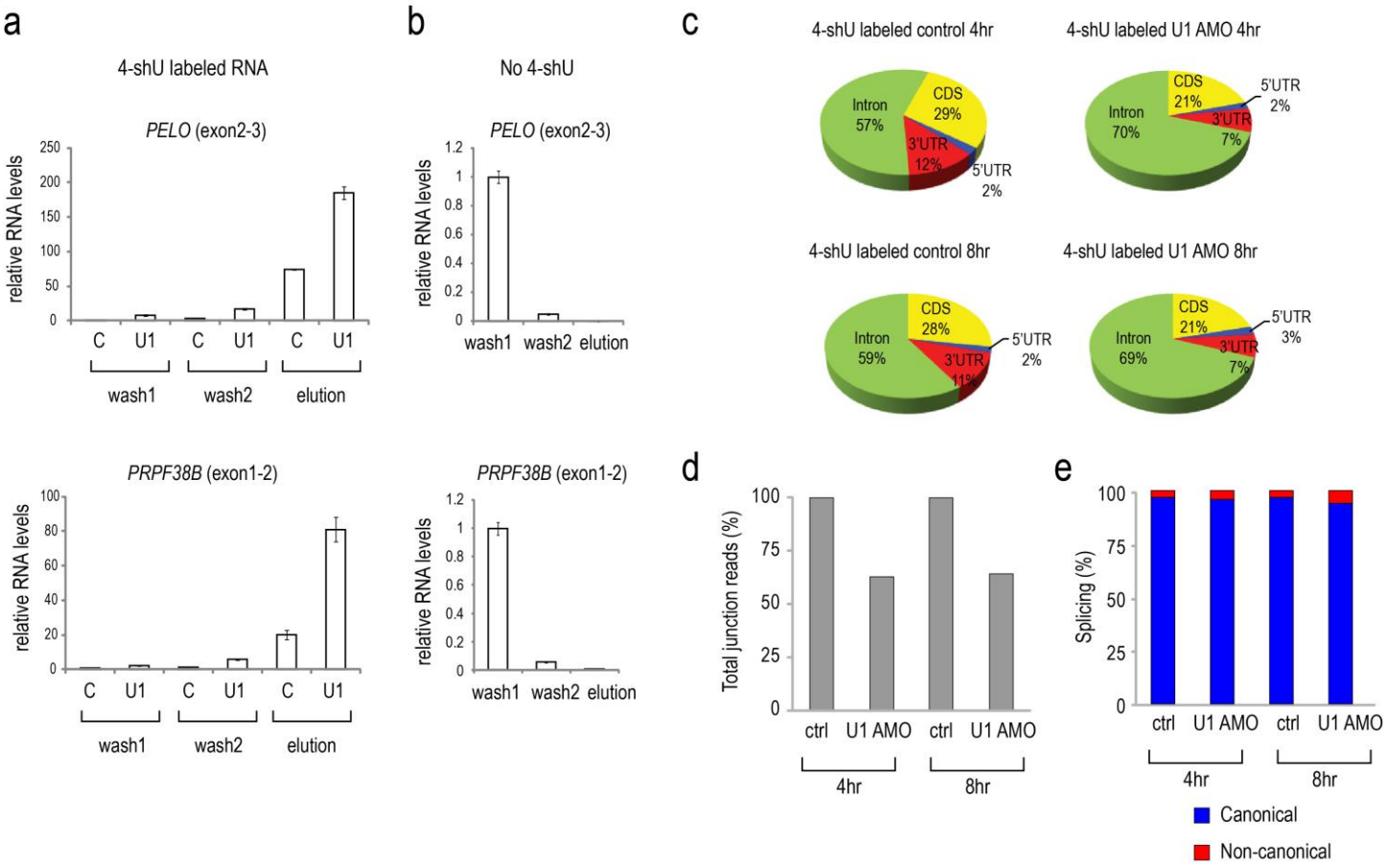
