCAG
***gst-4* Coding**
F_gst-4ex4
TGAAGTTGTTGAACCAGCC
R_ gst-4ex4
CCCAAGCCAATGAGTCTCCA

**Primers used for RT-qPCR:**
**mRNA primers:**
***act-3***
F_act-3
CACGAGACTTCTTACAACTCC
R_act-3
GCATACGATCAGCAATTCCT
**18S rRNA**

F_18S
CTTAATTTGACTCAACACGGGA
R_18S
GCTCGTTATCGGAATAAACCA
***pdk-1***
F_pdk-1
AGCCATCAACACCGTCTAAC
R_pdk-1
CGAATTGGCGCGTGGTGC
***ncl-1***
F_ncl-1
TACCTCCACCTCAACAAGAC
R_ncl-1
GATTCCCTCGATAACCACCT
***egl-30***
F_egl-30
GCCATACAGTCTATGATACGA
R_egl-30
GCCTCCTATCATAACATTCTT
***zfp-1***
F_zfp-1
TTCAGAATCACAGAGCAACAC
R_zfp-1
TGCGATACATGTTCAGAAGAG
***cye-1***
F_cye-1
TGGTCTCTGATGGTCAAACGA
R_cye-1
CCAAATGAAATGTTTCTCGGTG
***ain-2***
Fain-2
AACAACAAGCCGAAGAAGAG
Rain-2
CGAATTCTGAATCACTCTTGTC
***bet-2* (F57C7.1)**
Fbet-2
GAAGTTGTTTCCTCGTCCAG
Rbet-2
GCTCGTCCTCAATAAGATCAG
***unc-60***
F_unc-60
CAAGATCGACGAGAACAAGG
R_unc-60
ATTATCGGTTCGAGACTTCAC
***pqn-51***
F_pqn-51

CTTGGGATCACTATTGAAAGA
R_pqn-51
CTTCAACTTCACCATTTCCC
***rfp-1***
F_rfp-1
GCCGAATTAACACGTCTCAC
R_rfp-1
CATTTGACCTTCACATCGATCC
***aqp-1***
F_aqp-1
GCTAGGATGTCAGGGGGTC
R_aqp-1
CCAAGAGAAGCCACGAAAGC
**pre-mRNA primers:**
***pdk-1***
F_pdk-1pre
GGACCACCCTGTATAATTATGGA
R_pdk-1pre
TTGAAGGTGTTCTAAGTTTGGG
***egl-30***
F_egl-30pre
CCCTGGAGAAAGTACTAATTACGA
R_egl-30pre
TCTCCGTATCTCCGTATCTC
***zfp-1***
F_zfp-1pre
TCATGATCGCAGTTACATCTG
R_zfp-1pre
AATCTGTGCAATACAGTTCG
***aqp-1***
F_aqp-1pre
TGATCGGAACTGGATTTGGA
R_aqp-1pre
GAATTGACACTGTGAACTAACTGG

**Supplemental References**

Brown, N.P., Leroy, C., and Sander, C. (1998). MView: a web-compatible database search or multiple alignment viewer. Bioinformatics *14*, 380-381.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.
Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.
Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711-723.
Taylor, W.R. (1986). The classification of amino acid conservation. J Theor Biol *119*, 205-218.
.