Supplemental material for:

CPSF3 inhibition blocks pancreatic cancer cell proliferation through disruption of core histone mRNA processing

Abdulrahman. A. Alahmari, Aditi H. Chaubey, Venkata S. Jonnakuti, Arwen A. Tisdale, Carla D. Schwarz, Abigail C. Cornwell, Kathryn E. Maraszek, Emily J. Paterson, Minsuh Kim, Swati Venkat, Eduardo Cortes Gomez, Jianmin Wang, Katerina V. Gurova, Hari Krishna Yalamanchili, and Michael E. Feigin*

*Corresponding author:

Email: michael.feigin@roswellpark.org

This PDF file includes:

- Supplemental Figures S1 to S13
- Supplemental Figure Legends S1 to S13
- Materials and Methods
- Tables S1
- Tables S2
- References

Supplemental Fig S1

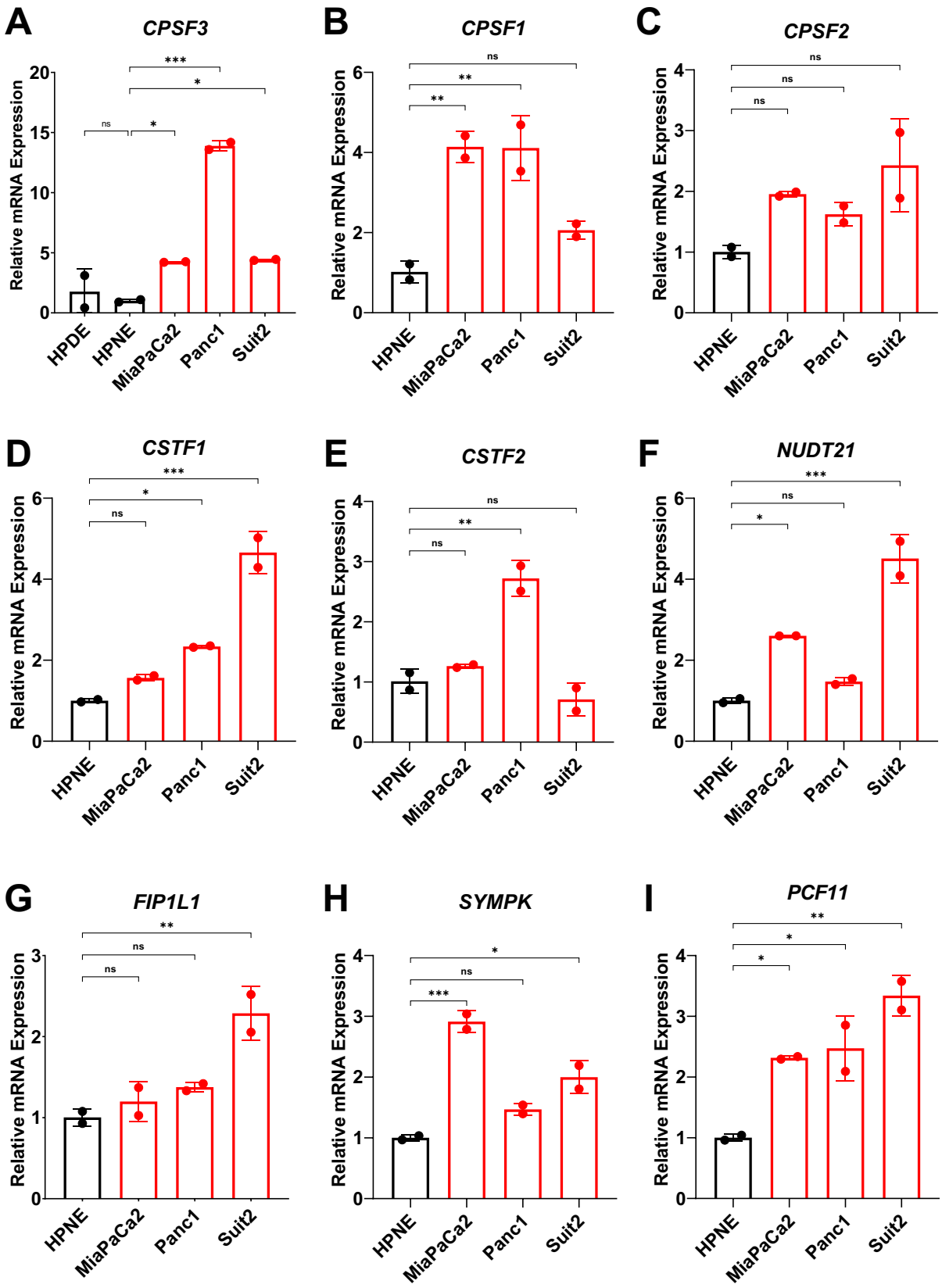


Fig. S1. Multiple CPA factors are upregulated in PDAC cells.
(A-I) Quantitative RT-PCR showing mRNA expression levels of multiple CPA factors in non-transformed pancreatic epithelial (black) and PDAC (red) cells. Data are shown as mean ± SD. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Ordinary one-way ANOVA with Dunnett's multiple comparisons test.

Supplemental Fig S2

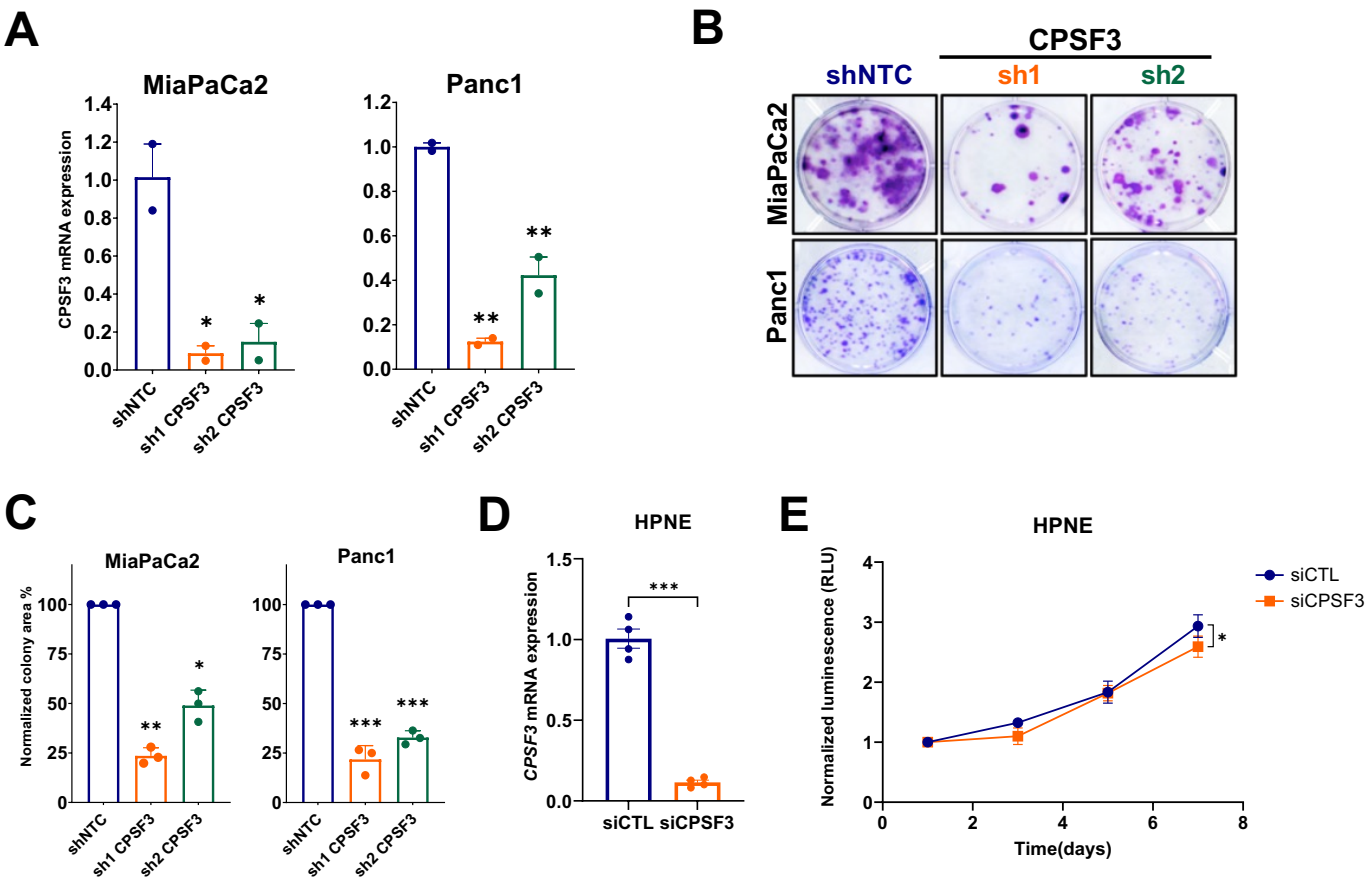


Fig. S2. PDAC cells, but not immortalized control cells, are affected by CPSF3 knockdown.

(A) mRNA expression of *CPSF3* in shNTC, sh1 CPSF3 and sh2 CPSF3 knockdown cells by qPCR. Graphs are representative of at least two independent experiments. Data are shown as mean ± SEM of technical duplicates. *****, $P < 0.01$, Ordinary one-way ANOVA with Dunnett's multiple comparisons test. **(B)** Clonogenic growth assay of shNTC (blue), sh1 (orange) and sh2 (green) CPSF3 knockdown cells. **(C)** Normalized colony area percentage of shNTC, sh1 and sh2 CPSF3 knockdown cells from **(B)**. *, $P < 0.05$; *****, $P < 0.01$; ******, $P < 0.001$; Ordinary one-way ANOVA with Dunnett's multiple comparisons test. **(D)** Quantitative RT-PCR showing *CPSF3* mRNA expression levels in HPNE cells upon *CPSF3* transient knockdown by siRNA. Data are shown as mean ± SEM. **, $P < 0.001$, unpaired t test. **(E)** Proliferation rates at days 0, 2, 4 and 6 of immortalized control HPNE cells upon *CPSF3* knockdown. Data are shown as mean ± SD. *, $P < 0.05$; 2way ANOVA with Sidak's multiple comparisons test.

