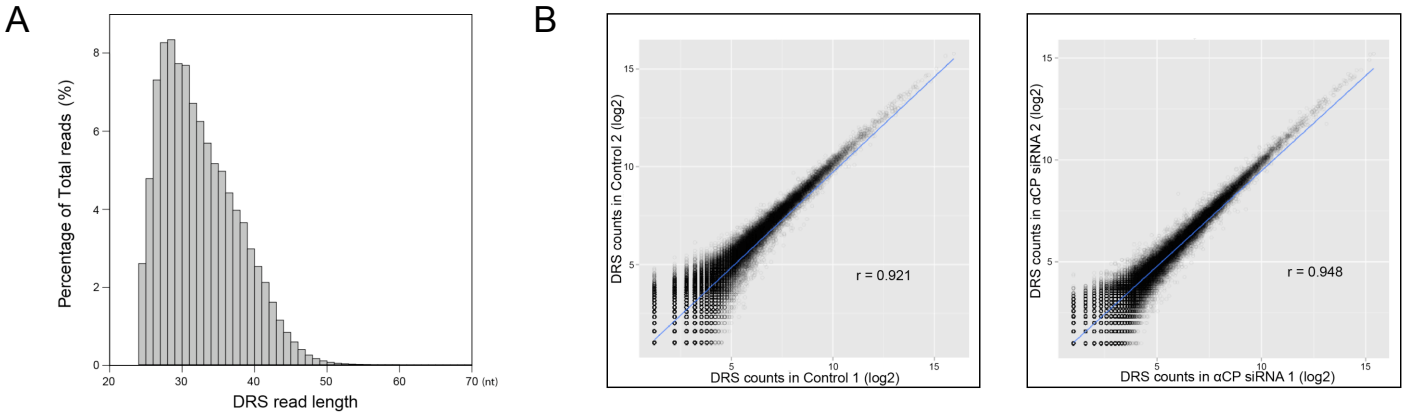


Figure S1.

A: Histogram of the DRS read lengths of all the four samples (pooled). Based on the high level of data reproducibility, siRNA control data sets from the two control samples and αCP siRNA data set from the two knockdown samples were pooled together respectively and used for this analysis.

B: Reproducibility of DRS reads. DRS read counts, normalized to the minimum number of the total non-internal priming reads of two DRS samples, are graphed in log2 scale. The DRS data were highly reproducible with Pearson correlation coefficient higher than 0.92 and 0.94 for two siRNA control samples and two αCP siRNA samples, respectively

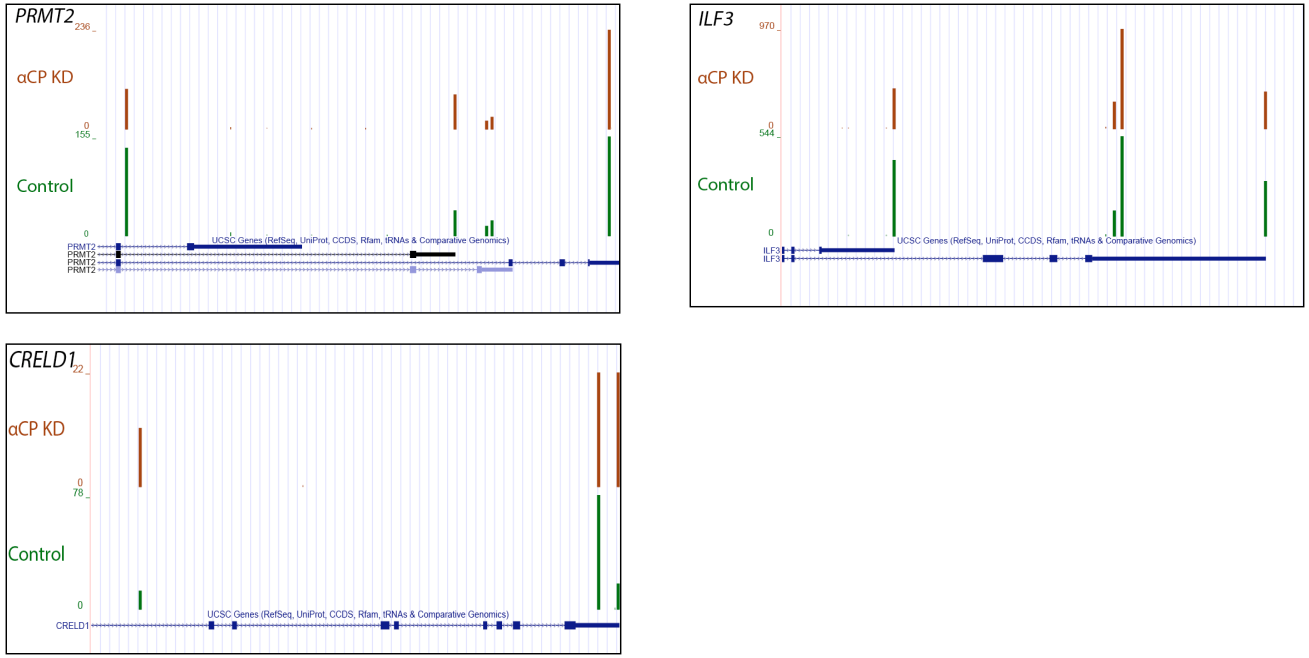
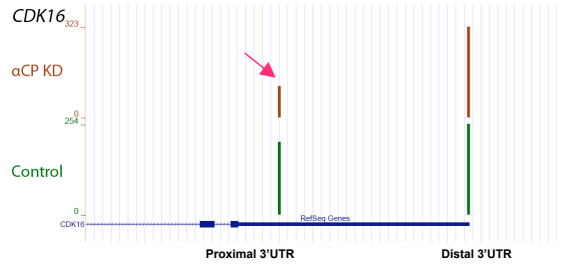


