Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6

Sooin Jang[1,2], Nicola J. Cook[3], Valerie E. Pye[3], Gregory J. Bedwell[1,2], Amanda M. Dudek[1,2], Parmit K. Singh[1,2], Peter Cherepanov[3,4,*] and Alan N. Engelman [1,2,*]

[1]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, [2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA, [3]Chromatin Structure and Mobile DNA, The Francis Crick Institute, London, NW1 1AT, UK and [4]Department of Medicine, Imperial College London, St-Mary's Campus, Norfolk Place, London, W2 1PG, UK

ABSTRACT

Cleavage factor I mammalian (CFIm) complex, composed of cleavage and polyadenylation specificity factor 5 (CPSF5) and serine/arginine-like protein CPSF6, regulates alternative polyadenylation (APA). Loss of CFIm function results in proximal polyadenylation site usage, shortening mRNA 3′ untranslated regions (UTRs). Although CPSF6 plays additional roles in human disease, its nuclear translocation mechanism remains unresolved. Two β-karyopherins, transportin (TNPO) 1 and TNPO3, can bind CPSF6 *in vitro*, and we demonstrate here that while the TNPO1 binding site is dispensable for CPSF6 nuclear import, the arginine/serine (RS)-like domain (RSLD) that mediates TNPO3 binding is critical. The crystal structure of the RSLD-TNPO3 complex revealed potential CPSF6 interaction residues, which were confirmed to mediate TNPO3 binding and CPSF6 nuclear import. Both binding and nuclear import were independent of RSLD phosphorylation, though a hyperphosphorylated mimetic mutant failed to bind TNPO3 and mislocalized to the cell cytoplasm. Although hypophosphorylated CPSF6 largely supported normal polyadenylation site usage, a significant number of mRNAs harbored unnaturally extended 3′ UTRs, similar to what is observed when other APA regulators, such as CFIIm component proteins, are depleted. Our results clarify the mechanism of CPSF6 nuclear import and highlight differential roles for RSLD phosphorylation in nuclear translocation versus regulation of APA.

INTRODUCTION

Posttranscriptional regulation of gene expression by alternative pre-mRNA splicing and alternative polyadenylation (APA) greatly increases the coding capacities of metazoan genomes, and several members of the serine/arginine (SR) superfamily of proteins play important roles in these critical metabolic processes. Canonical SR proteins, which are referred to as SR-rich splicing factors (SRSFs), harbor one or two RNA recognition motif (RRM) domains followed by a C-terminal arginine/serine domain (RSD) that is enriched in Arg-Ser dipeptide repeats [reviewed in (1,2)]. SR proteins are prominent components of nuclear speckles, which are sites for splicing factor storage and modification [reviewed in (3,4)]. In particular, phosphorylation of Ser residues within RSDs regulates the intranuclear movement of SR proteins (5,6) and their pre-mRNA splicing functionalities (7–9). RSD phosphorylation also plays a key role in the nuclear import of SR proteins by providing a binding platform for the β-karyopherin nuclear transport factor transportin-SR2, which is also known as transportin 3 (TNPO3) (10–12).

The SR superfamily also consists of several related proteins, which will be referred to herein as SR-like proteins, that differ from the canonical SRSF proteins in domain organization, RRM content, and/or peptide repeat signature within the RSD (13,14). The SR-like protein cleavage and polyadenylation specificity factor (CPSF) 6 in particular harbors a Pro-rich domain (PRD) situated between an N-terminal RRM and C-terminal RS-like domain (RSLD) that is enriched in Arg-Asp/Glu repeats (15) (Figure 1A).

Polyadenylation of pre-mRNA is mediated by the cleavage and polyadenylation (CPA) complex that is composed of four multiprotein complexes, cleavage factor I mammalian (CFIm), CFIIm, CPSF and cleavage stimulatory factor (CstF), as well as additional subunits including

*To whom correspondence should be addressed. Tel: +1 617 632 4361; Fax: +1 617 632 4338; Email: alan_engelman@dfci.harvard.edu
Correspondence may also be addressed to Peter Cherepanov. Tel: +44 20 3796 1798; Fax: +44 20 73875745; Email: Peter.Cherepanov@crick.ac.uk
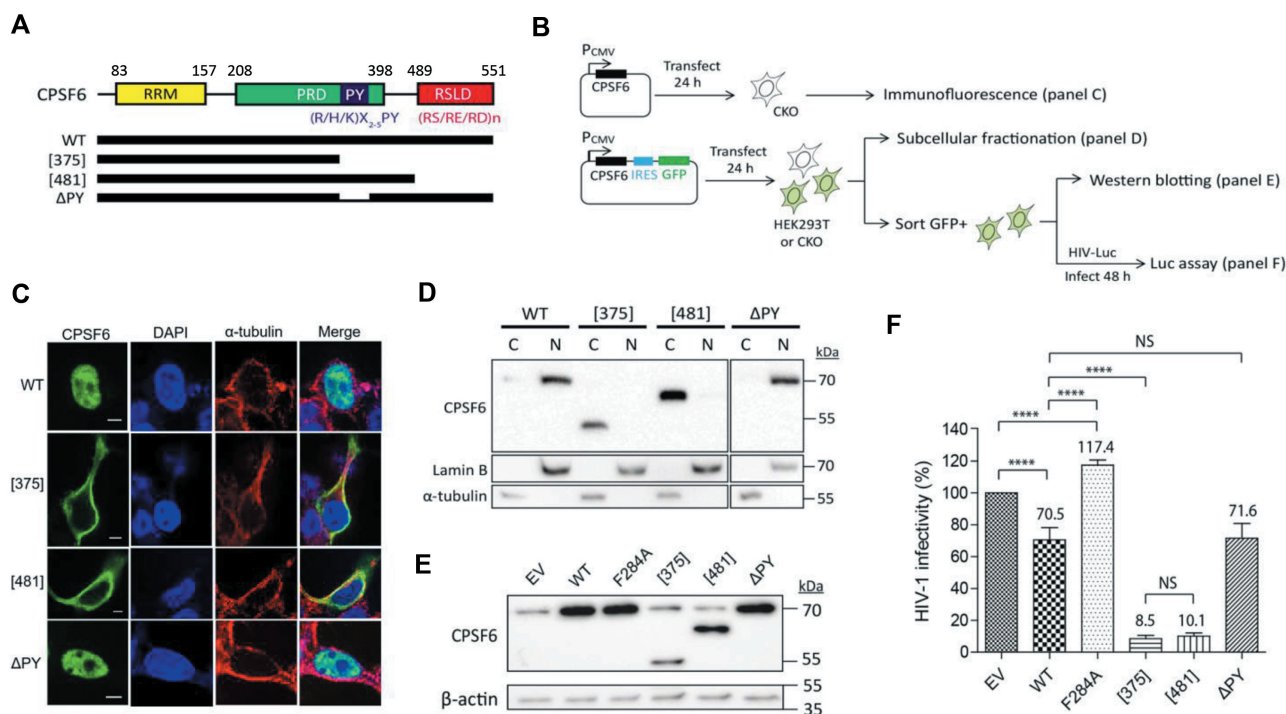
Figure 1. NLS function maps to the RSLD of CPSF6. (**A**) CPSF6[551] schematic and analyzed mutants. CPSF6 domains are noted by colored box, with numbers indicating amino acid positions of domain boundaries. The blue box within the PRD marks the position of the PY-NLS. (**B**) Experimental strategy to assess CPSF6 nuclear localization. CKO cells were transfected with CPSF6 variants based in pLB(N)CX or pIRES2-eGFP, and 24 h later cells were analyzed by immunofluorescence staining and confocal microscopy (panel C) or subcellular fractionation and western immunoblotting (panel D). Alternatively, GFP+ HEK293T cells transfected with pIRES2-eGFP-based vectors were sorted by flow cytometry and either lysed to monitor total CPSF6 expression level by western blot (panel E) or infected with HIV-1 (panel F). (**C**) Confocal microscopy of WT and mutant CPSF6 proteins. Cell nuclei and cytoplasm are demarcated by staining with DAPI and anti-αtubulin antibody, respectively. The results are representative of those observed in a minimum of three independent experiments. Bars indicate 5 μm. (**D**) Cytoplasmic (C) and nuclear (N) fractions were immunoblotted for CPSF6; α-tubulin and Lamin B were blotted as representative cytoplasmic and nuclear markers, respectively. Results are representative of those observed in minimally three independent experiments. See Supplementary Figure S1 for quantitation of WT and mutant CPSF6 nucleocytoplasmic distributions and Supplementary Table S5 for associated statistics. (**E**) WT and mutant CPSF6 expression levels from whole cell extracts. As a protein loading control, β-actin was blotted after removal of the CPSF6 signals by strip washing. Results are representative of those observed in minimally three independent experiments. (**F**) WT HEK293T cells expressing the indicated CPSF6 variants were infected with HIV-Luc, and levels of virus infection were graphed relative to cells transfected with the empty expression vector (EV) control, which was set at 100%. Mean values are written on the top of individual bar graphs. Error bars represent standard error of the mean for $n \geq 3$ independent experiments, with infections conducted in duplicate within each experiment. NS, not significant ($P > 0.05$); ****$P < 0.0001$ (two-tailed Student's *t* test). Numbers to the right of blots (panels D, E) mark migration positions of mass standards.

poly(A) polymerase [reviewed in (16)]. CFIm is a heterotetramer composed of an obligate dimer of CPSF5 and two copies of either CPSF6 or CPSF7 (15,17). Most human genes are regulated by APA and depletion of either CPSF5 or CPSF6, yet curiously not CPSF7, greatly enhances the use of proximal polyadenylation sites (PASs) within 3′ end regions and, consequently, the shortening of mRNA 3′ untranslated regions (UTRs) (18,19). Depletion of other CPA components such as small nucleolar RNA SNORD50A, which in turn interferes with the interaction of FIP1L1 (factor interacting with poly(A) polymerase and CPSF1; a component of the CPSF complex) with PASs (20), or CFIIm component cleavage and polyadenylation factor subunit (PCF11) (21), yields the opposing phenotype of distal PAS overutilization and the lengthening of mRNA 3′ UTRs.

CPSF6 is implicated in various human ailments including cancer (22–26) and human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) (27). CPSF6 interacts with the HIV-1 capsid protein (28,29) to regulate preintegration complex (PIC) nuclear import

(30–32) and the targeting of PICs to active chromatin for integration (33–35). CPSF6 at steady-state is nuclear (36) and its predicted ~60 kDa mass places it near the upper limit for passive diffusion into the nucleus (37). Although this is consistent with an active nuclear import mechanism (36), the mechanistic basis for CPSF6 nuclear import remains unresolved. Two different members of the β-karyopherin superfamily of nuclear import factors, TNPO1 and TNPO3, have been shown to interact with CPSF6 sequences in vitro. TNPO1 can bind a part of the PRD that harbors its associated interaction site, the nuclear localization signal (NLS) motif (R/H/K)X$_{2-5}$PY (PY-NLS) (38), while TNPO3 can bind the C-terminal RSLD (12). Although RSLD-TNPO3 binding occurred irrespective of RSLD phosphorylation (12), a *Drosophila melanogaster* CPSF6 mutant that harbored Ser>Ala substitutions at predicted RSLD phosphoacceptor sites mislocalized to the cytoplasm of fly cells (39), highlighting a potential role for β-karyopherin proteins other than TNPO3 in *Drosophila* CPSF6 nuclear import. Analysis of green fluorescent protein (GFP)-CPSF6

fusion proteins previously revealed that while the RSLD of human CPSF6 harbored a functional NLS, deletion of the PRD and most of the downstream linker region rendered the RSLD-containing fusion protein pan-cellular, suggesting that sequences within the central region of CPSF6, such as the PY-NLS, also contribute to nuclear import (36). While C-terminal truncation mutants that remove both the PY-NLS and RSLD of human CPSF6 are predominantly cytoplasmic and potently restrict HIV-1 infection (28,40), such constructs fail to address the individual contributions of previously identified β-karyopherin interaction sites to CPSF6 nuclear import.

In this study, we analyzed mutant CPSF6 proteins deleted specifically for the PY-NLS or RSLD, which revealed that human CPSF6 nuclear import is mediated by the RSLD without any apparent contribution from the PY-NLS. The X-ray crystal structure of the RSLD–TNPO3 complex was solved to ascertain critical CPSF6 interacting residues, and multiple approaches were subsequently used to pinpoint four positively charged amino acids that are conserved in SRSF1 as critical for TNPO3 binding and CPSF6 nuclear import. Unlike the SRSF1-TNPO3 interaction, which we show depends critically on phosphoserine, nonphosphorylated RSLD protein efficiently bound TNPO3 *in vitro*, and hypophosphorylated mutant CPSF6 was efficiently imported into cell nuclei. By contrast, hyperphosphorylated mimetic RSLD mutant protein failed to bind TNPO3 *in vitro*, and the corresponding full-length CPSF6 mutant was defective for nuclear import in cells. Although hypophosphorylated CPSF6 functioned in APA, a significant number of mRNAs harbored longer 3′ UTRs, similar to what was previously observed when other CPA components, such as PCF11, were depleted. Our results clarify the mechanism of CPSF6 nuclear import and highlight different roles for RSLD phosphorylation in nuclear translocation and the regulation of pre-mRNA APA.

MATERIALS AND METHODS

Plasmid DNAs

CPSF6 RSLD was expressed in *Escherichia coli* as a fusion protein containing either an N-terminal glutathione *S*-transferase (GST) tag from pGEX6P3 (GE Healthcare) or a hexahistidine-SUMO (His$_6$SUMO) tag from pSMT3, which was derived from pET-28b (41). CDC like kinase 1 (CLK1) was expressed in bacteria using pCDF-CLK1 (12). Full-length CPSF6 was expressed in human cells from the retroviral transduction vector pLB(N)CX-CPSF6[551] (33) or from pIRES2-eGFP-CPSF6[551], which was created by inserting CPSF6[551] coding sequences between the PstI and BamHI restriction sites of bicistronic expression vector pIRES2-eGFP (42). Single-round HIV-1 carrying the gene for firefly luciferase (HIV-Luc) was expressed from pNLX.Luc.R-.ΔAvrII (43) while the G glycoprotein from vesicular stomatitis virus was expressed using pCG-VSV-G (44). RSLD fusions with bacterial β-galactosidase (β-gal) were based in pGM-LacZ (45). Plasmid DNAs were built and mutagenized using standard PCR-based methodologies, and all protein coding regions that were amplified by PCR were verified by DNA sequencing. Supplementary Table S1 lists all plasmids used in this study.

Expression and purification of recombinant proteins

Details of the expression and purification of TNPO3 as well as GST- or His$_6$SUMO-tagged proteins were described previously (12). Briefly, GST-CPSF6–RSLD variants were expressed in *E. coli* strain BL21-CodonPlus(DE3)-RP (Agilent technologies) at 30°C for 4 h in the presence or absence of pCDF-CLK1 using 0.01% (w/v) isopropyl-β-D-thiogalactoside. Cell pellets were washed once with phosphate-buffered saline (PBS) and lysed by sonication in lysis buffer (500 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 0.25 mg/ml lysozyme, 25 mM Tris–HCl, pH 7.4). Cell debris was removed by centrifugation at 27 000 × g for 30 min at 4°C, and soluble fractions were loaded onto Glutathione-Sepharose 4 Fast Flow columns (GE Healthcare). Following extensive washing with 500 mM NaCl, 25 mM Tris–HCl, pH 7.4, the columns were developed with elution buffer (25 mM reduced glutathione, 500 mM NaCl, 25 mM Tris–HCl, pH 7.4). As an exception, GST-CPSF6–RSLD-WT expressed in the absence of pCDF-CLK1 was isolated and purified in 1M NaCl-containing buffers. Eluates concentrated by ultrafiltration using Amicon Ultra-4 centrifugal concentrators (Millipore) were further purified by size exclusion chromatography (SEC) over a HiLoad 16/60 Superdex-200 column in buffer containing 200 mM NaCl, 25 mM Tris–HCl, pH 7.4. His$_6$SUMO-CPSF6RSLD expressed in the presence of pCDF-CLK1 was purified by Ni-nitrilotriacetate agarose beads, followed by SEC on a Superdex-200 column in the presence of 500 mM NaCl. GST protein was expressed and purified as described (46).

