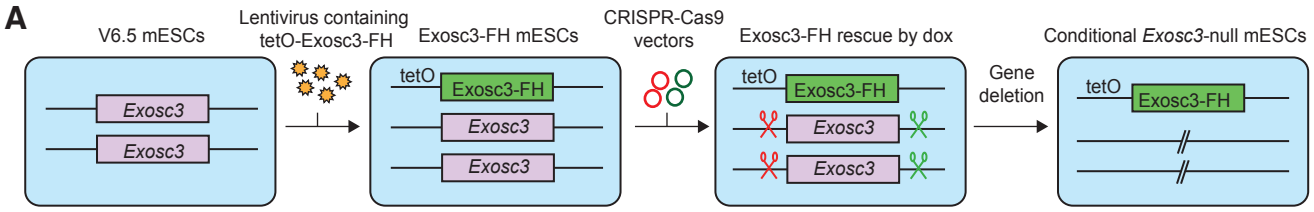
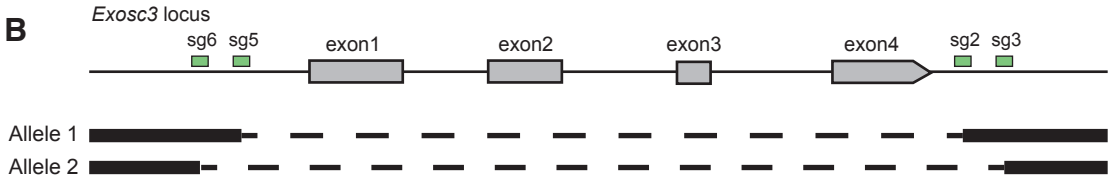
Molecular Cell, Volume *69*

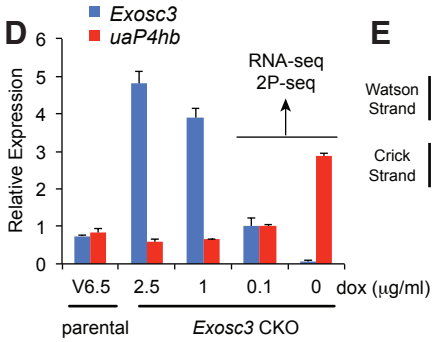
Supplemental Information

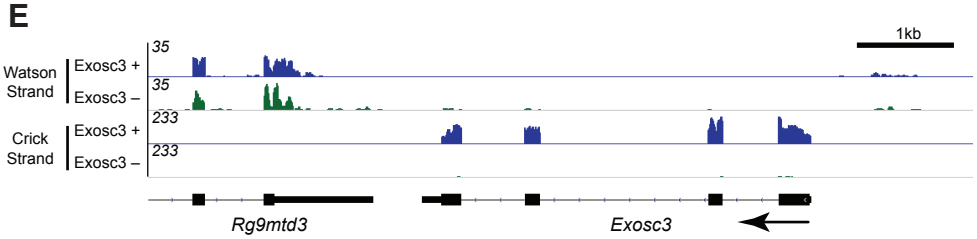
Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP

Anthony C. Chiu, Hiroshi I. Suzuki, Xuebing Wu, Dig B. Mahat, Andrea J. Kriz, and Phillip A. Sharp

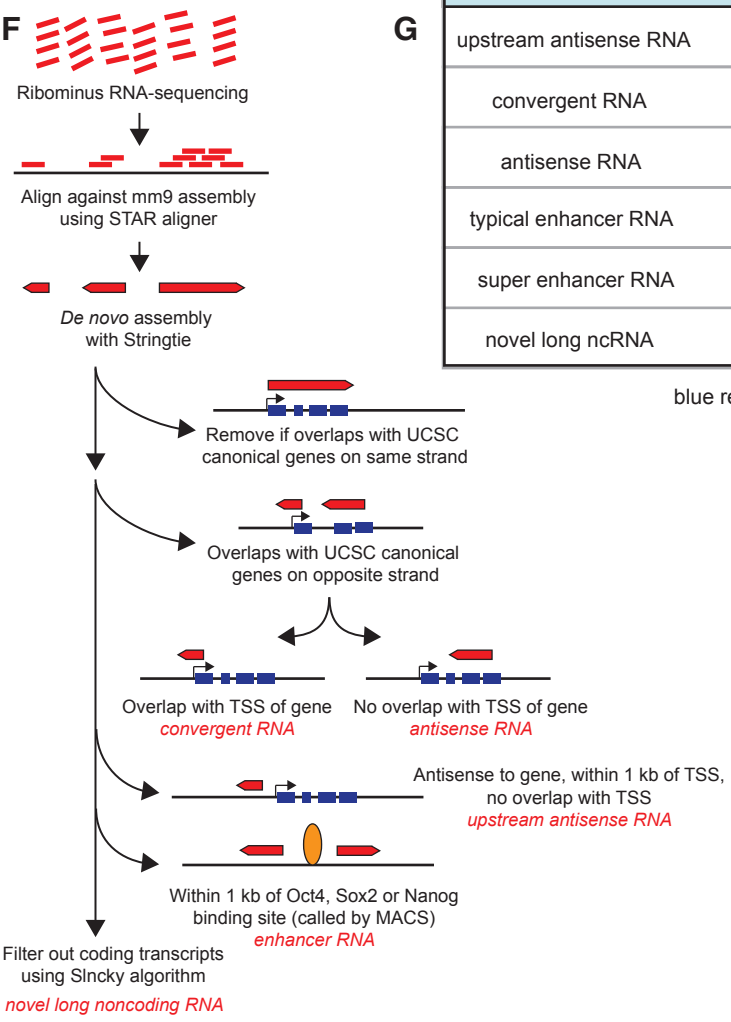








F



G

Transcript Type	Structure	Number of Intervals	Median Size
upstream antisense RNA		3336	845 bp
convergent RNA		1277	1233 bp
antisense RNA		4568	532 bp
typical enhancer RNA		904	702 bp
super enhancer RNA		393	936 bp
novel long ncRNA		7427	502 bp

blue represents UCSC canonical gene transcript

Figure S1. Development of *Exosc3* conditional deletion system and *de novo* transcriptome assembly, related to Figure 1.