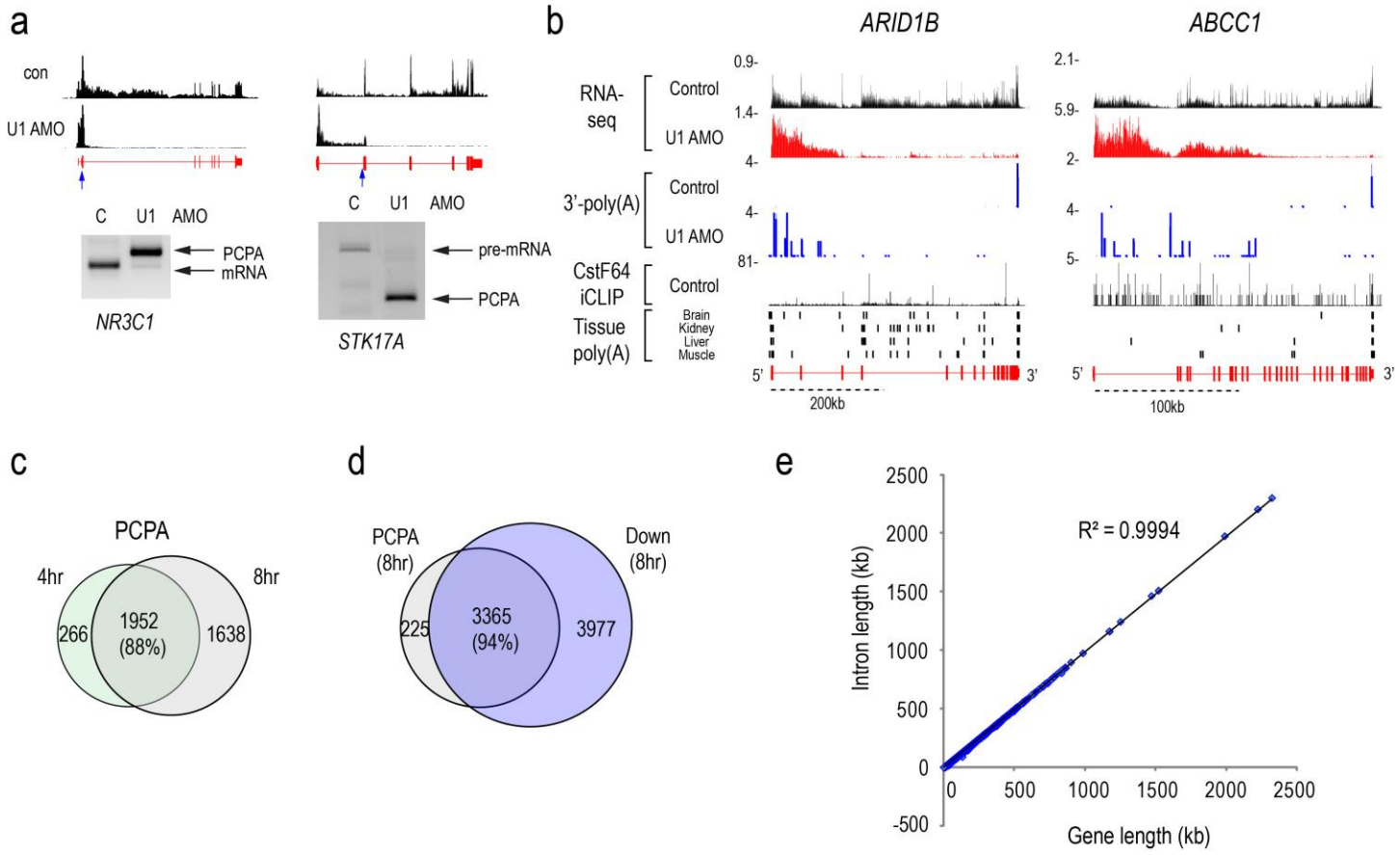