TNPO3–RSLD complex formation, X-ray crystallography and structure determination

Purified TNPO3 protein was supplemented with 2.3-fold molar excess of His$_6$SUMO-CPSF6 RSLD protein in the presence of 150 mM NaCl, and the mixture was incubated overnight at 4°C with His$_6$-tagged SUMO protease. To deplete the protease and the released His$_6$SUMO tag, the material was supplemented with 20 mM imidazole and passed through a 1-ml Ni HiTrap column (GE Healthcare) equilibrated in 20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole. Following concentration of the flow-through by ultrafiltration, the complex was purified by SEC through a Superdex-200 column in the presence of 25 mM Tris–HCl, pH 7.5, 150 mM NaCl. Peak fractions, supplemented with 5 mM dithiothreitol (DTT), were concentrated by ultrafiltration to ∼9 mg/ml.

Crystals of the TNPO3–RSLD complex were grown in hanging drops by vapor diffusion by mixing 1 μl of protein and 1 μl reservoir solution containing 11% (v/v) polyethylene glycol (PEG) methylether (MME) 500, 5.5% (w/v) PEG 20 000, 2% (w/v) benzamidine–HCl, 2% (v/v) 1,6-hexanediol, 50 mM MgCl$_2$, 0.1 M Tris-bicine pH 8.0 (Buffer system 3, Molecular Dimensions). Crystals, cryoprotected in 10% (v/v) glycerol, 12.8% PEG MME 500, 6.4% PEG 20 000, 4% 1,6-hexandiol, 1.5% benzamidine, 75 mM NaCl, 40 mM MgCl$_2$, 6 mM DTT, 80 mM Tris-bicine, pH 8.0, were frozen by plunging in liquid nitrogen. X-ray diffraction data were collected on the Diamond Light Source beamline I03 (Oxfordshire, UK) at 100K and wavelength 0.97625 Å, and the data were processed using Mosflm (47) and Scala

(48); diffraction data from two crystals were merged to obtain the final dataset. The structure was solved by molecular replacement in Phaser (49) using TNPO3 structure as a search model (12,50). The structure was built in Coot (51) and refined using Phenix (52). The model comprised two, nearly identical, copies of the TNPO3–RSLD complex in the asymmetric unit. While both TNPO3 chains were nearly complete, only octapeptides of CPSF6 could be built in the electron density. As discussed in the Results section, there is a redundancy where multiple sequences within the RSLD are capable of interacting with TNPO3. The D3 sequence (CPSF6 residues 520–528; see below) was used in the final model. X-ray data collection and model refinement statistics are given in Supplementary Table S2. Diffraction data used to detect anomalous scattering from sulfur and phosphorus atoms was collected using low energy X rays (λ = 2.41 Å) at the Diamond beamline I04; to obtain a high redundancy set, data from five crystals were merged to obtain a multiplicity of 28 and the anomalous map was calculated using model phases to 5.8 Å resolution.

GST pull down assay

Details of the GST pull down procedure were previously described (46). Briefly, GST or GST-CPSF6–RSLD variants were pre-bound to glutathione sepharose (1 μg of protein per 1 μl resin). Ten μl loaded resin was incubated with 4 μg TNPO3 and 3 μg bovine serum albumin (BSA) in GST pull down buffer [GST-PDB; 500 mM NaCl, 0.5% (3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate) (CHAPS), 2 mM DTT, 20 mM Tris–HCl, pH 7.4]. Following 2 h of rocking at 4°C, the beads were washed with three changes of 1 ml GST-PDB. Bound proteins, eluted in sodium dodecyl sulfate (SDS)-containing sample buffer, were separated by SDS-polyacrylamide gel electrophoresis (PAGE), and proteins were detected by staining with Coomassie brilliant blue. Percent TNPO3 recovery was quantified using ImageJ software.

Human cells and HIV-1 restriction assay

Wild type (WT) and *CPSF6* knockout (CKO) HEK293T cells were previously described (33). HeLa cells and Jurkat T cell clone E6-1 were purchased from America Type Culture Collection. HEK293T and HeLa cells were cultured in Dulbecco's modified Eagle's medium supplemented to contain 10% fetal bovine serum (FBS), 100 IU/ml penicillin, and 0.1 mg/ml streptomycin. Jurkat T cells were cultured with these same supplements in RPMI 1640 medium. CD4+ T cells purchased from Lonza were cultured in RPMI 1640 medium containing 20% FBS, 100 IU/ml penicillin, 0.1 mg/ml streptomycin, 1X non-essential amino acids, 1× sodium pyruvate and 5 mM HEPES. Where indicated, CD4+ T cells were propagated in the presence of 5 μg/ml phytohemagglutinin (PHA).

For HIV-Luc production, WT HEK293T cells were plated (4×10^6/10 cm dish) 1 day prior to co-transfection with 7.5 μg pNLX.Luc.R-.ΔAvrII and 1.5 μg pCG-VSV-G using PolyJet (SignaGen Laboratories). After 48 h, virus supernatants were filtered through 0.45 μm filters and concentrated ∼50-fold by ultracentrifugation at 53 000 × g

for 2 h at 4°C. HIV-Luc yield was assessed by p24 ELISA (Beckman Coulter). For the restriction assay, HEK293T cells were plated in six-well plates at $6–8 \times 10^5$ cells per well 1 day prior to transfection with 2 μg pIRES2-eGFP or CPSF6-expressing derivative using Effectene (Qiagen). At 24 h post-transfection, GFP-positive cells were selected in basic sorting buffer (1 mM EDTA, 25 mM HEPES, pH 8.0, 1% FBS, Mg^{2+}/Ca^{2+}-free PBS) by fluorescence-activated cell sorting (FACS). Approximately $3–5 \times 10^5$ sorted cells were lysed for immunoblotting while ∼2×10^5 cells were infected in duplicate with HIV-Luc (0.1 pg p24 per cell) in the presence of 4 μg/ml polybrene. At 48 h post-infection, cells were lysed and luciferase activity was determined as previously described (53). Luciferase values were normalized to the level of total protein in cell lysates as determined by the BCA assay (Pierce).

Western immunoblotting

Cell pellets were resuspended in lysis buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitor cocktail (Roche), incubated on ice for 30 min with rigorous vortex at 10 min intervals, and then centrifuged at 14 000 × g at 4°C for 20 min. Protein concentrations in supernatants were determined using the BCA assay, and samples (5–10 μg of total protein) were separated by SDS-PAGE, transferred to polyvinylidene difluoride (PVDF) membranes, and reacted with primary antibody, followed when indicated by horse radish peroxidase (HRP)-conjugated secondary antibody. Membranes were developed using ECL prime reagent (Amersham Biosciences) and imaged with a ChemiDoc MP imager (Bio-Rad). Quantification of western blot signal intensities was done by ImageJ software. Supplementary Table S3 lists all antibodies used in this study as well as associated dilution factors.

Subcellular fractionation and immunoprecipitation

Cells were washed twice with ice-cold PBS and gently resuspended in hypotonic buffer (20 mM Tris–HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) supplemented with protease inhibitor cocktail followed by incubation on ice for 15 min. IGEPAL CA-630 was subsequently added to the final concentration of 0.5%, and mixtures were immediately flicked by hand 3–5 times. Supernatants (cytoplasmic fractions) were isolated following centrifugation at 900 × g for 10 min at 4°C. Pellets were resuspended in nuclear extraction buffer (NE buffer; 10 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 10% glycerol, 0.1% SDS, 0.5% deoxycholate) supplemented with protease inhibitor cocktail and PhosSTOP phosphatase inhibitor cocktail (Sigma-Aldrich), incubated on ice for 30 min with intermittent rigorous vortex as above, and then centrifuged at 14 000 × g for 20 min at 4°C. Resulting supernatants (nuclear fractions) and cytoplasmic fractions were analyzed by western immunoblotting as described above. CPSF6 levels in nuclear and cytoplasmic fractions were assessed using ImageJ software and calculated as relative percentage of their western blot signal intensity from total, which was set as 100%.

For immunoprecipitation, nuclear fractions were isolated from $5–10 \times 10^7$ WT HEK293T or CKO cells that were transfected with hemagglutinin (HA)-tagged CPSF6 expression vectors based in pIRES2-eGFP for 48 h at 50–70% transfection efficiency as assessed by FACS. Isolated nuclear fractions (200–400 μg of total protein) were incubated with 1–2 μg anti-CPSF5 or anti-HA antibody for 2 h at 4°C, followed by addition of 20 μl protein G-Dynabeads (Invitrogen). To assess CPSF6 phosphorylation status by mass spectrometry (MS), nuclear fractions (0.5–1 mg total protein) were reacted with 1–2 μg of rabbit monoclonal anti-CPSF6 antibody (Abcam). Following 2 h of rocking at 4°C, the beads were washed with three changes of 1 ml NE buffer. Bound proteins eluted in SDS-containing sample buffer were analyzed by SDS-PAGE and western immunoblotting.

Cellular fractions (10–20 μg total protein) treated with 200 U of λ phosphatase protein at 30°C for 30 min were separated by SDS-PAGE in gels containing 50 μM Phos-tag (Wako) and 50 μM MnCl$_2$. Phos-tag-containing gels were equilibrated at room temperature in transfer buffer containing 1 mM EDTA for 20 min to remove Mn^{2+}, followed by 20 min with transfer buffer lacking EDTA before transfer to PVDF membranes for western immunoblotting.

Immunofluorescence-based detection of CPSF6 expression

CKO cells ($\sim6 \times 10^5$) were plated on glass coverslips in 12-well plates prior to transfection with 0.5–1 μg pLB(N)CX-CPSF6[551] or CPSF6 mutant expressing derivative using Effectene or PolyJet. At 24 h post-transfection, cells were fixed in 4% paraformaldehyde (PFA) for 15 min at room temperature, washed three times with PBS, permeabilized with 0.5% triton X-100 for 15 min at room temperature, and rewashed three times with PBS. The cells were then blocked by adding 5% nonfat dry milk in PBS containing 0.1% Tween 20 (PBST) for 1 h at room temperature, followed by staining with rabbit monoclonal anti-CPSF6 antibody and mouse monoclonal anti-α-tubulin antibody (Santa Cruz). After overnight incubation at 4°C, cells were washed three times with PBST, and then stained with secondary Alexa Fluor 488 conjugated-anti-rabbit IgG (Abcam) and Alexa Fluor 594 conjugated-anti-mouse IgG (Abcam) antibodies (see Supplementary Table S3) for 1 h at room temperature. After rinsing three times with PBST, the cells were covered with mounting Prolong Gold antifade reagent containing 4′,6-diamidino-2-phenylindole (DAPI; Life Technologies) and analyzed using a 100x objective lens under oil with lasers (405 nm, DAPI; 488 nm, Alexa Fluor 488; 561 nm, Alexa Fluor 594) on a Yokogawa spinning disk confocal microscope at the Dana-Farber Cancer Institute (DFCI) Confocal and Light Microscopy Core Facility.

Assay for transferable NLS function

HeLa cells were cultured in 12-well plates ($\sim2–4 \times 10^4$ per well) 1 day prior to transfection with 1 μg pGM-LacZ-based plasmids using Effectene. At 24 h post-transfection, cells were fixed in 4% PFA for 15 min at room temperature, rinsed twice with PBS, and stained with 0.4 mg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-galactopyranoside) in PBS containing 4 mM potassium ferrocyanide, 4 mM potassium ferricyanide, and 2 mM MgCl$_2$. After 1 h incubation at 37°C, cells were rinsed twice with PBS and observed through a 20x objective lens for transmitted light microscopy using a Nikon Eclipse-Ti microscope.

Fluorescence polarization (FP) spectroscopy

CPSF6 peptides labeled with FITC-Ahx were synthesized by GenScript. Increasing concentrations of TNPO3 were mixed with 50 nM labeled peptide in 10 μl buffer containing 200 mM NaCl, 20 mM HEPES pH 7.5, 0.5% CHAPS, 2 mM DTT for 10 min at room temperature in a black 384-well non-binding surface coated microplate (Corning). The level of fluorescence polarization (mP) was measured using a SpectraMax M5 microplate reader (Molecular Devices). Excitation and emission wavelengths were fixed at 488 and 526 nm (cut off, 515 nm), respectively. Unlabeled peptides were used as competitors by increasing their concentrations into pre-mixtures of fixed concentrations of 50 nM labeled SRSF1 RSP peptide and ~1.0 μM TNPO3, the concentration of which was determined based on the equilibrium dissociation constant (K_d) value of 1.3 ± 0.19 μM for the RSP-TNPO3 complex. K_d was estimated by nonlinear regression analysis with an equation for one-site specific binding using GraphPad Prism 7 software.

RNA-Seq and bioinformatics analyses

Approximately $0.5–1 \times 10^6$ cells that had been transfected for 24 h with pIRES-eGFP or derivatives expressing HA-tagged CPSF6 protein were sorted by FACS for GFP expression level and plated in 12-well plate and allowed 48 h for cell recovery prior to cell lysis. Total RNA was isolated using the Quick-RNA miniprep (Zymo Research), prepped for 75-bp paired end Illumina HiSeq and sequenced on a NextSeq500 sequencer by the DFCI Molecular Biology Core Facilities as described (54). Two independent experimental sample sets were sequenced on separate NextSeq500 sequencer runs.

Percent distal polyadenylation usage index (PDUI) analysis was performed by DaPars software as described previously (55) using aligner hisat2 version 2.1.0 with default options (56) and human genome build hg19. Results parsed for significance across biological replicates (see below) were compared with prior reports documenting mRNAs with significantly longer 3′ UTRs following knockdown of SNORD50A (20), PCF11 (21), or CstF64/64τ (57) as follows. For SNORD50A, 95 genes with significant shift from distal to proximal PAS and 68 genes with significant change from proximal to distal PAS usage were assessed. PCF11 knockdown datasets were parsed in two separate ways to yield two gene lists. Three independent knockdown experiments identified 647 significantly affected genes (631 and 16 with longer and shorter 3′ UTRs, respectively). Analysis of stage 4S neuroblastoma samples revealed an additional 721 genes that we parsed into shorter and longer 3′ UTR bins based on P value (<0.05) and longer-to-shorter mRNA isoform ratio (if >0, mRNAs were partitioned to longer 3′ UTR bin; vice versa for <0). CstF64 datasets were likewise parsed in two different ways depending on whether

the knockdown experiment targeted CstF64 (85 genes at high confidence with altered PASs) or both CstF64 and CstF64τ (201 genes at high confidence). For both datasets, genes were assigned shorter or longer 3′ UTRs based on ratio of proximal-to-distal PAS usage compared to control conditions. In all, this provided 10 gene lists, five each composed of genes with shorter versus longer 3′ UTRs. The lists were then compared in pairwise fashion for overlap with our experimental CPSF6 mutant dataset.

