Science
AAAS

Supplementary Materials for

A ubiquitous disordered protein interaction module orchestrates transcription elongation

Katerina Cermakova†, Jonas Demeulemeester†, Vanda Lux†, Monika Nedomova, Seth R. Goldman, Eric A. Smith, Pavel Srb, Rozalie Hexnerova, Milan Fabry, Marcela Madlikova, Magdalena Horejsi, Jan De Rijck, Zeger Debyser, Karen Adelman, H. Courtney Hodges*, and Vaclav Veverka*

*Correspondence to: chodges@bcm.edu and veverka@uochb.cas.cz
†These authors contributed equally to the manuscript

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Data S1 to S3 (.xlsx)
MDAR Reproducibility Checklist

Materials and Methods

Analysis of domain enrichment and interactomes

We obtained a complete list of proteins in human cells and a list of structural folds associated with each of the proteins as defined by their PFAM (Protein Families database) number in the UniProt database (*44*, *45*). We calculated the expected frequency of appearance of each of the known structural folds (represented by unique PFAM number) in the human proteome by counting the total number of proteins harboring each of the structural folds divided by total number of proteins.

We obtained lists of proteins associated with different processes from GO terms (**Table S1**) and a list of unique domains (unique PFAM numbers) associated with each of these proteins. We calculated the frequency of occurrence of each structural fold in each GO term as well as the enrichment over expected frequency (hypergeometric test). Experimentally validated interactions between studied factors were obtained from the STRING database (*46*). All data are summarized in **Data S1**. All databases were accessed on 2020-04-05.

Identification of TFIIS N-terminal domain Interacting Motifs (TIMs)

We recently reported structural conservation of motifs recognizing the LEDGF TND in LEDGF interaction partners (*27*, *30*, *31*). Comparing structural similarities between the human LEDGF TND in complex with binding partners (*27*, *30*, *31*, *47*) and Iws1–Spt6 complexes from distantly related species (*32*, *33*) (**Fig. S3A-B**) allowed us to define the following characteristic features of TND-interacting motifs (TIMs): (a) localization within an intrinsically disordered region of a nuclear protein; (b) a region containing bulky hydrophobic residues with high helical propensity; followed by (c) a flexible stretch enriched in acidic (E or D) residues and/or residues that could be phosphorylated to yield a negative charge (S or T, **Fig. S3**).

We aligned our previously identified LEDGF-interacting motifs with the published TIMs in yeast Iws1, Spt18 and Med13 (*32*) to construct a [DENLA] [VILQF] [FLV] [DESNG] [DESNG] [DESNG] . [DESNG] [DESNG] consensus TIM (**Fig. S3**). Subsequently, we used this consensus to discover putative novel human TIM instances in a proteome-wide search using SLiMSearch4 (*48*), which takes into account the relative conservation of the identified motif compared to its surroundings (Relative Localization Conservation (*49*)) and its accessibility (IUPred2 disorder and ANCHOR2 binding scores (*50*)). A total of 376 motif hits were obtained

representing 0.0034% of the possible sequences in the human proteome. Putative TIMs were further filtered based on their presence outside of globular domains (yielding 300 putative TIMs in 285 proteins) and nuclear localization of the proteins (150 putative TIMs in 140 proteins, see **Data S2**). Altogether, this search yielded list of 81 putative TIMs present in 71 different transcription regulators which was further curated by ensuring (1) conservation of the putative TIM across species, (2) location in unstructured protein sequences (*51*), (3) predicted presence of helical propensity (*52*) and (4) importance for transcription elongation and regulation of RNAP2 activity in available literature.

Cloning

Constructs for recombinant bacterial expression. For expression in *E. coli*, the coding sequences for the TNDs of ELOA (aa 26-108), TFIIS (aa 1-80), LEDGF (aa 345–431), HRP2 (aa 469-549), MED26 (aa 1-87), PPP1R10 (aa 1-147) and IWS1(aa 550-692) were amplified and cloned into a modified T7 promoter-driven pMCSG7 vector described previously (*31*). This vector encodes an N-terminal His_6 affinity tag followed by the tobacco etch virus (TEV) protease recognition site. Upon TEV cleavage, the encoded proteins retained a cloning artifact of 5 amino acid residues (SNAAS) at their N-terminus and 2 amino acid residues (GS) at the C-terminus. For structural determination of the TFIIS(TND)-IWS1(TIM) complex, a chimeric construct was prepared similarly as described previously for other complexes (*27*). The chimera was constructed by attaching IWS1 (aa 446-571) to the C-terminal end of TFIIS TND. Upon expression, the TND and IWS1 portions of the chimera are connected by $(GS)_5SG$ flexible linker through BamHI restriction site.

Constructs for human cells. The plasmids for the fluorescent two-hybrid protein-protein interaction assay were constructed by inserting EGFP-LacI-NLS amplified from pKG215 (a gift from Iain Cheeseman (Addgene plasmid #45110; http://n2t.net/addgene:45110; RRID: Addgene_45110)), and TagRFP-NLS into pRRL-CAG, pRRL-CAG-IRES-puro and pRRL-CAG-IRES-neo lentiviral transfer vectors. Subsequently, InFusion was used to prepare the library of TNDs and TIMs in these vectors. In particular, insertion of IWS1-TIM (aa 445-510) and SPT6-TIM (aa 2-248) at the N-terminal end of EGFP-LacI-NLS sequence resulted in pRRL-CAG vector harboring TIM-EGFP-LacI-NLS coding sequence. The library of TNDs including ELOA (aa 26-108), TFIIS (aa 1-80), LEDGF (aa 345–431), HRP2 (aa 469-549), MED26 (aa 1-

87), PPP1R10 (aa 1-147) and IWS1(aa 550-692), was cloned as C-terminal fusions of RFP-NLS, resulting in pRRL-CAG vectors harboring RFP-NLS-TND coding sequences. Full-length IWS1 was inserted as an N-terminal fusion into pRRL-CAG-EGFP-LacI-NLS-IRES-puro vector and full-length TFIIS as a C-terminal fusion into pRRL-CAG-RFP-NLS-IRES-neo vector, enabling selection of transduced cells. For expression of full-length IWS1 in mammalian cells, a 3x-FLAG tag was inserted into the pRRL-CAG-IRES-puro backbone using adaptor ligation. Subsequently, IWS1 was amplified and inserted by InFusion (Takara Bio) resulting in an 3xFLAG-IWS1-IRES-puro expression cassette. All point mutations were introduced into these plasmids by site directed ligase-independent mutagenesis (*53*). The integrity of all plasmids was verified by Sanger sequencing.

NMR spectroscopy and structure determination

Recombinant protein expression and purification. Competent BL21(DE3) *E. coli* cells were transformed with the bacterial expression plasmids described above. The transformed colonies were grown in LB or minimal medium, supplemented with $(^{15}NH_4)_2SO_4$ alone or with ^{13}C glucose for double labeled proteins, at 37 °C to an OD_{600} of 0.8. Cultures were then transferred to 18 °C and protein production was induced by 0.25 mM of ETG (Ethyl β-D-thiogalactopyranoside). After overnight protein expression for 15 h at 18 °C, bacterial cultures were pelleted by centrifugation (5,000 g, 4 °C, 20 min) and stored at -20 °C. Each gram of bacterial pellet was lysed using 10 ml of lysis buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 2 mM β-mercaptoethanol, 10 μM EDTA) supplemented with 1× protease inhibitor cocktail-EDTA free (Roche), DNase I (DN25 Sigma-Aldrich, 2 units/ml) and sonicated (Branson Sonifier 250) until a homogeneous lysate was obtained. The lysates were centrifuged at 30,000 g for 30 min and the supernatant was loaded on Ni-chelate agarose resin. Bound fractions were eluted with elution buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, 10 μM EDTA). Fractions containing the protein of interest were pooled, TEV protease was added, and cleavage of the His_6-tag was performed during dialysis to 50-100-fold excess of lysis buffer overnight at 4 °C. The protein of interest retained in the flow-through fraction on Ni chelate agarose was concentrated using an Amicon Ultra centrifugal filtration unit. The final purification by size exclusion chromatography was performed on Superdex 75 or 200 10/300 GL (GE Healthcare Life Sciences) in 25 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol,

10 μM EDTA. Pooled fractions were concentrated and exchanged into 25 mM deuterated Tris-HCl pH 7.5, 200 mM NaCl, 1 mM TCEP, 10 μM EDTA, 0.02% NaN_3. The concentration of the purified proteins was measured with a NanoDrop spectrophotometer and confirmed by amino acid analysis. The purity was determined by SDS-PAGE followed by Coomassie Brilliant Blue and silver staining. The proteins were either used directly or they were shock-frozen in liquid nitrogen and stored at -80 °C until use.

Peptide synthesis. Peptides were synthesized by solid phase synthesis in the Laboratory of Medicinal chemistry, IOCB, ASCR v. v. i., Prague, Czech Republic. IWS1-pTIM1 peptide (KEL SDK KNE EKD LFG XDX EXG NEE EN; X=Phosphoserine) was synthesized by Peptide Specialty Laboratory, Heidelberg, Germany.