Supplementary Figure 1

U1 inhibition causes a shift of RNA-seq reads from exons to introns.

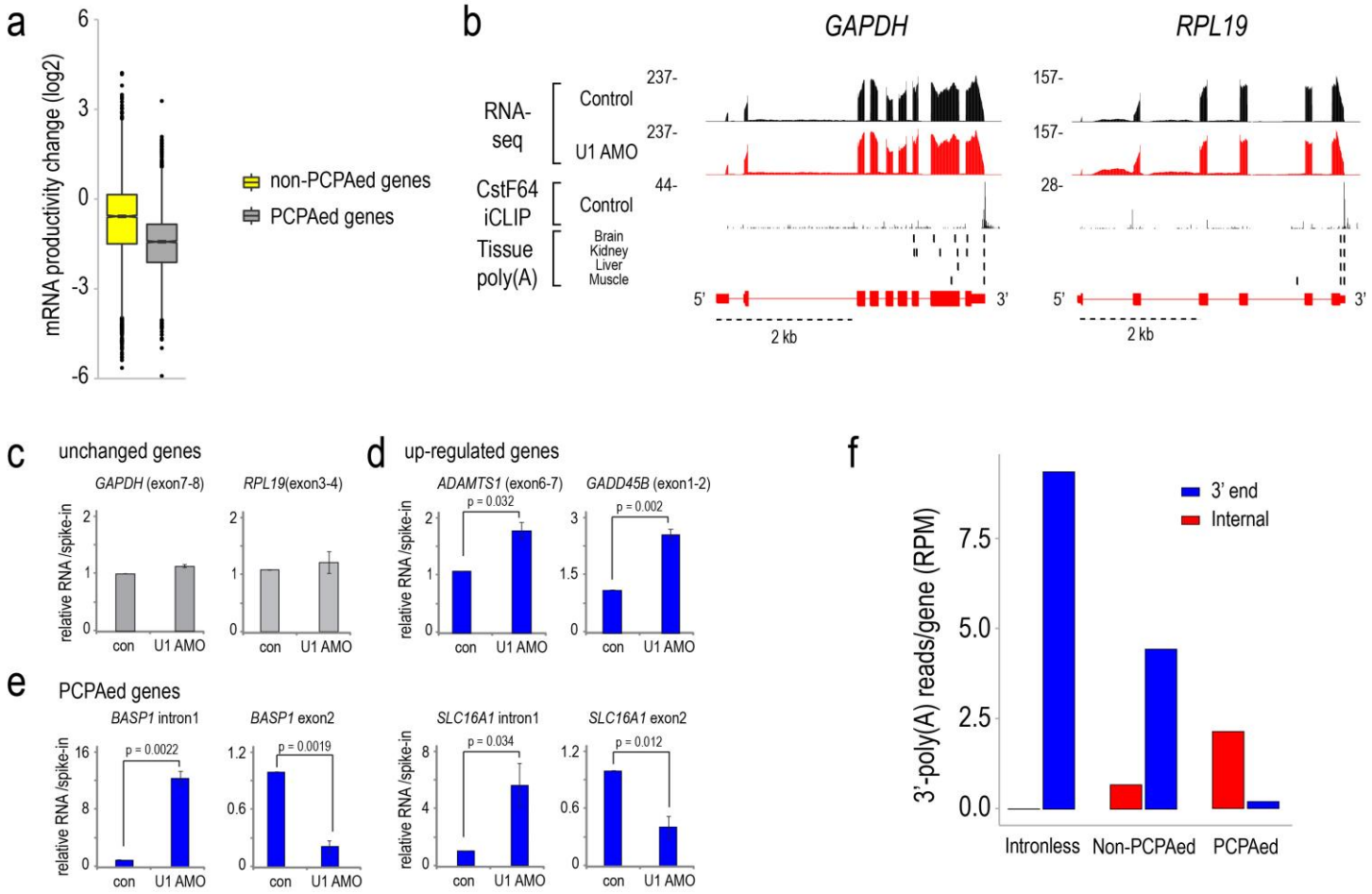
(a) Evidence for the high purity of 4-shU-labeled RNAs used for RNA-seq. HeLa cells transfected with control or U1 AMO either not labeled or metabolically labeled with 4-shU and RNA purified as described in Methods. RT-qPCR was used to quantify two regions in *PELO* (exon 2-3 junction) and *PRPF38B* (exon 1-2 junction). After 2 cycles of 4-shU-labeled RNA selection, significant enrichment of these regions were detected in the elution, in contrast to separate, unlabeled RNA samples where they were exclusively detected in the flow through (wash) (b). Data are represented as mean ± standard deviation (n=3, independent cell cultures). The high stringency procedure achieved very strong enrichment (80-200 fold) of 4-shU RNAs over unlabeled. (c) Pie charts representing the mapping distribution of 4-shU labeled RNA-seq reads distribution in CDS, 5' UTR, 3' UTR and introns of control and U1 AMO treated HeLa cells, as indicated. Histogram showing the (d) total number of junction reads as a percent of the total sequencing depth, and (e) percentage of these junction reads spanning canonical (previously annotated) exon-exon splice junctions (blue) versus aberrant (non-canonical or *de novo*) spliced reads (red) in control and U1 AMO samples. Shown are both 4 and 8 hours 4-shU labeled RNA-seq.



Supplementary Figure 2

U1 inhibition causes multiple, moderate PCPAs.