Alignment of sequencing reads from ref. (21) was performed as follows: raw data in fastq format was accessed (SRR5271230_2. fastq, SRR5271427_2. fastq, SRR5271428_2. fastq, and SRR5271429_2. fastq for PCF11 knockdowns; SRR5271226_2. fastq for CPSF6 knockdown, and SRR5271412_2. Fastq and SRR5271413_2. fastq for mock1.7 and mock2 knockdowns) and reads with five or more consecutive A residues were selected and trimmed. Trimmed reads were aligned to human hg19 using bowtie2 aligner in default option (58) and aligned reads were normalized by bamCoverage (59). Density plots were visualized using the WashU Epigenome Browser (60).

MS and *in silico* analyses of CPSF6 phosphorylation

GST-RSLD proteins purified from *E. coli* and full-length CPSF6 derivatives immunoprecipitated from nuclear extracts of WT HEK293T or CKO cells were fractionated by SDS-PAGE and stained with Coomassie blue. Protein-containing gel slices were excised and digested by Asp-N protease (proteins produced in bacteria) or trypsin followed by peptide extraction (61). The eluting peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide by the Taplin Mass Spectrometry Facility at Harvard Medical School using an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Phosphorylation assignments were determined by the Ascore algorithm (62).

Sites of predictive CPSF6 phosphorylation (Ser, Thr, Tyr) were tabulated from different web-based databases including PhosphoSitePlus (63), dbPAF (64), and Phosida (65). Supplementary Table S4 lists the results of these analyses. Kinases predicted to interact with specific CPSF6 sequences were assessed using two different computational tools including Netphorest version 2.1 (66) using a minimum score cutoff of 0.18 and Scansite version 4.0 (67) with medium stringency option.

Statistical analyses

Results of HIV-1 restriction assays and nucleocytoplasmic CPSF6 localization by subcellular fractionation were compared using two-tailed Student's *t*-test assuming equal variances in Microsoft Excel. Results of GST pulldown and coimmunoprecipitation assays were analysed by unpaired *t*-test calculated using two-tailed *P* value (critical *P* value < 0.05) in GraphPad Prism 7 software; results of FP spectroscopy K_d values were similarly analysed by paired *t*-test.

The DaPars algorithm (55) uses Fisher's exact test adjusted for multiple testing by Benjamini–Hochhberg at 5% false discovery rate to assess significance differences of mean PDUI values. Messenger RNAs with significant changes were scored using the following criteria: absolute ΔPDUI value (|ΔPDUI|) ≥0.2 and absolute value of $\log_2$ change in PDUI ≥ 1 (|$\log_2$[PDUI$_{sample}$/PDUI$_{WT}$]| ≥ 1), which equates to a 2-fold change. The hypergeometric test was used to assess statistically significant differences in numbers of overlapping genes between samples assuming a total population of 25 000 human genes (68).

RESULTS

The RSLD confers CPSF6 nuclear import without any contribution from PY-NLS

PY-NLS and RSLD sequences within CPSF6 mediate binding to respective β-karyopherin proteins TNPO1 and TNPO3 *in vitro* (12,38). However, the contribution of each of these sequences to CPSF6 nuclear import is unclear (28,36,39,40). Human CPSF6 is expressed as two major isoforms, the larger of which contains 588 residues; the smaller splice variant, composed of 551 amino acids, lacks internal residues upstream from the PY-NLS and RSLD (28). We have analyzed NLS function in the context of CPSF6[551] (bracketed information denotes C-terminal amino acid position or mutant name; single missense mutants are named alongside [551]) because it is the predominant isoform expressed in cells (33) (also see below). Two site-specific deletion mutants, CPSF6[ΔPY], which lacked residues 376–390 but retained the RSLD, and CPSF6[481], which removed the RSLD but retained PY-NLS (Figure 1A), were compared to the previously described CPSF6[375] mutant that lacked both the PY-NLS and RSLD. CPSF6[375] aberrantly localized to the cytoplasm and potently inhibited HIV-1 infection (40).

Figure 1B summarizes the different assays that were used to assess CPSF6 nuclear localization. To avoid complications from co-detection of endogenous protein, confocal microscopy (Figure 1C) and biochemical fractionation (Figure 1D) were analyzed in *CPSF6* knockout (CKO) cells that were transiently transfected with test constructs. These techniques were performed 24 h post-transfection to minimize the number of cell divisions and hence potential nuclear accumulation of NLS-defective mutants due to binding to cellular components other than β-karyopherins. To be consistent with prior reports, restriction of HIV-1 infection, which was used as an indirect readout of cytoplasmic CPSF6 accumulation, was analyzed in WT cells (28,40).

As expected (36,40), CPSF6[551] and CPSF6[375] localized to the cell nucleus and cytoplasm, respectively (Figure 1C, D; see Supplementary Figure S1 and Supplementary Table S5 for quantitation of biochemical fractionation data and associated statistics, respectively). While CPSF6[ΔPY] displayed the WT phenotype, CPSF6[481], like CPSF6[375], was retained in the cytoplasm. Thus, under these conditions, CPSF6 NLS function mapped to the RSLD, with no discernible contribution from PY-NLS. The results of HIV-1 infection were completely consistent with this interpretation. Cells expressing WT CPSF6[551] supported marginally (∼30%) less infection than cells transfected with the empty vector (EV) control (Figure 1E, F), which was likely due to a modicum of WT protein that localized to the cell cytoplasm under these conditions of protein overexpression (Supplementary Figure S2). Expression

of the CA binding-defective mutant CPSF6[551]F284A ([33]) failed to elicit the infection reduction, revealing, as expected, that the WT phenotype depended on the interaction of CPSF6 with HIV-1 CA. Also as expected ([40]), CPSF6[375] expression significantly restricted HIV-1 infection (Figure 1E, F). Cells expressing CPSF6[481] restricted infection to a level that was indistinguishable from the CPSF6[375] control, whereas cells expressing CPSF6[ΔPY] supported the same level of infection as cells expressing WT CPSF6[551] (Figure 1E, F; see Supplementary Table S6 for study-wide statistical comparisons of constructs analyzed in the HIV-1 restriction assay).

Structure of the CPSF6–RSLD complex

To gain insight into the structural basis for CPSF6 NLS function, we crystalized the TNPO3–RSLD complex. Although we showed previously that nonphosphorylated RSLD protein (CPSF6 residues 404–551) interacted with TNPO3, phosphorylation greatly enhanced the binding of SRSF1 to TNPO3 ([12]). The RSLD construct used here (residues 481–551; Figure 2A) was accordingly co-expressed in bacteria with human CLK1 ([12]). Purified TNPO3 and RSLD proteins were combined, and the complex was purified by SEC (Supplementary Figure S3). Crystals of the TNPO3–RSLD complex diffracted X-rays to 2.7 Å resolution (Supplementary Table S2). As with prior crystal structures ([12,50]), the vast majority of full-length TNPO3 could be built into the electron density map (Figure 2B). Similar to the TNPO3-SRSF1 co-crystal structure, only a fraction of the input CPSF6 RSLD (9 of the 70 residues) was ordered in the density map. The core interacting SRSF1 RSD sequence [204]RSRpSRpSRSR[212] was unambiguously assigned in the prior structure because the protein construct contained the upstream RRM2 domain and featured a well-ordered RRM2-RS linker region (Figure 2B) ([12]). The input CPSF6 fragment by contrast comprised the RSLD with only 9 upstream linker residues. Not unexpected due to the repetitive nature of the RSLD sequence (see Figure 4A below), three different, though sequence-related stretches, termed D1, D2 and D3, could be modeled as phosphorylated peptides to fit the omit *Fo-Fc* difference map (Figure 2C, D). The electron density strongly suggested phosphorylation at position 6 of the bound peptide, and the presence of a phosphorous atom was supported by anomalous X-ray scattering at a wavelength of 2.41 Å (Figure 2C; Supplementary Figure S4). Alignment of the D1, D2 and D3 sequences with the core SRSF1 NLS revealed that positions 5–7 and 9 were invariant across the four sequences, whereas positions 3, 4 and 8 were chemically related considering the functional similarity of Asp/Glu residues with phosphoserine ([69,70]). Position 2 of the CPSF6 sequences, the first position with resolvable side chain, harbored a relatively bulky amino acid (Arg, His, or Tyr) as compared to Ser in SRSF1. The resolved CPSF6 RSLD region importantly engaged TNPO3 helices 15–17, including the critical R-helix (orange in Figure 2A–C) that mediated direct contacts with the SRSF1 RSD ([12]).

As expected, the interactions involving the residues of TNPO3 and CPSF6 are highly similar to those observed in the TNPO3-SRSF1 complex ([12]) (see below). However, the

overall conformation of the β-karyopherin in the TNPO3–RSLD complex is less open, allowing the terminal carboxyl group of TNPO3 (residue Arg923) to make a salt bridge with Arg166 (Supplementary Figure S5). In all prior TNPO3 structures, the distance between the C-terminus and Arg166 ranged from 21 to 27 Å (Supplementary Figure S5), underscoring adaptability of the karyopherin fold.

RSLD sequences critical for CPSF6 NLS function

Due to the inherent ambiguity in assessing a specific CPSF6 sequence to the electron density map, a series of cell-based and biochemical assays were performed to ascertain RSLD sequences that contributed to nuclear import and TNPO3 binding, paying particular attention to modeled D1–D3 regions. A set of internal deletion and C-terminal truncation mutants that spanned the RSLD was analyzed initially. Sequences downstream from D3 were removed in the C-terminal truncation mutant CPSF6[531], the CPSF6[520] truncation additionally lacked the D3 region, and mutant CPSF6[503] lacked both D2 and D3 regions (Figure 3A). Three internal deletion mutants ΔD1-D3 collectively removed all RSLD sequences upstream from residue Arg532.

Biochemical fractionation revealed that CPSF6[503] was largely cytoplasmic, indicating that sequences downstream from D1 played a critical role in nuclear import (Figure 3B; Supplementary Figure S1 and Table S5). Indeed, the level at which CPSF6[503] restricted HIV-1 infection was indistinguishable from the CPSF6[481] control that lacked the entire RSLD (Figure 3C, 3D). In repeat experiments, about 26% of CPSF6[520] fractioned to the nucleus while ∼33% of CPSF6[531] was nuclear (Supplementary Figure S1). Thus, while D2 and D3 sequences play important roles in CPSF6 nuclear import, they themselves are insufficient to impart full nuclear import function in the context of the CPSF6[531] deletion. The three internal deletion ΔD1–ΔD3 mutants behaved similarly to one another in biochemical fractionation (Figure 3B; Supplementary Figure S1 and Table S5) and HIV-1 restriction assays (Figure 3C, D and Supplementary Table S6). We conclude that different regions of the RSLD, including sequences downstream from D3, can contribute to CPSF6 nuclear import.

A series of 10 Ala substitution mutant proteins was constructed to further delineate RSLD sequences critical for CPSF6 nuclear import. CPSF6[D1allA], CPSF6[D2allA] and CPSF6[D3allA] contained eight contiguous Ala residues in place of the WT D1, D2 and D3 sequences, whereas CPSF6[D12allA], CPSF6[D13allA] and CPSF6[D23allA] harbored all pairwise combinations of the 8-mer mutant constructs. To test the contributions of residues downstream from D3, 5-Ala block substitution mutants 1A–4A, which targeted residues 532 through 551, were analyzed (Figure 4A). Due to the throughput nature of the approach, the Ala substitution mutants were assessed initially in the HIV-1 restriction assay.

CPSF6[D1allA], CPSF6[D2allA] and CPSF6[D3allA] restricted HIV-1 infection ∼2- to 3-fold as compared to WT CPSF6[551]. Whereas all three 16-mer combination mutant proteins restricted infection to greater extents than their 8-mer counterparts, the two mutants with altered D3 sequences, CPSF6[D13allA] and CPSF6[D23allA], were
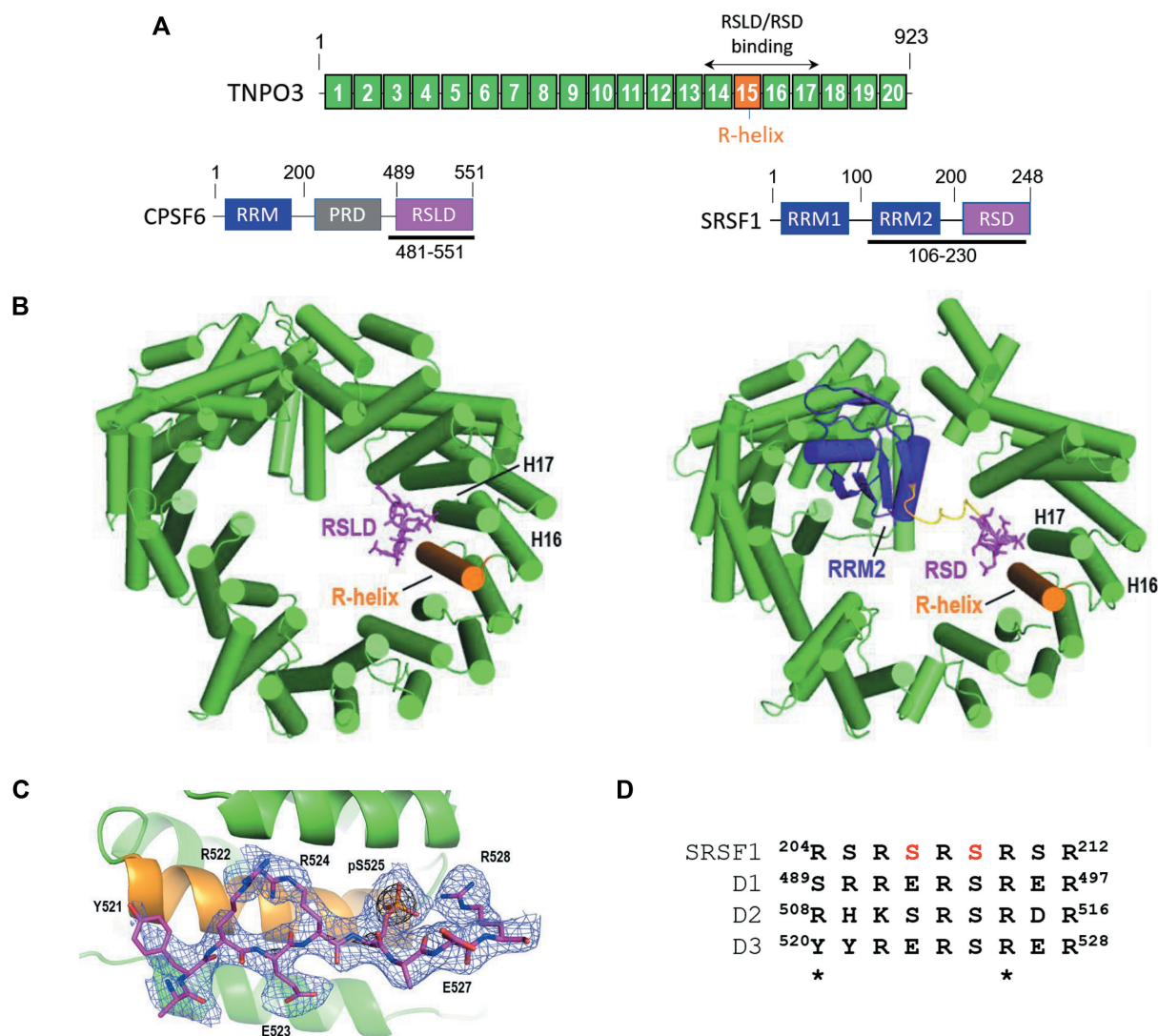
Figure 2. Structure of the TNPO3-CPSF6 RSLD complex. (**A**) Crystallized proteins and fragments are shown schematically. Full-length TNPO3 (923 residues) is segmented to highlight α helices; the R-helix (helix 15), which interacts intimately with cargo substrate, is highlighted in orange. CPSF6 and SRSF1 fragments are indicated by horizonal line and amino acid coordinates. (**B**) Left, TNPO3 is green, with helices 15–17 highlighted. The modeled RSLD D3 region is shown as magenta sticks. The structure of the TNPO3-SRSF1 complex (pdb code: 4c0o), shown to the right, highlights positions of the RRM2 domain (blue) and RSD (magenta). The RRM2-RSD linker is in yellow. (**C**) Close up view of omit difference map, with CPSF6 residues 520–528 modeled as sticks. Light blue and black meshes display the omit $F_o - Fc$ map contoured at 2σ and 8σ, respectively. (**D**) Sequence alignment of modeled D1, D2 and D3 peptides with analogous SRSF1 RS repeat region. Ser residues modified by phosphorylation in the SRSF1-TNPO3 structure are in red (12). Asterisks mark unresolved side-chains in the RSLD-TNPO3 structure.

consistently more potent, blocking HIV-1 to levels that were indistinguishable from the control RSLD deletion CPSF6[481] mutant protein (Figure 4B). Of the downstream 5-Ala mutant proteins, CPSF6[1A] and CPSF6[3A] restricted infection to levels similar to the 8-mer D1-D3allA substitution mutants, where CPSF6[2A] and CPSF6[4A] were significantly less potent, behaving more like WT CPSF6[551] (Figure 4B and Supplementary Table S6). From this analysis we inferred that D3 sequences contributed significantly to CPSF6 NLS function.