Spectroscopy and structural calculations. NMR spectra were acquired at 25 °C on an 850 MHz Bruker Avance spectrometer using a triple-resonance ($^{15}N/^{13}C/^{1}H$) cryoprobe. The sample was prepared at 0.3-0.8 mM concentration in a volume of either 0.16 or 0.35 ml in buffer (25 mM deuterated Tris-HCl pH 7.5, 200 mM NaCl, 1 mM TCEP, 10 μM EDTA, 0.02% NaN_3), 5-10% D_2O/90-95% H_2O. A series of double- and triple-resonance spectra (*54*, *55*) were recorded to obtain sequence-specific backbone resonance assignments for the studied constructs. Aromatic side-chain resonance assignments and ^{1}H-^{1}H distance restraints were derived from 3D $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ NOESY-HMQC, which were acquired using a NOE mixing time of 100 ms.

To follow changes in the chemical shifts of a protein upon partner protein binding, we calculated chemical shift perturbations (CSPs). The CSP of each assigned resonance in the 2D $^{15}N/^{1}H$ HMQC or 3D HNCO spectra of the protein in the free state is calculated as the geometrical distance in ppm to the peak in the 2D $^{15}N/^{1}H$ HMQC or 3D HNCO spectra acquired under different conditions using the formula: $\mathrm{CSP} = \sqrt{\Delta\delta_H{}^2 + (\Delta\delta_N \cdot \alpha)^2 + (\Delta\delta_C \cdot \beta)^2}$, where α and β are weighing factors of 0.2 and 0.1 used to account for differences in the proton and nitrogen/carbon spectral widths (*56*). For these interaction studies, recombinant ^{15}N-labeled TNDs of TFIIS, ELOA, LEDGF, HRP2, MED26, PPP1R10 and IWS1 were prepared and 2D $^{15}N/^{1}H$ HMQC experiments in the presence of increasing amounts of each of the TIMs were measured. Unlabeled TIMs were prepared as synthetic peptides with the exception of IWS1 and SPT6, which were prepared as recombinant proteins due to their length. 2D $^{15}N/^{1}H$ HMQC

experiments were sufficient for monitoring of TIMs binding to structured TNDs, as well as TFIIS/ELOA competition for IWS1 binding, while 3D HNCO experiments were used for resolving CSPs in intrinsically disordered regions of IWS1 TIMs. In titration experiments the 30 μM of ^{15}N-labeled TNDs were titrated with various concentrations (1:0.5 to 1:8 molar ratio) of unlabeled peptide or DMSO as a control. For each titration point, the chemical shift perturbations in ^{15}N/^{1}H HMQC spectra measured in the SOFAST fashion were calculated and the dissociation constant was determined by a nonlinear least squares analysis for the 10 most perturbed residues using the equation

$$\mathrm{CSP}_{obs} = \mathrm{CSP}_{max} \times \frac{[\mathrm{L}] + [\mathrm{P}] + \mathrm{K}_D - \sqrt{([\mathrm{L}] + [\mathrm{P}] + \mathrm{K}_D)^2 - 4 \times [\mathrm{L}] \times [\mathrm{P}]}}{2 \times [\mathrm{P}]}$$

where CSP_{obs} is the observed combined CSP at the given total ligand concentration [L], CSP_{sat} is the CSP at saturation, and [P] is the total protein concentration. In the TFIIS/ELOA competition experiments, ^{15}N-labeled TFIIS or ELOA TNDs (10 μM) were mixed with 40 μM IWS1-derived peptide containing the TIM1 binding site (KEL SDK KNE EKD LFG XDX EXG NEE EN; X=Phosphoserine). Binding of ^{15}N-labeled protein was outcompeted with unlabeled TFIIS or ELOA TNDs and LEDGF TND as negative control. All measurements were performed in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 2% v/v DMSO. To determine dissociation constants, titration spectra were processed in MestReNova (MestreLab Inc.), changes in the position of individual peaks were tracked and plotted against the corresponding concentration in GraphPad Prism 9.0. To compare the strength of individual binary interactions, the mean CSP value was calculated for the 10% most affected residues. This value negatively correlates with measured K_D values (**Fig. S8**).

For structural determination, ^{13}C/^{15}N-labeled recombinant proteins were prepared and complete ^{15}N, ^{13}C, and ^{1}H resonance assignments were obtained. The complete resonance assignments allowed automated assignment of the NOEs identified in the 3D ^{15}N- and ^{13}C-edited NOESY spectra. In particular, the structural calculations for TNDs and the TFIIS-IWS1 single-chain complex were carried out in Cyana (*57*) using NOESY data in combination with backbone torsion angle restraints, generated from the assigned chemical shifts using the program TALOS+ (*58*). First, the combined automated NOE assignment and structure determination protocol (CANDID) was used for automatic NOE cross-peak assignment. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle restraints were used to calculate a

set of converged structures with no significant restraint violations (distance and van der Waals violations <0.5 Å and dihedral angle constraint violations <5°). They were further refined in explicit solvent using the YASARA software with the YASARA forcefield (*59*). The 40 structures with the least restraint violations were selected for further analysis using the Protein Structure Validation Software suite (www.nesg.org). The statistics for the resulting structures are summarized in **Table S3**. Analysis of the previously published MED26 structure (PDB: 5ODD) (*60*, *61*), revealed the presence of an unusual void volume in the protein core. Therefore, we acquired and thoroughly analyzed new data providing 60% more NOE-derived long-range restraints and calculated a structure that could be reliably used for comparison to other TNDs.

The structure of the PPP1R10 TND in complex with TIM2 (PPP1R10-IWS1) was obtained in Haddock (*62*) using a standard docking protocol in combination with the NMR data acquired for $^{13}C/^{15}N$-labeled PPP1R10 in the presence of a synthetic peptide. We combined the free PPP1R10 TND structure and the peptide structure calculated in Cyana using the $^{13}C/^{15}N$-filtered TOCSY and NOESY spectra. For construction of the IWS1-SPT6-TFIIS-PPP1R10-LEDGF quinary complex, we used the solution structures of binary complexes obtained for TFIIS-IWS1, PPP1R10-IWS1, LEDGF-IWS1 (*27*) and a homology model of the IWS1-SPT6 complex based on our IWS1 TND structure and the *Encephalitozoon cuniculi* X-ray structure (*32*). Initial energy minimization was followed by 100 ns of molecular dynamics in YASARA.

Visualization. The obtained structures were rendered using The PyMOL Molecular Graphics System, Version 2.3.4 (Schrödinger, LLC).

AlphaScreen assays

Protein expression. *E. coli* expressing GST-IWS1-3xFLAG were lysed with an Emulsiflex C3 (ATA Scientific) in 50 mM Tris pH 7.5, 200 mM NaCl, 2 mM DTT and loaded on Glutathione Sepharose 4B (Cytiva, GE17-0756-05) and eluted with the same buffer supplemented with 40 mM glutathione. The eluted fractions were concentrated and applied to Superdex 200 Increase 10/300 GL (Cytiva, GE28-9909-44) 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP. His_6-tagged proteins for AlphaScreen were prepared as described above, but the TEV protease cleavage step was omitted. The purest fractions after Ni-chelate chromatography were concentrated and further purified on Superdex 75 10/300 GL (Cytiva, GE29-1487-21) in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP.

Interaction assay. All reactions were performed in 96-well ProxiPlates (white, opaque, shallow, Perkin Elmer, 6006290) using a Perkin Elmer Enspire plate reader. All assays were performed in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.05% Tween 20, 0.1% BSA. The protein concentrations for each experiment were determined by cross-titration to perform the assay at high signal-to-noise ratio and to avoid perturbations at high concentrations. GST-tagged IWS1-3xFLAG fragment (150 nM in reactions, IWS1 residues 352-548) was incubated for 1 h at RT in the dark with His_6-tagged IWS1 binding proteins (0-100 nM LEDGF or HRP2). In case of competition experiments, 0-10 μM untagged binding proteins were added to the mixture of 150 nM GST-tagged IWS1 and 20 nM His-tagged binding proteins (LEDGF or HRP2). In parallel, glutathione-coated acceptor beads (Perkin Elmer, AlphaLISA, AL109C, 5 mg/ml, Lot 2781549) were mixed with nickel chelate donor beads (Perkin Elmer, AlphaScreen, AS101D, Lot 2750736). The two mixtures (proteins and beads) were incubated separately for 1 h at RT in the dark and subsequently mixed together on the plate and incubated for another 1.5 h while shaking. The final concentration of each bead type was 2 ng/μl. Data were evaluated using GraphPad Prism 9.0. After background subtraction, data were fitted using nonlinear regression and a one site specific binding model ($Y = \mathrm{Bmax} * x(K_D + x)$; Bmax is the maximum specific binding).