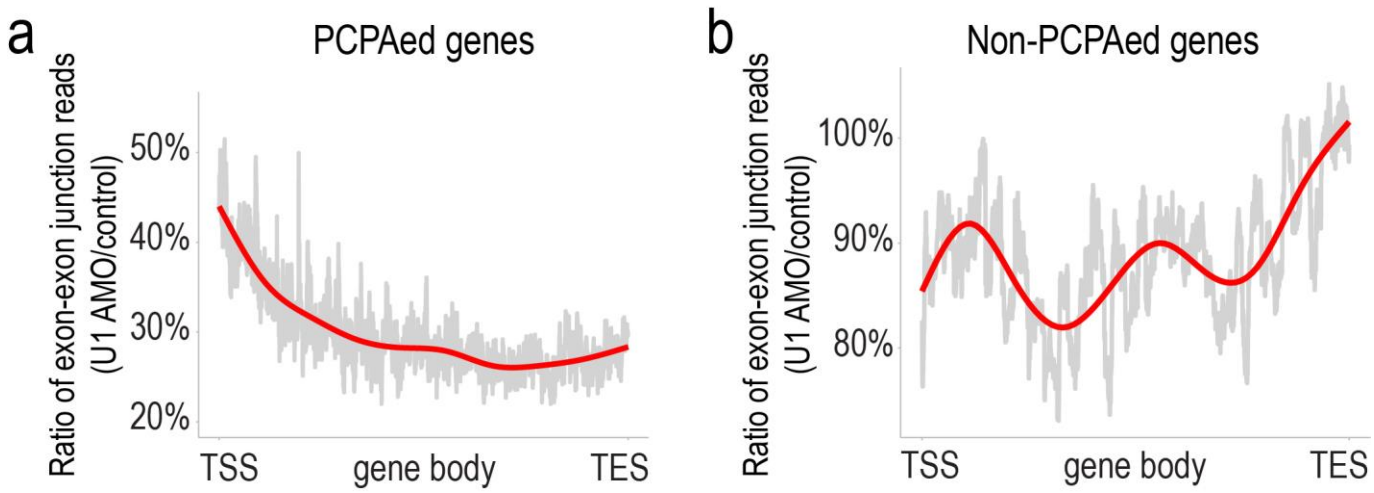
(a) PCPA validation by 3' RACE. Cells transfected with control or U1 AMO for 8 hours and metabolically labeled with 4-shU were used for analysis. After 2 cycles of 4-shU-labeled RNA selection, RNA was converted into cDNA and 3'RACE was performed as described previously[1]. Blue arrow indicates the forward primer location of 3' RACE. (b) Genome browser views of RNA-seq of representative genes showed multiple moderate PCPAs in several introns. Venn diagrams showing the overlaps between PCPAed genes detected in 4 and 8 hours post transfection (c), and between PCPAed genes and down-regulated genes 8h post transfection sample (d). (e) Gene size highly correlates with size of introns. Scatterplots showing the Spearman correlation between total gene size and total intron size in all expressed genes (RPKM ≥ 1) in HeLa cells.



Supplementary Figure 3

PCPAed genes are more down-regulated than non-PCPAed genes.