(A) Generation of *Exosc3* CKO mESCs.

(B) Design of CRISPR-Cas9 sgRNAs targeting endogenous *Exosc3* gene.

(C) Sanger sequencing of PCR products across CRISPR target sites validating deletion of endogenous *Exosc3*.

(D) qRT-PCR of spliced *Exosc3* mRNA and uaRNA *uaP4hb*. Plotted is the mean from 3 biological replicates; error bars represent standard error.

(E) Genome browser shot of RNA-seq reads at *Exosc3* upon removal of dox showing confirmation of specific loss of *Exosc3* transcripts.

(F) Strategy for *de novo* transcriptome assembly and interval classification.

(G) Properties of intervals defined by *de novo* transcriptome assembly.

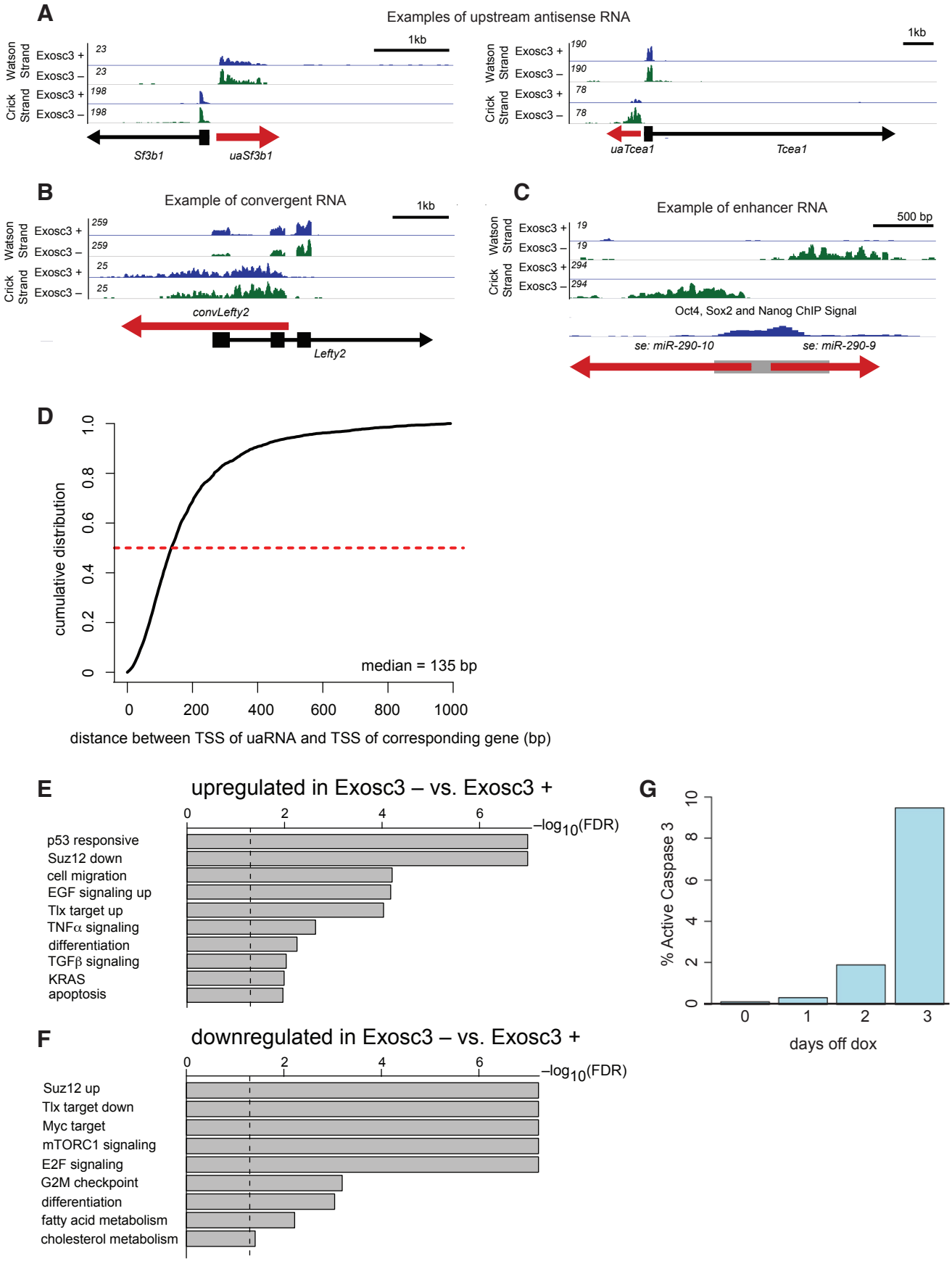

A
Examples of upstream antisense RNA
1kb
Watson Strand
Crick Strand
Exosc3 +
Exosc3 –
23
23
198
198
Sf3b1
uaSf3b1
190
190
78
78
uaTcea1
Tcea1
B
Example of convergent RNA
259
259
25
25
convLefty2
Lefty2
C
Example of enhancer RNA
500 bp
19
19
294
294
Oct4, Sox2 and Nanog ChIP Signal
se: miR-290-10
se: miR-290-9
D
cumulative distribution
median = 135 bp
distance between TSS of uaRNA and TSS of corresponding gene (bp)
E
upregulated in Exosc3 – vs. Exosc3 +
–log10(FDR)
p53 responsive
Suz12 down
cell migration
EGF signaling up
Tlx target up
TNFα signaling
differentiation
TGFβ signaling
KRAS
apoptosis
F
downregulated in Exosc3 – vs. Exosc3 +
Suz12 up
Tlx target down
Myc target
mTORC1 signaling
E2F signaling
G2M checkpoint
differentiation
fatty acid metabolism
cholesterol metabolism
G
% Active Caspase 3
days off dox

Figure S2. Characterization of Exosc3-targeted RNA classes, related to Figure 1.

(A-C) Genome browser shots of representative examples of newly defined intervals (red) for upstream antisense RNAs (A), convergent RNA (B), and enhancer RNA (C). RNA-seq reads are illustrated as Exosc3 + (blue) and Exosc3 – (green). For the enhancer RNA, the grey box represents the ChIP peaks of the combined binding profiles of Oct4, Sox2, and Nanog transcription factors defined by MACS.