Culturing of human cell lines

Acquisition and culture conditions. Lenti-X 293T (#632180) and IMR-32 (RRID:CVCL_0346) cell lines used in this study were acquired from Takara Bio and ATCC, respectively. U2OS 2-6-3 cells were a generous gift from David L. Spector (CSHL, Cold Spring Harbor, NY, USA) and KELLY cells (RRID:CVCL_2092) were a generous gift from Charles Lin (BCM, Houston, TX, USA). The deep deletion of IWS1 in KELLY cells was reported in cBioPortal (*63*, *64*). HEK293T cells (RRID:CVCL_0045) were a generous gift from J. Konvalinka (IOCB, Prague, CZ). Lenti-X 293T and KELLY cells were maintained in high-glucose DMEM (Gibco), U2OS 2-6-3 cells in McCoy's media (Sigma) and IMR-32 cells in MEM (Gibco). HEK293T cells were maintained in high-glucose DMEM (Gibco) supplemented with 10% heat-inactivated FBS (Gibco). U2OS 2-6-3 cells were cultured under selective conditions with 150 μg/ml Hygromycin B (GE Healthcare). DMEM and McCoy's media were both supplemented with 10% heat inactivated FBS (Corning), 10 mM HEPES (Gibco), 1× GlutaMAX (Gibco), 1× non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco) and

1× Pen/Strep (Gibco). MEM media was supplemented with 10% heat inactivated FBS (Corning), 1× non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco) and 1× Pen/Strep (Gibco). All cell lines were confirmed to be free of mycoplasma by PCR testing, and cultured in a humidified incubator maintained at 37 °C and 5% CO_2

Lentiviral vectors and transduction. Lentiviral vectors were produced by co-transfection of Lenti-X 293T cells with packaging plasmids (psPAX2, pMD2.G) and a transfer plasmid (derived from pRRL backbone) using linear polyethylenimine hydrochloride (PEI, Polysciences, Inc. 247651, MW 40,000). The culture medium was changed to DMEM prepared as described above but with 2% FBS 16 h after transfection. Lentiviral particles containing culture medium was harvested from the cells 72 h after transfection. The medium was first filtered through a 5 μm filter (Millipore), concentrated by ultrafiltration on a 100-kDa Amicon filter (Millipore) and stored at −80 °C. All cell lines were transduced at 30% confluency on 6-well plates using a spinfection protocol (1,000 g, 30 min, 32 ºC). KELLY cells were selected with 1 μg/ml puromycin and U2OS 2-6-3 cells were selected with 800 μg/ml G418. Selection drugs were added to the cell lines 48 h after transduction.

Fluorescent two-hybrid (F2H) protein-protein interaction assay

Expression of constructs. U2OS 2-6-3 cells were transduced with lentiviral vector for delivery of TIM-LacI-EGFP-NLS library or LacI-EGFP-NLS control and TagRFP-NLS-TND library or TagRFP-NLS control and imaged 7 days post transduction. U2OS 2-6-3 cells were first transduced with lentiviral vector for delivery of TagRFP-NLS-TFIIS WT, TagRFP-NLS-TFIIS RK-AA mutant or TagRFP-NLS control. Cells were selected with 800 μg/ml G418, subsequently transduced with IWS1-LacI-EGFP-NLS WT, IWS1-LacI-EGFP-NLS M1 mutant or LacI-EGFP-NLS control and imaged 4 days post infection.

Live-cell microscopy. U2OS 2-6-3 cells were seeded into a 12-well tissue culture plate at 60% confluency. The next day, cells were stained with Hoechst 33342 (Thermo Fisher) according to the manufacturer's instructions. Live cells were imaged on an inverted epifluorescence microscope (Nikon Ti2E stand) with environmental control (37 °C with humidified 5% CO_2) using 20x or 40x objectives. EGFP, TagRFP and Hoechst 33342 fluorescence was imaged using FITC (470/40 ex, 525/50 em), Texas Red (560/40 ex, 630/70 em), and DAPI (350/50 ex, 460/50 em) filter sets respectively. Image analysis to determine the

mean EGFP and TagRFP intensity in the fluorescent puncta and in the nucleus was performed in Nikon Elements AR.

Treatment with 1,6-hexanediol. U2OS 2-6-3 cells were transfected with plasmid DNA encoding IWS1-TIM-EGFP-LacI in combination with individual TagRFP-TND constructs according to the manufacturer's instructions (Thermo Fisher, #L3000008). As a positive control for liquid-liquid phase separation, we employed EGFP-polyQ-ATXN1 (a kind gift from Huda Zoghbi; Addgene plasmid # 32492; http://n2t.net/addgene:32492; RRID: Addgene_32492). 48 h after transfection, cells were permeabilized with 1% Tween 20 for 10 min and stained with Hoechst 33342. Cells were imaged as described above prior to and 1 min after treatment with 5% 1,6-hexanediol (Sigma).

Fluorescence recovery after photobleaching (FRAP) assay

U2OS 2-6-3 cells were cultured on chambered μ-Dish 35 mm coverslips (Ibidi). Experiments were performed on an inverted microscope (Nikon Ti2E stand) with environmental control at 37 °C. The fluorescent foci of U2OS 2-6-3 cells were photobleached using a 405-nm laser. EGFP and TagRFP fluorescence of the cell was monitored as a time course. The degree of fluorescence recovery within the region of interest (ROI, the photobleached portion of the nucleus) was monitored in NIS Elements software. The ratio of fluorescence intensity within the photobleached ROI over the fluorescence intensity in the ROI before the photobleaching was plotted and fit to exponential curves using non-linear least squares fitting to obtain the recovery half-lives.

Pull-down studies

Cellular fractionation. Following transfection of FLAG-IWS1 constructs at 70% confluency according to the manufacturer's instructions (Thermo Fisher, #L3000008), HEK293T cells were selected with puromycin for 48 h and then harvested at 96 h. Cell pellets from 15 cm dishes were resuspended in 1.5 fold volume of fractionation buffer A (20 mM Tris pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM Na_3VO_4, 10 mM NaF, 10 μM MG132 and 1x cOmplete Mini protease inhibitor cocktail (EDTA free, Merck Roche) and incubated on ice for 15 min prior to addition of Igepal CA-630 (12.5% solution in sterile MilliQ water) to a final concentration of 0.3125% v/v. Samples were immediately gently mixed and incubated on ice for

5 min. After 10 min of centrifugation at 2,000 g, the supernatant was transferred to a new microfuge tube and centrifuged again at 17,500 g for 10 min. The supernatant fraction was used for bait immobilization.

Pull-down assay. Pull-down experiments were performed using independent cell culture replicates (n = 3) of wild-type or mutant variants (M3, M123) of 3xFLAG-tagged full-length IWS1 protein bait or empty control. Each bait sample was mixed with 35 µl of anti-DYKDDDDK affinity resin beads (Pierce, Thermo Scientific, A36801) in a ratio of 0.8 mg total protein per 5 µl of beads in capped spin columns (Pierce, Thermo Scientific, 69725) and incubated for 1.5 h at 4 °C on an end-over-end rotator. Resin beads were washed three times with 200 µl of fractionation buffers A and B, both without MG132. All flow-through, wash and elution fractions were collected by centrifugation of the spin column (1,000 g, 2 min, RT). Nuclear fractions were then added to the beads at a ratio of 0.48 mg of total protein per 5 µl of beads and incubated for 1.5 h at 4 °C on an end-over-end rotator. Beads were washed three times in 200 µl of fractionation buffer B without MG132 and three times in 200 µl of TBS buffer. Bound proteins were eluted with 245 µl of 0.8 mg/ml 3x DYKDDDDK peptide (Pierce, Thermo Scientific) in TBS.

Protein concentration of load, flow-through and wash fractions was determined using the Bradford assay (Bio-Rad, Protein Assay Dye reagent Concentrate, 500006). Protein concentration of eluted fractions was determined by measuring absorbance at 280 nm using NanoDrop Microvolume UV-Vis Spectrophotometer (Thermo Scientific). 1x cOmplete Mini protease inhibitor cocktail (EDTA free, Merck Roche) was added to elution fractions before processing for mass spectrometry and western blot analysis. Excess of the eluting peptide was removed by filtration (Microcon-10 kDa Centrifugal Filter) and proteins were digested with trypsin overnight directly on the filter (*65*).

Co-immunoprecipitation studies

Cellular fractionation. Following transfection of FLAG-IWS1 constructs at 70% confluency according to the manufacturer's instructions (Thermo Fisher, #L3000008), HEK293T cells were selected with puromycin for 48 h and then harvested at 96 h. Cell pellets from 15 cm dishes were resuspended in 1.5 fold volume of fractionation buffer A (20 mM Tris pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM Na_3VO_4, 10 mM NaF, 10 µM MG132 and 1x

cOmplete Mini protease inhibitor cocktail (EDTA free, Merck Roche) and incubated on ice for 15 min prior to addition of Igepal CA-630 (12.5% solution in sterile MilliQ water) to a final concentration 0.3125% v/v. Samples were immediately gently mixed and incubated on ice for 5 min. After 10 min of centrifugation at 2,000 g, the supernatant was transferred to a new micro test tube and centrifuged again at 17,500 g for 10 min.

Pellets were washed three times with buffer A and centrifuged each time at 6,000 g at 4 °C for 10 min. The washed pellet was then resuspended in a volume of buffer B (20 mM Tris pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 1 mM Na_3VO_4, 10 mM NaF, 10 μM MG132, and 1x protease inhibitor cocktail, EDTA free) equal to volume of the cell pellet. The suspension was sonicated 4x30 s in a sonication bath on ice and then 1 μl of Benzonase nuclease (250 U/μL, Merck, E-1014) was added per each 200 μl of buffer B. The suspension was incubated on ice for 45 min and it was mixed gently every 10 min. Igepal CA-630 was added to the nuclear suspension to a final concentration of 0.2% v/v. During 30 min incubation, the suspension was mixed gently every 5 min. The suspension was centrifuged at 14,500 g at 4 °C for 10 min and the soluble (nuclear) fraction was transferred to a new tube and used for downstream use. Protein concentrations were determined using the Bradford assay (BioRad, Protein Assay Dye Reagent Concentrate, 500006).