Supplemental Fig S3

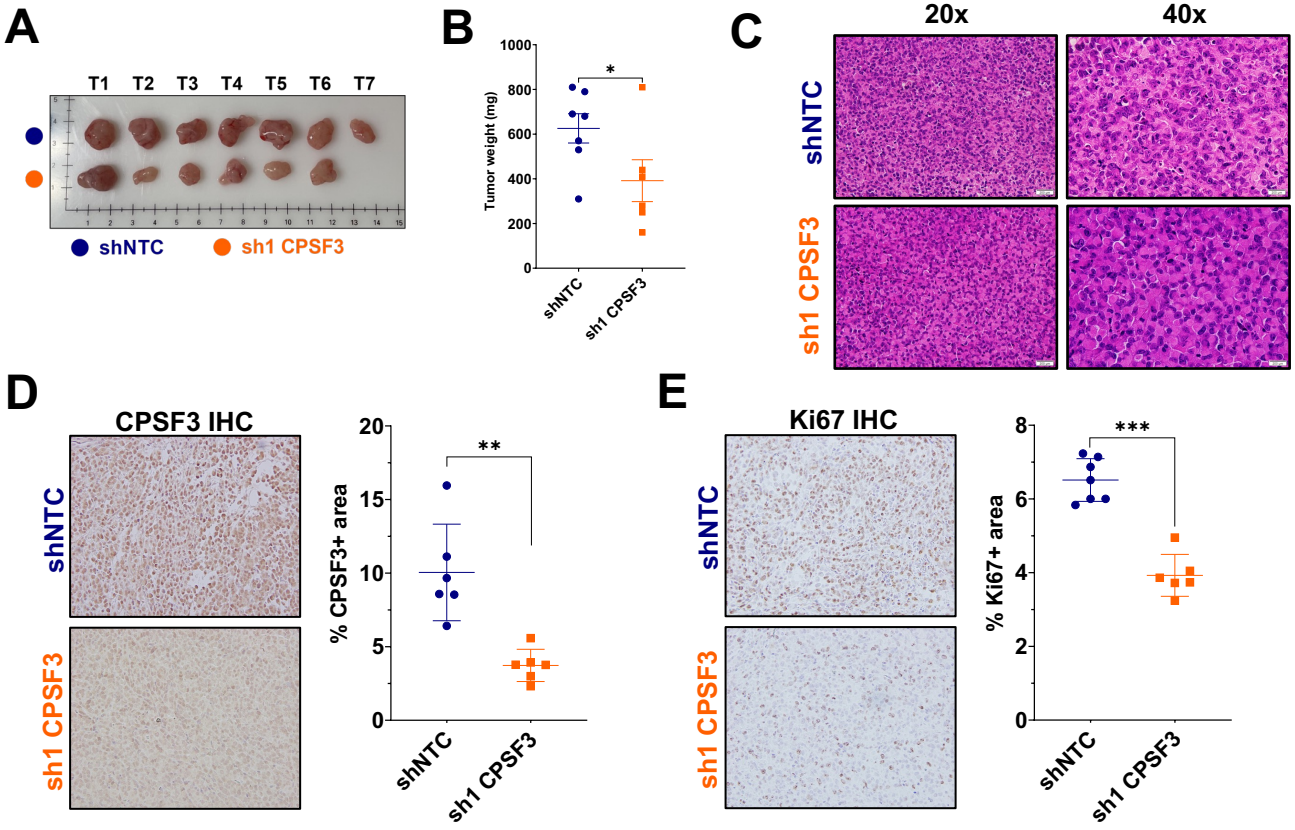


Fig. S3. CPSF3 knockdown decreases proliferation in vivo.
(A) Gross images of shNTC (n=7, blue) and shCPSF3 (n=6, orange) dissected tumors. **(B)** Endpoint tumor weight. *, $P < 0.05$, unpaired t test with Welch's correction. **(C)** Hematoxylin and Eosin (H&E) staining of shNTC (left) and sh1 CPSF3 (right) xenograft tumors (MiaPaCa2 cells). **(D, E)** Representative IHC for CPSF3 and Ki67, respectively. Box and whisker plots indicate the percentage of CPSF3- and Ki67-positive areas in the tumors. *****, $P < 0.01$; ******, $P < 0.001$; unpaired t test.

Supplemental Fig S4

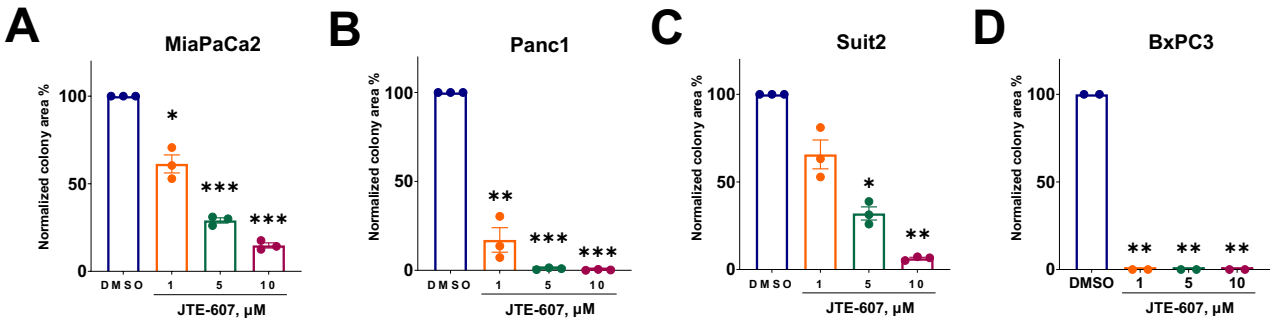


Fig. S4. JTE-607 decreases colony formation of PDAC cells.
(A-D) Normalized colony area percentage of PDAC cell lines from **Figure 2F**. *, $P < 0.01$; *****, $P < 0.001$; ******, $P < 0.0001$; Ordinary one-way ANOVA with Dunnett's multiple comparisons test. Data are shown as mean ± SEM.

Alahmari, AA; et al.

Supplemental Fig S5

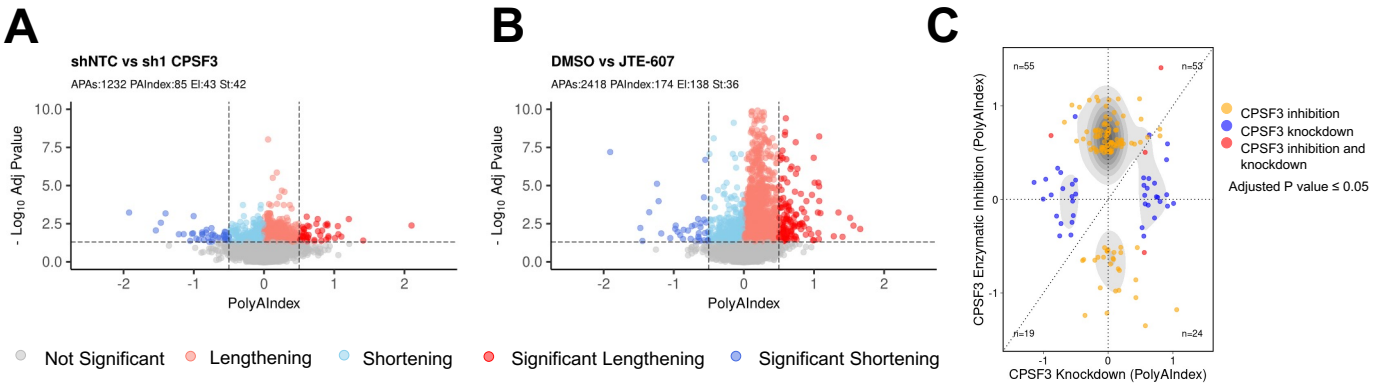
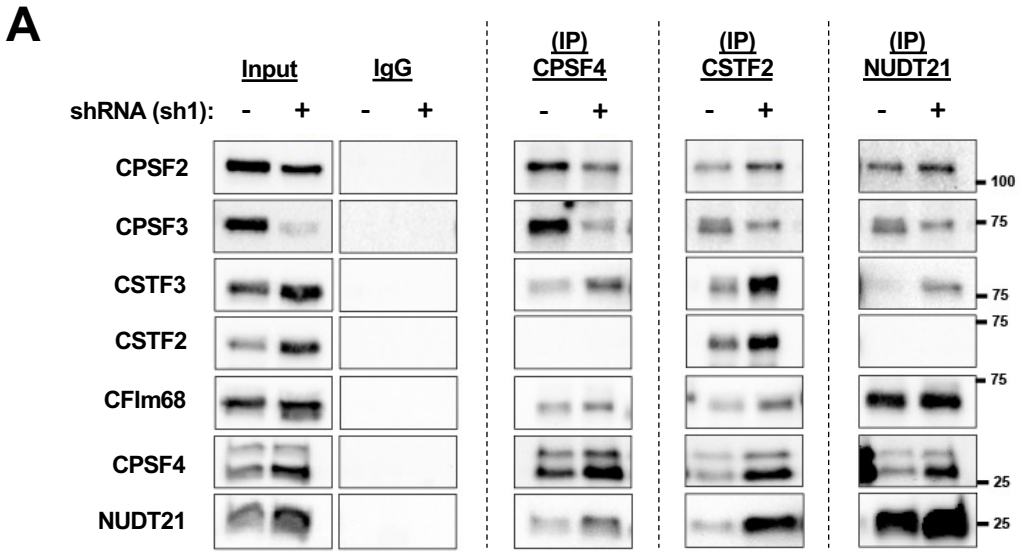
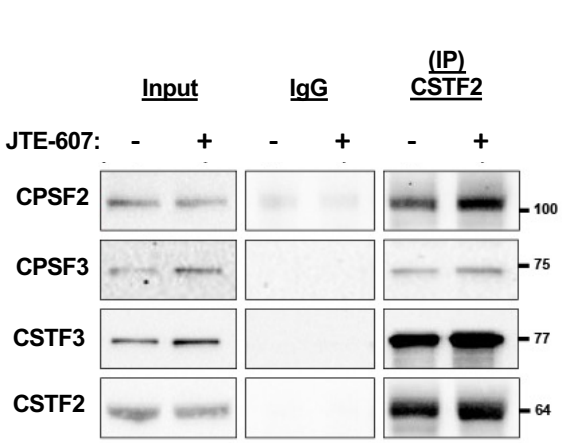