(D) Cumulative distribution of the distance between the TSS of uaRNAs and the TSS of their corresponding mRNAs. Red dashed line represents where the cumulative distribution would be the median.

(E, F) Boxplot showing selected statistically significant pathways identified by gene set enrichment analysis (GSEA) for genes upregulated (E) or downregulated (F) upon Exosc3 loss. Dotted line represents false discovery rate of 0.05.

(G) Percent of cells that are positive for active caspase 3 by FACS analysis.

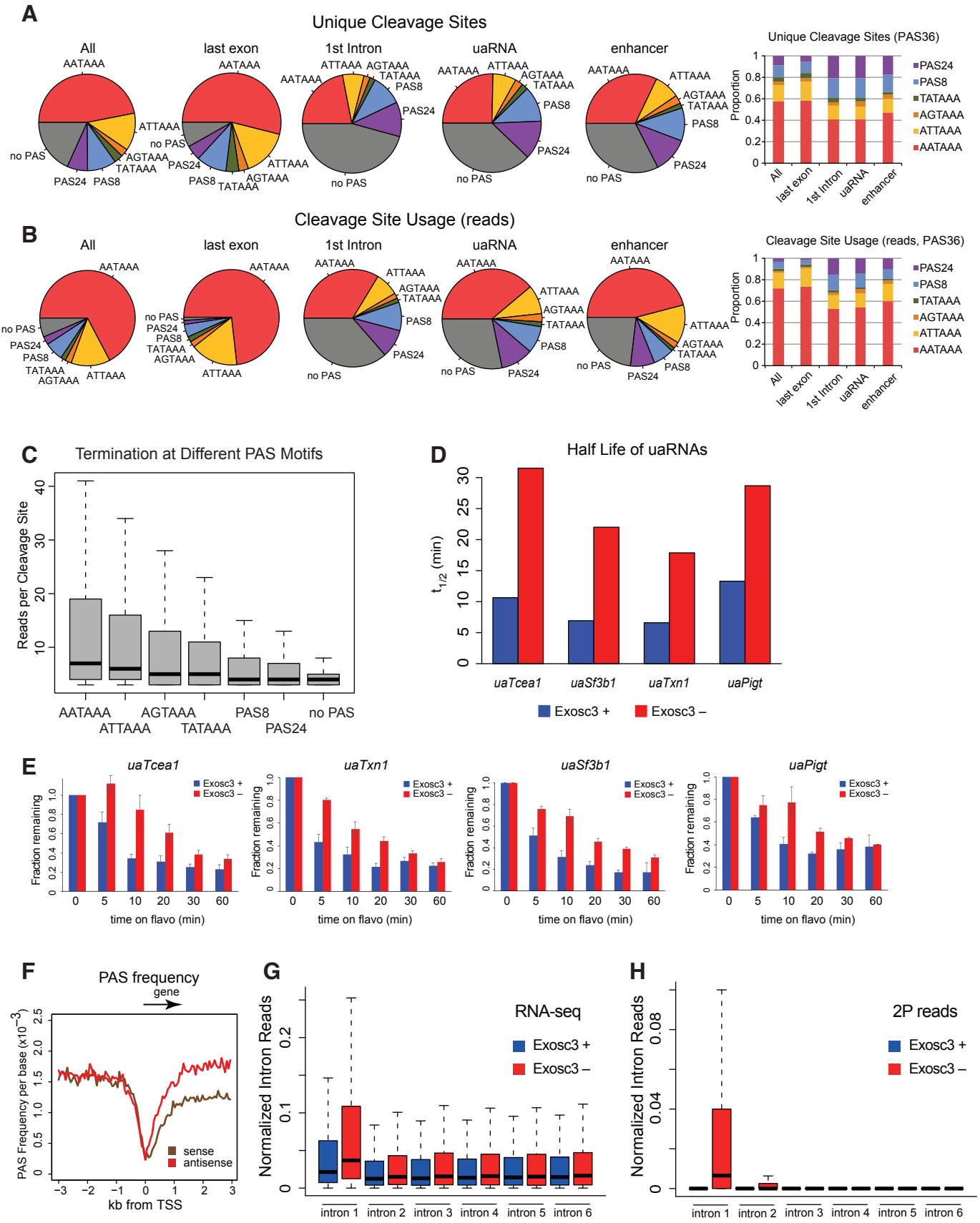

A
Unique Cleavage Sites
All
last exon
1st Intron
uaRNA
enhancer
AATAAA
ATTAAA
AGTAAA
TATAAA
PAS8
PAS24
no PAS
Unique Cleavage Sites (PAS36)
Proportion
B
Cleavage Site Usage (reads)
Cleavage Site Usage (reads, PAS36)
C
Termination at Different PAS Motifs
Reads per Cleavage Site
AATAAA
ATTAAA
AGTAAA
TATAAA
PAS8
PAS24
no PAS
D
Half Life of uaRNAs
t1/2 (min)
uaTcea1
uaSf3b1
uaTxn1
uaPigt
Exosc3 +
Exosc3 –
E
Fraction remaining
time on flavo (min)
F
PAS frequency
gene
PAS Frequency per base (x10−3)
sense
antisense
kb from TSS
G
RNA-seq
Normalized Intron Reads
intron 1
intron 2
intron 3
intron 4
intron 5
intron 6
H
2P reads

Figure S3. PAS termination at TSS proximal region, related to Figures 1 and 2.

(A) Pie chart showing distribution of various PAS motifs for all uniquely detected cleavage sites (left). Right panel shows distribution of 36 PAS motifs.

(B) Pie chart showing distribution of various PAS motifs for all detected cleavage site reads (left). Right panel shows distribution of 36 PAS motifs.

(C) Boxplot showing number of reads for each cleavage sites sorted by type of PAS motif. Tails reflect 95^{th} percentile.

(D) Half-lives of uaRNAs, determined by qRT-PCR analysis of cells with or without dox, further treated with 1 μM flavopiridol, as shown in panel (E).