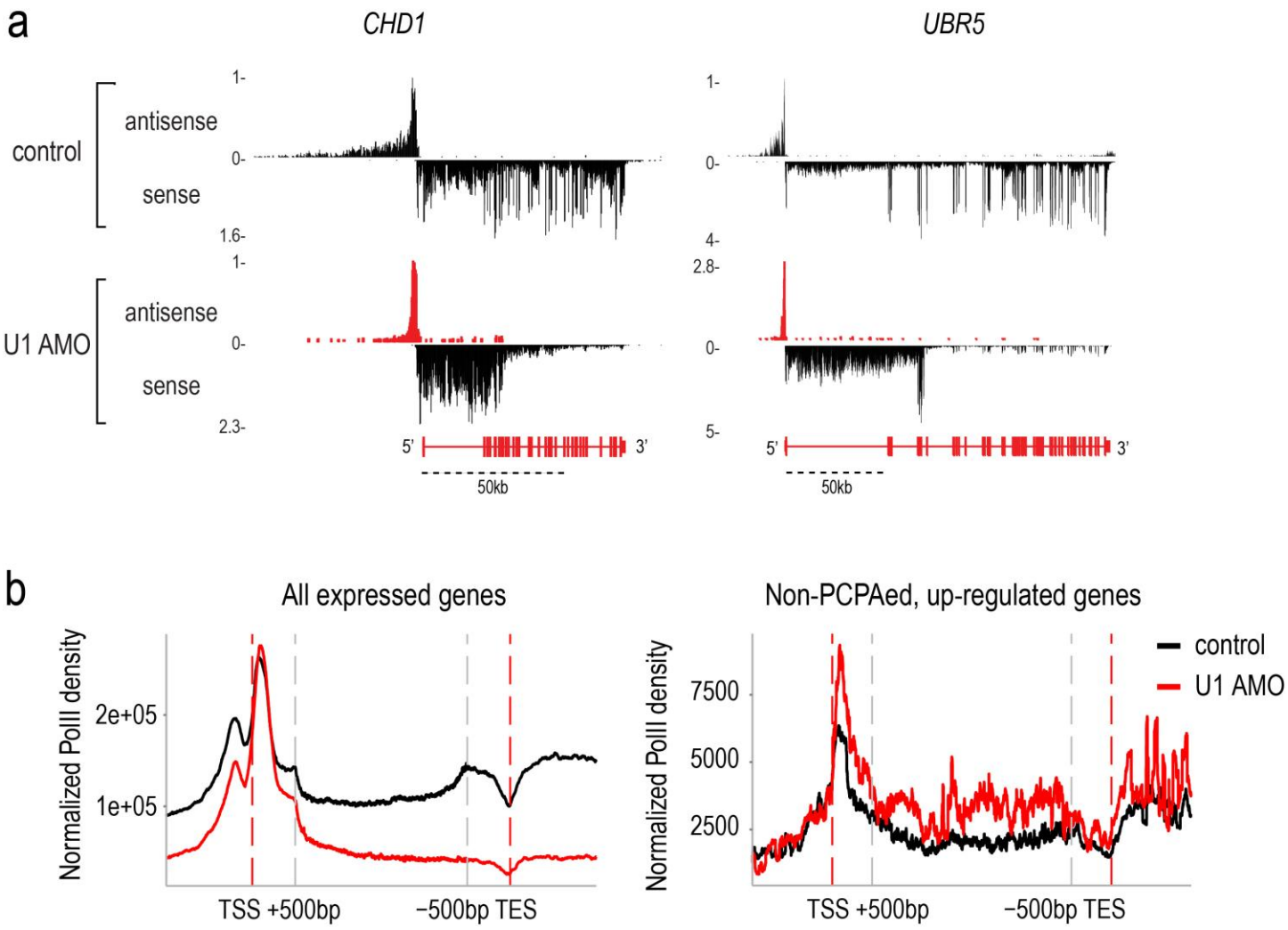
(a) Boxplots showing the gene expression changes in non-PCPAed genes (n=5,052) and PCPAed genes (n=3,590) in the 8 h 4-shU labeled RNA-seq from cells treated with U1 AMO. For boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5x IQR; points, outliers. Statistical tests used are described in Methods. (b) Genome browser view of Non-PCPAed and small genes with no expression change. (c-e) RT-qPCR confirms gene expression in RNA-seq. HeLa cells transfected with control or U1 AMO and metabolically labeled with 4-shU were used for RT-qPCR analysis. ERCC RNA spike-in controls were added to each sample before the rRNA depletion process and used for normalization. Data are represented as mean ± standard deviation (n=3, independent cell cultures). *P* value was calculated with two-tailed Student's t-test. A Poisson test measuring RNA-seq reads in exons normalized to the total mapped reads (*P* value < 0.01) confirmed the RT-qPCR results. (f) Intronless genes are PCPA resistant in U1 AMO. Histogram showing the 3'-poly(A) reads in gene body (internal) and 3' end region in intronless genes (n = 143), non-PCPAed genes (n = 3,254) and PCPAed genes (n = 2,692). For all genes expressed in HeLa cells (RPKM ≥ 1), only those with 3'-poly(A) reads in either their gene body or 3' end were selected for each group in this analysis.



Supplementary Figure 4

PCPAed genes lose more exon-exon junctions near the TES than non-PCPAed genes.