Co-immunoprecipitation. IWS1-FLAG variants were immunoprecipitated as follows: Each nuclear fraction of HEK293T cells expressing wild-type or M123 mutant of 3xFLAG-tagged full-length IWS1 protein was mixed with 30 μl of anti-DYKDDDDK affinity resin beads (Pierce, Thermo Scientific, A36801) in a ratio 0.5 mg of total protein per 5 μl of beads in capped spin columns (Pierce, Thermo Scientific, 69725) and incubated for 1.5 h at 4 °C on an end-over-end rotator. All flow through, wash and elution fractions were collected by centrifugation of the spin column (1,000 g, 2 min, RT). Beads were washed three times in 150 μl of fractionation buffer B without MG132 and three times in 150 μl of Tris-buffered saline (TBS). Bound proteins were eluted from resin beads with 210 μl of 0.5 mg/ml 3x DYKDDDDK peptide (Pierce, Thermo Scientific) in TBS. Protein concentration of eluted fractions was determined by measuring absorbance at 280 nm using NanoDrop Microvolume UV-Vis Spectrophotometer (Thermo Scientific) and analyzed by western blotting.

Liquid Chromatography-Tandem Mass Spectrometry Analysis

Digested peptides were injected on a nanoLC (UltiMate 3000 RSLC) coupled to an Orbitrap Fusion Lumos Mass spectrometer (Thermo Fisher) as described previously(*66*). Raw data were processed using MaxQuant (version 1.6.3.4, Max-Planck-Institute of Biochemistry, Planegg, Germany) with release 2021_03 of the reviewed UniProtKB human sequence database. Proteins annotated as common contaminants, reverse hits, or identified only by site were excluded. MaxQuant label-free quantification (LFQ) intensities were used for comparisons of independent biological replicates obtained for samples with and without bait FLAG-IWS1 variants in the Perseus software (*67*). Only data consisting of at least two valid values in at least one experimental group were considered. All intensities were log2-transformed and statistical analysis (Student's *t*-test) was performed using Perseus. Differentially enriched proteins were identified by requiring a fold change >2 in either direction and $p < 0.05$ (Student's *t*-test).

Western blotting

Immunoblotting. For analysis of proteins from KELLY cells, total protein was harvested using RIPA lysis buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, cOmplete ULTRA protease inhibitors (Roche). Harvested cells were washed 2 times in cold PBS, subsequently resuspended and incubated in cold RIPA lysis buffer for 20 min on ice. Samples were sonicated using a Bioruptor (5 cycles of 30 seconds ON/OFF on high power) and cleared by centrifugation at 18,000 g for 15 min at 4 °C. Protein concentration in the cleared supernatant was determined using the Pierce Detergent Compatible Bradford Assay Kit (Thermo Fisher) according to manufacturer's instructions. We loaded 20 µg of total protein per well into Novex NuPAGE Bis-Tris 4-12% polyacrylamide gels in MOPS SDS running buffer (Thermo Fisher) and separated the protein bands according to manufacturer's instructions. Bands from the gel were transferred to 0.45 µm PVDF membrane (Millipore) over night. For analysis of co-IP and pull-down samples 10 µg of total protein for inputs and 4 µg of total protein for elution fractions was separated on 4-20% Tris-glycine Precast Gels (Bio-Rad) at 250 V for 35 min and transferred onto nitrocellulose membrane (Merck Millipore) at 100 V for 1.5 h using a wet blotting system (Bio-Rad). Western blots were subsequently blocked in 5% BLOTTO (Santa Cruz) or in 1% casein blocker in Tris-buffered saline with 0.1% Tween 20 Detergent (TBS-T), incubated in primary antibody overnight at 4 °C, washed 3 times in TBS-T, and probed with secondary antibody for 1-2 h at RT.

Antibodies. The following antibodies were used in this study: IWS1 (Cell Signaling Technology, 5681S, RRID:AB_10694503), beta-actin (GeneTex, GTX629630, RRID:AB_2728646), FLAG M2 (Sigma Aldrich, F3165, RRID:AB_259529), Histone H3 (Cell Signaling Technology, 4499S, RRID:AB_10544537), HRP2 (Bethyl Laboratories, A304-314A, RRID:AB_2620510), LEDGF (Bethyl Laboratories, A300-848A, RRID:AB_2171223), PNUTS (Bethyl Laboratories, A300-439A, RRID:AB_420948), TCEB3 (Bethyl Laboratories, A300-942A, RRID:AB_805846), TFIIS (Bethyl Laboratories, A302-240A, RRID:AB_1730887), IRDye 800CW Goat anti-Rabbit IgG (LI-COR Biosciences, 926-32211, RRID:AB_621843), IRDye 680RD Goat anti-Mouse IgG (LI-COR Biosciences, 926-68070, RRID:AB_10956588), HRP conjugated mouse anti-rabbit IgG-HRP Secondary antibody (Santa Cruz Biolabs, sc-2357, RRID:AB_628497) and m-IgGk BP-HRP Secondary antibody (Santa Cruz Biolabs, sc-516102, RRID:AB_2687626).

Imaging. Detection of HRP-conjugated antibodies was performed using Clarity Western or Clarity Max ECL substrate (Bio-Rad) on a Bio-Rad ChemiDoc MP imager. Membranes probed with secondary fluorescent antibodies were imaged using an Odyssey CLx Infrared Imaging System (LI-COR Biosciences).

RNA sequencing (RNA-seq)

Sample preparation. KELLY cells were harvested from 50% confluent 10 cm tissue culture plates (n = 2 independent cell culture replicates for each condition). Cells were washed once in fresh culture medium to remove dead cells and residual RNase activity before harvesting with TRIzol reagent (Invitrogen), with total RNA isolated according to the manufacturer's instructions. mRNA was obtained using polyA enrichment and converted to cDNA before fragmentation and 2x 150bp paired-end sequencing on an Illumina NovaSeq 6000.

RNA-seq analysis. Analysis of RNA-seq reads was performed as described previously (*68*). Briefly, RNA-seq reads obtained were processed by mapping to the hg38 reference human genome using HISAT2 2.1.0 and reads with mapping quality <10 were discarded. Reads within genes were counted using HTSeq and these counts were processed using DESeq2 with default parameters. Log2 fold changes were calculated using maximum a posteriori estimation using a zero-mean normal prior (Tikhonov-Ridge regularization). FDR-corrected P values were calculated using the Benjamini-Hochberg procedure. Differential calls were made by requiring a

fold change >1.5 in either direction and a FDR-corrected P < 0.10. Alternative splicing was assessed using rMATS 4.0.2 using default parameters (*69–71*).

Chromatin immunoprecipitation sequencing (ChIP-seq)

Sample preparation. KELLY cells were harvested from 50% confluent 15 cm tissue culture plates (n = 2 independent cell culture replicates for each condition). Cells were double cross-linked as follows. The first cross-linking step was performed with 2 mM disuccinimidyl glutarate (DSG) for 45 min followed by 2x PBS washes. The second cross linking step was done in 1% formaldehyde for 10 min. Excess formaldehyde was quenched by the addition of glycine to 125 mM. Fixed cells were washed, pelleted, and snap-frozen in liquid nitrogen. Chromatin immunoprecipitation was performed as described previously(*68*) with modifications below. Cell nuclei were isolated, resuspended in shearing buffer and sonicated in a Bioruptor Plus (Diagenode) at high power setting for 50 cycles (30 s on, 30 s off), yielding a modal fragment size <600 bp. ChIP assays were performed using the Zymo-Spin ChIP kit (D5210) following the manufacturer's instructions with anti-FLAG antibody (Millipore Sigma, F3165, RRID:AB_259529). After PCR amplification, DNA quantified by Qubit fluorometric quantitation. Sequencing was performed using single-end reads on the Illumina HiSeq platform.

ChIP-seq analysis. Analysis of high-throughput sequencing data was performed as described previously (*68*). Briefly, single-end ChIP-seq reads were processed by mapping to the hg38 reference human genome using Bowtie 2.1.0 (*72*), rejecting reads that contain more than a single mismatch. Reads were extended to 200 bp before quantification using bedtools 2.28.0. Metagene plots were derived using bwtool (*73*). Read coverage values in genome tracks are the mean values across both replicates from each condition. Calculation of mean genome track densities was performed using bwtool (*73*).

Precision nuclear run-on sequencing (PRO-seq)

Harvesting and permeabilization. KELLY cells were harvested (n = 2 independent cell culture replicates), washed 2x with ice cold PBS and resuspended in 250 µl of wash buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 250 mM sucrose, 5 mM $MgCl_2$, 1 mM EGTA, 0.5 mM DTT, 10% (v/v) glycerol) supplemented with 10 µl SUPERase-In RNase inhibitor per 50 ml. Subsequently, 10 ml of permeabilization buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 250 mM

sucrose, 5 mM $MgCl_2$, 1 mM EGTA, 0.1% (v/v) Igepal CA-630, 0.5 mM DTT, 0.05% (v/v) Tween-20, 10% (v/v) glycerol) supplemented with 1 cOmplete protease inhibitor tablet (EDTA-free) and 10 µl SUPERase-In RNase inhibitor per 50 ml) was added to the cells and samples were incubated on ice for 5 min. Cells were pelleted and washed 1x with wash buffer before they were resuspended in freezing buffer (50 mM Tris-HCl pH 8.0, 40% (v/v) glycerol, 5 mM $MgCl_2$, 1.1 mM EDTA, 0.5 mM DTT) supplemented with 10 µl of SUPERase-In RNase inhibitor per 50 ml) and snap frozen in liquid nitrogen.