(E) Relative abundance of uaRNAs from oligo-dT primed cDNA after treatment with 1 μM flavopiridol in Exosc3 + (blue) or Exosc3 – (green) mESCs, determined by qRT-PCR analysis. Plotted is the mean from 3 biological replicates; error bars represent standard error.

(F) Frequency of predicted canonical PAS motifs (AATAAA/ATTAAA) flanking TSS on sense (brown) or antisense strand (red).

(G) Boxplot showing normalized RNA-seq reads ($FPKM_{\text{intron N reads}}/FPKM_{\text{mature transcript reads}}$) for UCSC canonical genes with at least 6 introns.

(H) Boxplot showing normalized intronic cleavage site reads (Sum of 2P reads in intron N/sum of 2P reads in last exon) for UCSC canonical genes with at least 6 introns.

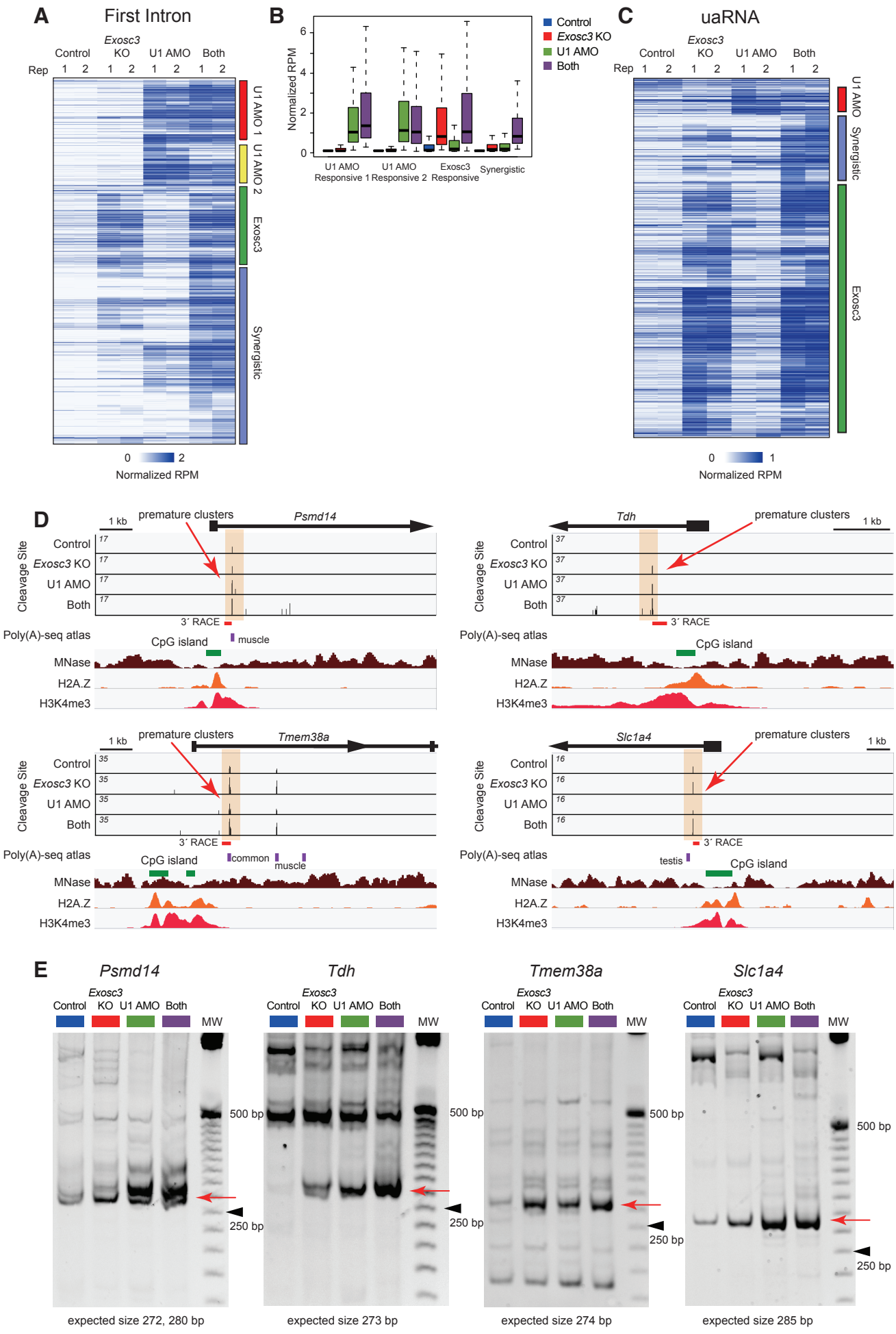

A
First Intron
Exosc3
Control KO U1 AMO Both
Rep 1 2 1 2 1 2 1 2
U1 AMO 1
U1 AMO 2
Exosc3
Synergistic
0 2
Normalized RPM
B
Normalized RPM
Control
Exosc3 KO
U1 AMO
Both
U1 AMO Responsive 1
U1 AMO Responsive 2
Exosc3 Responsive
Synergistic
C
uaRNA
Exosc3
Control KO U1 AMO Both
Rep 1 2 1 2 1 2 1 2
U1 AMO
Synergistic
Exosc3
0 1
Normalized RPM
D
Psmd14
premature clusters
1 kb
Cleavage Site
Control
Exosc3 KO
U1 AMO
Both
3′ RACE
Poly(A)-seq atlas
CpG island
muscle
MNase
H2A.Z
H3K4me3
Tdh
Tmem38a
common
Slc1a4
testis
E
Psmd14
Tdh
Tmem38a
Slc1a4
Exosc3
Control KO U1 AMO Both
MW
500 bp
250 bp
expected size 272, 280 bp
expected size 273 bp
expected size 274 bp
expected size 285 bp

Figure S4. Effects of U1 inhibition and Exosc3 depletion on PAS termination, related to Figure 3.