D2 and D3 sequences were further targeted by mutagenesis in an attempt to define the smallest number of amino acid changes that disrupted CPSF6 NLS function. SRSF1 residues Arg206 and Arg208 mediate multiple contacts with

TNPO3 residues Asp750 and Glu660/Asp751, respectively (12). As Arg or Lys occur at analogous D1–D3 positions (Figure 2D) and mediate similar contacts with TNPO3 (Figure 5A), targeted Glu substituents were made, yielding a set of constructs that ranged in complexity from the single missense mutant CPSF6[551]K510E to CPSF6[4Glu], with all four D2 and D3 positions changed (Figure 5A). Whereas CPSF6[551]K510E restricted HIV-1 infection less than two-fold compared to WT CPSF6[551], double mutants CPSF6[D2EE] and CPSF6[D3EE] behaved similarly to their respective CPSF6[D2allA] and CPSF6[D3allA] counterparts (compare Figure 5B with Figure 4B). Importantly, CPSF6[4Glu] behaved indistinguishably from CPSF6[481] and CPSF6[D23allA] controls in assays for
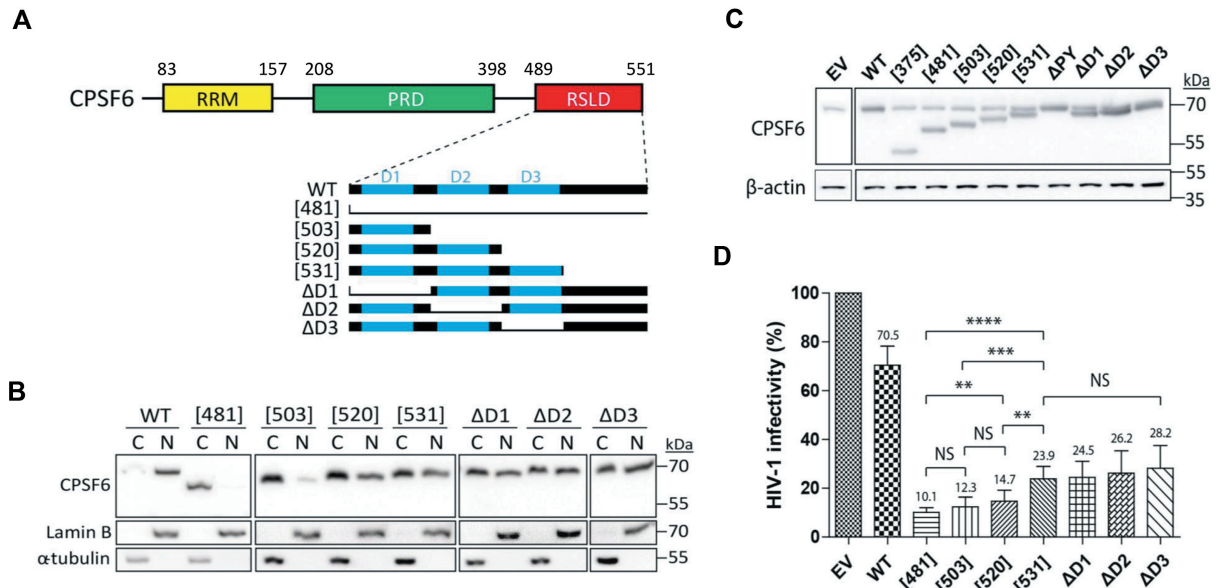
Figure 3. Deletion analysis of CPSF6 RSLD function. (**A**) Schematic map showing the constructs analyzed; positions of CPSF6 D1–D3 sequences are noted in blue. (**B**) Fractionation of WT and CPSF6 mutant expressing cells. Other descriptions as in Figure 1D; see Supplementary Figure S1 for associated quantitation. (**C**) Expression levels of WT and mutant CPSF6 proteins. Other descriptions as in Figure 1E. (**D**) Results of HIV-1 infection. NS ($P > 0.05$); **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. Other descriptions as in Figure 1F.

both HIV-1 restriction (Figures 4B and 5B) and CPSF6 nuclear import (Figure 5C).
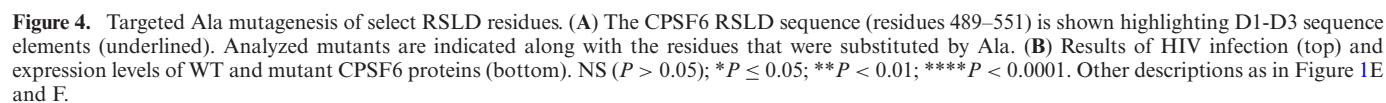
Transferable CPSF6 NLS function

Several types of NLSs, including the relatively compact basic element from simian virus (SV) 40 large T antigen (71) as well as the RSDs of SRSF1 and SRSF3 (72), are transferrable, as they confer karyophilic behavior to otherwise non-nuclear proteins. To test whether the CPSF6 RSLD was sufficient to target a large heterologous protein into the nucleus, we fused it to either the N-terminus or C-terminus of bacterial β-galactosidase (gal), and examined the cellular localization of the fusion proteins in transfected HeLa cells. As expected (45), the unmodified bacterial enzyme localized throughout the cell. Also as expected (45), fusing the SV40 large T antigen NLS to the N-terminus of β-gal conferred nuclear import (Supplementary Figure S6). Regardless of which terminus received the CPSF6 RSLD fusion, it too conferred nuclear localization to β-gal, revealing relatively context independent transferrable NLS function. These findings are in line with the prior report of GFP-RSLD nuclear localization (36). Importantly, the 4Glu amino acid substitutions negated transferrable RSLD NLS function (Supplementary Figure S6).

Ser phosphorylation and CPSF6 nuclear import

As Ser phosphorylation is critical for the interaction between SRSF1 and TNPO3 (12) (see below), we next tested to what extent phosphorylation might contribute to CPSF6 NLS function. Due to the importance of D2 and D3 sequences (Figures 4, 5, Supplementary Figure S6), respective Ser residues Ser511, Ser513, and Ser525 were mutated individually to Ala or in S2A or S3A combinations (Figure

6A). Fractionation of transfected cells revealed that each single S>A mutant protein as well as the multi-mutated CPSF6[S2A] and CPSF6[S3A] proteins were, like the WT CPSF6[551] control, predominantly nuclear (Figure 6B; Supplementary Figure S1 and Table S5). We therefore extended the mutagenesis to a total of 9 predicted phosphoacceptor sites encompassing residues Ser487 to Ser525, which included all predicted sites within the RSLD (Figure 6A; Supplementary Table S4). The resulting CPSF6[S8YA] mutant protein also localized to the nucleus (Figure 6B, C, Supplementary Figure S1), indicating that CPSF6 nuclear import may occur independent of RSLD phosphorylation. To mimic the effect of constitutive RSLD phosphorylation (69,70), we assessed the S8YD mutant that carried aspartic acid in place of the predicted Ser and Tyr phosphoacceptor sites (Figure 6A). In stark contrast to the CPSF6[S8YA] mutant, CPSF6[S8YD] was defective for nuclear localization (Figure 6B, C, Supplementary Figure S1 and S7).

Due to the recent report that the nuclear translocation of *Drosophila* CPSF6 required RSLD phosphorylation (39), we next monitored the phosphorylation status of WT human and mutant CPSF6 proteins. Initially, SDS-PAGE patterns of predicted phosphoacceptor site mutants were analyzed in the presence of co-polymerized alkoxide-bridged dinuclear Mn^{2+}, also known as Phostag (73). This reagent interacts with phosphate groups of modified Thr, Ser or Tyr residues, retarding the mobility of phosphoproteins through the polyacrylamide matrix. Under baseline SDS-PAGE conditions, CPSF6[551] co-migrated with a 70 kDa molecular mass standard (Figures 1, 3–6). HEK293T cell CPSF6 as well as CPSF6[551] expressed in CKO cells however appeared as slowly migrating species in the presence of Phos-tag, indicating that most if not all cellular CPSF6 is phosphorylated (Figure

4672*Nucleic Acids Research, 2019, Vol. 47, No. 9*

Figure 4. Targeted Ala mutagenesis of select RSLD residues. (**A**) The CPSF6 RSLD sequence (residues 489–551) is shown highlighting D1-D3 sequence elements (underlined). Analyzed mutants are indicated along with the residues that were substituted by Ala. (**B**) Results of HIV infection (top) and expression levels of WT and mutant CPSF6 proteins (bottom). NS ($P > 0.05$); *$P \leq 0.05$; **$P < 0.01$; ****$P < 0.0001$. Other descriptions as in Figure 1E and F.

6D). Ser513 appeared to be the dominant phosphoacceptor site within D2 and D3, as the migration pattern of the CPSF6[551]S513A mutant protein in the presence of Phos-tag differed more from WT CPSF6[551] than did either the CPSF6[551]S511A or CPSF6[551]S525A pattern. As the migration pattern of CPSF6[S3A] in turn differed more from the CPSF6[551]S513A pattern than did CPSF6[S2A], it seemed that Ser525 could also be phosphorylated under these conditions. Because the migration patterns of CPSF6[S8YA] and CPSF6[S8YD] mutant proteins were the same in the presence and absence of Phos-tag (Figure 6D), we concluded that RSLD phosphorylation determined the SDS-PAGE mobility shift of endogenous CPSF6 protein. Consistent with this interpretation, WT CPSF6[551] co-migrated with the 70 kDa marker in the presence of Phos-tag if nuclear extracts were pretreated with λ phosphatase,

and the migration pattern of CPSF6[481], which lacked the RSLD, was independent of phosphatase treatment or Phos-tag reagent (Supplementary Figure S8A). Patterns of CPSF6 expression across cell types, including Jurkat T cells as well as resting and stimulated primary CD4+ T cells, mirrored those observed in HEK293T cells in the absence and presence of Phos-tag (Supplementary Figure S8B–D).

We next utilized MS to assess the phosphorylation status of CPSF6. Anti-CPSF6 antibodies recovered endogenous CPSF6[551] from HEK293T cell nuclear extracts and exogenously expressed CPSF6[551] and mutant derivatives from CKO cell extracts at similar levels (Figure 7A). Immunoblot analysis prior to immunoprecipitation importantly revealed expected migration patterns of WT and mutant proteins in the presence of Phos-tag (Figure 7B). We note that the assessment of potential CPSF6 phosphory-
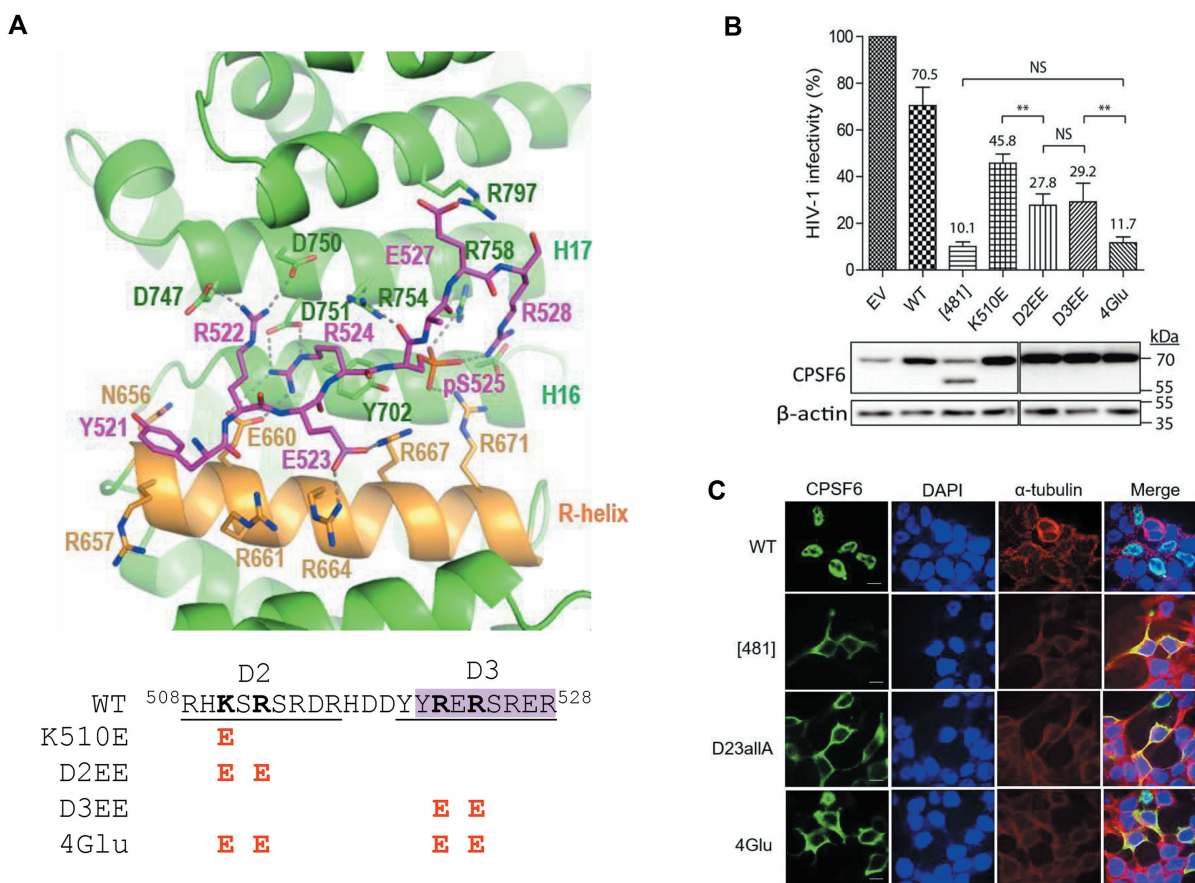
Figure 5. Charge reversal of four electropositive residues (K510, R512, R522, and R524) within the RSLD negates CPSF6 NLS function. (**A**) Details of the TNPO3–RSLD D3 region interaction. Dotted lines represent hydrogen bonds. See Figure 2 for additional labeling. Below is the amino acid sequence of the D2–D3 region (targeted basic residues in bold), color coded to denote the CPSF6 residues in the structure. The analyzed Glu substitution mutants are noted. (**B**) Results of HIV-1 infection (top) and expression levels of WT and mutant CPSF6 proteins (bottom). NS ($P > 0.05$); **$P < 0.01$. Other descriptions as in Figure 1E and F. (**C**) Confocal microscopy analysis of WT and indicated CPSF6 mutant proteins. Bars, 10 μm. Other descriptions as in Figure 1C.

lation sites varied widely across prediction program (Supplementary Table S4). Of 31 total sites, 8 were predicted by two programs and just two sites, Thr404 and Thr407, were predicted by all three (Supplementary Table S4). Indeed, Thr404 and Thr407 at virtual absolute certainty were identified by MS as phosphorylated across experimental samples (Figure 7C). Thr157 in the RRM domain was also phosphorylated in all samples, while Thr235 in the PRD was phosphorylated only in exogenously expressed proteins. Four sites in the RSLD, Ser494, Ser500, Ser511 and Ser513, were phosphorylated in both endogenous CPSF6 and exogenously expressed CPSF6[551] protein. Thus, while conserved Ser residues within D1 (Ser500) and D2 (Ser511 and Ser513) were identified at high confidence as modified, phosphoserine at residue Ser525 in D3 was not detected. The S513A substitution in CPSF6[551]S513A counteracted phosphorylation of both Ser511 and Ser513, indicating potential cooperativity in the modification of these two sites. As predicted from the Phos-tag analysis, RSLD phosphorylation was fully abrogated in the CPSF6[S8YA] mutant protein (Figure 7C).