Nuclear run-on. Aliquots of frozen (-80 °C) permeabilized cells were thawed on ice and pipetted gently until fully resuspended. Aliquots were removed and permeabilized cells were counted using a Luna II instrument (Logos Biosystems). For each sample, 1 million permeabilized cells were used for nuclear run-on, with 50,000 permeabilized Drosophila S2 cells added to each sample for normalization. Nuclear run-on assays and library preparation were performed essentially as described (*74*) with modifications noted: 2× nuclear run-on buffer consisted of 10 mM Tris pH 8, 10 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 40 µM/ea biotin-11-NTPs (Perkin Elmer), 0.8 U/µl SuperaseIN (Thermo) and 1% sarkosyl. Run-on reactions were performed at 37 °C. Adenylated 3' adapter was prepared using the 5' DNA adenylation kit (NEB) and ligated using T4 RNA ligase 2, truncated KQ (NEB, per manufacturer's instructions with 15% PEG-8000 final) and incubated at 16 °C overnight. 180 µl of betaine blocking buffer (1.42 g of betaine brought to 10 ml with binding buffer supplemented to 0.6 µM blocking oligo (TCC GAC GAT CCC ACG TTC CCG TGG /3InvdT/)) was mixed with ligations and incubated 5 min at 65 °C and 2 min on ice prior to addition of streptavidin beads. After T4 polynucleotide kinase (NEB) treatment, beads were washed once each with high salt, low salt, and blocking oligo wash (0.25× T4 RNA ligase buffer (NEB), 0.3 µM blocking oligo) solutions and resuspended in 5' adapter mix (10 pmol 5' adapter, 30 pmol blocking oligo, water). 5' adapter ligation was as described (*74*) but with 15% PEG-8000 final concentration. Eluted cDNA was amplified in 5 cycles (NEBNext Ultra II Q5 master mix (NEB) with Illumina TruSeq PCR primers RP-1 and RPI-X) following the manufacturer's suggested cycling protocol for library construction. A portion of the pre-PCR library was serially diluted for test amplification to determine optimal amplification of final libraries. Pooled libraries were sequenced using an Illumina NovaSeq 6000. Because of lower than expected read depth of one WT replicate, a third independent

replicate was added to ensure adequate read coverage. All replicates were in agreement and included for data analysis.

Read processing and spike-in mapping. All custom scripts described herein are available as described in the Code Availability statement. Using a custom script (trim_and_filter_PE.pl), FASTQ read pairs were trimmed to 41 bp per mate and read pairs with a minimum average base quality score of 20 retained. Read pairs were further trimmed using cutadapt 1.14 to remove adapter sequences and low-quality 3' bases (--match-read-wildcards -m 20 -q 10). R1 reads, corresponding to RNA 3' ends, were then aligned to the Drosophila genome index (dm3) using Bowtie 1.2.2 (-v 2 -p 6 --best --un), with those reads not mapping to the spiked-in genome serving as input to the primary genome alignment step using Bowtie 1.2.2 (-v 2 --best).

Primary genome mapping. Paired-end reads were mapped to the hg38 reference genome via HISAT2 v2.2.1 (--known-splicesite-infile). Reads mapping to the hg38 reference genome were then sorted, via samtools 1.3.1 (-n), and subsequently converted to bedGraph format using a custom script (bowtie2stdBedGraph.pl). Because R1 in PRO-seq reveals the position of the RNA 3' end, the "+" and "–" strands were swapped to generate bedGraphs representing the 3' end position at single nucleotide resolution.

Transcript selection and gene-level features. Annotated transcripts were obtained from human (GRCh38.99) GTF files from Ensembl. We assigned a single, dominant TSS and transcription end site (TES) to each active gene using a custom script (get_gene_annotations.sh), which uses RNA-seq read abundance and PRO-seq R2 reads (RNA 5' ends) to identify dominant TSSs, and RNA-seq profiles to define the most commonly used TESs. RNA-seq and PRO-seq data from WT and mutant IWS1 cells were used for this analysis, to capture gene activity under both conditions. Transcripts with {immunoglobulin, Mt_tRNA, Mt_rRNA} biotypes were removed. To compare position-dependent effects at promoter-proximal regions and gene bodies, overlapping transcripts, transcripts <400 nt in length, and those lacking PRO-seq reads were filtered out. Transcripts filtered in this manner reflected <5% of both increased and decreased transcripts from bulk RNA-seq hits. PRO-seq signal in each sample was calculated in two windows: (1) from the annotated +50 to +150 nt downstream, and (2) from +250 nt to the transcription end site (TES), using a custom script (make_heatmap.pl). Given good agreement between replicates and a similar return of spike-in reads, bedGraph files were merged within

conditions and depth-normalized to generate bigWig files binned at 10 bp. Genome browser tracks were generated using PyGenomeTracks (*75*).

Metagene plots and pausing indices. Average metagene plots were generated by summing reads within 20-bp bins centered at each indicated position with respect to the TSS and dividing by the number of TSSs included in each group. For PRO-seq, statistical significance for comparisons was assessed by Mann-Whitney tests. For comparisons among gene groups, a set of definitively unaffected transcripts were used that had P_{adj} >0.1 and log2 fold changes between +0.3 and –0.3. A single transcript model was selected to represent each of the increased or decreased genes from RNA-seq analyses meeting the criteria described above, yielding 412 decreased, 675 increased, and 7,767 unaffected genes. For calculation of changes in gene body PRO-seq signal and pausing indices, the region from +50 to +150 nt downstream was considered the promoter region and from +250 downstream of the TSS to the TES was considered the gene body. Pausing indices were calculated as the number of promoter region reads divided by the number of gene body reads.

Statistical analysis

To test for significant enrichment, the hypergeometric test for over-representation was used. All other statistical tests were performed as two-sided tests using R or Prism. For multiple hypothesis testing, P values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) correction procedure to obtain P_{adj} values. GraphPad Prism 9.0 software (GraphPad Software, San Diego, CA, USA) and R (3.6.1) were used for statistical analysis. P values smaller than the 64-bit double precision machine epsilon (2^{-52} = 2.22e-16) are reported as P < 2.2e-16.

Data availability

The structures, NMR restraints and resonance assignments were deposited in the Protein Data Bank (PDB, accession codes: 6ZUY, 6ZUZ, 6ZV0, 6ZV1, 6ZV2, 6ZV3 and 6ZV4) and Biological Magnetic Resonance Bank (BMRB, accession codes: 34535, 34536, 34537, 34538, 34539, 34540 and 34541). All high-throughput sequencing data used in this project have been deposited in the Gene Expression Omnibus (GEO) database with SuperSeries accession number GSE179917.

Code availability

All analyses were performed using publicly or commercially available software (*41–43*). Scripts used for genome-wide analyses are available at the following links: https://github.com/hodgeslab/workflows (doi:10.5281/zenodo.5511049), https://github.com/AdelmanLab/NIH_scripts (doi:10.5281/zenodo.5519915), and https://github.com/AdelmanLab/GetGeneAnnotation_GGA (doi:10.5281/zenodo.5519928).

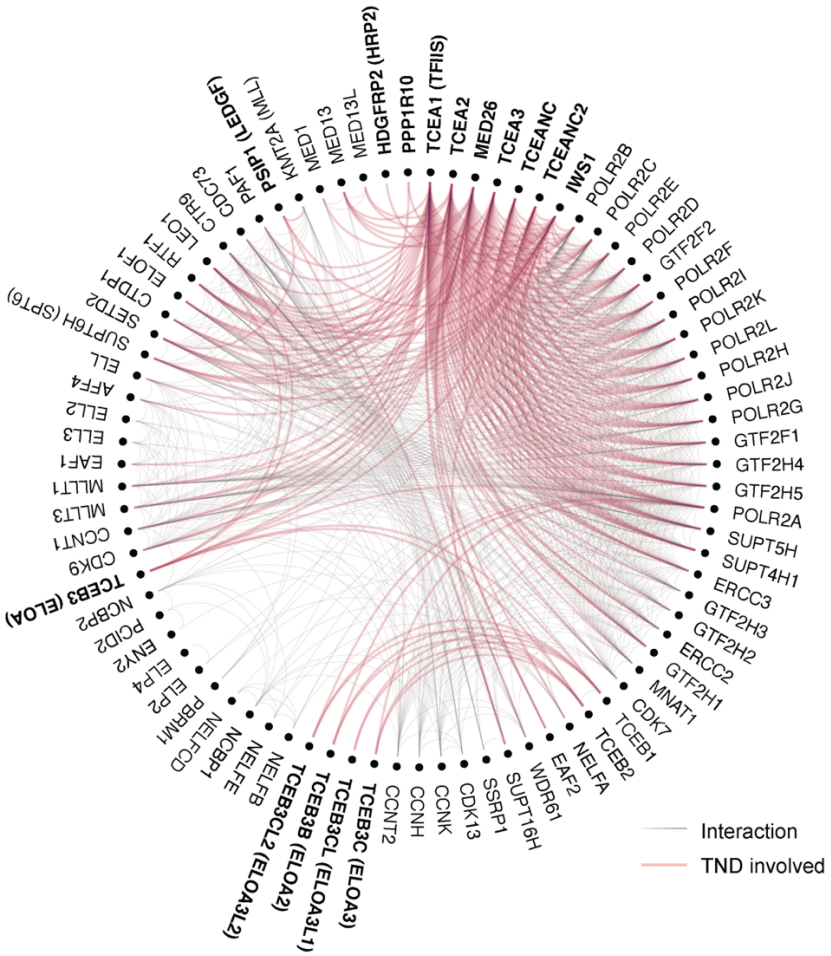

Fig. S1: Interaction network of TND-containing proteins and other proteins known to participate in transcription elongation. The interaction network in Fig. 1D, enlarged and augmented to contain all proteins presented in this study. This interaction network is defined by the GO term "transcription elongation from RNAP2 promoters" (GO:0006368), and is supplemented to include all TND- and TIM-containing proteins analyzed in this study. Interactions confirmed in this study are included in this network.