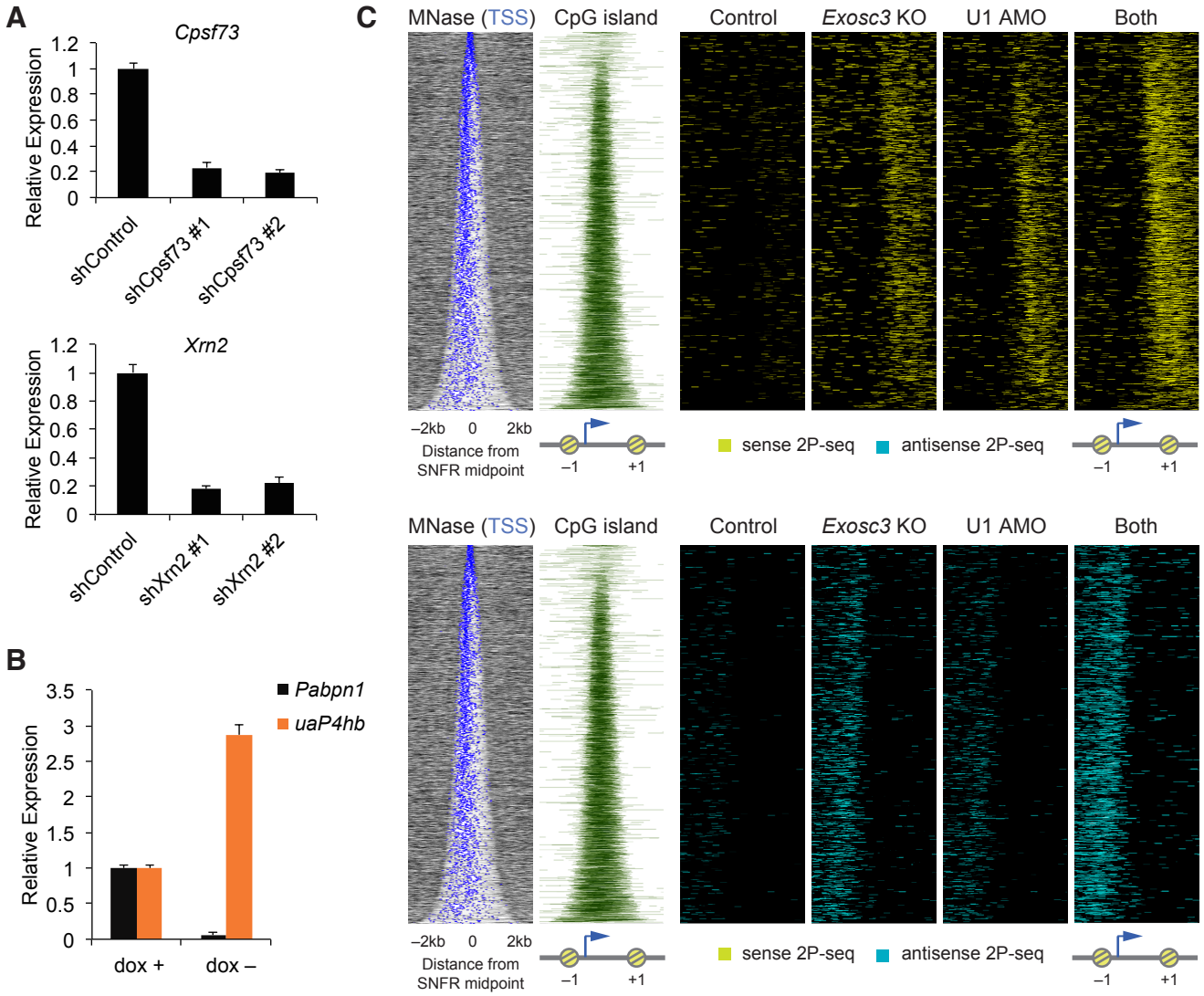
(A) Heatmap of library-size normalized RPM for hierarchically clustered PAS-linked cleavage clusters within the first intron.

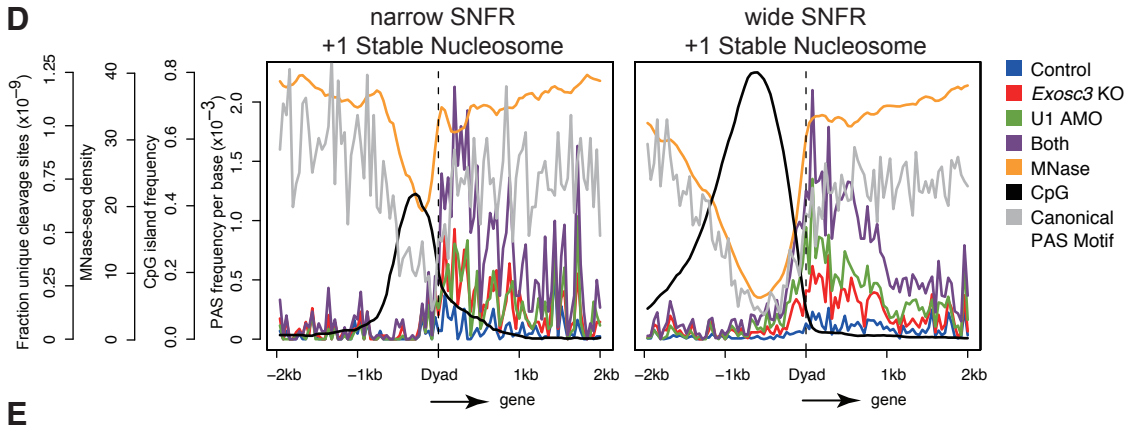
(B) Boxplot showing normalized RPM for 4 identified clusters in panel (A).

(C) Heatmap of library-size normalized RPM for hierarchically clustered PAS-linked cleavage clusters within uaRNAs.

(D) Genome browser shot of *Psmd14*, *Tdh*, *Tmem38a*, and *Slc1a4* showing PAS-linked cleavage sites (top, orange shade), annotated CpG island (green), MNase-seq (brown), H2A.Z ChIP-seq (orange), and H3K4me3 ChIP-seq (red). Previously reported PAS sites in mouse tissues (purple) and PCR products are also shown.