To assess if the PY-NLS might cryptically activate and assume NLS function in the absence of RSLD phospho-

rylation, the ΔPY and S8YA mutations were combined. CPSF6[ΔPYS8YA] protein was nuclear as assessed by both confocal microscopy and cellular fractionation (Supplementary Figure S1 and S7).

Assessment of direct TNPO3–CPSF6 RSLD binding

GST pulldown assay. The results of previous experiments suggested that whereas the interaction between TNPO3 and the CPSF6 RSLD would depend on D2 and D3 sequences, it would be largely phosphorylation independent. To test these ideas, GST-RSLD fusion proteins containing the WT, D23allA or 4Glu mutant sequences were expressed in *E. coli* in the absence or presence of co-expressed CLK1 kinase. Following purification by glutathione sepharose and SEC, the proteins were analyzed by SDS-PAGE in the presence of Phos-tag, which confirmed that proteins expressed in the presence of CLK1 were indeed phosphorylated (Figure 8A). MS analysis of the WT protein expressed in the absence of CLK1 confirmed the lack of RSLD phosphorylation. By contrast, CLK1 co-expression resulted in the phosphorylation of four RSLD residues at ≥90% certainty, including Ser494, which was phosphorylated in endogenous

Figure 6. Analysis of predicted RSLD phosphoacceptor site mutant proteins. (**A**) CPSF6 residues 487–528 are shown highlighting predicted phosphoacceptor sites (red; Supplementary Table S4) and D1-D3 regions (underlined). Residues targeted by Ala or Asp substitution mutagenesis are indicated below. Residues determined by MS to be phosphorylated in HEK293T cell protein are indicated by asterisk (also see Figure 7C). (**B**) Fractionation of WT and mutant CPSF6 proteins. Results are representative of those observed in two independent experiments. Other descriptions as in Figure 1D. See Supplementary Figure S1 for quantitation of nuclear and cytoplasmic fractions and Supplementary Table S5 for associated statistics. (**C**) Confocal microscopy detection of WT and the indicated CPSF6 mutants. Results representative of those observed in two independent experiments. Other descriptions as in Figure 1C. (**D**) Phosphorylation status of CPSF6 variants was assessed by including Phos-tag (+) in polyacrylamide gels, followed by immunoblotting for CPSF6 (upper panels). Electrophoretic patterns in the absence of Phos-tag (–) were included for comparison (lower panels). Results are representative of those observed in two independent experiments. NE, nuclear extract; CE, cytoplasmic extract.

CPSF6 (Figures 7C and 8B). Phosphorylation of the GST-RSLD-WT protein at residues Ser487, Ser489 and Ser525, which was not detected in the endogenous protein sample, could be due to several reasons, including the lack of protein phosphatases in the bacterial system, expression of just the RSLD fragment of CPSF6, and/or co-expression of a non-physiologically relevant kinase. We do note that CLK2, which was not studied here, was predicted by the Scansite program (67) to modify CPSF6 RSLD residues in cell nuclei (Figure 7C).

TNPO3 binding was assessed by GST pulldown in the presence of BSA to address binding specificity. In repeat experiments, the non-phosphorylated GST-RSLD-WT protein recovered ∼30% of input TNPO3 without any evidence for BSA pulldown. The levels of TNPO3 recovery

with GST-RSLD-D23allA and GST-RSLD-4Glu were importantly reduced to the background level observed with GST alone control protein (Figure 8C, D). TNPO3 binding moreover was independent of RSLD phosphorylation, as the recovery of TNPO3 by GST-RSLD-S8YA and phosphorylated GST-RSLD-WT was indistinguishable from the level recovered by the non-phosphorylated GST-RSLD-WT protein (Figure 8C-F). In sharp contrast, GST-RSLD-S8YD failed to bind TNPO3 (Figure 8E, F). We conclude that TNPO3 binding profiles of recombinant GST-RSLD fusion proteins correlate succinctly with the nuclear import profiles of full-length CPSF6 counterpart proteins in HEK293T cells.

Peptide binding as assessed by FP. Our results highlight the contribution of D2 and D3 sequences in binding to
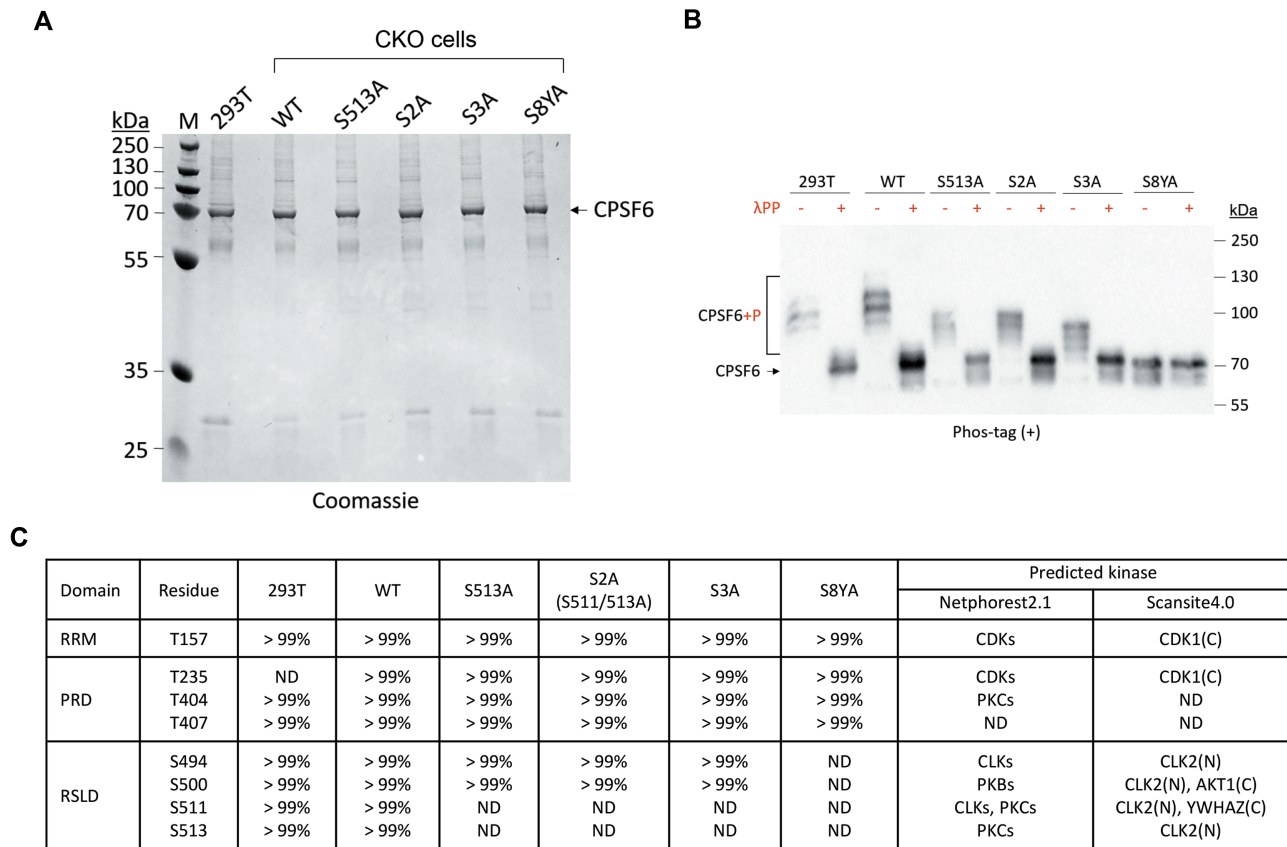
Figure 7. Quantitative assessment of CPSF6 phosphorylation. (**A**) Products of anti-CPSF6 immunoprecipitation from HEK293T cell nuclear extract as well as CKO cell extracts following transfection with WT CPSF6[551] or indicated phosphoacceptor site mutant expression vector. The gel was stained with Coomassie brilliant blue. (**B**) Prior to immunoprecipitation, extracts were fractionated by SDS-PAGE in the presence of Phos-tag and analyzed for CPSF6 content by immunoblotting; samples treated with λ protein phosphatase (PP) prior to SDS-PAGE are indicated by '+'. Panel A and B results are representative of those observed in two independent experiments; abutting numbers indicate migration positions of mass standards in kDa. (**C**) MS results for immunoprecipitated samples shown in panel A. Percent values indicate the level of certainty at which the indicated residue was identified as phosphorylated in at least one of two experimental replicates. ND, not detected. Netphorest 2.1 (66) and Scansite 4.0 (67) were used to predict potential interacting kinases; C and N indicates cytoplasm and nuclear, respectively. PKC, protein kinase C; PKB, protein kinase B; AKT1 (a.k.a. PKB-α), RAC-α serine/threonine-protein kinase; YWHAZ, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta; CDK, cyclin-dependent kinase.

TNPO3 and CPSF6 nuclear import (Figures 4, 5 and 8). While D2 residues Ser511 and Ser513 were phosphorylated in HEK293T cell nuclei, Ser525 in D3 scored as unmodified (Figure 7C). Binding affinities of synthetic RSLD peptides to recombinant TNPO3 were next measured by FP spectrometry to further investigate the contributions of phosphorylation, as well as specific D3 amino acid residues, to the protein-protein interaction. The canonical SRSF1 RS sequence was synthesized as an unmodified 8-mer peptide or with key Ser positions (analogous to Ser207 and Ser209) phosphorylated (RSP). While RS peptide failed to bind TNPO3 ($K_d > 1$ mM), RSP revealed a K_d of 1.3 ± 0.19 μM (Figure 9A; $P = 0.002$—see Supplementary Table S7 for study-wide comparisons of K_d value significance). CPSF6 peptides were initially synthesized as 8-mers that initiated with the first fully resolvable residue across D1–D3 sequences (Figure 2C). A wide range of binding affinities, which spanned 8.7 μM to >1 mM, were recorded for these unmodified peptides. Phosphorylation modestly enhanced the binding affinities of D1 (from 141 μM to 28 μM; $P = 0.03$) and D3 (from 8.7 μM to 3.4 μM; $P = 0.005$) peptides,

while phosphorylation significantly enhanced the binding affinity of the D2 peptide, from undetectable to 2.3 μM (Figure 9A). Thus, the presence of Glu in D1 and D3 at the position analogous to Ser207 in SRSF1 reduced the dependency on phosphorylation for binding to TNPO3 from all-or-none, in the cases of RS and D2 peptides, to ~5- and 2.5-fold for D1 and D3, respectively.

We next analyzed binding affinities of larger D2-D3 region peptides. Extending the N-terminus of D3 to include the amino acids that lie between D2 and D3 did not improve binding affinity over that seen with starting D3 peptide (Figure 9B). The further extended D23 peptide however did bind TNPO3 at a significantly higher affinity than starting D3 ($P = 0.039$), indicating context-dependent influence of unmodified D2 sequence to the protein-protein interaction. The affinity of D23 peptide-TNPO3 binding was importantly reduced significantly (~25-fold) by the 4Glu amino acid substitutions in the D23-4Glu peptide (Figure 9B; Supplementary Table S7). Results of competition assays for binding of the positive control SRSF1 RSP peptide also demonstrated the specificity of CPSF6 RSLD peptide bind-
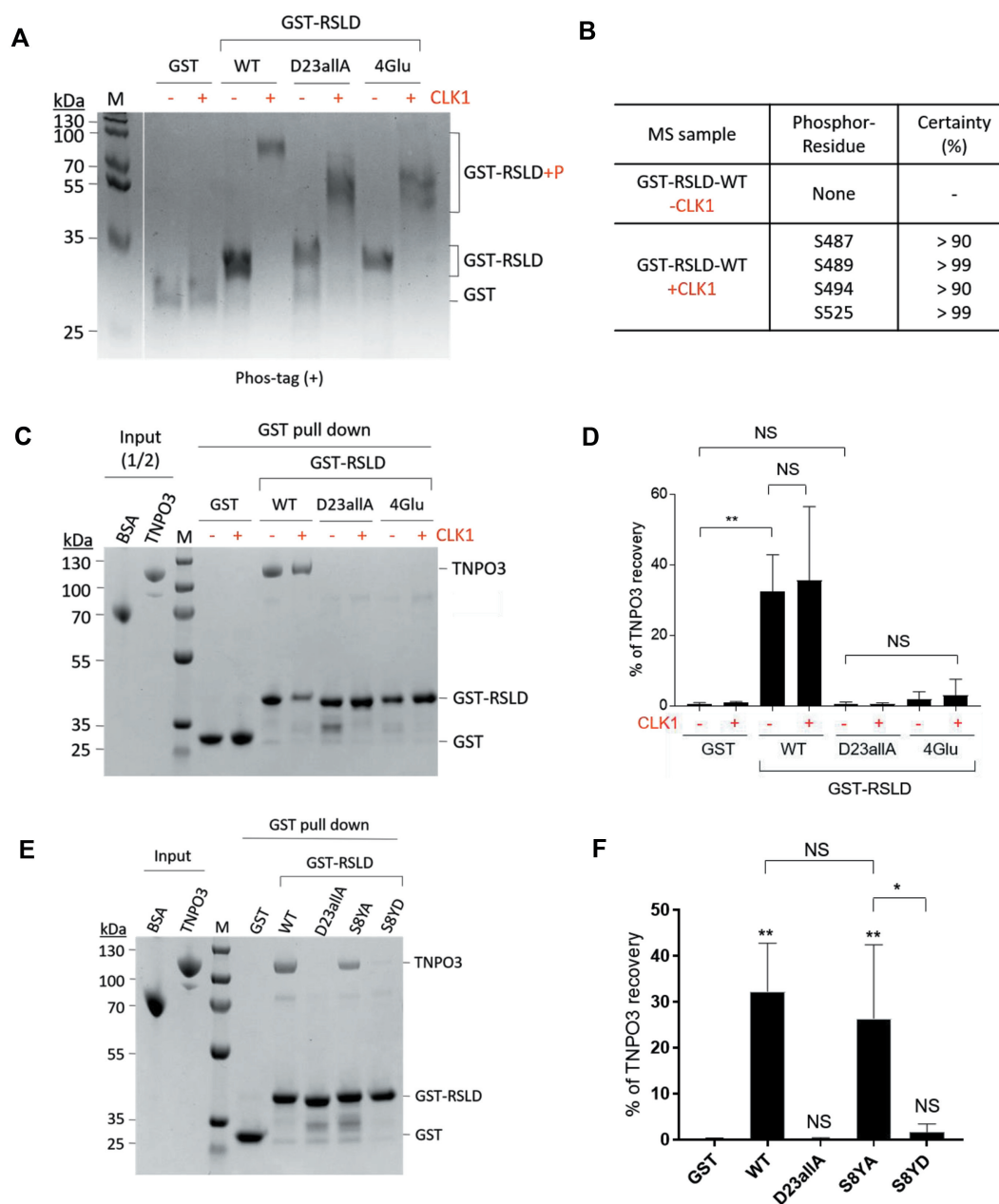
Figure 8. CPSF6 RSLD-TNPO3 binding *in vitro*. (**A**) GST and the indicated GST-RSLD fusion proteins, expressed in the absence or presence of CLK1 as indicated, were analyzed by SDS-PAGE in the presence of Phos-tag. (**B**) Phosphorylation sites identified by MS in GST-RSLD-WT protein expressed in absence or presence of CLK1. (**C**) Pulldown of TNPO3 with indicated WT or mutant RSLD protein that had been expressed in the presence or absence of CLK1. GST alone served as negative control. (**D**) Quantification of % TNPO3 recovery (mean ± standard deviation (SD) of two independent experiments) normalized to the amount of input GST-CPSF6 RSLD. (**E**) GST pulldown results; the indicated GST proteins were expressed in *E. coli* in the absence of CLK1 co-expression. (**F**) Quantitative analysis of TNPO3 recovery (mean ± SD of two independent experiments) normalized to the amount of input GST-CPSF6 RSLD. Gels were stained with Coomassie blue; numbers to the left of the gels are migration positions of mass standard (kDa). Panel D and F indicators: NS ($P > 0.05$); **$P < 0.01$; *$P < 0.05$.

ing to TNPO3 (Supplementary Figure S9). Phosphorylation of the D23 peptide enhanced TNPO3 binding affinity ∼3–4 fold, though this affect was independent of whether D2, D3, or both sequences were modified (compare D2P-D3, D2-D3P, and D2P-D3P binding affinities in Figure 9B).