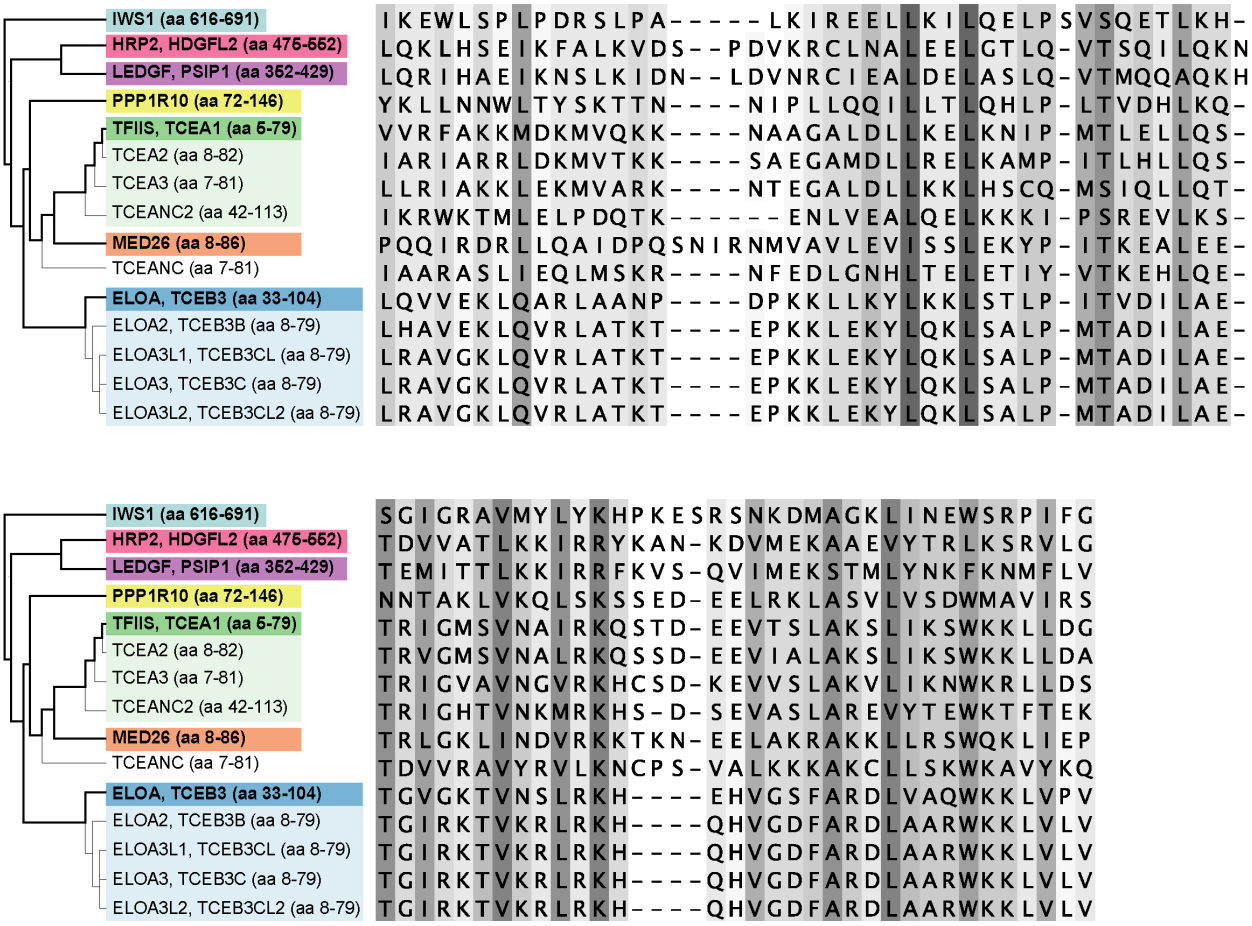

Fig. S2: Multiple sequence alignment of all known human TNDs. Degree of conservation for each amino acid residue is highlighted in multiple sequence alignment. The conservation tree was calculated based on the percentage identity between the sequences at each aligned position. The representative domains that were used for structural determination and integration studies are presented in bold. The HEAT subdomains of IWS1 (aa 543-616) and PPP1R10 (aa 1-71) are excluded from the alignment but are situated N-terminal to the TND sequence presented here.

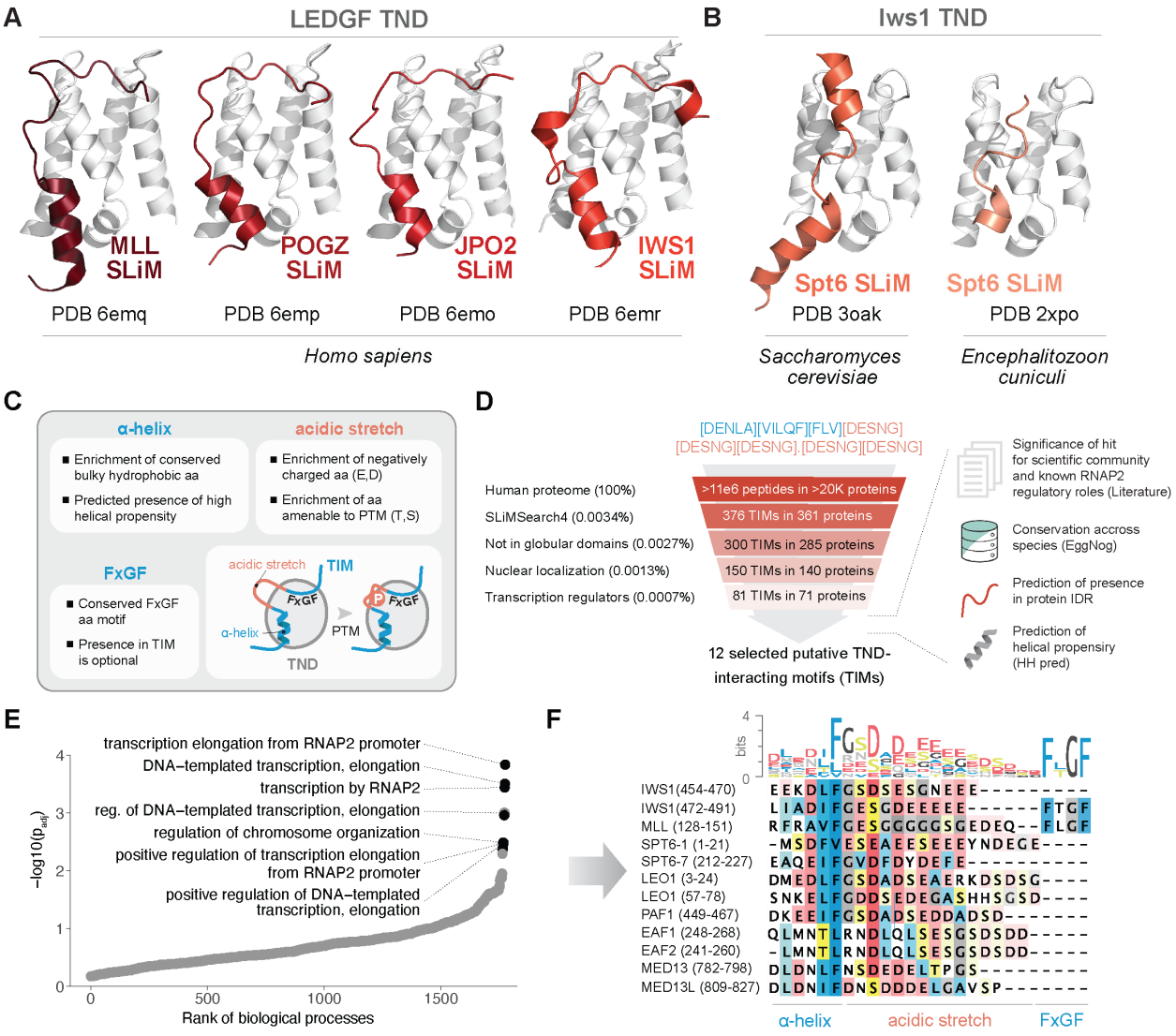

Fig. S3: Properties and bioinformatic search of TND-interacting motifs (TIMs). (**A**) Available solution structures of human LEDGF TND in complex with structurally conserved short linear motifs (SLiMs) of MLL, POGZ, JPO2 and IWS1 (PDB ID codes 6emq, 6emp, 6emo and 6emr). (**B**) The TND subdomain of Iws1 homologues from different species interacting with Spt6 (PDB ID codes 3oak and 2xpo) highlights similarity to LEDGF/p75 TND-SLiM interfaces. (**C**) General molecular features found in known SLiMs recognizing different TNDs allow prediction of TND-interacting motifs (TIMs). (**D**) Workflow of proteome-wide search for putative TIMs. (**E**) GO terms enriched in SLiMSearch4 hits (361 proteins) are ranked based on adjusted P-values (Hypergeometric test, Benjamini–Hochberg correction for multiple hypothesis testing). Relevant hits are highlighted and labeled. (**F**) Search for conserved TIM sequences harboring patterns highlighted in panel (C) yielded putative human TIMs in different transcription regulators.