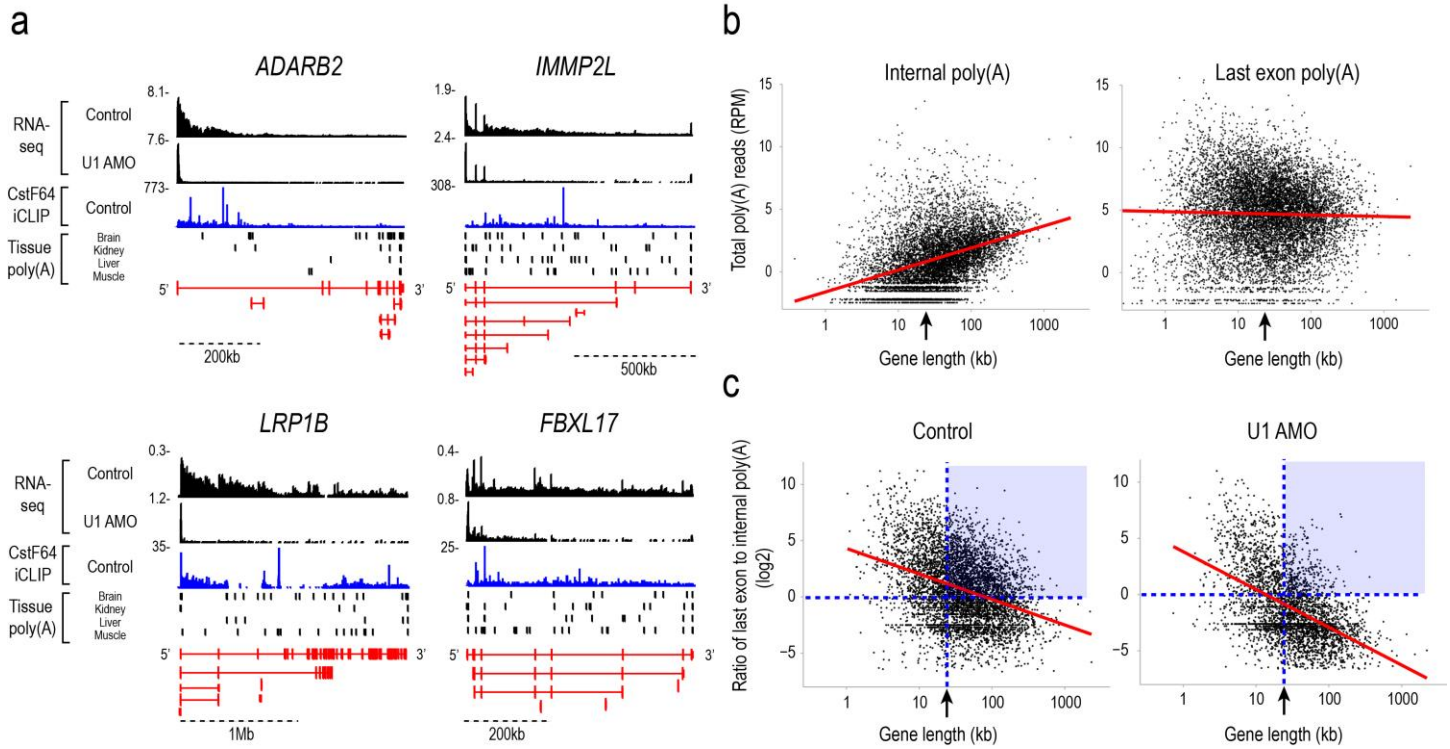
Metagene plot showing the ratio of exon-exon junction reads (U1 AMO/control, grey line) binned along the gene body, 5' to 3', in the PCPAed genes (a) and the non-PCPAed genes (b). All genes used for this analysis, from TSS to TES, were scaled to the same length (3 kb). The thick red line represents a smoothed fit line for each data point.



Supplementary Figure 5

Pol II metagenes of all expressed genes and non-PCPAed genes.

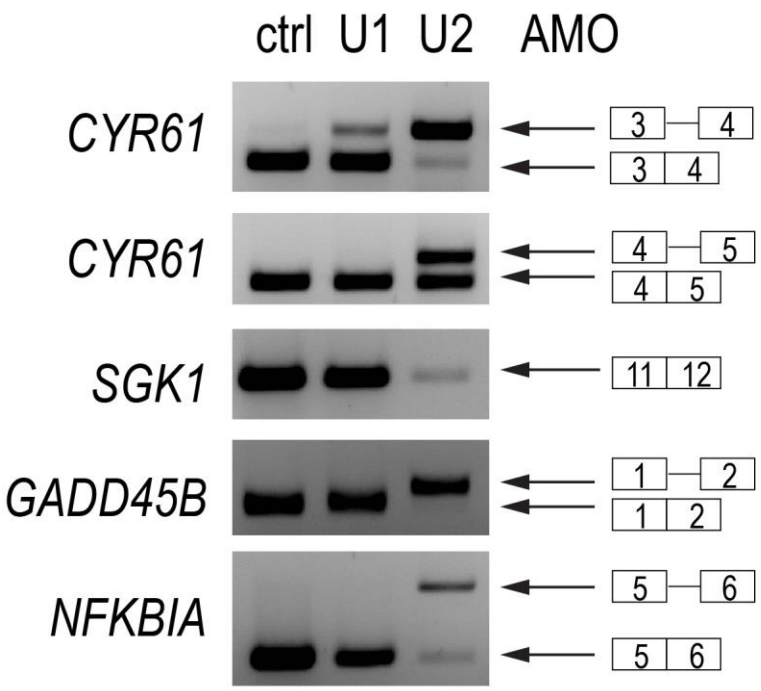
(a) U1 inhibition's effect on upstream, antisense transcription. Genome browser view of each gene and its upstream, antisense transcript showing PCPA in both directions by U1 AMO. (b) Metagene plot of pol II ChIP-seq reads for all expressed genes (n = 9,744) and highly up-regulated, non-PCPAed genes (n = 115) in control (black) or with U1 AMO (red), relative to TSS regions (TSS -1000 bp to +500 bp) and TES (500 bp upstream of the annotated mRNA 3' ends and 1000 bp downstream) in control and U1 AMO. Each gene's body, between TSS + 500 bp and TES - 500 bp, was scaled to 2 kb.