Figure S2. Examples of complex APA pathways impacted by depletion of αCP

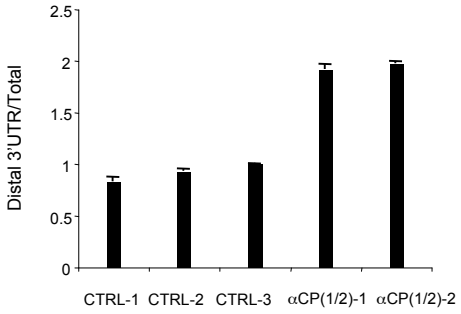
3 transcripts were shown here as examples of the complex APA. On these transcripts, the alternative polyadenylation events occur both on introns and exons and can not be catogarized as SE-APA or DE-APA.

Fig.S3

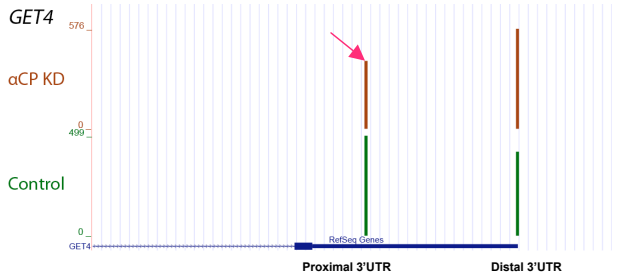
A



Proximal: ATTGGCCTGTCAACCCACCCATTGGCCTGTCTGCTGGGTGCTAACAAAGC
Distal: GGGGGTTGGGTTCCTGAATGCACCATAATCGCTGTATGAAATATTAAAAA



B



Proximal: TGCTGGCCGCCGCGTCCCCCGAGATTGACCCACAATAAAGCACAGGCCTTACCG
Distal: ATTTCAGGAAGGCTTGTGTGAACCGTTGCGCATAAATAAACCCTTTCTACCGGAAA