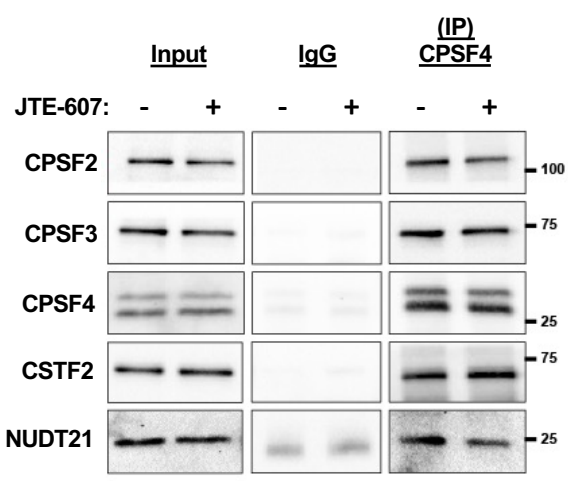
Fig. S5. mRNA 3'-end processing is distinct between CPSF3 knockdown and inhibition.

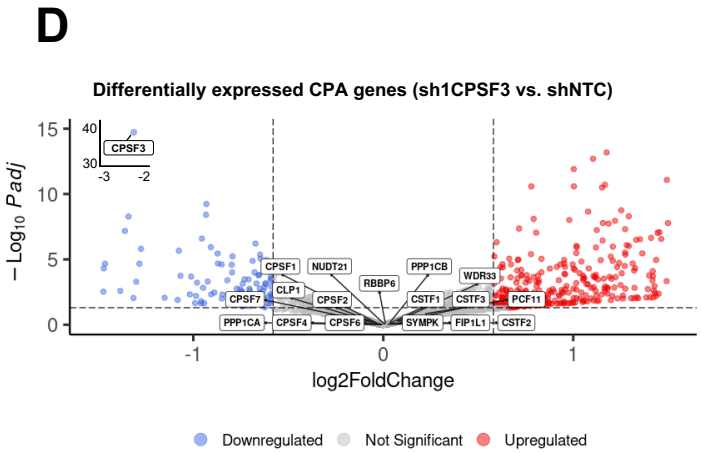
(A, B) Volcano plots showing differential CPA genes (-0.5 > PolyAIndex > 0.5; P<0.05). Blue dots indicate 3′-UTR shortening while red dots indicate 3′-UTR lengthening. Significant shortening genes = 42 and 36 upon *CPSF3* knockdown and inhibition, respectively. Significant lengthening genes = 43 and 138 upon *CPSF3* knockdown and inhibition, respectively. **(C)** Dot plot with quadrants representing APA differences in both *CPSF3* knockdown (72hr post transduction) and inhibition (24hr post JTE-607 treatment). Only 2 genes are uniquely altered in both conditions (Red dots, top-right quadrant); DAGs = differential APA genes.

Supplemental Fig S6









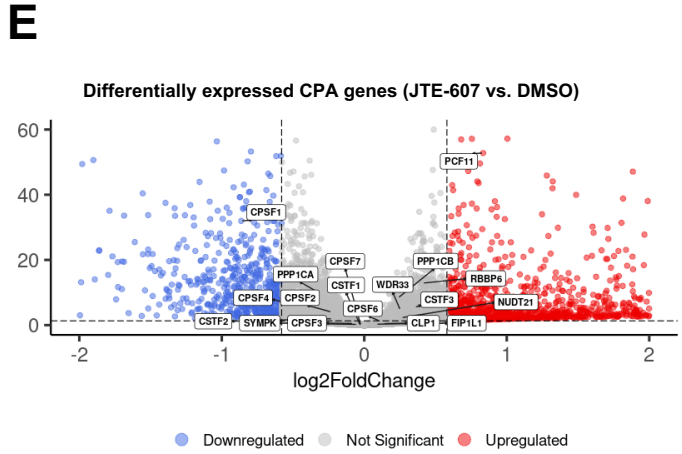


Fig. S6. CPSF3 knockdown, but not inhibition, destabilizes the CPA complex. **(A – C)** Western blots of the immunoprecipitation (IP) assay using (left to right) anti-CPSF4, anti-CSTF2 and anti-NUDT21 to co-immunoprecipitate different components of the CPA complex. IP experiments were performed using *CPSF3* knockdown **(A)** and JTE-607 treated cells **(B and C)**. **(D, E)** Volcano plots of differentially expressed genes upon *CPSF3* knockdown and inhibition, respectively. Blue and Red dots indicate downregulated and upregulated genes, respectively (FC>1.5, Adjusted *P*-value < 0.05). Grey dots are not differentially expressed. Multiple CPA genes are labeled.

Supplemental Fig S7

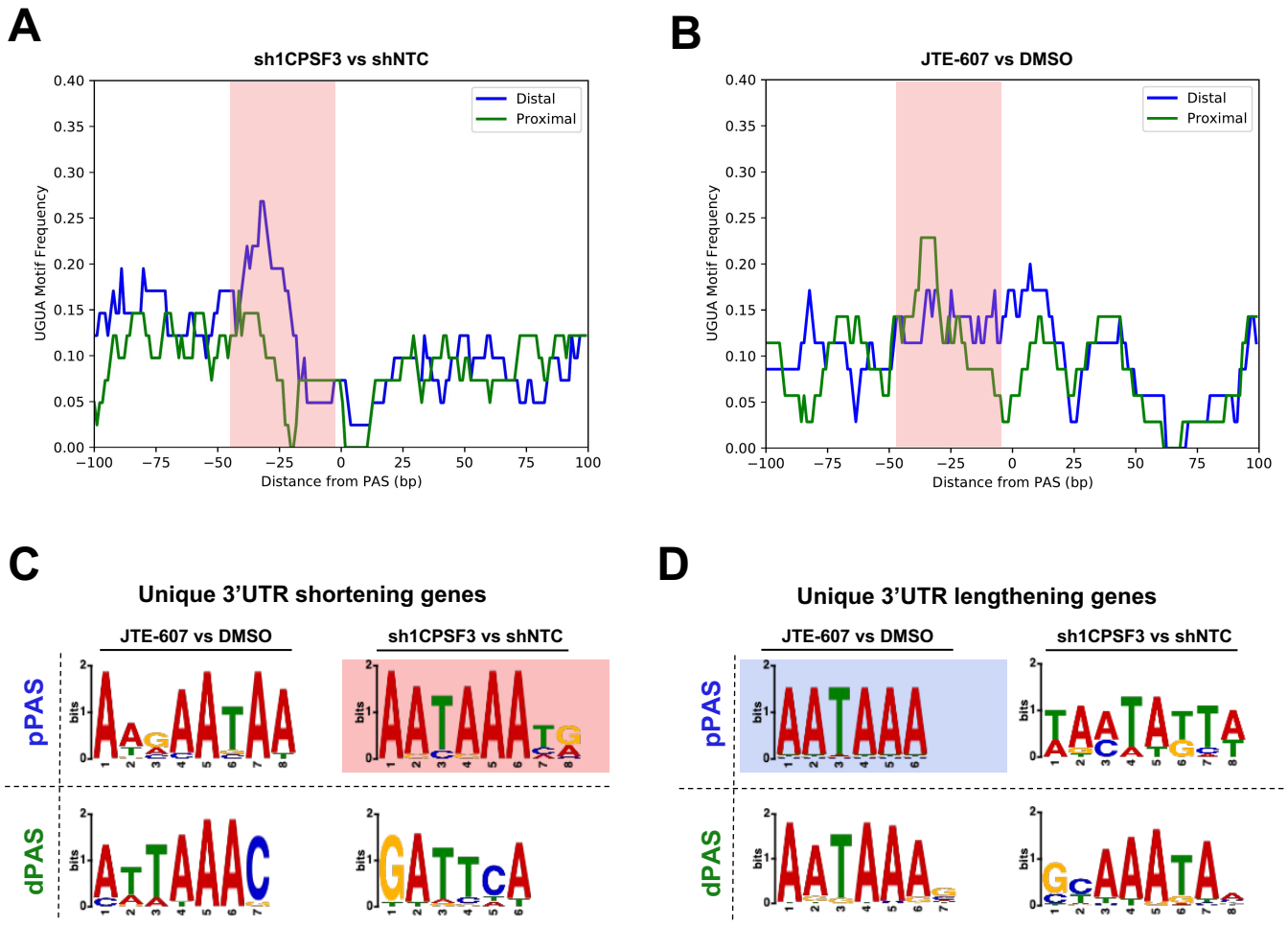


Fig. S7. PAS surrounding motifs of DAGs are distinct between knockdown and inhibition.
(A, B) The distribution of the NUDT21 binding motif, UGUA, surrounding PASs of genes undergoing shortening events upon *CPSF3* knockdown (A) and inhibition (B). Green and blue lines represent the distribution of UGUA around the proximal and distal PAS, respectively. **(C, D)** Motif enrichment analyses of Top sequences surrounding PASs of APA short **(C)** and long **(D)** genes upon CPSF3 inhibition (left) and knockdown (right).