Supplementary Figure 6

U1 AMO increases transcription attrition in large genes.

(a) Transcription attrition naturally occurs in large genes. Genome browser views of RNA-seq of representative genes with transcription attrition. RefSeq gene structures along with any additional isoforms from AceView are shown underneath the panels (RefSeq is the top track). (b) Internal and last exon 3'-poly(A) reads distribution versus gene size. Scatter plot of each gene's 3'-ploy(A) reads in either the gene body (from TSS up to, but excluding, the last exon) or only the last exon in several human tissues[12]. Regression lines of internal poly(A) reads is dependent on gene size (left panel, $R^2 = 0.15$, P value < 0.05) while last exon poly(A) reads is not (right panel, $R^2 = 5e\text{-}4$, P value > 0.05). Arrow in x-axis represents the location of median gene length of all expressed genes, 22.8 kb. (c) Scatter plot of ratio of the total number of 3'-poly(A) reads found in the last exon compared to those in the gene body (from TSS up to, but excluding, the last exon) in control (left panel) and U1 AMO (right panel) RNA-seq. Arrow and vertical blue dashed line in x-axis represents the location of median gene length of all expressed genes, 22.8 kb. Fewer genes in the upper right blue colored zone represents increased transcription attrition (less full-length mRNA) in large genes by U1 AMO treatment.



Supplementary Figure 7

U2 AMO induces splicing inhibition.

Cells transfected with control, U1 or U2 AMO for 8 h and metabolically labeled with 4-shU were used for analysis. The decreases in spliced products with U2 AMOs show that splicing is dependent on the U2 snRNP.

	4-shU labeled RNA 4hr		4-shU labeled RNA 8hr	
	Control	15 nmole U1 AMO	Control	15 nmole U1 AMO
Total reads	75,253,970	142,789,530	235,555,670	125,807,542
Mapped reads	53,142,918	106,294,882	166,850,091	82,358,105
% Mapped reads	70.6%	74.4%	70.8%	65.5%
	Poly(A) RNA 4shU		pol II ChIP	
	Control	15 nmole U1 AMO	Control	15 nmole U1 AMO
Total reads	91,948,220	96,459,254	54,576,053	45,081,305
Mapped reads	82,197,810	85,491,282	52,673,238	42,746,731
% Mapped reads	89.4%	88.6%	96.5%	94.8%

Supplementary Table 1. Summary of RNA-seq data. RNA-seq and Pol II ChIP-seq samples were prepared as described in Methods. Total number of input reads, mapped reads and the percentage of mapped reads for each RNA-seq sample are presented.

	4hr 4-shU labeled U1 AMO		8hr 4-shU labeled U1 AMO	
	number of events	percentage (%)	number of events	percentage (%)
Alternative 5' Splice Site	145	0.28	243	0.47
Alternative 3' Splice Site	146	0.24	271	0.45
Alternative First Exon	977	1.20	1088	1.33
Alternative Last Exon	816	3.50	881	3.78
Mutually Exclusive Exon	39	0.24	60	0.37
Intron Retention	393	2.19	594	3.31
Exon Skipping or Inclusion	682	0.35	1436	0.73
Tandem UTR	39	0.73	68	1.28

Supplementary Table 3. Alternative splicing changes caused by U1 level change. The various classifications of splicing events were identified using the MISO software with the following parameters: $|\Delta\Psi| \geq 0.3$, Bayes factor ≥ 10, at least 1 inclusion reads, at least 1 exclusion read, and the sum of exclusion and inclusion reads ≥ 10.