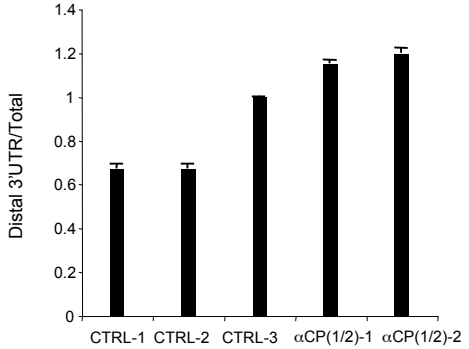
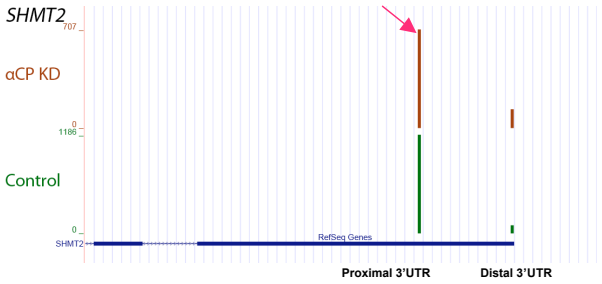
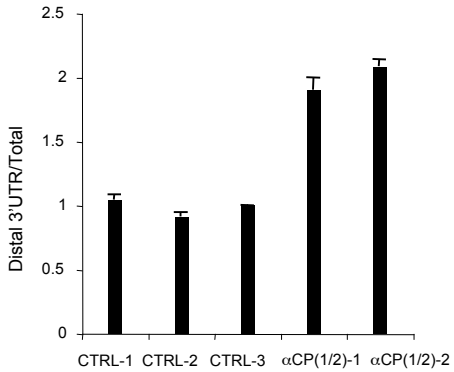


Fig.S3 (cont)

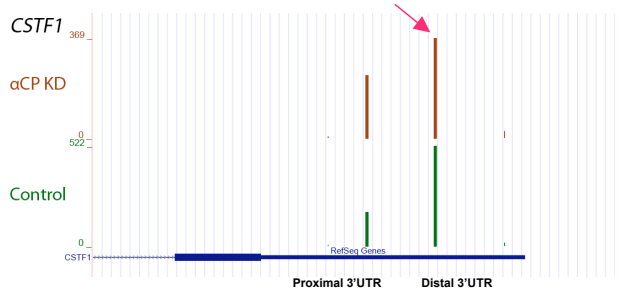
C



Proximal: GCAGGCAGGGGTGGGTAGGCACCCTCCTTCCTGTTTTTATCTAATAAAATGCTAA
Distal: ACACCAGACGTGATTAGCAGGCGCAGCAAATTCAATTTGTTAAATGAAATTGTAT



D



Proximal: TAAGTACAGGACTTGCCGTTTCTTTTGATCTCTTGATTGAAGGAGGATAGGGCATTAAAGT
Distal: CAAAAGCATTTTCAAGAATCCCTCAATTCTATCCTGAAATGATGTTATTCTGATAATAAAGC

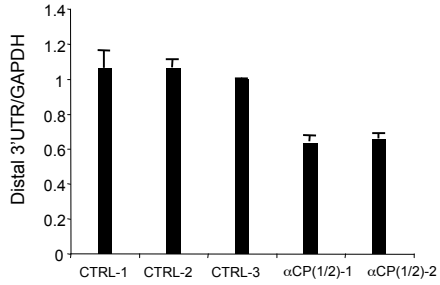
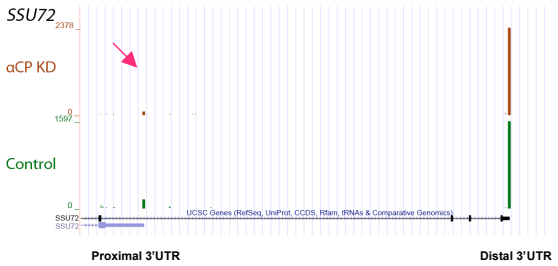
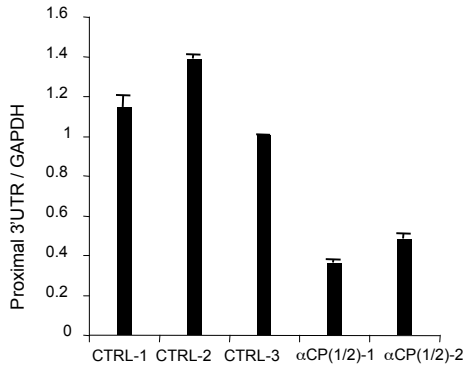


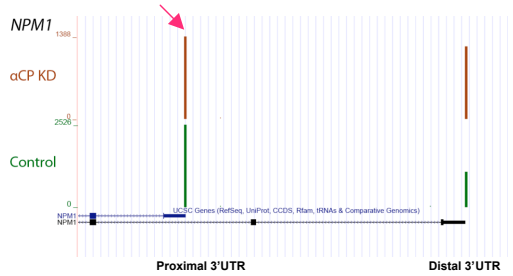
Fig.S3 (cont)