Supplemental Fig S8

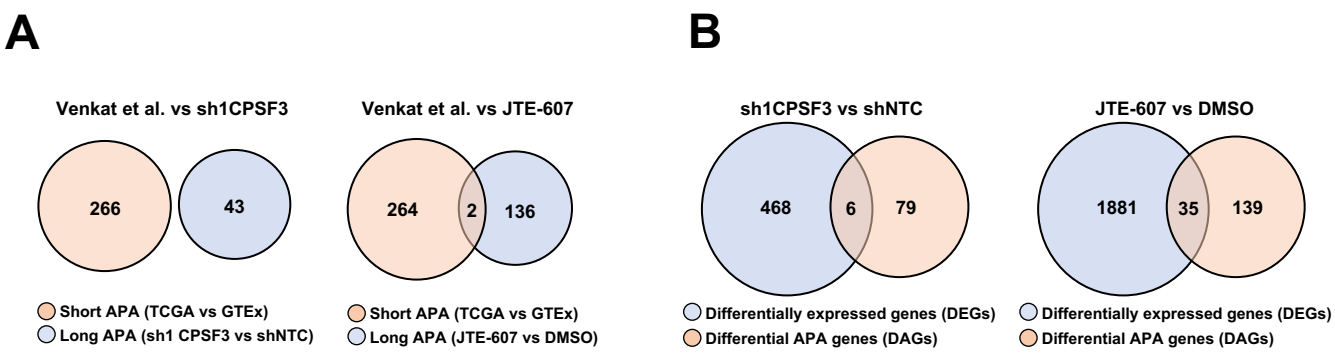
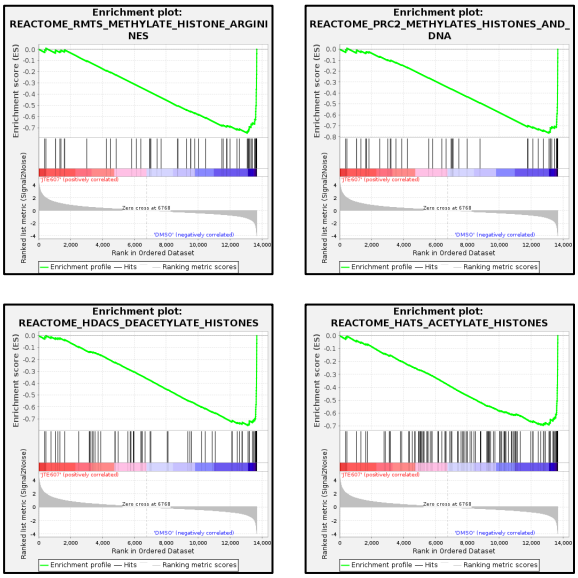


Fig. S8. Clinically APA altered genes are not affected by CPSF3 disruption.
(A) Venn diagrams of short APA genes from TCGA-PAAD (Venkat et al.) data and Long APA genes upon *CPSF3* knockdown and inhibition, respectively. **(B)** Venn diagrams of differential APA genes (DAGs) and differentially expressed genes (DEGs) upon *CPSF3* knockdown and inhibition, respectively.

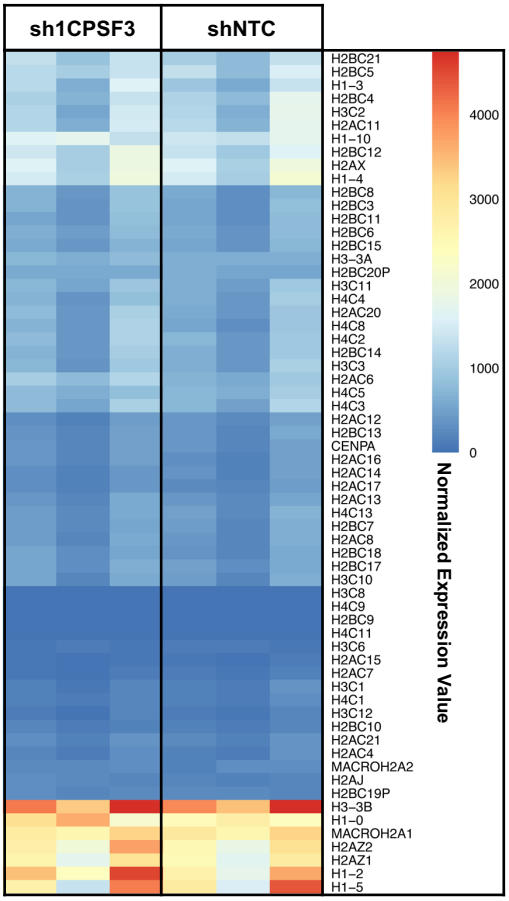
Alahmari, AA; et al.

Supplemental Fig S9

A



B



C

Fig. S9. Replication-dependent histones are not affected by *CPSF3* knockdown. **(A)** Gene set enrichment analysis (GSEA) from RNA-seq data upon 10μM JTE-607 treatment. **(B)** Heatmap of normalized expression values from DESeq2 results in Panc1 cells upon *CPSF3* knockdown. Expression is plotted as normalized expression values. **(C)** Venn diagrams of differentially expressed genes upon *CPSF3* knockdown (orange) and inhibition (blue). Only 119 genes are differentially expressed in both conditions.

Supplemental Fig S10

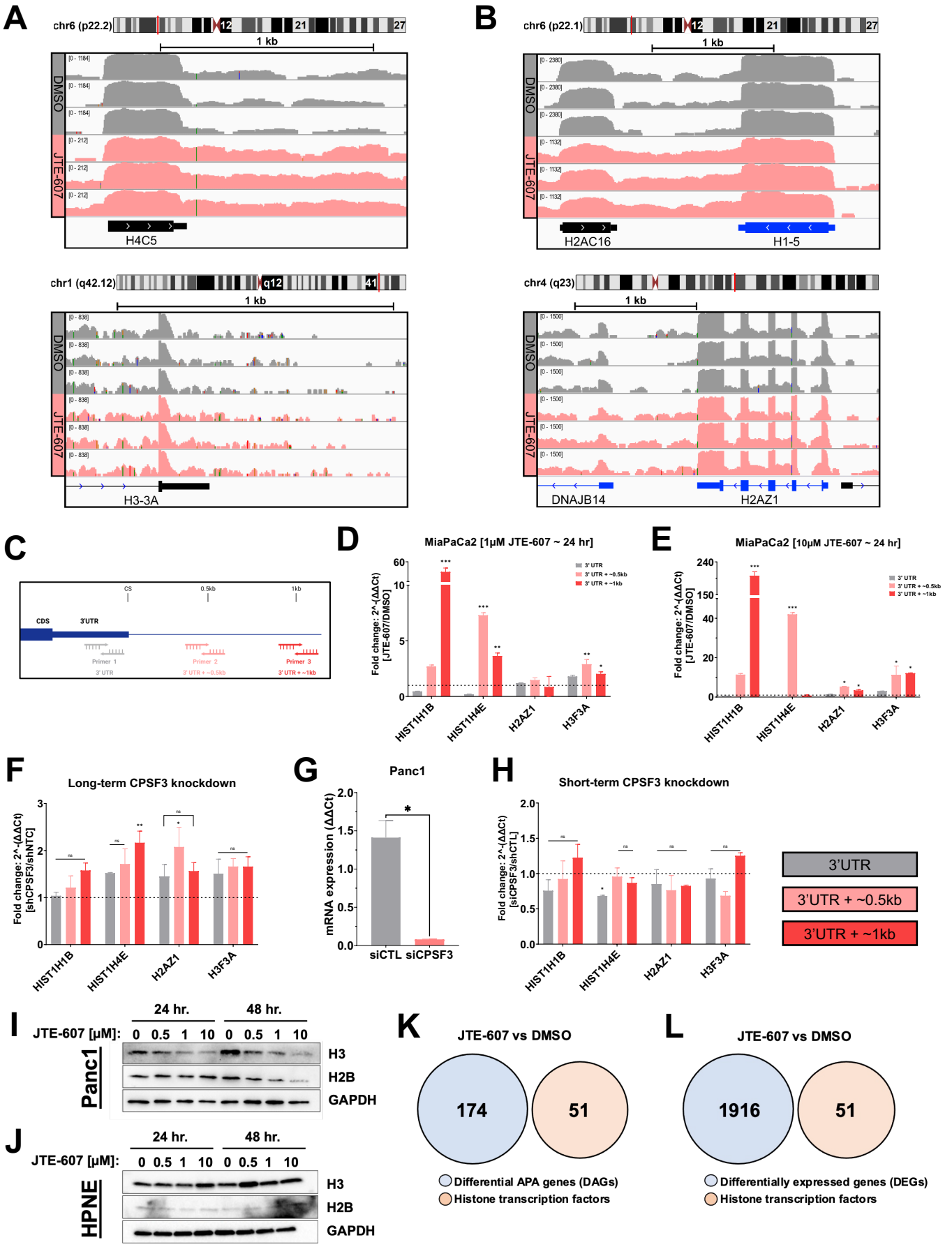
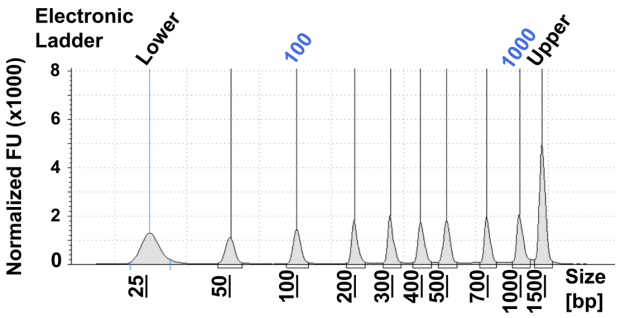