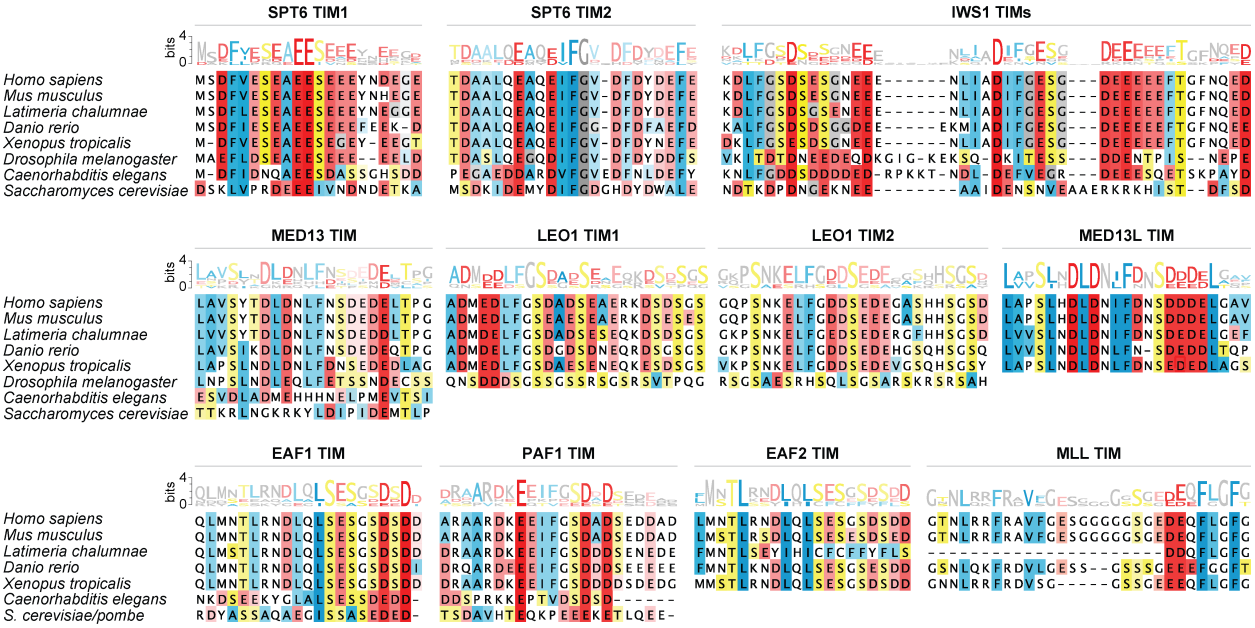

Fig. S4: Alignment of individual TIMs reveals extensive conservation. Multiple sequences alignment of TIMs found in homologues of SPT6, IWS1, MED13, LEO1, MED13L, EAF1, PAF1, EAF2 and MLL from different species. Conservation of features characteristic for TND recognizing TIMs is highlighted in blue (bulky aliphatic residues), red (negatively charged residues) and yellow (residues amenable to phosphorylation).

Fig. S5: SPT6 interacts with TNDs through its TIM1 and TIM2. (**A**) Multiple sequence alignments of seven conserved unstructured sequences with amino-acid composition consistent with a potential TIM. (**B**) Minimal chemical shift perturbations (CSPs) in the SPT6 backbone amide signals (aa 1-240) in the presence of different TNDs are presented. Two of the seven conserved sequences interact with TNDs, which underscores the selectivity and specificity of motif features by TND binding partners beyond non-specific charge and hydrophobicity. TIM1 and TIM2 and are highlighted in the figure.

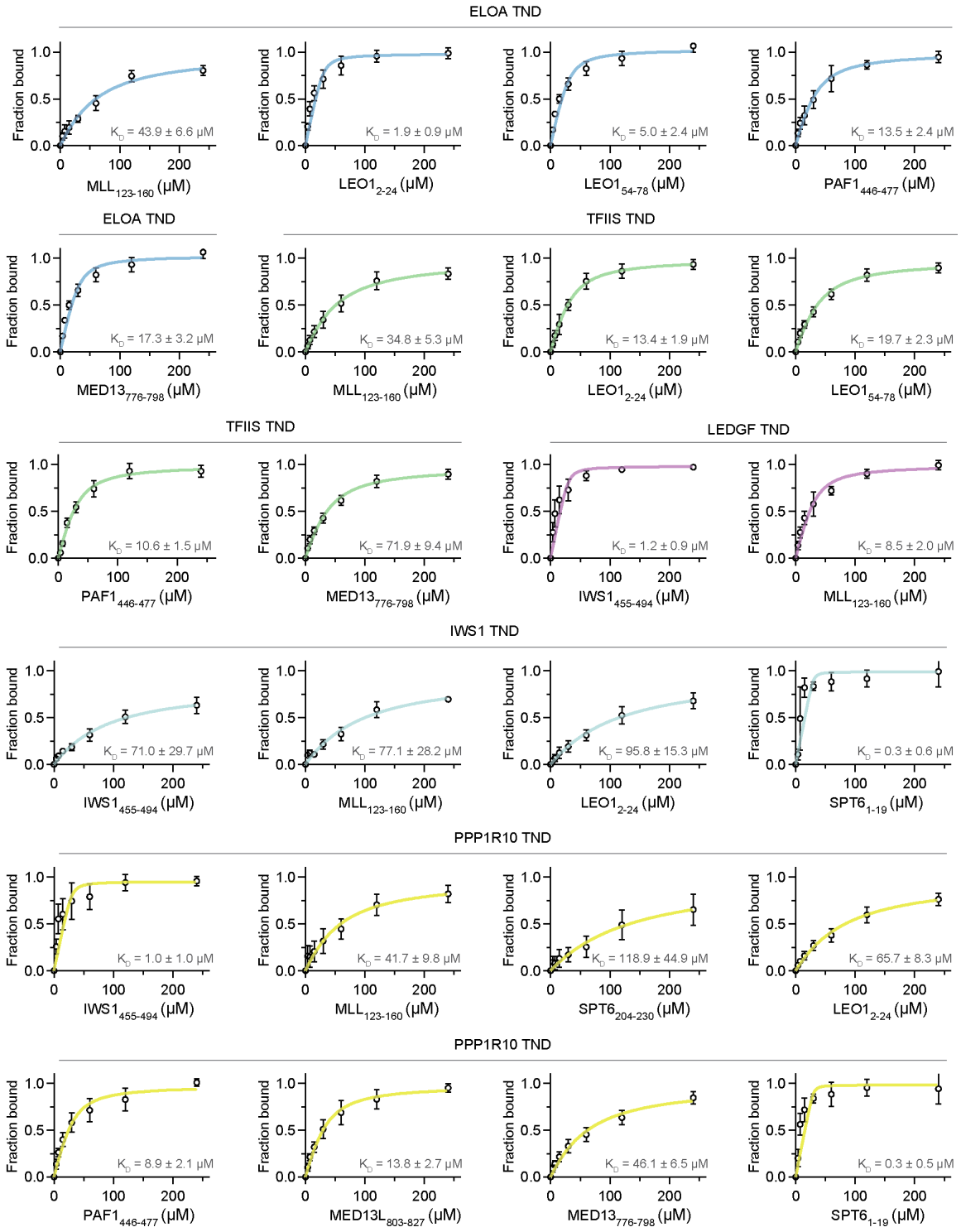

Fig. S6: Dissociation constants (K_D) for TND-TIM complexes. Curve fits from NMR titrations of different TNDs with TIMs. Dissociation constants (K_D) were determined by following the chemical shift perturbations (CSP) of the TND backbone amide signals induced upon titration with TIMs. Plotted values are mean ± standard deviation across the most perturbed residues (n = 10). The error of K_D measurements presented is the error of the fit.