E





F



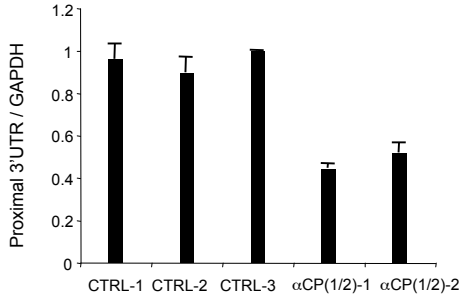


Figure S3. QPCR validations of a spectrum of APA events triggered by αCP depletion.

A subset of APA events identified in the DRS analysis of αCP depleted cells was independently assessed by targeted RT-QPCR. The diagram at the top of each set of studies represents the DRS data set in the context of the genome browser diagram of the respective locus. The positions of the proximal and the distal 3' UTR are marked below the browser view; the sequences encompassing and 5' to each of the competing APA sites are shown for **A-D**; the C-rich sequences (putative αCP binding sites) and the alternative polyA signals (AAUAAA) are highlighted in red. The data in red corresponds to number of sequence counts in the αCP-depleted sample and the data in green is from sequence counts from control siRNA sample. The red arrow in each display indicates the position of polyadenylation site whose activity is decreased secondary to αCP depletion. Below each figure is a histogram that summarizes the same set of RNAs assayed by QPCR. Primers were designed to quantify the RNA levels at the impacted APA sites. Frames **A-C** show three transcripts in which a C-rich motif precedes the proximal polyA site, resulting in a decrease in relative usage of the proximal polyA site subsequent to αCP depletion (αCP KD) when compared to the that in cells treated with control siRNA (Control); in each case these data documents an elevated ratio of the distal 3'UTR isoform (ie., use of the distal polyA site) relative to the total polyA site usage. **D** shows a transcript with a C-rich motif preceding the more distal alternative polyA sites; in this case the depletion of αCP results in decreased usage of the distal polyA site. Frames **E-F** are examples of APA involving PA sites located in different exons (DE-APA). In these two cases, the decreased usage of the proximal polyA site was confirmed and presented in the individual histogram as the ratio of the proximal 3'UTR isoform relative to control GAPDH mRNA. All the real-time RT-PCR quantifications were normalized to GAPDH mRNA and presented as a ratio versus the CTRL-3 (cyclophilin siRNA) defined as 1. Standard deviation for each sample is shown (n=3).

A-D: Examples of APA involving competing PA sites within the same terminal exon (SE-APA).

A. Cyclin-dependent kinase 16 (CDK16) gene

B. Golgi to ER traffic protein 4 homolog (GET4) gene

C. Serine hydroxymethyltransferase 2 (SHMT2) gene

D. Cleavage stimulation factor (CSTF1) gene (also known as CSTF50).

E-F: Examples of APA involving PA sites located in different exons (DE-APA).

E, SSU72 RNA polymerase II CTD phosphatase homolog (Ssu72) gene

F, Nucleophosmin (NPM1)gene.