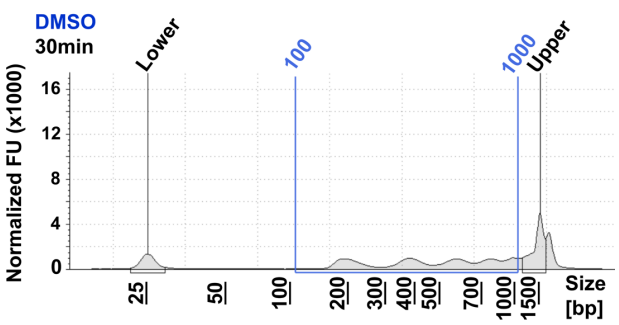
Fig. S10. JTE-607 induces transcriptional read-through in RD histones.
(A) UCSC genome browser-generated density plots of two representative replication-dependent histones highlighting the differences of read coverage beyond the 3'UTR boundaries upon 24hr of JTE-607 treatment (pink) relative to DMSO control group (grey) using total RNA-seq data. **(B)** UCSC genome browser-generated density plots of two representative replication-independent histones highlighting the differences of read coverage beyond the 3'UTR boundaries upon 24hr of JTE-607 treatment (pink) relative to DMSO control group (grey) using total RNA-seq data. Y-axes indicates Log(read density). **(C)** Schematic showing primer design for read-through quantification by RT-qPCR. **(D, E)** Quantification of replication-dependent histone read-through in MiaPaCa2 cells after 24 hr of **(D)** 1μM and **(E)** 10μM of JTE-607 treatment by RT-qPCR. Data were normalized to DMSO controls (Dashed horizontal line). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; 2-way ANOVA with Sidak's multiple comparisons test. **(F)** Quantification of histone read-through in Panc1 upon long-term *CPSF3* knockdown by shRNA by RT-qPCR. **(G)** Quantitative RT-qPCR showing *CPSF3* mRNA expression levels in Panc1 cells upon CPSF3 transient knockdown by siRNA. Data are shown as mean ± SEM.*, $P < 0.05$, unpaired t test. **(H)** Quantification of histone read-through in Panc1 upon short-term *CPSF3* knockdown using siRNA by RT-qPCR. Data were normalized to non-targeting controls (Dashed horizontal line). *, $P < 0.05$; **, $P < 0.01$; 2way ANOVA with Sidak's multiple comparisons test. **(I, J)** Protein levels of H3 and H2B histones upon JTE-607 treatment in Panc1 (I) and HPNE (J) cells. **(K, L)** Venn diagrams showing the overlap between histone transcription factors and significant APA-altered genes (DAGs) **(K)** or differentially expressed genes (DEGs) **(L)** upon JTE-607 treatment.

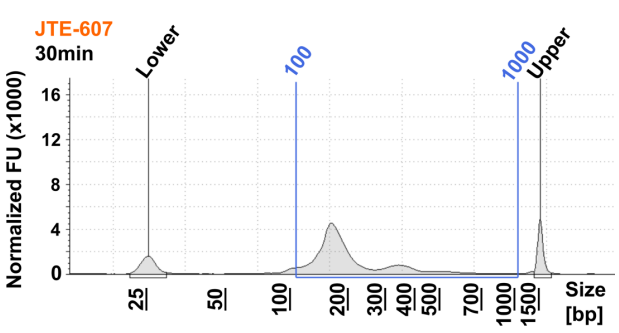
Supplemental Fig S11

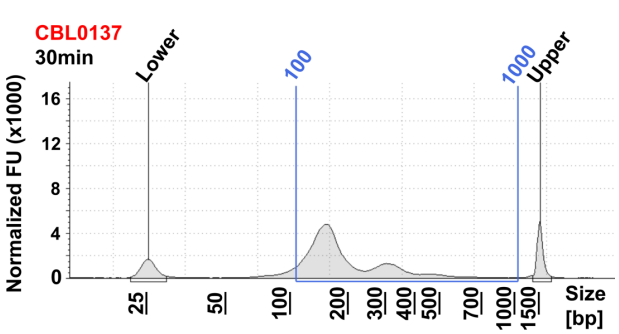
A

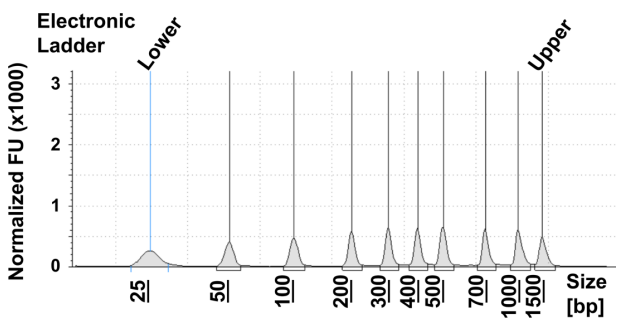
Term	P-value	FDR	Combined Score
Chromatin assembly (GO:0031497)	7.36E-06	0.006693444	130.922877
Nucleosome assembly (GO:0006334)	2.17E-05	0.00884124	128.7851932
Mitochondrial cytochrome c oxidase assembly (GO:0033617)	3.83E-05	0.00884124	262.0222728
Nucleosome organization (GO:0034728)	3.89E-05	0.00884124	85.24289775
Establishment of protein localization to chromatin (GO:0071169)	5.01E-05	0.009110558	609.5011178

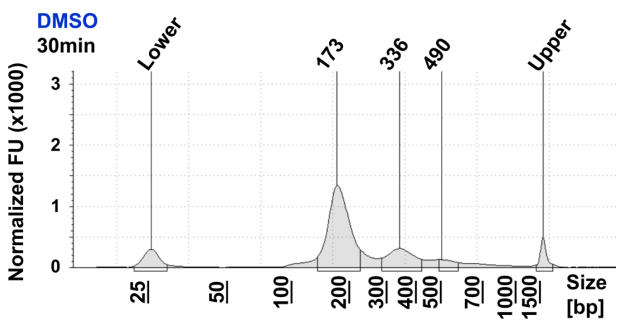


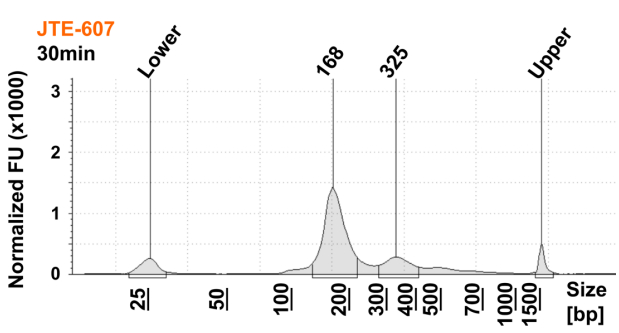












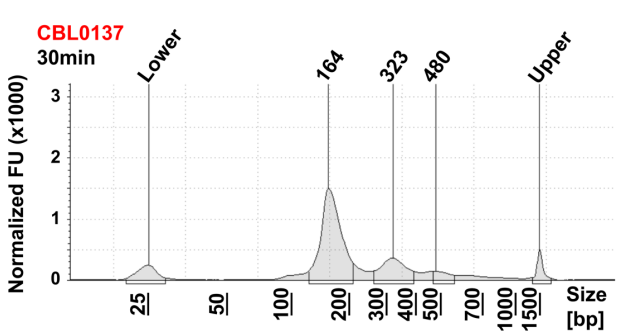


Fig. S11. JTE-607 increases genomic DNA digestion in Panc1 cells.
(A) Gene ontology analysis of significantly downregulated genes upon JTE-607 treatment. Enrichr was used to perform the analysis. **(B - E)** Nucleosome intensity plots of Panc1 cells upon 10μM JTE-607 treatment for 24 hr. Reference ladder shown in **(B)**. Lower base pair sizes indicate more digestion at that specific size. The intensity of digestion is represented by the Y-axis. **(F - I)** Nucleosome intensity plots of HPNE cells upon 10μM JTE-607 treatment for 24 hr. Reference ladder shown in **(F)**. Lower base pair sizes indicate more digestion at that specific size. The intensity of digestion is represented by the Y-axis.