Residue-by-residue alteration of the D3 peptide sequence highlighted the relative importance of Tyr at position 1 over phosphoserine at position 5 for TNPO3 binding. Substitut-

ing Glu for Tyr in the baseline D3 sequence significantly reduced binding, from 8.7 μM to >1 mM for the ER/S peptide ($P = 0.03$), while the same substitution in the modified D3P peptide reduced binding ∼5-fold, from 3.4 μM to 17.1 μM ($P = 0.002$). Similarly, incorporating Tyr at position 1 or phosphoserine at position 5 into the iterative ERERERER backbone increased affinity for TNPO3 binding by ∼10 and 5-fold, respectively, which was a statistically
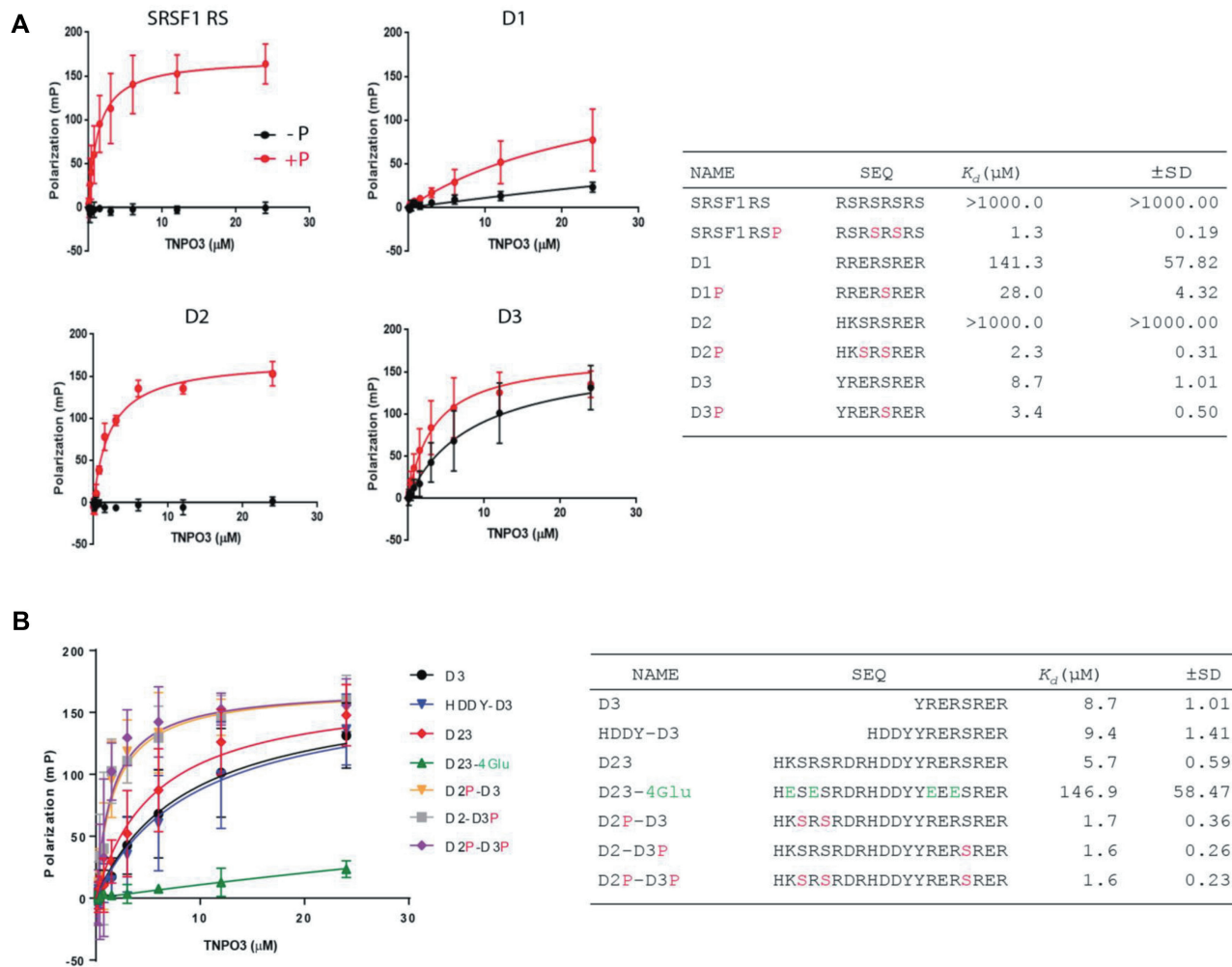
Figure 9. CPSF6 RSLD peptide binding to TNPO3 by FP spectrometry. (**A**) Binding profiles of SRSF1 RS and CPSF6 D1, D2 and D3 peptides to TNPO3; +P, phosphorylated peptide (red trace); -P, unphosphorylated peptide (black trace). Peptide sequences (red S, phosphoserine) and estimated equilibrium dissociation constants (K_d; mean ± SD of three independent experiments) of peptide-TNPO3 complexes are shown to the right. (**B**) Binding profiles of extended peptides to TNPO3. The Glu substitutions in the D23-4Glu peptide sequence are highlighted in green; other labeling as in panel A. See Supplementary Table S7 for statistical analysis of K_d value difference comparisons.

significant difference (Supplementary Figure S10 and Table S7).

Regulation of APA function by CPSF6 RSLD phosphorylation

The CFIm complex plays a key role to regulate APA. In the absence of CPSF5 or CPSF6, polyadenylation occurs at upstream sites that are proximal to translational stop codons, effectively shortening mRNA 3′ UTRs (19,55). To assess the role of CPSF6 phosphorylation in APA, RNA was isolated from CKO cells expressing HA-tagged CPSF6[551] or HA-CPSF6[S8YA]; added HA tags facilitated detection of potential CPSF6 interacting proteins by immunoprecipitation (see below). Negative controls included CKO cells transfected with EV or with vector expressing the RRM domain deletion mutant HA-CPSF6[Δ116–122], which is defective for CPSF5 binding (34,74). Filtered RNA-Seq reads were analyzed for percent distal polyadenylation usage index (PDUI) to assess 3′ UTR length changes. PDUI values

near zero indicate preferential use of proximal PASs while values closer to 1 indicate distal site usage (75). Test case PDUI values were compared to WT HEK293T cells transfected with EV. Two independent experiments were performed, and resulting data were parsed to highlight significant findings in common to both experimental replicates.

As expected (33,76), numerous genes with significant PDUI differences were evident in CKO cells transfected with EV, including 327 mRNAs with significantly shorter 3′ UTRs ($P < 10^{-300}$) and 2 mRNAs with longer 3′ UTRs ($P = 0.0002$) from an average of 9,884 messages analyzed (Figure 10A and Supplementary Figure S11A; see Supplementary Table S8 for list of all mRNAs with significant 3′ UTR changes). Also as expected (33), HA-CPSF6[551] expression in large part restored the normal APA pattern to CKO cells, reducing the number of mRNAs with shorter 3′ UTRs by approximately 10-fold (from 327 to 32; Figure 10). Despite efficient HA-CPSF6[Δ116–122] expression (Supplementary Figure S12A), the CPSF5 binding defective mutant by contrast failed to rescue APA function, as 321 mR-
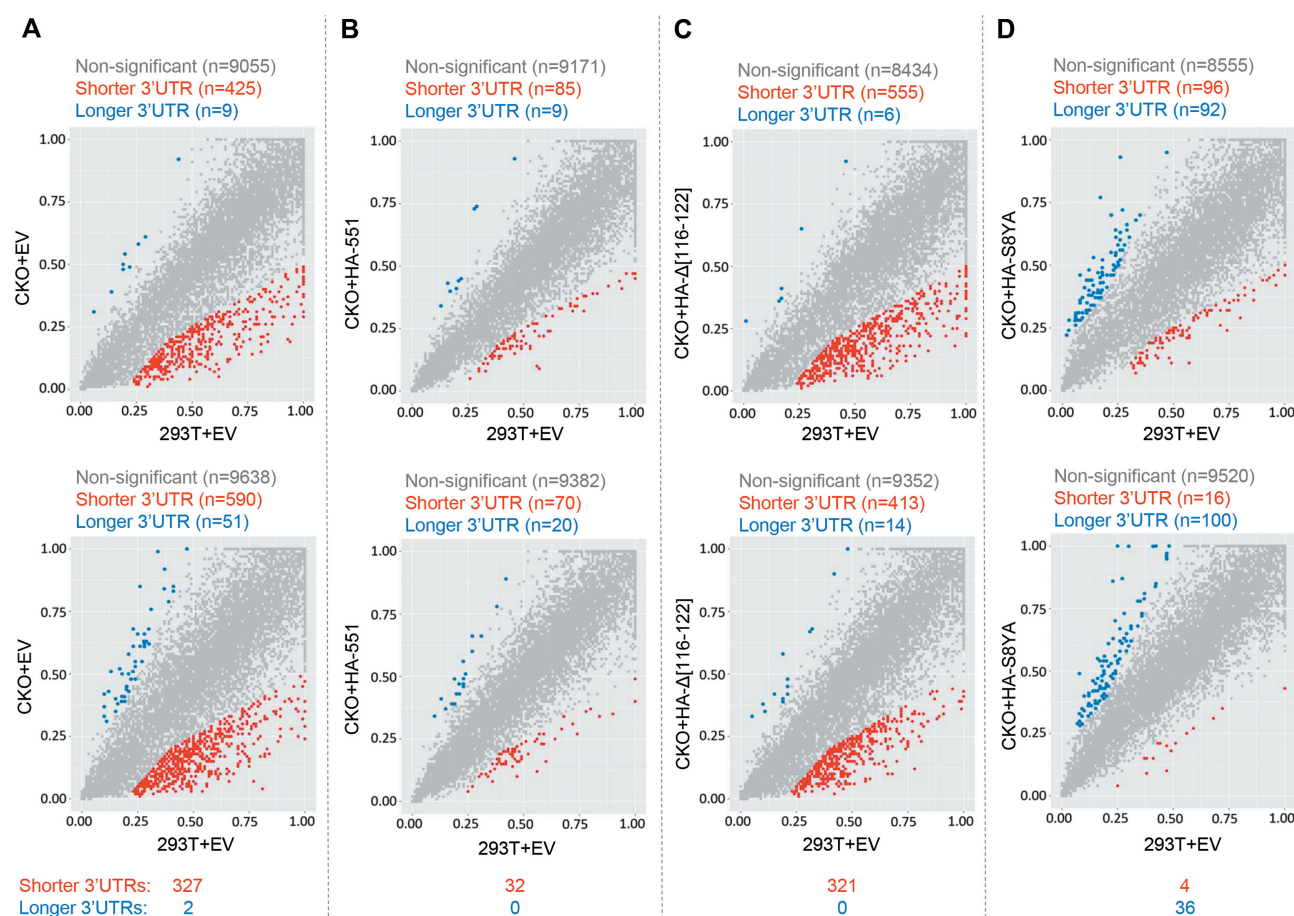
Figure 10. Regulation of APA by CPSF6 RSLD phosphorylation. (**A–D**) Scatter plots of PDUI values for WT HEK293T cells transfected with empty vector (EV; x-axes) versus CKO cells transfected with EV (A), WT HA-CPSF6[551] expression vector (B), mutant HA-CPSF6[Δ116–122] expression vector (C), or HA-CPSF6[S8YA] expression vector (D). Each dot represents one gene. Genes with significantly different PDUI values compared with WT HEK293T cells+EV are shown in blue (upper left quadrant) or in red (lower right). Results of each experimental replicate are shown as independent scatter plot; the number of genes with significantly different PDUI values compared with WT+EV in common to both replicates is shown below the plots (see Supplementary Figure S11A for associated *P* values).

NAs in this case harbored significantly shortened 3′ UTRs ($P < 10^{-300}$) (Figure 10C, Supplementary Figure S11A). We note significant overlap among mRNAs with shorter 3′ UTRs between samples (Supplementary Table S8): 240 of the genes identified in CKO cells transfected with EV versus HA-CPSF6[Δ116–122] expression vector were the same ($P < 10^{-300}$), and most of the messages (28 of 32) identified in CKO cells expressing WT HA-CPSF6[551] were likewise in common with either the EV or HA-CPSF6[Δ116–122] sample ($P = 7 \times 10^{-41}$; Supplementary Figure S11B).

Expression of HA-CPSF6[S8YA] in CKO cells yielded the unexpected phenotype of preferential utilization of distal PASs and, as a consequence, extended 3′ UTRs. In this case only 4 mRNAs harbored shorter 3′ UTRs, which, although a significant degree of overlap ($P = 3.6 \times 10^{-7}$), paled in comparison to the 36 genes between samples with extended 3′ UTRs ($P = 1.5 \times 10^{-64}$) (Figure 10D, Supplementary Figure 11A, C). To investigate this phenotype in detail, mRNAs with extended 3′ UTRs (Supplementary Table S8) were compared to those noted previously through the depletion of CPA components SNORD50A (20), PCF11 (21), or CstF64/CstF64τ (57), which revealed

one ($P > 0.05$), nine ($P = 9.6 \times 10^{-5}$), and two ($P = 0.02$) respective genes in common (Supplementary Figure S11D). Two genes, *POLR2E* and *MKLN1*, were moreover common to three of the four datasets (Supplementary Figure S11D). Alignment of select 3′ UTR regions between studies revealed similarities in PAS usage under the different experimental conditions (Supplementary Figure S11E).