(E) Nested 3′ RACE analysis of four other premature termination events (*Psmd14*, *Tdh*, *Tmem38a*, and *Slc1a4*) on a nondenaturing polyacrylamide gel. Red arrows indicate most frequent termination sites, which have been sequence validated. Molecular weight ladder is 25 bp ladder, and black arrowheads indicate 250 bp.


A
Cpsf73
Relative Expression
shControl
shCpsf73 #1
shCpsf73 #2
Xrn2
shControl
shXrn2 #1
shXrn2 #2
B
Pabpn1
uaP4hb
Relative Expression
dox +
dox –
C
MNase (TSS)
CpG island
Control
Exosc3 KO
U1 AMO
Both
–2kb 0 2kb
Distance from SNFR midpoint
sense 2P-seq
antisense 2P-seq
–1
+1


D
narrow SNFR
+1 Stable Nucleosome
wide SNFR
+1 Stable Nucleosome
Fraction unique cleavage sites (x10−9)
MNase-seq density
CpG island frequency
PAS frequency per base (x10−3)
–2kb
–1kb
Dyad
1kb
2kb
gene
Control
Exosc3 KO
U1 AMO
Both
MNase
CpG
Canonical PAS Motif

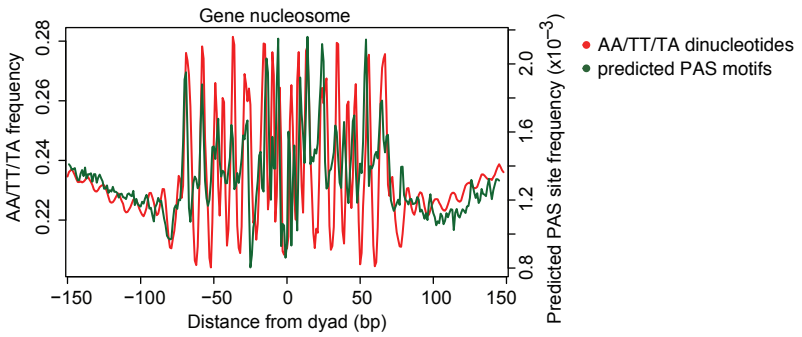

E
Gene nucleosome
AA/TT/TA frequency
Predicted PAS site frequency (x10−3)
Distance from dyad (bp)
AA/TT/TA dinucleotides
predicted PAS motifs

Figure S5. PAS termination and –1/+1 stable nucleosomes, related to Figures 3 and 4.

(A) Knockdown of Cpsf73 and Xrn2, validated by qRT-PCR analysis. Plotted is the mean from 3 biological replicates; error bars represent standard error.

(B) Generation of conditional *Pabpn1* knockout mESCs. qRT-PCR results of *Pabpn1* mRNA and uaRNA *uaP4hb* are shown. Plotted is the mean from 3 biological replicates; error bars represent standard error.

(C) Heatmap of MNase-seq, CpG islands, and PAS-linked cleavage sites (top: sense 2P-seq reads, bottom: antisense 2P-seq reads) around a 2 kb window flanking the SNFR midpoint for non-overlapping expressed genes with 2P clusters, ranked by increasing SNFR width.

(D) Metaplots of PAS-linked cleavage sites, MNase-seq, CpG islands, and predicted canonical PAS motifs around the dyad axis of the +1 stable nucleosome of genes with narrow SNFR (< 750 bp, left) and wide SNFR (> 750 bp, right).

(E) AA/TT/TA dinucleotide frequency (red) and frequency of predicted canonical PAS motifs (green) per gene body nucleosome in a 150 bp window from chemical mapping-defined dyad axis. Gene body nucleosomes are between TSS and 2 kb upstream from the TES of genes.