Supplemental Fig S12

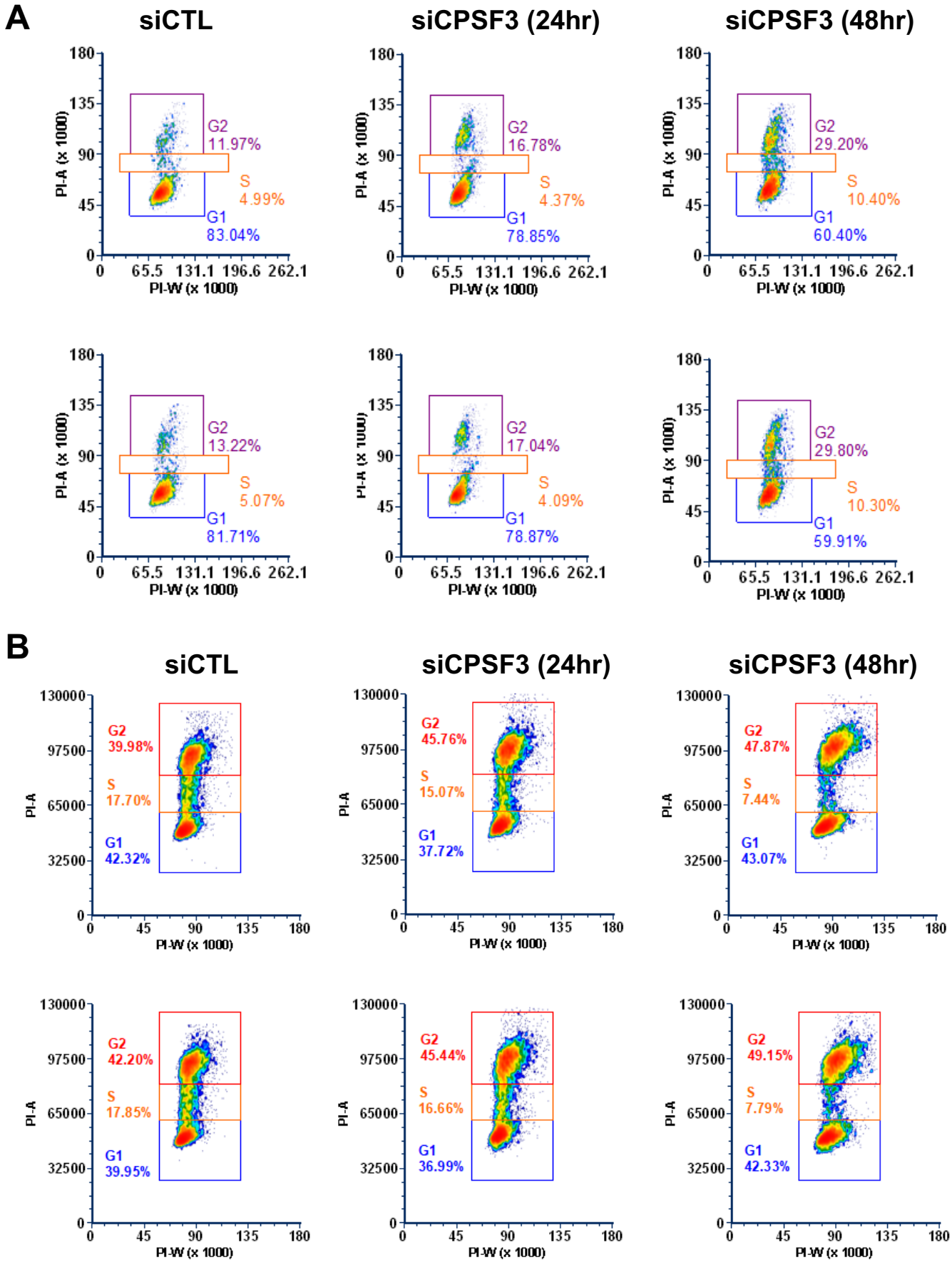


Fig. S12. S-phase population upon CPSF3 knockdown.
(A-B) Quantification of cells in the S-phase upon transient knockdown of CPSF3 by siRNA in HPNE cells **(A)** and Panc1 cells **(B)** after 24hr (middle panels) and 48hr (right panels) as compared to non-targeting controls (siCTL, left panels).

Supplemental Fig S13

Alahmari, AA; et al.

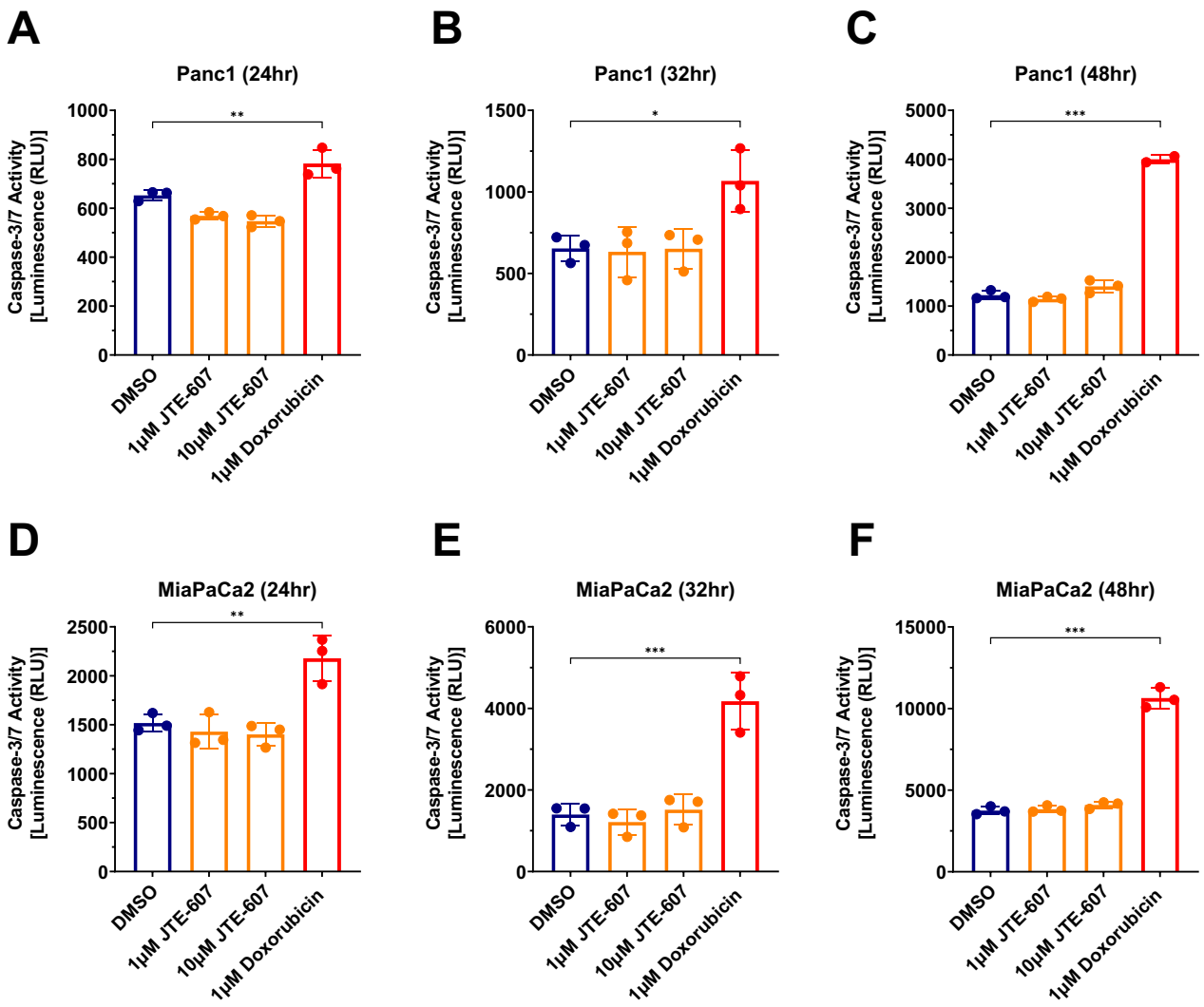


Fig. S13. JTE-607 does not induce significant levels of apoptosis in PDAC cells. **(A-F)** Caspases 3/7 activity upon 1µM (yellow bars) and 10µM of JTE-607 treatment as compared to 0.1% DMSO (blue bars) and 1µM Doxorubicin as a positive control (red bars) for 24hr **(A and D)**, 32hr **(B and E)** and 48hr **(C and F)** in Panc1 (top panels) and MiaPaCa2 (bottom panels) cells. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Ordinary one-way ANOVA with Dunnett's multiple comparisons test.

Materials and Methods

Cell lines and *in vitro* culture

HEK293T, MiaPaCa2, Panc1, Suit2, Human immortalized C7 CAFs and PancPat CAFs cells were cultured in DMEM (DMEM [+] 4.5 g/L glucose, L-glutamine, sodium pyruvate) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Non-transformed pancreatic cell line HPNE cells were cultured in 75% DMEM+25% Medium M3 Base supplemented with 2mM L-glut, 1.5g/L sodium bicarbonate, 5% FBS, 10ng/mL hEGF and 5.5mM D-glucose. Non-transformed pancreatic cell line HPDE cells were cultured in Keratinocyte SFM (serum-free media) supplemented with 25mg BPE, 2.5µg EGF, 1X anti-anti and 50µg/mL Gentamicin. HeLa-TI cells were cultured in phenol red-free FluoroBrite DMEM complete Media. All cell lines were cultured at 37°C with 5% CO_2 and tested negative for Mycoplasma.

Generation of stable and transient CPSF3 knockdown cells

For stable knockdown, two vectors expressing short-hairpin RNA (shRNA) targeting CPSF3 were purchased from Sigma-Aldrich. Cells were infected with lentivirus harboring pLKO.1-shNTC (non-targeting control) and pLKO.1-shCPSF3 at 40% confluency. Polybrene was used to increase the efficacy of infection. After 72 hours, cells were selected with 2.5µg/ml puromycin. Knockdown was confirmed by qPCR and immunoblotting. For transient knockdown, cells were plated at ~$1.5x10^5$ – $2.5x10^5$ cells per well in a 6-well plate one day prior to transfection. The next day cells were transfected with 60nM of siRNA against CPSF3 and a non-targeting control. Transfection media were replaced with complete media 6h after transfection. After 24h and 48h of transfection, cells were collected for subsequent experiments.

RNA isolation and quantitative PCR

Cells were lysed with TRIzol reagent. RNA was isolated using Direct-zol RNA Miniprep. cDNA was synthesized using iScript cDNA Synthesis Kit. qPCR was conducted with SYBR Green pre-designed PCR primers (IDT or Bio-Rad) mixed with iTaq Universal SYBR Green Supermix and run on CFX connect systems (Bio-Rad). For read-through

qPCR quantification, primers were designed and ordered from IDT and are listed in Table S1. Data were analyzed in Microsoft Excel and graphed using GraphPad Prism (v 9.3.0).