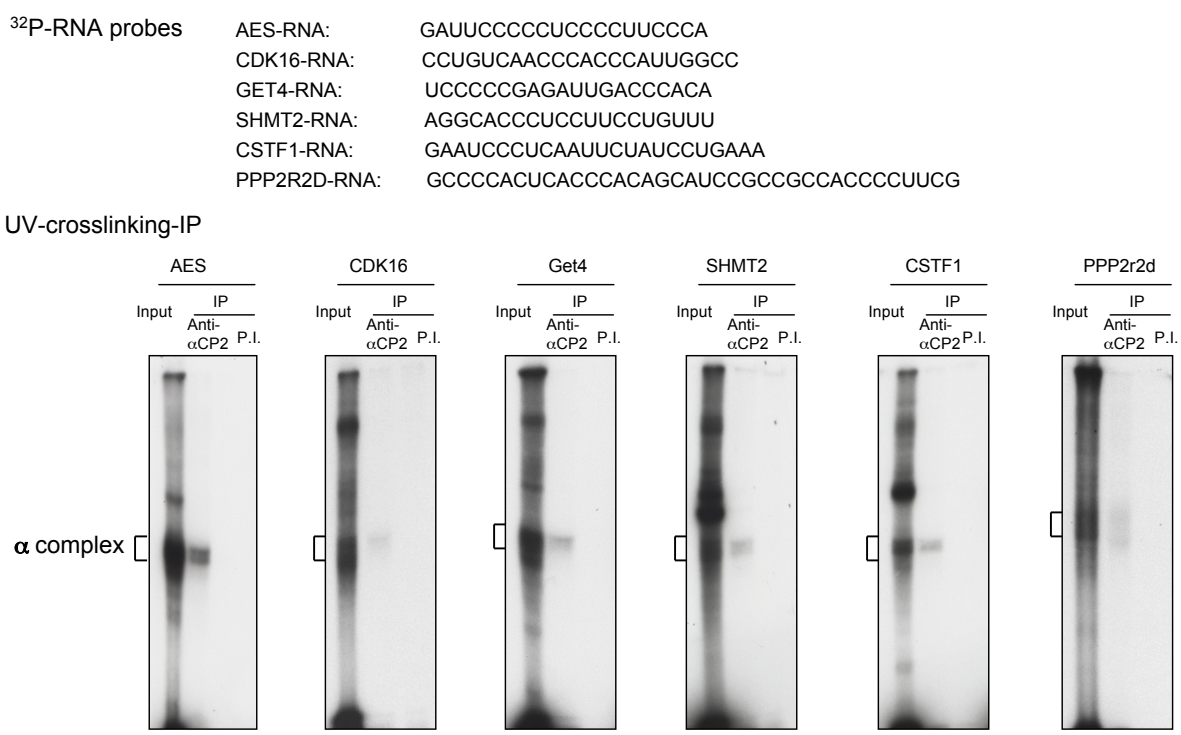


Fig. S4 *In vitro* RNA-protein interaction assay. The 32P-labeled RNA probes corresponding to each C-rich segments in the 3'UTRs of the noted genes (sequence shown at the top of the figure) was incubated with HeLa cell nuclear extract and subjected to UV-cross-linking and immunoprecipiation with affinity purified antibody to αCP2 or preimmune (PI) serum. Each ^{32}P-labeled RNA substrate was incubated with HeLa cell nuclear extract under cleavage conditions, UV cross-linked, RNase digested, and the residual labeled complex was visualized on a denaturing SDS-PAGE gel. The identity of the α-complex was confirmed by immunoprecipitation of the cross-linked products with antibody specific to αCP2/KL (the bracket).
The designations above each lane indicate the identity of the 3' UTR sequence used in the reaction.

Table S1. Processing of direct RNA sequencing (DRS) data sets

RNA Samples	Control siRNA -1	Control siRNA -2	αCP siRNA-1	αCP siRNA-2
Sequenced reads	9.9 million	18.5 million	17.4 million	16.4 million
Uniquely mapped reads	3,056,826	7,069,003	6,399,451	5,863,787
PolyA reads after removing internal priming (Non-internal priming reads)	2,811,092	6,557,478	5,919,346	5,420,574
Percentage of non-internal priming reads	28.4%	35.5%	34.0%	33.0%
Non-internal priming reads overlapping known annotation	1,564,685	4,027,412	3,480,473	3,185,739
Overlapping percentage of DRS reads to known gene/polyA_DB2	55.7%	61.4%	58.8%	58.8%

Table S2. Summary of confirmation studies by RT-PCR

Gene symbol	Fold-change KD/Control by DRS	P-value	FDR
HBZ	3.43	0	0
αCP1	0.13	0	0
αCP2	0.38	0	0
DDIT4	6.24	6.71E-287	1.96E-283
ACADVL	0.21	6.43E-78	2.77E-75
PHGDH	4.37	1.39E-62	4.86E-60
ACSM3	0.45	1.47E-58	4.20E-56
PRG2	0.26	3.35E-42	5.33E-40
ALAS2	4.13	2.27E-28	2.28E-26
CFLAR	0.36	4.25E-27	4.07E-25
SLC7A11	3	4.36E-27	4.15E-25
CHAC1	8.21	3.08E-26	2.77E-24
COMTD1	0.42	7.89E-19	4.51E-17
SCO2	0.46	8.19E-13	2.77E-11
FAHD2B	0.31	9.77E-12	2.99E-10

mRNAs with significant alterations in steady state levels subsequent to the αCP depletion were validated by real time RT-PCR assays. The gene symbol is indicated in the first column and the fold change in overall mRNA representation (the ratio of αCP depletion samples/ control samples) is shown in the second column, while the last two columns show the P-values and false discovery rates (FDR) for each of the data shown.