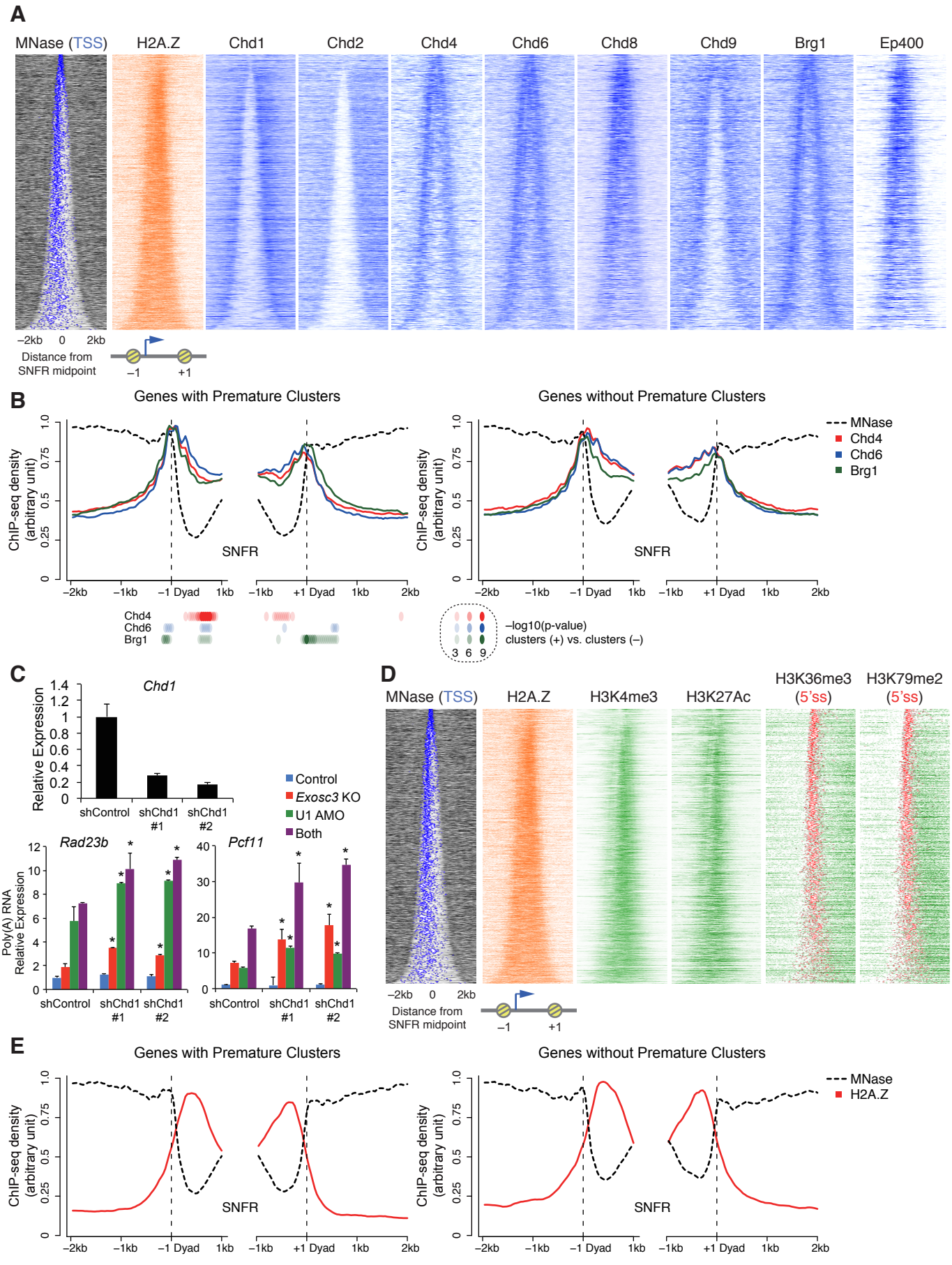

A
MNase (TSS)
H2A.Z
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Brg1
Ep400
–2kb 0 2kb
Distance from SNFR midpoint
–1 +1
B
Genes with Premature Clusters
Genes without Premature Clusters
ChIP-seq density (arbitrary unit)
SNFR
–2kb –1kb –1 Dyad 1kb
–1kb +1 Dyad 1kb 2kb
MNase
Chd4
Chd6
Brg1
Chd4
Chd6
Brg1
–log10(p-value) clusters (+) vs. clusters (–)
3 6 9
C
Chd1
Relative Expression
shControl shChd1 #1 shChd1 #2
Control
Exosc3 KO
U1 AMO
Both
Rad23b
Pcf11
Poly(A) RNA Relative Expression
shControl shChd1 #1 shChd1 #2
D
MNase (TSS)
H2A.Z
H3K4me3
H3K27Ac
H3K36me3 (5'ss)
H3K79me2 (5'ss)
–2kb 0 2kb
Distance from SNFR midpoint
–1 +1
E
Genes with Premature Clusters
Genes without Premature Clusters
ChIP-seq density (arbitrary unit)
SNFR
MNase
H2A.Z

Figure S6. Distribution of chromatin remodelers and histone modification around –1/+1 stable nucleosomes, related to Figure 5.

(A) Heatmap of ChIP-seq signal for various chromatin remodelers in a 2 kb window flanking the SNFR midpoint for non-overlapping expressed genes, ranked by increasing SNFR width. Chd1 prefers +1 stable nucleosomes, whereas several factors such as Ep400 and Chd4 prefer –1 stable nucleosomes.

(B) Read coverage of MNase-seq and chromatin remodelers ChIP-seq in a –2 kb to +1 kb window around the –1 stable nucleosome dyad axis and –1 kb to +2 kb window around the +1 stable nucleosome dyad axis, separated for genes with premature intron clusters (left) and expression-matched genes without premature intron clusters (right). Chd4, Chd6, and Brg1 do not show clear differences between two gene sets. P values with K-S test at each bin for each factor are displayed.

(C) Top: knockdown of Chd1, validated by qRT-PCR analysis. Bottom: effects of knockdown of Chd1 on induction of premature polyadenylated transcripts of *Rad23b* and *Pcf11*, determined by qRT-PCR using the fusion primer covering gene-specific sequence and poly(A) tail. Plotted is the mean from 3 biological replicates; error bars represent standard error. * $P < 0.05$, compared to shControl.

(D) Heatmap of ChIP-seq signal for various histone marks in a 2 kb window flanking the SNFR midpoint for non-overlapping expressed genes, ranked by increasing SNFR width. H2A.Z is enriched over SNFR region upstream of cleavage signals. H3K4me3 and H3K27ac were also enriched within the SNFR. Histone marks of productive elongation (H3K36me3 and H3K79me2) were found mostly in the downstream sense orientation.

(E) Read coverage of MNase-seq and H2A.Z ChIP-seq in a –2 kb to +1 kb window around the –1 stable nucleosome dyad axis and –1 kb to +2 kb window around the +1 stable nucleosome dyad axis, separated for genes with premature intron clusters (left) and expression-matched genes without premature intron clusters (right). H2A.Z do not show clear differences between two gene sets.

Sources of the datasets are described in Table S2.