RNA-sequencing

For each condition, three biological samples were sequenced. All the sequencing was done on total RNA (ribosomal RNA depleted) and primed with random primers. Briefly, cells were trypsinized, washed with 1X PBS and sent frozen (-80°C) for RNA sequencing (RNA-seq). 500ng total RNA was used to prepare the sequencing libraries using KAPA RNA HyperPrep Kit with RiboErase (HMR) (Roche Sequencing Solutions) following manufacturer's protocol. Briefly, ribosomal RNA (rRNA) was depleted from total RNA and DNase-digested to remove gDNA contamination. RNA was purified, fragmented and first strand cDNA was synthesized using random primers. cDNA:RNA hybrids were converted into double-stranded cDNA (dscDNA) using dUTP incorporation. Adapters were added to the 3' ends, ligated to library insert fragments and the library amplified in a strand-specific manner. Data were analyzed by the Bioinformatics Shared Resource (Roswell Park Comprehensive Cancer Center).

Nuclear isolation

Cells were grown and collected at 80-90% confluence for nuclei isolation. Briefly, Cells were washed with cold 1X PBS and then incubated with hypotonic buffer (10mM Tris pH 8, 1.5mM MgCl2, 2mM KCl) for 15 minutes. Cells were scraped and nuclei release was assessed under the microscope using trypan blue staining. Cell suspension was centrifuged at 2000 rpm for 5 minutes at 4^{o}C. Pellets were washed with hypotonic buffer and centrifuged at 700 x g for 5 minutes at 4^{o}C. Pellets (nuclei) were lysed using NP40 lysis buffer for 1 hour while rotating at 4^{o}C and sonicated once at the end. Nuclear lysate was centrifuged at max speed for 10 minutes at 4^{o}C. Supernatant was collected as the final nuclear lysate.

Immunoprecipitation

Immunoprecipitation was performed using protein G beads. Briefly, beads were mixed thoroughly and washed twice with 1X TBS-T buffer (25mM Tris, 0.15M NaCl, 0.05%

Tween-20 Detergent) containing protease and phosphatase inhibitors. Beads were collected using magnetic stand to remove wash buffer. To preclear the samples, 1 mg of sample was added to the beads and incubated with rotation for 1 hour at 4^{o}C. Beads were separated from the lysate using magnetic stand, and precleared samples were transferred to a clean tube. New magnetic beads (~25ul/1ml lysate) were washed twice, and precleared samples were added to the beads. Antibodies (IgG control or target antibody) were added to lysates and incubated overnight on rotator at 4^{o}C. Beads were collected using magnetic stand and supernatant (unbound proteins) was discarded. Magnetic beads were washed twice with cold 1X TBS-T buffer and once with ultrapure water then collected on magnetic stand. Finally, beads were incubated with 100ul of SDS-PAGE Sample Buffer containing reducing agent at 95ºC for 5-10 minutes. Magnetic beads were separated, and the supernatant was used for western blot analysis.

Immunoblotting

For whole cell lysates, samples were lysed using RIPA lysis buffer (50mM Tris. HCl pH 7.5, 150mM NaCl, 5mM EDTA pH 8, 1% Triton X-100, 0.5% NP-40) in the presence of 10ug/ml protease inhibitors (Aprotinin, Leupeptin, PMSF), boiled at 95°C for 5min and resolved by SDS-PAGE. Proteins were transferred to nitrocellulose membranes at a constant voltage of 100V for 70 minutes at 4°C using Mini Trans-Blot® Cell (Bio-Rad). Membranes were blocked in TBST (Tris-buffered saline (TBS) with 0.1% v/v TWEEN-20) and 5% w/v nonfat dry milk. Primary antibodies (listed in Table S2) were diluted in 3% BSA in TBST and incubated overnight at 4°C. Membranes were incubated with horseradish peroxidase-conjugated secondary antibodies at 1:2,000 for 1 hour at room temperature. Pierce ECL Western Blotting Substrate or Supersignal West Femto Maximum Sensitivity substrate were used for chemiluminescent detection. Signals were visualized and imaged using the ChemiDoc XRS+ System and Image Lab Software (Bio-Rad).

Proliferation and clonogenicity assays

For proliferation experiments, cells were seeded at a density of 250 cells/well (MiaPaCa2 and Panc1 cells) or 1000 cells/well (HPNE cells) into a white 96-well plate in triplicate.

Cell proliferation was measured using CellTiter-Glo Luminescent Cell Viability Assay Kit at days 0, 2, 4 and 6. For clonogenicity assays, cells were seeded at a density of 500 cells/well into a 6-well plate in triplicate. After 11 days, cells were fixed with 4% formalin, stained with 0.2% Crystal Violet and images were obtained for analysis. Colony area was measured using ImageJ software. Data were normalized to control data points.

Caspase 3/7 activity assay

Panc1 and MiaPaCa2 cells were seeded at a density of 1000 cells per well in 384-well plate. Cells were then treated with 1μM and 10μM of JTE-607 as well as 1μM of Doxorubicin as a positive control. Caspase3/7 activity were then detected by measuring luminescence using the Caspase-Glo® 3/7 Assay System (Promega; Cat# G8091).

Cell cycle analysis

Cells were trypsinized and resuspended in 1X PBS, then fixed with ice-cold 70% ethanol for 1 hour at -20°C. Cells were then washed with cold 1X PBS and incubated with RNaseA (200μg/ml) at 37°C for 1 hour. Propidium iodide (40μg/ml) was then added, incubated for 1 hour in the dark and analyzed by FACS at 488nm. Data were analyzed by FCS express software (v7.06.0015).

BrdU incorporation assay

Cells were cultured under optimum conditions and incubated with 50μM BrdU (5-Bromo-2′-deoxyuridine; Sigma-Aldrich; B5002) for 4 hours. Cells were then rinsed with 1X PBS, trypsinized, permeabilized in 70% ice cold ethanol with gentle vortexing and stored at -20°C overnight. Next, cells were pelleted and DNA was hydrolyzed by incubating with 500μl of 2N HCl, 0.5% Triton X-100 in 1X PBS, incubated for 30 minutes at room temperature and then neutralized by adding 1.5ml of 0.1 M sodium tetraborate (pH 8.5) for 2 minutes. Cells were then pelleted, washed once with 1% BSA in 1X PBS and resuspended in 50μl 0.5% Tween 20, 1% BSA in 1X PBS. Next, 10^6 cells were incubated with 1μg Anti-BrdU-FITC for 1 hour at room temperature. Cell pellets were washed once with 150μl 1% BSA in 1X PBS, resuspended in 500μL 1X PBS with RNaseA (200 μg/ml) and PI (40 μg/ml) and incubated at room temperature for 30 minutes in the dark. Cells

were analyzed by flow cytometry immediately and a compensation step was performed. Data were analyzed by FCS Express software (v7.06.0015).

Xenograft experiments

Animal experiments were approved by the Roswell Park Institutional Animal Care and Use Committee. MiaPaCa2 cells infected with shNTC and sh1 CPSF3 constructs were trypsinized, washed with 1X PBS and counted. $5x10^5$ cells were resuspended in 50µl of 1X PBS/Matrigel in a 1:1 ratio and injected subcutaneously into the flanks of 8-week old NOD/SCID/IL2Rγ$^{-/-}$ (NSG) mice. When palpable, tumor volume was determined by caliper measurements obtained in 2 dimensions and calculated as width^2 x length/2 twice a week. Mice were euthanized when the first tumor reached 1400 mm^3, tumors were dissected, and tumor volumes were measured.

Histology and Immunohistochemistry

Freshly dissected tumors were fixed in 10% neutral buffered formalin solution (Sigma-Aldrich, Cat. # HT501128) for 24 hours prior to processing. Tumor processing was performed by the Experimental Tumor Model (ETM) Shared Resource team at RPCC. Briefly, Formalin-fixed paraffin embedded (FFPE) blocks were sectioned (5 µm), rehydrated, stained with hematoxylin and eosin, dehydrated, dried, and mounted. For immunohistochemistry, after deparaffinization and rehydration, antigen unmasking was performed using Retriever 2100 pressure cooker (Aptum). Slides were then immunostained overnight at 4°C, washed, treated with ImmPRESS Polymer reagent (Vector Laboratories) and counterstained. Slides were visualized using OLYMPUS BX41 microscope, imaged using CellSens Standard software and analyzed using ImageJ.

JTE-607 studies

For dose-response measurements, cells were seeded at a density of 1000 cells per well in a 96-well white plate. The next day, JTE-607 was titrated over a range of concentrations using the Tecan D300e Digital Dispenser and cell viability was measured 72 hours post drug titration using a CellTiter-Glo Luminescent Cell Viability Assay Kit (Promega). For cell proliferation experiments, cells were seeded at a density of 250 cells per well in a 96-

well white plate. DMSO control or JTE-607 was dispensed at varying concentrations and proliferation was measured using CellTiter-Glo Assay at days 0, 2, 4 and 6. For clonogenicity experiments, cells were seeded at a concentration of 500 cells per well and treated with different concentrations of JTE-607. Cells were allowed to grow over a period of 11-14 days after which they were fixed in 10% formalin, stained with 0.2% Crystal Violet and images were obtained for analysis. Colony area was measured using ImageJ software. Data were normalized to DMSO control data points.

Micrococcal digestion

Micrococcal Nuclease (MNase) was performed as previously described (Safina et al. 2017). Briefly, cells were trypsinized, washed with 1X RSB buffer (10mM Tris HCL, pH7.6; 15mM NaCl; 1.5mM MgCl2) and pelleted at 1000rpm for 4 minutes at room temperature. Cell pellets were resuspended in 1X RSB buffer with 1% TritonX-100, homogenized with a loose pestle (5 strokes) and centrifuged for 5 minutes at 2000rpm at 4°C. Pellets were washed two times with 1ml of buffer A (10mM Tris HCL, pH7.6; 15mM NaCl; 60mM KCl; 0.34M Sucrose; 0.1% B-mercaptoethanol; 0.15mM Spermine; 0.5 mM Spermidine; 0.25mM PMSF) and nuclei were pelleted at 160g for 10 minutes at 4°C. Nuclei were resuspended in 1.5ml of buffer A supplemented with 1mM of $CaCl_2$. Nuclear suspensions (500µl) were digested with 200U/ml Micrococcal nuclease at 37°C at different time points. Digestion was inactivated by 15mM EDTA. 10%SDS and 1M NaCl were added to extract genomic DNA. DNA was run and visualized using TapeStation 4200 system.