Species	Exon number	Intron / exon length in PCPAed genes	Intron / exon length in non-PCPAed and up-regulated genes	PCPAed genes / non-PCPAed genes
Fly	2	0.059	0.162	0.365
Pufferfish	2	0.573	0.419	1.368
Chicken	2	1.707	0.660	2.585
Mouse	2	3.939	0.843	4.674
Human	2	6.944	1.321	5.257
Fly	3	0.135	0.427	0.317
Pufferfish	3	1.121	1.370	0.818
Chicken	3	2.572	1.799	1.430
Mouse	3	7.207	2.646	2.724
Human	3	11.061	1.993	5.550
Fly	4	0.328	0.858	0.383
Pufferfish	4	1.569	1.657	0.947
Chicken	4	2.776	2.252	1.233
Mouse	4	8.123	5.413	1.501
Human	4	15.611	6.208	2.515
Fly	5	0.457	1.077	0.425
Pufferfish	5	1.789	2.018	0.886
Chicken	5	3.661	2.286	1.602
Mouse	5	7.694	3.526	2.182
Human	5	9.110	4.260	2.138
Fly	6	0.592	0.551	1.074
Pufferfish	6	2.183	2.411	0.905
Chicken	6	5.770	2.993	1.928
Mouse	6	10.630	2.401	4.428
Human	6	19.911	3.593	5.541

Supplementary Table 6. Introns expanded more in PCPAed genes than non-PCPAed and up-regulated genes in vertebrate evolution. Orthologous human genes for each species were downloaded from Ensembl[33], and the intron expansion (intron/ exon ratio) was compared between PCPAed and non-PCPAed, up-regulated genes that have same number of exons. For every exon number tested, 2 through 6, there was a progressive and much greater intron expansion in PCPAed genes.

	Forward primer (5' to 3')	Reverse primer (5' to 3')
PELO	GAC AAC AAA CTG CTC CTG GAA AA	TGT CCG GAG GAG GCA TGT AC
PRP38B	CGA GGT GGT GGA CGA GAT CTA	TGC TTC CTT TCT CCC ATG GT
ADAMTS1 exon6-7	AAA CGA GTG CGC TAC AGA TCC T	TGT TCC TCT CTA AAG GTT TTT CCA TTA TT
CYR61 exon4-5	CCT TGT GGA CAG CCA GTG TAC A	GGA TTT CTT GGT CTT GCT GCA T
GADD45B exon1-2	TGG AGC CGC ATC CAC TGT	GAG CGG CCA CCA AAA GCT
CYR61 ex3-4	AGT GGG TCT GTG ACG AGG ATA GTA	ACC AGT TCC ACA GGT CTT TGA GCA
CYR61 ex4-5	TGA GTG CCG CCT TGT GAA AGA AAC	AAC AGC CTG TAG AAG GGA AAC GCT
SGK1	AGC CTT ATG ACA GGA CTG TGG ACT	AGC CGC TTT GTC CTG TCC TTC T
GADD45B	GTT GGT TTC CGC AAC TTC CTG GAT	CAT TCA TCA ACT TGG CCG ACT CGT
NFKBIA	ACC TGG TGT CAC TCC TGT TGA AGT	TGG CCT CCA AAC ACA CAG TCA TCA
MYC exon1-2	GCT GCT TAG ACG CTG GAT TT	GAG TCG TAG TCG AGG TCA TAG TT
GAPDH exon7-8	ATG ACA ACT TTG GTA TCG TGG A	GGC CAT CCA CAG TCT TCT G
RPL19 exon3-4	GGC ACA TGG GCA TAG GTA A	GAT TCA CGG TAT CTT CTG AGC A
BASP1 intron1	GCG TGG CAG CAA TGG AA	AAA CGT GGT CAA TAT CAA GTT CTC AT
BASP1 exon2	CGC AGC CTC TGC AGA AGA G	TCT TTCACG GTT ACG GTT TGG
SLC16A1 intron1	GGG CTG CTC AGC TTG TAA GG	GCC AAC AAA AAC TGA AAA CAA CAG
SLC16A1 exon2	TAC ACC CCC CCA GAT GGA	AAG CCG ATG GAA ATG AAA GCT
ERCC 003	CTG GTC TTC CAT CTG GTG TTT	GCT ACA AAG AAG TGG GCT AAG A
ERCC 017	GCT AGC ACT CCC TGG AAT ATC	GGC TTA CTT CCC TCT GAT TGT
ERCC 025	TGG ATA GGA GCG ACC GAT TA	GCT TCC ACG GTA GAC CAT AAC
ERCC 034	ATG TCT GCG GAC TAT TGG TAA G	ACA CGG TCG TTG GGT TAT TC
ERCC 062	CCT GAG GTT GGA GAG ACA TTA T	CCT GCA CAG TGA GTT TCT TTC

Supplementary Table 7. Primers used for Real time RT-PCRs and RT-PCR to validate 4-shU selection specificity (Supplementary Fig. 1), small gene increase (Fig. 3 and Supplementary Fig. 3) and splicing (Supplementary Fig. 7).