We reasoned that CPSF6 RSLD phosphorylation could be important for interaction with APA factors that when depleted shift PAS site usage to downstream regions. To test this hypothesis, nuclear extracts of WT HEK293T cells, CKO cells, or CKO cells transfected with test constructs were immunoprecipitated using antibodies against CPSF5 or the HA tag. Anti-CPSF5 antibody recovered similar levels of CPSF7 across samples, though, as expected, failed to coimmunoprecipitate CPSF6 from CKO cells or from CKO cells expressing HA-CPSF6[Δ116–122] (Supplementary Figure S12A). Similarly, anti-HA antibodies failed to coimmunoprecipitate CPSF5 from extracts of HA-CPSF6[Δ116–122]-expressing cells. Similar levels of FIP1L1 protein were recovered from cells expressing HA-CPSF6[551] and RSLD mutant HA-CPSF6[S8YA] using
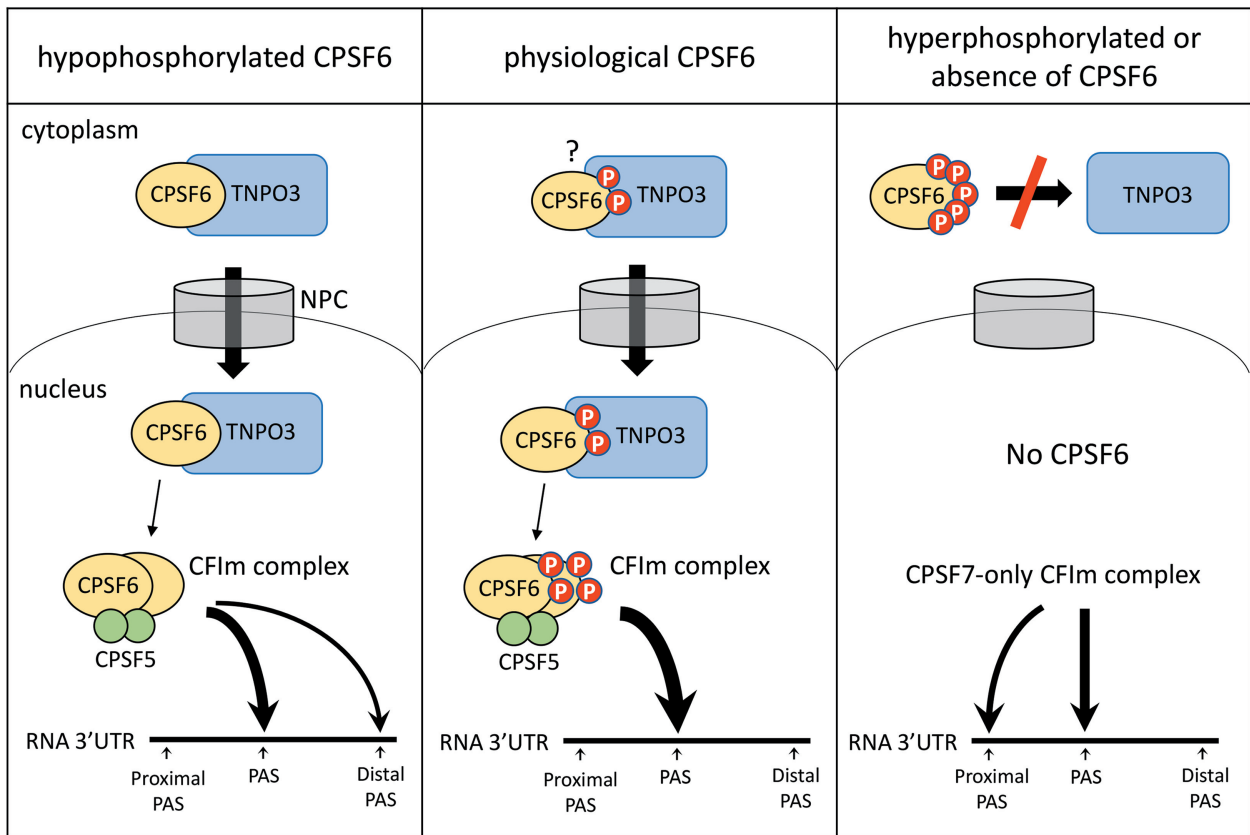
Figure 11. The role of RSLD phosphorylation in CPSF6 nuclear import and APA. The physiological condition is shown in the middle panel. As CPSF6 at steady-state is a nuclear protein, it is unclear if the RSLD is phosphorylated when TNPO3 is engaged in the cytoplasm for nuclear transport (?). Following CFIm complex formation in the nucleus, RSLD phosphorylation regulates the choice of pre-mRNA PAS. In the absence of RSLD phosphorylation (left panel), nuclear import proceeds efficiently, but RSLD hypophosphorylation results in a significant number of unnaturally distal PASs. When CPSF6 is absent from the nucleus due to gene knockout or expression of mutant protein unable to bind TNPO3 (right panel), PASs shift dramatically to proximal positions. NPC, nuclear pore complex.

both anti-CPSF5 and anti-HA antibodies, as well as with anti-HA antibody from extracts of cells expressing HA-CPSF6[Δ116–122] (Supplementary Figure S12B, C). These results are consistent with the observation that the interaction between CPSF6 and FIP1L1 is mediated by the RSLD in a phosphorylation independent manner (76). Despite repeated attempts, we were unable to convincingly coimmunoprecipitate PCF11 or CstF64/CstF64τ from nuclear extracts using either anti-CPSF5 or anti-HA antibodies.

DISCUSSION

Nuclear import mechanisms of SR and SR-like proteins

The nexus of the RSD-TNPO3 interaction and its role in protein nuclear import is best understood for canonical SR superfamily members. In these cases, the protein interaction relies on RSD phosphorylation (10–12) (Figure 9). However, the roles of RSDs/RSLDs in SR-like protein nuclear import are far less clear. Indeed, until this report it was not known if an interaction with TNPO3 would underlie the mechanism of SR-like protein nuclear import. The SR-like protein Acinus, as one example, is expressed as three isoforms, all with a common C-terminal region harboring an RRM domain and RSLD (77). A mutant of the S' isoform

with its RSLD deleted was transported efficiently into cell nuclei (78), indicating that Acinus nuclear import may very well be TNPO3-independent. Although the RSLD within SNRNP70 is karyophilic, nuclear import activity mapped to basic amino acid residues within the RRM domain of this splicing factor (79). Such observations highlight inherent complications of analyzing localization properties of modular molecules such as SR proteins that contain RRM and RSDs, each of which can influence nuclear targeting as well as subnuclear localization (72,78,80).

Prior work mapped human CPSF6 nuclear import function to the RSLD as well as an internal region between the RRM and RSLD (36). As purified TNPO3 and TNPO1 proteins were shown to bind these respective regions (12,38), we systematically assessed their contributions. Nuclear import function mapped to the CPSF6 RSLD without any discernable role for PY-NLS. Substituting Glu for four basic residues in the RSLD that are conserved within the core contact region of the phosphorylated SRSF1 RSD rendered CPSF6 cytoplasmic (Figure 5), disrupted transferrable NLS function (Supplementary Figure S6), and ablated the RSLD-TNPO3 interaction in vitro (Figure 8). Taken together, we conclude that the interaction detected in vitro between TNPO1 and residues 364–

392 within the PRD is irrelevant to CPSF6 nuclear import in cells. Although we have not examined the same construct that was analyzed by Dettwiler and colleagues (36), we suspect that the pan-cellular phenotype reported for this GFP fusion could be due to protein folding issues related to the removal of a relatively large internal chunk of CPSF6.

Differing roles for phosphorylation in SR protein biology

The roles of phosphorylation in SR protein biology are best understood for canonical family members such as SRSF1 (also known as ASF/SF2) and SRSF2 (previously called SC35), where the posttranslational modification regulates TNPO3 binding and hence nuclear import (10,11), localization to nuclear speckles (4), movement from speckles to sites of transcription (5,81) and assembly into the pre-mRNA spliceosome (7). Roles of phosphorylation in assembly include suppression of non-specific interactions with RNA as well as to orchestrate protein-protein interactions with other SR protein slicing factors, such as SNRNP70, through their RSDs [reviewed in (82)]. Catalysis of pre-mRNA splicing by contrast requires SR protein dephosphorylation (7), after which the proteins can recycle through speckle storage, spliceosome recruitment and assembly and catalysis of pre-mRNA splicing by additional rounds of phosphorylation/dephosphorylation (82).

Potential roles for phosphorylation in the biology of SR-like proteins are by comparison underdeveloped. The Acinus RSLD can be modified by different kinases including AKT (83) and SR protein-specific kinase 2 (SRPK2) (84). While the S453A/S604A phosphoacceptor site mutant like WT Acinus localized to nuclear speckles, the contribution of these two changes to the overall phosphorylation status of the WT or mutant protein was not reported (78). Prior findings differ as to whether the phosphomimetic S453D mutant localized to speckles (78) or dispersed throughout the nucleoplasm (84), which may have been influenced by the different Acinus isoforms analyzed (78).

The substitution of 14 predicted Ser/Thr phosphoacceptor sites in the PRD-RSLD linker and RSLD of *Drosophila* CPSF6 rendered the mutant protein defective for nuclear import (39). As that study discovered a link between CPSF6 and autophagy, nuclear import was conducted under conditions of nutrient deprivation, which could account for the differences between studies. We also note fairly different RSLD sequences between species: whereas 63 residues compose the human RSLD, the *Drosophila* domain harbors 132 residues. Sites of *Drosophila* RSLD phosphorylation identified from in vitro kinase reactions mapped to the segment of the protein that lacked homology to human CPSF6 (Supplementary Figure S13), which could also contribute to differences between studies. Treatment of nutrient-deprived human cells with 50 μM TG003, which at this concentration is a pantropic inhibitor of CLK kinases as well as dual specificity tyrosine-phosphorylated-regulated kinase 1A and 1B (85), reportedly inhibited human CPSF6 nuclear import (39). As such treatment almost certainly dysregulated the phosphorylation of numerous cellular proteins, additional work is required to determine if RSLD phosphorylation is specifically required for nuclear import of human CPSF6 under limited nutrient conditions.

MS identified CPSF6 RSLD D1 and D2 Ser residues as phosphorylated in HEK293T cell nuclei, while Ser525 in D3 was apparently unmodified. Phosphorylation of Ser525 in recombinant GST-RSLD-WT protein that was synthesized in the presence of CLK1 (Figure 8B) accounts for the anomalous X-ray scattering observed at position 6 of the bound peptide (Supplementary Figure S4). Surely, phosphorylation of D2 residues Ser511 and Ser513 can significantly increase the affinity for binding of CPSF6 RSLD sequences to TNPO3. However, our data clarify this affect as highly context dependent. While all-or-none in the context of the 8-mer D2 peptide, D2 region phosphorylation increased the binding affinity of the 20-mer D23 peptide circa 3.3-fold (Figure 9). We speculate that increased local concentration of separate TNPO3 binding sites accounts for the apparent avidity of D2 and D3 sequences in peptide binding (Figure 9) and CPSF6 nuclear import (Figures 4, 5). While we cannot rule out with absolute certainty some contribution from RSLD phosphorylation to TNPO3 binding in the context of the complete domain and/or full-length CPSF6, we conclude that human CPSF6 is imported into the nucleus in a largely RSLD phosphorylation-independent manner under normal growth conditions.

Our data suggests caution may be warranted when interpreting results of protein mobility shift in Phos-tag gels. Though an excellent predictor of some modifications, RSLD mutant CPSF6[S8YA] seemed fully dephosphorylated based on migration pattern and phosphatase sensitivity (Figures 6D and 7B). However, MS identified three upstream phosphothreonine residues in this protein (Figure 7C). Perhaps the relatively high Pro content in the central portion of CPSF6 masked these residues from effectively engaging the Phos-tag. Gel mobility patterns also suggested that Ser525 could be phosphorylated (Figures 6D and 7B), though this modification was not confirmed by MS in human CPSF6.

Regulation of APA by CPSF6 RSLD phosphorylation

The hypophosphorylated CPSF6[S8YA] mutant supported significant distal PAS utilization (Figures 10, 11 and Supplementary Figure S11), suggesting that RSLD phosphorylation may regulate the functionality of other CPA complex components. The interaction between CPSF6 and FIP1L1, which is mediated through their RSLDs (76), occurs independent of phosphorylation (76) (Supplementary Figure S12). Because RSD phosphorylation is known to regulate interactions between SR proteins (8,82,86), it seemed possible that phosphorylation of the CPSF6 RSLD could mediate its interaction with other nuclear factors that regulate APA. Due to the overlap observed in some 3′ UTR signatures in the presence of CPSF6[S8YA] or when CstF64/CstF64τ or PCF11 expression is dysregulated, we focused effort on potential interactions with these proteins. Considering their participation in APA as CstF and CFIIm complex components, it is not surprising that CPSF6 was previously reported to co-fractionate with CstF64 and PCF11 (87,88). However, we are unaware of any reports of direct interactions between these proteins and CPSF6, and, despite extensive effort, we were unable to coimmunoprecipitate PCF11, CstF64, or CstF64τ from nuclear extracts

using antibodies direct against CPSF5 or CPSF6. Although our results highlight a role for RSLD phosphorylation in the regulation of APA, additional work is required to decipher the precise mechanism that is provided by the post-translational modification. The CPSF6[S8YA] mutant provides a valuable tool to further probe functionally relevant intermolecular interactions among the myriad of players that regulate APA of cellular pre-mRNA.