Bioinformatics Analyses

Differential gene expression analysis

Raw sequencing reads passed quality filter from Illumina RTA were first pre-processed by using FASTQC (v0.11.8) for sequencing base quality control. The reads were mapped to GRCh38 human reference genome and GENCODE (v38) (Kent et al. 2002) annotation database using STAR (v2.7.9a) (Dobin et al. 2013). Alignment files are indexed using samtools (v1.14) (Li et al. 2009). A Second pass QC was done using alignment output with RSeQC (v4.0.0) (Wang et al. 2012) to examine abundances of genomic features, splicing junction saturation and gene-body coverage. Gene expression was quantified

using featureCounts (v2.0.0) (Liao et al. 2014) with --fracOverlap 1 option and then formatted in a raw counts data matrix. Differential expression analyses were performed with DESeq2 (v1.36.0) (Love et al. 2014), a variance-analysis package developed to infer statistically significant differences in RNA-seq data using a Negative Binomial GLM. Genes are called differentially expressed (DE) when having a Fold-Change (FC) > 1.5 and FDR < 0.05 (using Benjamini-Hochberg method to control false discovery rate). Downstream and visualization plots were done using regularized-log2 transformation implemented by DESeq2. Heatmaps were generated using the pheatmap (v1.0.8) R package.

3'UTR alternative polyadenylation analysis

For each sample, PCR duplicates were identified using UMI bar codes. UMI tools extract was used to extract the UMI nucleotides from the reads and append them to the read names. Initial quality control was performed using fastp. Reference human genome and annotations of the build GRCh38 release 33 were downloaded from the GENCODE portal. UMI marked and trimmed reads were aligned to respective reference genomes using STAR version 2.7.3a (Dobin et al. 2013). Sample-wise alignments were saved as Sequence Alignment Map (SAM) files. SAMtools V0.1.19 'view', 'sort' and 'index' modules were used to convert the SAM files to Binary Alignment Maps (BAM) files, sort by chromosomal coordinates, and index respectively (Danecek et al. 2021). Mapped reads are deduplicated using UMI tools dedup based on the UMI marked read names and the mapping coordinates (Smith et al. 2017). Alternative polyadenylation (APA) analysis was performed using PolyAMiner-Bulk. Samples were processed with the following parameters: -a 0.65 -pa_p 1 -pa_a 5 -pa_m 5 -outPrefix 3UTROnly -expNovel 1 -s 2 -ignore UTR5,CDS,Intron,UN -apriori_annotations -modelOrganism human. APA analysis focusing only on 3'UTR regions was performed by excluding the polyA sites mapped to other genomic regions with the parameter -ignore UTR5,CDS,Intron,UN. Total reads per feature were computed by adding all individual C/PAS site counts mapped to respective features and total gene counts were computed as the sum of all individual feature C/PAS counts. A beta-binomial test was performed using the proportions of individual feature counts to the total gene counts to infer differential C/PAS usage changes. Our RNA-seq

was based on random priming which does not specifically enrich for polyadenylated RNAs as oligo(dT) priming would. PolyAMiner-Bulk overcomes the challenges associated with identifying polyadenylation sites in random priming-based RNA-seq. PolyAMiner-Bulk is adept at inferring changes in alternative polyadenylation (APA) from bulk RNA-Seq datasets. It enhances the detection process by integrating both established reference polyadenylation sites and newly detected Cleavage and Polyadenylation Sites (C/PASs). It further refines the candidate C/PASs by applying a deep learning model called C/PAS-BERT and utilizing vector projections to analyze alternative polyadenylation (APA) patterns within genes. It is important to note that PolyAMiner-Bulk does not solely rely on RNA primed with oligo-dT (also known as softclipped reads) to pinpoint alternative polyadenylation sites (C/PASs). While some existing tools in this field depend exclusively on softclipped reads to identify candidate C/PASs for subsequent APA analysis, our approach is more comprehensive. PolyAMiner-Bulk combines both established reference C/PASs from databases like PolyASite and PolyADB with newly identified C/PASs to create a larger candidate C/PAS library and retain only high-confidence C/PASs from our candidate library. This approach offers several advantages. It allows us to leverage existing C/PAS annotations and sequencing biochemistry to (i) detect a greater number of C/PASs, (ii) distinguish C/PASs that are closely located, and (iii) overcome sensitivity limitations associated with relying solely on softclipped or oligo (dT) reads.

Motif enrichment analysis

To discover novel, ungapped motifs of recurring, fixed-length patterns in our sequence datasets, we employed the STREME methodology (Bailey et al. 2015). STREME (Simple, Thorough, Rapid, Enriched Motif Elicitation) discovers ungapped motifs that are enriched in our sequence datasets. The enriched motifs were derived by first identifying unique 3'UTR shortened and lengthened genes for both the *CPSF3* knockdown and JTE-607 conditions. Subsequently, we extracted sequences that are 100 nucleotides in length, flanking the 50 base pairs upstream and downstream of the most proximal and distal cleavage sites, referencing the sequence database Human hg38 (UCSCMammal/hg38.fna). Consequently, we obtained four distinct sets of sequences: a pair of proximal and distal sequences for both the *CPSF3* knockdown and JTE-607

conditions. Lastly, we used the STREME program to find enriched motifs. The program shuffles each set to create a corresponding control set and uses a Fisher's Exact Test to determine significance of each motif found in the positive set as compared with its representation in the control set, using an adjusted p-value significance threshold of 0.05. Only significant motifs were shown in the figures.

Quantification and Statistical Analysis

Image analysis

Raw images were obtained and processed using ImageJ, CellSens Standard and Image Lab software.

Statistical analyses

Experimental findings were obtained from at least two independent experiments. Statistics were performed in GraphPad Prism 9. In general, $P < 0.05$ was considered statistically significant. All statistical methods and P-values are provided in the figure legends. Asterisks in graphs denote statistically significant differences as described in figure legends.

Table S1. Primer pairs used to quantify histone read-through by RT-qPCR.

Target	Primer pair	Primer direction	5’ – 3’ sequence
H1-5 (HIST1H1B)	3’UTR (P1)	Forward primer	TGTTGCGGTTTTCACACGC
		Reverse primer	CTAAAGCTGCAAAGGCCAAGAA
	3’UTR+~0.5kb (P2)	Forward primer	AGACCCCATCTTGAAACTTGC
		Reverse primer	AGCCATTTAGGCACCAGCTA
	3’UTR+~1kb (P3)	Forward primer	AGAGCTGGGCCACTGGTTA
		Reverse primer	TTCCTAAAAGTCATTCCTAAGCTCT
H4C5 (HIST1H4E)	3’UTR (P1)	Forward primer	ACAGGGACGCACTCTTTACG
		Reverse primer	TGGGAAGTCGAGATGCTGAG
	3’UTR+~0.5kb (P2)	Forward primer	GCTCACGCAAGGAGAGGTT
		Reverse primer	ACTTCTAAGGGACAACTGGGT
	3’UTR+~1kb (P3)	Forward primer	TCCACAGTTATGCCCCAGATG
		Reverse primer	CACAACGGAAGTTATGGCTGG
H2AFZ (H2AZ1)	3’UTR (P1)	Forward primer	ACTGGAATCACCAACACTGGA
		Reverse primer	ACTGTCTAAAGGATGCCTGGA
	3’UTR+~0.5kb (P2)	Forward primer	ATTTCACCTTTTCCGTCCCA
		Reverse primer	CCGCGAAGACTAACAAGACAC
	3’UTR+~1kb (P3)	Forward primer	GAAGGGGACACTCGTTTTCA
		Reverse primer	CGCATCCTCCCTCGCTTG
H3F3A (H3-3A)	3’UTR (P1)	Forward primer	AGAATCCACTATGATGGGAAACA
		Reverse primer	TCCCCTATTTTTCCACTCGC
	3’UTR+~0.5kb (P2)	Forward primer	AACTGCCCTAGAAGTGATACGA
		Reverse primer	GCAGTAAGAATGCAAGCCCA
	3’UTR+~1kb (P3)	Forward primer	ACTATGTGCTCACTGTCCAGGT
		Reverse primer	GTACATACGTTGAGTGCCAAGT

Table S2. Primary antibodies list.

Primary antibody target	Application	Vendor	Antibody ID
GAPDH	WB	Proteintech	60004-1-Ig
CPSF2	IP & WB	Novus Biologicals	NB100-79823
CPSF3	WB	Abcepta	AT1610a
CPSF3	IHC	Atlas Antibodies	HPA034657
CPSF4	IP & WB	Novus Biologicals	NB100-79826
NUDT21	IP & WB	Santa Cruz	sc-81109
CFIm68	IP & WB	Bethyl Laboratories	A301-358A
CSTF2	IP & WB	Bethyl Laboratories	A301-092A
CSTF3	IP & WB	Santa Cruz	SC-376575
FITC anti-BrdU	FACS	BioLegend	364104
Ki67	IHC	Novus Biologicals	NB600-1209

Refrences

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* **43**: W39-49.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: giab008. https://doi.org/10.1093/gigascience/giab008.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.

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.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 1–21.

Safina A, Cheney P, Pal M, Brodsky L, Ivanov A, Kirsanov K, Lesovaya E, Naberezhnov D, Nesher E, Koman I, et al. 2017. FACT is a sensor of DNA torsional stress in eukaryotic cells. *Nucleic Acids Res* **45**: 1925–1945.

Smith T, Heger A, Sudbery I. 2017. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499.

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**: 2184–2185.