Table S3. Primers for DEG and APA studies

Primers for DEG studies			
Name	Sequence	Name	Sequence
ACADVL-F	TAGGAGAGGCAGGCAAACAGCT	ACADVL-R	CACAGTGGCAAACTGCTCCAGA
CHAC1-F	GTGGTGACGCTCCTTGAAGATC	CHAC1-R	GAAGGTGACCTCCTTGGTATCG
FAHD2B-F	TGGTGACCAAGGACAGTGTAGC	FAHD2B-R	AGGCTATCAGGTCCTCTGTCTTG
PHGDH-F	CTTACCAGTGCCTTCTCTCCAC	PHGDH-R	GCTTAGGCAGTTCCCAGCATTC
PRG2-F	CCTGGTGAGAAGTCTTCAGACG	PRG2-R	CGCTGACAGAACACTGGATTCG
DDIT4-F	GTTTGACCGCTCCACGAGCCT	DDIT4-R	GCACACAAGTGTTCATCCTCAGG
Primers for APA studies			
Name	Sequence	Name	Sequence
SSU72-3:	TCCAGGACCAGCTCCCGACAA	SSU72-4:	TGCACTGCACAGGGTGACACG
NPM1-V3-1:	TGCCACCCATGCCTCTTCAGG	NPM1-V3-2:	TGTCAGGTGAGGCAAATGCACA
AES-1:	CCCCACTCACCACCTCTGC	AES-2:	TCACCCTGCTCCTCCCAGC
AES-3:	AAAAGCAACACCCACACCTG	AES-4:	AAGCCCACCCTTGATACAGC
CDK16-1:	TCACAGGGCAGCCCCCAACT	CDK16-2:	GGCAGGCAGCAGGGGACAAG
CDK16-3:	CCGCCATCCCCAGTTGCAGG	CDK16-4:	GGACTGGGCTGGGGAGACCC
Get4-1:	TTCCTACGGGGGCCTGCTCG	GET4-2:	AGGAGCCACCCAGGACTCGC
GET4-3:	TCTGCCCGTCTCTTTCATAACG	GET4-4:	AGCGTCTCCCCTTTTGTCATCG
SHMT2-1:	TTGGTGCGGGAGGGAAGACCT	SHMT2-2:	AGGGTGCCTACCCACCCCTG
SHMT2-3:	TGTACACCGCTCCGCTCCCA	SHMT2-4:	TGCTGCGCCTGCTAATCACG
CSTF1-1:	GCGCGGTTTTGGTACCGGAGAT	CSTF1-2:	GCAGCTGAGACAGGACGTGGG
CSTF1-3:	TGCTGAACCCTCTGTGTATTTCCCT	CSTF1-4:	AGTCTCAGAGCTGCATGCCTGCTTA
PPP2r2d-1:	AGACGCGAACGTGAGGACCA	PPP2r2d-2:	GCGGAGGCGGCTTTCTCCTG
PPP2r2d-3:	CCGTGTCCTCTCGGCCTTCCT	PPP2r2d-4:	GCGGATGCTGTGGGTGAGTGG
PCBP2-F:	ATCTGCTGCCAGCATTAGCCTG	PCBP2-R:	GGTGGTGAACAGCAGAAAGGGA
PCBP2-a:	CTCCAGCCTCCCCAACCCTCTT	PCBP2-b:	TAGCTGAAGCCCCGGTACCCTT
PCBP2-c:	GGCCATTTCAGCTGAGTGCCTGTT	PCBP2-d:	GCCAAGAAGGCCTGCTGATAGCTT

In the DGE study primers set, F and R are forward and reverse primers for the corresponding genes respectively. For APA primers, amplicons of primer pairs 1 and 2 represent the total (proximal and distal) polyA usage, and amplicons of primer pairs 3 and 4 represent distal polyA usage. For Pcbp2 gene, the location of primers of F, R, a, b, c and d are as shown in **Fig. 6C.**