DATA AVAILABILITY

The coordinates and structure factors for the RSLD-TNPO3 complex were deposited with the Protein Data Bank under accession code 6GX9. The RNA sequences for the PDUI analyses have been deposited in the National Center for Biotechnology Sequence Read Archive (NCBI SRA) under accession number PRJNA520804. MS data were deposited to the ProteomeXchange Consortium via the Proteomics Identifications (PRIDE) (89) partner repository under database identifier PXD012713.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Shepard,P.J. and Hertel,K.J. (2009) The SR protein family. *Genome Biol.*, **10**, 242.
2. Long,J.C. and Caceres,J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
3. Spector,D.L. (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.*, **9**, 265–315.
4. Galganski,L., Urbanek,M.O. and Krzyzosiak,W.J. (2017) Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.*, **45**, 10350–10368.
5. Colwill,K., Pawson,T., Andrews,B., Prasad,J., Manley,J.L., Bell,J.C. and Dunca,P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.*, **15**, 265–275.
6. Misteli,T., Cáceres,J.F., Clement,J.Q., Krainer,A.R., Wilkinson,M.F. and Spector,D.L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J. Cell Biol.*, **143**, 297–307.
7. Mermoud,J.E., Cohen,P.T. and Lamond,A.I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.*, **13**, 5679–5688.
8. Xiao,S.H. and Manley,J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
9. Cao,W., Jamison,S.F. and Garcia-Blanco,M.A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA*, **3**, 1456–1467.
10. Lai,M.C., Lin,R.I., Huang,S.Y., Tsai,C.W. and Tarn,W.Y. (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J. Biol. Chem.*, **275**, 7950–7957.
11. Lai,M.C., Lin,R.I. and Tarn,W.Y. (2001) Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10154–10159.
12. Maertens,G.N., Cook,N.J., Wang,W., Hare,S., Gupta,S.S., Oztop,I., Lee,K., Pye,V.E., Cosnefroy,O., Snijders,A.P. *et al.* (2014) Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2728–2733.
13. Blencowe,B.J., Bowman,J.A., McCracken,S. and Rosonina,E. (1999) SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.*, **77**, 277–291.
14. Boucher,L., Ouzounis,C.A., Enright,A.J. and Blencowe,B.J. (2001) A genome-wide survey of RS domain proteins. *RNA*, **7**, 1693–1701.
15. Rüegsegger,U., Blank,D. and Keller,W. (1998) Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol. Cell*, **1**, 243–253.
16. Tian,B. and Manley,J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
17. Rüegsegger,U., Beyer,K. and Keller,W. (1996) Purification and characterization of human cleavage factor I involved in the 3′ end processing of messenger RNA precursors. *J. Biol. Chem.*, **271**, 6107–6113.
18. Gruber,A.R., Martin,G., Keller,W. and Zavolan,M. (2012) Cleavage factor Im is a key regulator of 3′ UTR length. *RNA Biol.*, **9**, 1405–1412.
19. Martin,G., Gruber,Andreas R., Keller,W. and Zavolan,M. (2012) Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Rep.*, **1**, 753–763.
20. Huang,C., Shi,J., Guo,Y., Huang,W., Huang,S., Ming,S., Wu,X., Zhang,R., Ding,J., Zhao,W. *et al.* (2017) A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.*, **45**, 8647–8660.
21. Ogorodnikov,A., Levin,M., Tattikota,S., Tokalov,S., Hoque,M., Scherzinger,D., Marini,F., Poetsch,A., Binder,H., Macher-Göppinger,S. *et al.* (2018) Transcriptome 3′end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.*, **9**, 5331.
22. Hidalgo-Curtis,C., Chase,A., Drachenberg,M., Roberts,M.W., Finkelstein,J.Z., Mould,S., Oscier,D., Cross,N.C.P. and Grand,F.H. (2008) The t(1;9)(p34;q34) and t(8;12)(p11;q15) fuse pre-mRNA processing proteins SFPQ (PSF) and CPSF6 to ABL and FGFR1. *Genes Chromosomes Cancer*, **47**, 379–385.
23. Naumann,N., Schwaab,J., Metzgeroth,G., Jawhar,M., Haferlach,C., Göhring,G., Schlegelberger,B., Dietz,C.T., Schnittger,S., Lotfi,S. *et al.* (2015) Fusion of PDGFRB to MPRIP, CPSF6, and GOLGB1 in three patients with eosinophilia-associated myeloproliferative neoplasms. *Genes Chromosomes Cancer*, **54**, 762–770.
24. Binothman,N., Hachim,I.Y., Lebrun,J.-J. and Ali,S. (2017) CPSF6 is a clinically relevant breast cancer vulnerability target: Role of CPSF6 in breast cancer. *EBioMedicine*, **21**, 65–78.
25. Liu,T., Wen,L., Yuan,H., Wang,Y., Yao,L., Xu,Y., Cen,J., Ruan,C., Wu,D. and Chen,S. (2018) Identification of novel recurrent

CPSF6-RARG fusions in acute myeloid leukemia resembling acute promyelocytic leukemia. *Blood*, **131**, 1870–1873.

26. Qin,Y.-Z., Huang,X.-J. and Zhu,H.-H. (2018) Identification of a novel CPSF6-RARG fusion transcript in acute myeloid leukemia resembling acute promyelocytic leukemia. *Leukemia*, **32**, 2285–2287.

27. Henning,M.S., Dubose,B.N., Burse,M.J., Aiken,C. and Yamashita,M. (2014) In vivo functions of CPSF6 for HIV-1 as revealed by HIV-1 capsid evolution in HLA-B27-positive subjects. *PLoS Pathog.*, **10**, e1003868.

28. Lee,K., Ambrose,Z., Martin,T.D., Oztop,I., Mulky,A., Julias,J.G., Vandegraaff,N., Baumann,J.G., Wang,R., Yuen,W. *et al.* (2010) Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*, **7**, 221–233.

29. Price,A.J., Fletcher,A.J., Schaller,T., Elliott,T., Lee,K., KewalRamani,V.N., Chin,J.W., Towers,G.J. and James,L.C. (2012) CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog.*, **8**, e1002896.

30. Peng,K., Muranyi,W., Glass,B., Laketa,V., Yant,S.R., Tsai,L., Cihlar,T., Müller,B. and Kräusslich,H.G. (2014) Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife*, **3**, e04114.

31. Chin,C.R., Perreira,J.M., Savidis,G., Portmann,J.M., Aker,A.M., Feeley,E.M., Smith,M.C. and Brass,A.L. (2015) Direct visualization of HIV-1 replication intermediates shows that capsid and CPSF6 modulate HIV-1 intra-nuclear invasion and integration. *Cell Rep.*, **13**, 1717–1731.

32. Bejarano,D.A., Peng,K., Laketa,V., Börner,K., Jost,K.L., Lucic,B., Glass,B., Lusic,M., Müller,B. and Kräusslich,H.-G. (2019) HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the Nuclear Pore Complex. *eLife*, **8**, e41800.

33. Sowd,G.A., Serrao,E., Wang,H., Wang,W., Fadel,H.J., Poeschla,E.M. and Engelman,A.N. (2016) A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1054–E1063.

34. Rasheedi,S., Shun,M.-C., Serrao,E., Sowd,G.A., Qian,J., Hao,C., Dasgupta,T., Engelman,A.N. and Skowronski,J. (2016) The cleavage and polyadenylation specificity factor 6 (CPSF6) subunit of the capsid-recruited pre-messenger RNA cleavage factor I (CFIm) complex mediates HIV-1 integration into genes. *J. Biol. Chem.*, **291**, 11809–11819.

35. Achuthan,V., Perreira,J.M., Sowd,G.A., Puray-Chavez,M., McDougall,W.M., Paulucci-Holthauzen,A., Wu,X., Fadel,H.J., Poeschla,E.M., Multani,A.S. *et al.* (2018) Capsid-CPSF6 interaction licenses nuclear HIV-1 trafficking to sites of viral DNA integration. *Cell Host Microbe*, **24**, 392–404.

36. Dettwiler,S., Aringhieri,C., Cardinale,S., Keller,W. and Barabino,S.M.L. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J. Biol. Chem.*, **279**, 35788–35797.

37. Mattaj,I.W. and Englmeier,L. (1998) Nucleoplasmic transport: The soluble phase. *Annu. Rev. Biochem.*, **67**, 265–306.

38. Lee,B.J., Cansizoglu,A.E., Suel,K.E., Louis,T.H., Zhang,Z. and Chook,Y.M. (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell*, **126**, 543–558.

39. Tang,H.-W., Hu,Y., Chen,C.-L., Xia,B., Zirin,J., Yuan,M., Asara,J.M., Rabinow,L. and Perrimon,N. (2018) The TORC1-Regulated CPA complex rewires an RNA processing network to drive autophagy and metabolic reprogramming. *Cell Metab.*, **27**, 1040–1054.

40. Hori,T., Takeuchi,H., Saito,H., Sakuma,R., Inagaki,Y. and Yamaoka,S. (2013) A carboxy-terminally truncated human CPSF6 lacking residues encoded by exon 6 inhibits HIV-1 cDNA synthesis and promotes capsid disassembly. *J. Virol.*, **87**, 7726–7736.

41. Yunus,A.A. and Lima,C.D. (2009) Purification of SUMO conjugating enzymes and kinetic analysis of substrate conjugation. *Methods Mol. Biol.*, **497**, 167–186.

42. Shun,M.-C., Raghavendra,N.K., Vandegraaff,N., Daigle,J.E., Hughes,S., Kellam,P., Cherepanov,P. and Engelman,A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.

43. Koh,Y., Wu,X., Ferris,A.L., Matreyek,K.A., Smith,S.J., Lee,K., KewalRamani,V.N., Hughes,S.H. and Engelman,A. (2013) Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. *J. Virol.*, **87**, 648–658.

44. Shun,M.-C., Daigle,J.E., Vandegraaff,N. and Engelman,A. (2007) Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular Emerin protein. *J. Virol.*, **81**, 166–172.

45. Maertens,G., Cherepanov,P., Debyser,Z., Engelborghs,Y. and Engelman,A. (2004) Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *J. Biol. Chem.*, **279**, 33421–33429.

46. Cherepanov,P., Devroe,E., Silver,P.A. and Engelman,A. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.*, **279**, 48883–48892.

47. Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271–281.

48. Evans,P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 282–292.

49. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

50. Tsirkone,V.G., Beutels,K.G., Demeulemeester,J., Debyser,Z., Christ,F. and Strelkov,S.V. (2014) Structure of transportin SR2, a karyopherin involved in human disease, in complex with Ran. *Acta Crystallogr. F Struct. Biol. Commun.*, **70**, 723–729.

51. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

52. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

53. Lu,R., Limon,A., Devroe,E., Silver,P.A., Cherepanov,P. and Engelman,A. (2004) Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. *J. Virol.*, **78**, 12735–12746.

54. Wang,H., Shun,M.-C., Dickson,A.K. and Engelman,A.N. (2015) Embryonic lethality due to arrested cardiac development in Psip1/Hdgfrp2 double-deficient mice. *PLoS one*, **10**, e0137797.

55. Masamha,C.P., Xia,Z., Yang,J., Albrecht,T.R., Li,M., Shyu,A.B., Li,W. and Wagner,E.J. (2014) CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, **510**, 412–416.

56. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.

57. Yao,C., Biesinger,J., Wan,J., Weng,L., Xing,Y., Xie,X. and Shi,Y. (2012) Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 18773–18778.

58. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

59. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.

60. Zhou,X., Maricque,B., Xie,M., Li,D., Sundaram,V., Martin,E.A., Koebbe,B.C., Nielsen,C., Hirst,M., Farnham,P. *et al.* (2011) The human epigenome browser at Washington University. *Nat. Methods*, **8**, 989–990.

61. Peng,J. and Gygi,S.P. (2001) Proteomics: the move to mixtures. *J. Mass Spectrom.*, **36**, 1083–1091.

62. Beausoleil,S.A., Villén,J., Gerber,S.A., Rush,J. and Gygi,S.P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.*, **24**, 1285–1292.

63. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
64. Ullah,S., Lin,S., Xu,Y., Deng,W., Ma,L., Zhang,Y., Liu,Z. and Xue,Y. (2016) dbPAF: an integrative database of protein phosphorylation in animals and fungi. *Sci. Rep.*, **6**, 23534.
65. Gnad,F., Gunawardena,J. and Mann,M. (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
66. Miller,M.L., Jensen,L.J., Diella,F., Jørgensen,C., Tinti,M., Li,L., Hsiung,M., Parker,S.A., Bordeaux,J., Sicheritz-Ponten,T. *et al.* (2008) Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.*, **1**, ra2.
67. Obenauer,J.C., Cantley,L.C. and Yaffe,M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.
68. Plaisier,S.B., Taschereau,R., Wong,J.A. and Graeber,T.G. (2010) Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.*, **38**, e169.
69. Trautwein,C., Caelles,C., van der Geer,P., Hunter,T., Karin,M. and Chojkier,M. (1993) Transactivation by NF-IL6/LAP is enhanced by phosphorylation of its activation domain. *Nature*, **364**, 544–547.
70. Maciejewski,P.M., Peterson,F.C., Anderson,P.J. and Brooks,C.L. (1995) Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *J. Biol. Chem.*, **270**, 27661–27665.
71. Yoneda,Y., Arioka,T., Imamoto-Sonobe,N., Sugawa,H., Shimonishi,Y. and Uchida,T. (1987) Synthetic peptides containing a region of SV 40 large T-antigen involved in nuclear localization direct the transport of proteins into the nucleus. *Exp. Cell Res.*, **170**, 439–452.
72. Caceres,J.F., Misteli,T., Screaton,G.R., Spector,D.L. and Krainer,A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.*, **138**, 225–238.
73. Kinoshita,E., Kinoshita-Kikuta,E., Takiyama,K. and Koike,T. (2006) Phosphate-binding tag, a new tool to visualize phosphorylated proteins. *Mol. Cell. Proteomics*, **5**, 749–757.
74. Yang,Q., Coseno,M., Gilmartin,G.M. and Doublié,S. (2011) Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure*, **19**, 368–377.
75. Xia,Z., Donehower,L.A., Cooper,T.A., Neilson,J.R., Wheeler,D.A., Wagner,E.J. and Li,W. (2014) Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat. Commun.*, **5**, 5274.
76. Zhu,Y., Wang,X., Forouzmand,E., Jeong,J., Qiao,F., Sowd,G.A., Engelman,A.N., Xie,X., Hertel,K.J. and Shi,Y. (2018) Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol. Cell*, **69**, 62–74.
77. Sahara,S., Aoto,M., Eguchi,Y., Imamoto,N., Yoneda,Y. and Tsujimoto,Y. (1999) Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. *Nature*, **401**, 168–173.
78. Wang,F., Wendling,K.S., Soprano,K.J. and Soprano,D.R. (2014) The SAP motif and C-terminal RS- and RD/E-rich region influences the sub-nuclear localization of Acinus isoforms. *J. Cell Biochem.*, **115**, 2165–2174.
79. Romac,J.M., Graff,D.H. and Keene,J.D. (1994) The U1 small nuclear ribonucleoprotein (snRNP) 70K protein is transported independently of U1 snRNP particles via a nuclear localization signal in the RNA-binding domain. *Mol. Cell. Biol.*, **14**, 4662–4670.
80. Cazalla,D., Zhu,J., Manche,L., Huber,E., Krainer,A.R. and Cáceres,J.F. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol. Cell. Biol.*, **22**, 6871–6882.
81. Gui,J.F., Lane,W.S. and Fu,X.D. (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*, **369**, 678–682.
82. Zhou,Z. and Fu,X.D. (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, **122**, 191–207.
83. Hu,Y., Yao,J., Liu,Z., Liu,X., Fu,H. and Ye,K. (2005) Akt phosphorylates acinus and inhibits its proteolytic cleavage, preventing chromatin condensation. *EMBO J.*, **24**, 3543–3554.
84. Jang,S.-W., Yang,S.-J., Ehlén,Å., Dong,S., Khoury,H., Chen,J., Persson,J.L. and Ye,K. (2008) Serine/arginine protein–specific kinase 2 promotes leukemia cell proliferation by phosphorylating Acinus and regulating Cyclin A1. *Cancer Res.*, **68**, 4559–4570.
85. Muraki,M., Ohkawara,B., Hosoya,T., Onogi,H., Koizumi,J., Koizumi,T., Sumi,K., Yomoda,J.-I., Murray,M.V., Kimura,H. *et al.* (2004) Manipulation of alternative splicing by a newly developed inhibitor of Clks. *J. Biol. Chem.*, **279**, 24246–24254.
86. Yeakley,J.M., Tronchere,H., Olesen,J., Dyck,J.A., Wang,H.Y. and Fu,X.D. (1999) Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J. Cell Biol.*, **145**, 447–455.
87. Havugimana,Pierre C., Hart,G.T., Nepusz,T., Yang,H., Turinsky,A.L., Li,Z., Wang,Peggy I., Boutz,Daniel R., Fong,V., Phanse,S. *et al.* (2012) A census of human soluble protein complexes. *Cell*, **150**, 1068–1081.
88. Wan,C., Borgeson,B., Phanse,S., Tu,F., Drew,K., Clark,G., Xiong,X., Kagan,O., Kwan,J., Bezginov,A. *et al.* (2015) Panorama of ancient metazoan macromolecular complexes. *Nature*, **525**, 339–344.
89. Perez-Riverol,Y., Csordas,A., Bai,J., Bernal-Llinares,M., Hewapathirana,S., Kundu,D.J., Inuganti,A., Griss,J., Mayer,G., Eisenacher,M. *et al.* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.*, **47**, D442–D450.