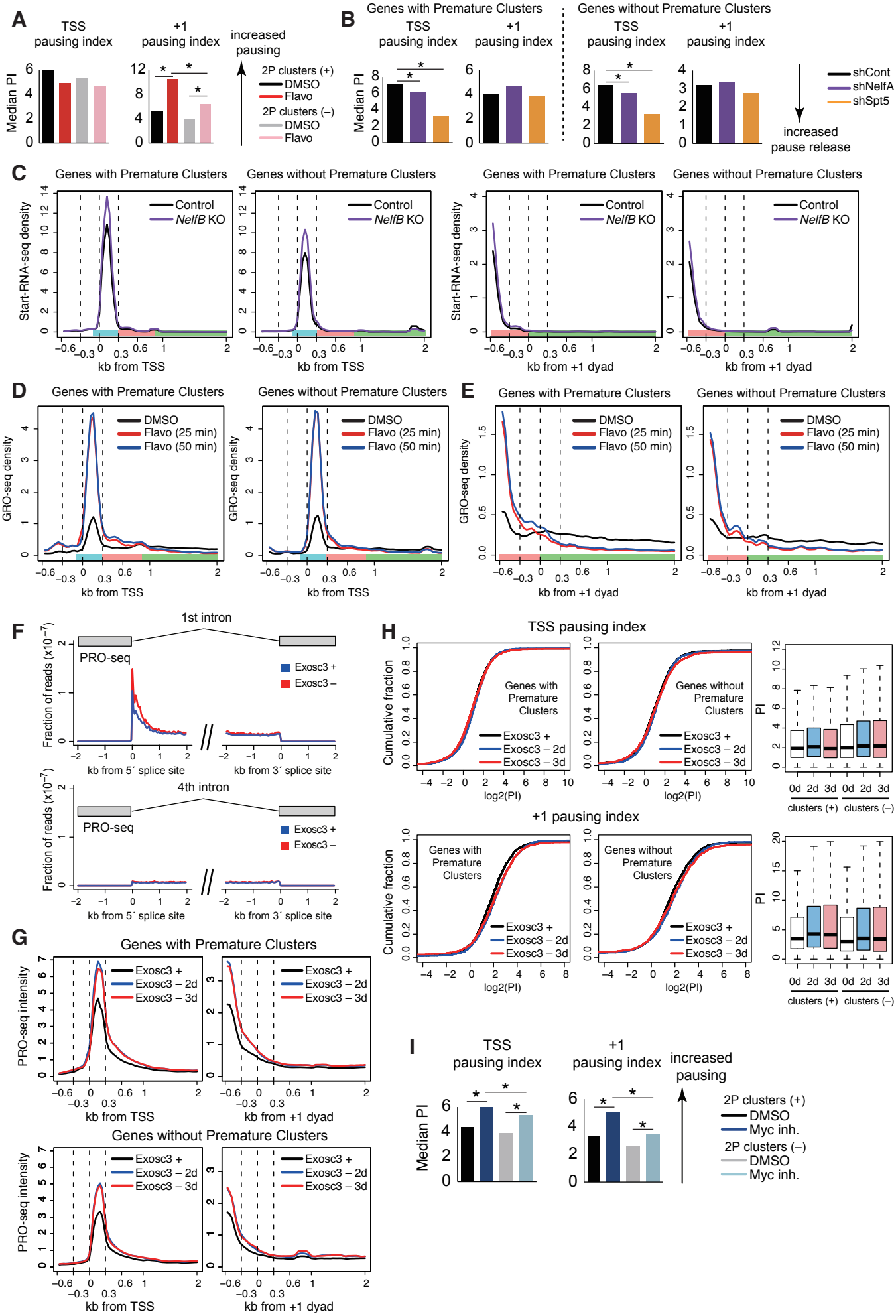

A
TSS pausing index
+1 pausing index
increased pausing
Median PI
2P clusters (+)
DMSO
Flavo
2P clusters (–)
DMSO
Flavo
B
Genes with Premature Clusters
Genes without Premature Clusters
TSS pausing index
+1 pausing index
shCont
shNelfA
shSpt5
increased pause release
C
Genes with Premature Clusters
Genes without Premature Clusters
Control
NelfB KO
Start-RNA-seq density
kb from TSS
kb from +1 dyad
D
GRO-seq density
DMSO
Flavo (25 min)
Flavo (50 min)
kb from TSS
E
kb from +1 dyad
F
1st intron
4th intron
PRO-seq
Exosc3 +
Exosc3 –
Fraction of reads (x10⁻⁷)
kb from 5´ splice site
kb from 3´ splice site
G
PRO-seq intensity
Exosc3 +
Exosc3 – 2d
Exosc3 – 3d
kb from TSS
kb from +1 dyad
H
TSS pausing index
+1 pausing index
Cumulative fraction
log2(PI)
PI
0d 2d 3d 0d 2d 3d
clusters (+) clusters (–)
I
TSS pausing index
+1 pausing index
increased pausing
Median PI
2P clusters (+)
DMSO
Myc inh.
2P clusters (–)
DMSO
Myc inh.

Figure S7. Analysis of TSS and +1 nucleosome pausing and pausing effects by Exosc3 depletion, related to Figures 6 and 7.

(A, B) Comparison of TSS pausing index and +1 nucleosome pausing index shown in Figures 6E and 6G. The median pausing indices are shown in a raw scale. * $P < 0.01$ with K-S test.

(C) Metaplots of Start-RNA-seq density around the TSS or +1 dyad of wide SNFR genes in control and *NelfB* KO mESCs. StartRNA-seq datasets (Williams et al., 2015) was reanalyzed. Start-RNA-seq captures nascent capped short, less than 200 nts, RNA species, and read coverage is restricted to the TSS proximal regions in contrast to GRO-seq. Thus, it is difficult to detect +1 nucleosome-associated pause for wide SNFR genes, where the distances between the TSS and nucleosome-related pause sites are longer than 200 nts, in Start-RNA-seq datasets. Even with this limitation, Start-RNA-seq reads are higher in genes with premature clusters than genes without. In addition, depletion of NelfB resulted in slight increase of Start-RNA-seq reads in both gene sets with/without premature clusters.

(D) Metaplots of density of GRO-seq (Jonkers et al., 2014) around the TSS with flavopiridol treatment.

(E) Metaplots of density of GRO-seq (Jonkers et al., 2014) around the +1 dyad with flavopiridol treatment.

(F) Mean exon-removed PRO-seq signal flanking 5′ or 3′ splice sites for the first or fourth intron for Exosc3 + or Exosc3 – (dox off 3 days) samples, normalized by library depth.

(G) Metaplots of PRO-seq density around the TSS (left) or +1 dyad (right) for genes with 2P clusters (top) and expression-matched genes without 2P cluster (bottom) in Exosc3 + and Exosc3 – mESCs (dox off 2 and 3 days).

(H) Cumulative distribution plot of $\log_2$(pausing index) of the TSS (top) or +1 stable nucleosome pause (bottom) for genes with 2P clusters (left) and expression-matched genes without 2P clusters (right) in Exosc3 + and Exosc3 – mESCs (dox off 2 and 3 days). Right boxplots show distribution of pausing indices.

(I) Comparison of TSS pausing index and +1 nucleosome pausing index shown in Figure 7C. The median pausing indices are shown in a raw scale. * $P < 0.01$ with K-S test.

SUPPLEMENTAL TABLES

Table S1, related to Figures 1, 2, 3, and 5.

sgRNA and primer information.

Table S2, related to Figures 1-7.

Information about the datasets used in this study.