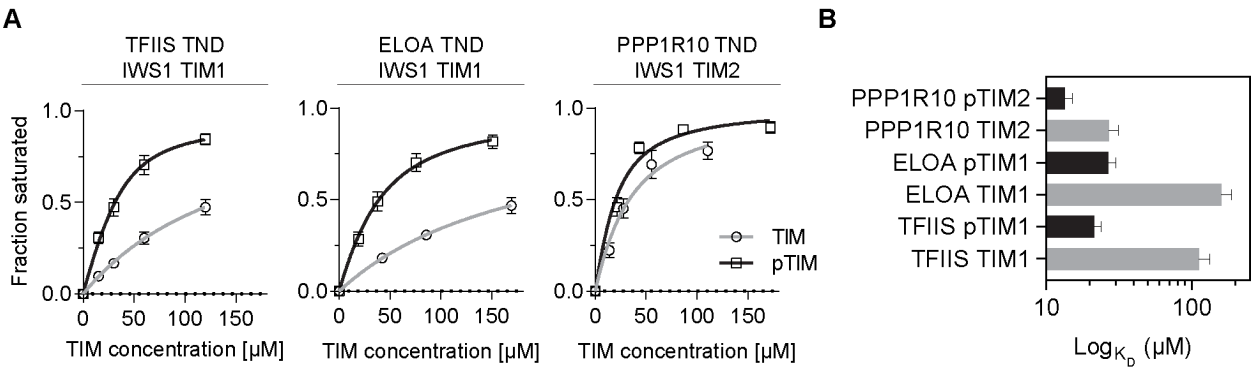

Fig. S7: Phosphorylation increases affinity of TIMs for TNDs. (**A**) K_D fits from NMR titrations of TFIIS, ELOA and PPP1R10 TNDs with IWS1 TIMs. Dissociation constants were determined by following the chemical shift perturbations (CSP) of the TNDs induced upon titration with non-phosphorylated and phosphorylated TIM (pTIM). Plotted values are mean ± standard deviation across the most perturbed residues (n = 10). (**B**) Comparison of K_D values determined in panel (A).

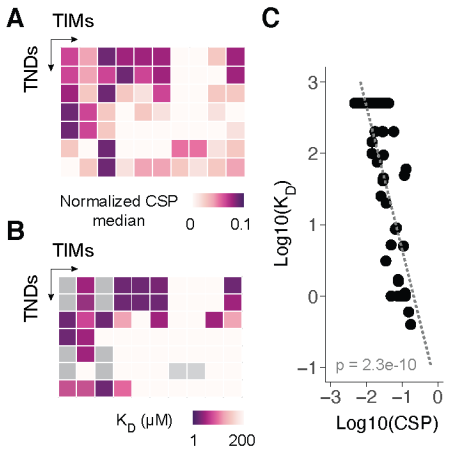

Fig. S8: The mean chemical shift perturbation is correlated with dissociation constants (K_D) determined by NMR. (**A**) Heatmap of the mean of chemical shift perturbation (CSP) calculated for 10% most affected residues for each TND-TIM binary interaction. Identical heatmap with row and column labels is presented in Fig. 2F. (**B**) Heatmap of dissociation constants (K_D) determined by NMR each TND-TIM binary interaction. (**C**) Correlation plot of mean CSP values and dissociation constants for each TND-TIM binary interaction.

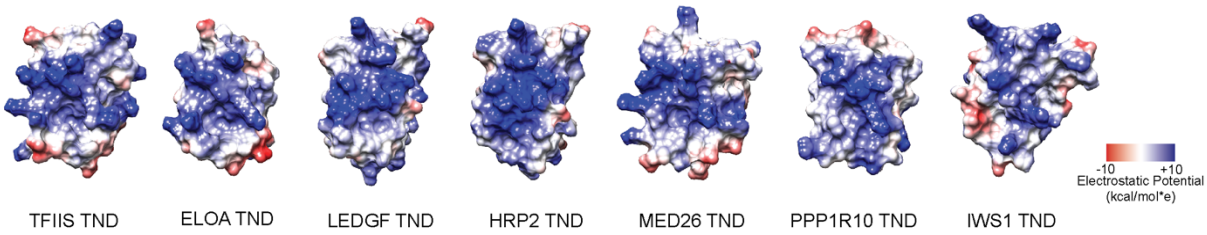

Fig. S9: The TND fold is characterized by the presence of a positively charged basic patch formed by α3, α4, and α5. Electrostatic potential is depicted on the surface of each TND.

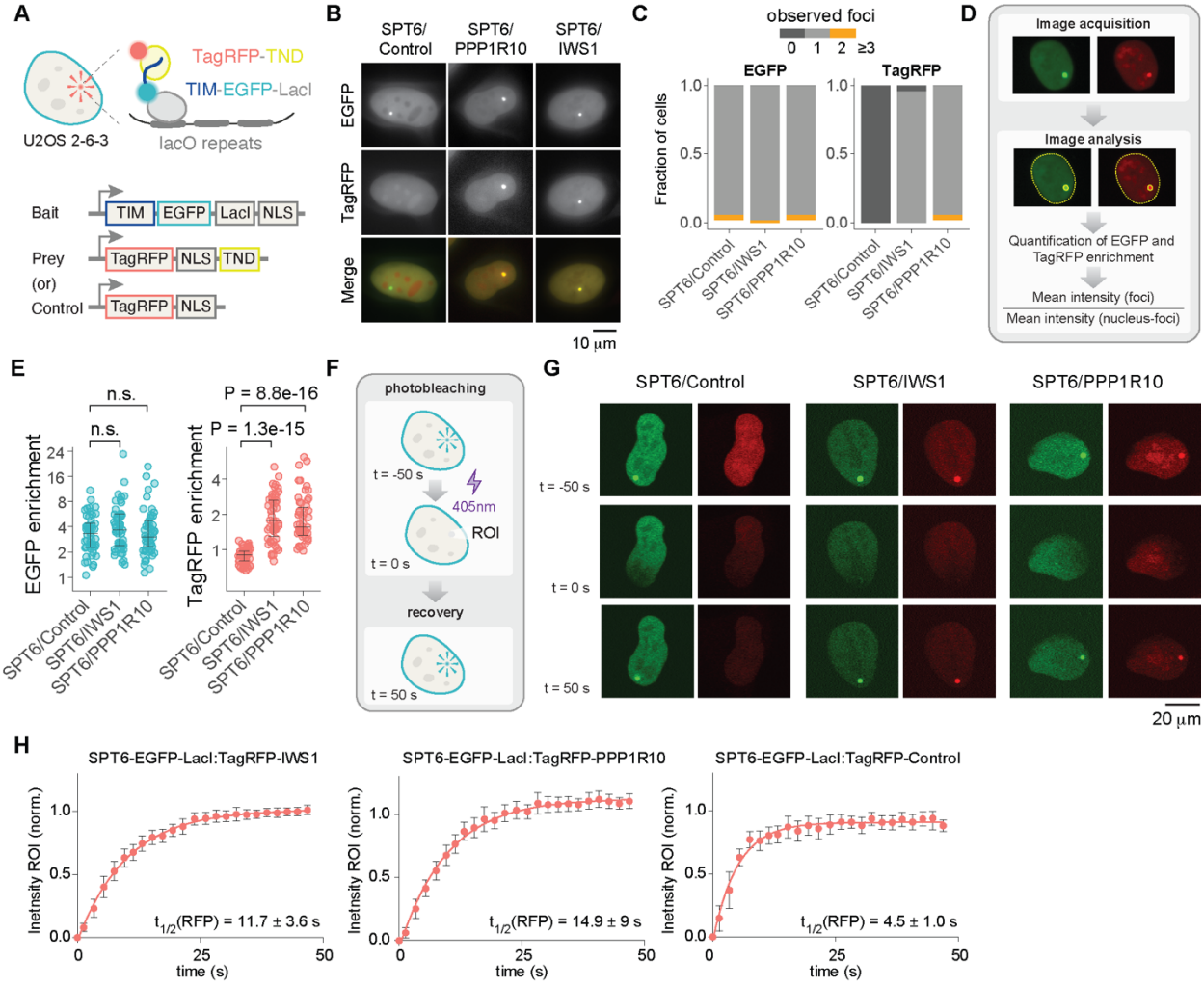

Fig. S10: Optimization of fluorescent two-hybrid (F2H) protein-protein interaction assay. (**A**) Schematic of the protein-protein interaction assay in U2OS 2-6-3 cells. Cells contain a single locus (target site) integrated in the genome consisting of lacO arrays where bait proteins fused to LacI are immobilized. TIM-EGFP-LacI fusions were used as a bait and TagRFP-TNDs were co-expressed as a prey. (**B**) Fluorescence images of SPT6-EGFP-LacI and of TagRFP-control, IWS1-TND or PPP1R10-TND. The merge of the two channels shows that both proteins are recruited at the target site, yielding yellow foci. (**C**) Number of fluorescent foci present in each of the analyzed cells (n = 30 cells). (**D**) Schematic of analytical workflow for determination of EGFP and TagRFP enrichment in fluorescent foci. (**E**) Quantification of EGFP and TagRFP enrichment in fluorescent foci in U2OS 2-6-3 cells (n = 30) expressing SPT6-TIM-EGFP-LacI bait and TagRFP-PPP1R10-TND, TagRFP-IWS1-TND or TagRFP control. (**F**) Workflow of fluorescence recovery after photobleaching (FRAP) experiments. (**G**) Example fluorescence images of U2OS 2-6-3 cells in different stages of FRAP experiment. SPT6-EGFP-LacI is presented in green and TagRFP-control, IWS1-TND or PPP1R10-TND are presented in red. (**H**) FRAP recovery curves for TagRFP-PPP1R10-TND and TagRFP-IWS1-TND. Nuclear localized TagRFP alone is presented as a control for free diffusion.

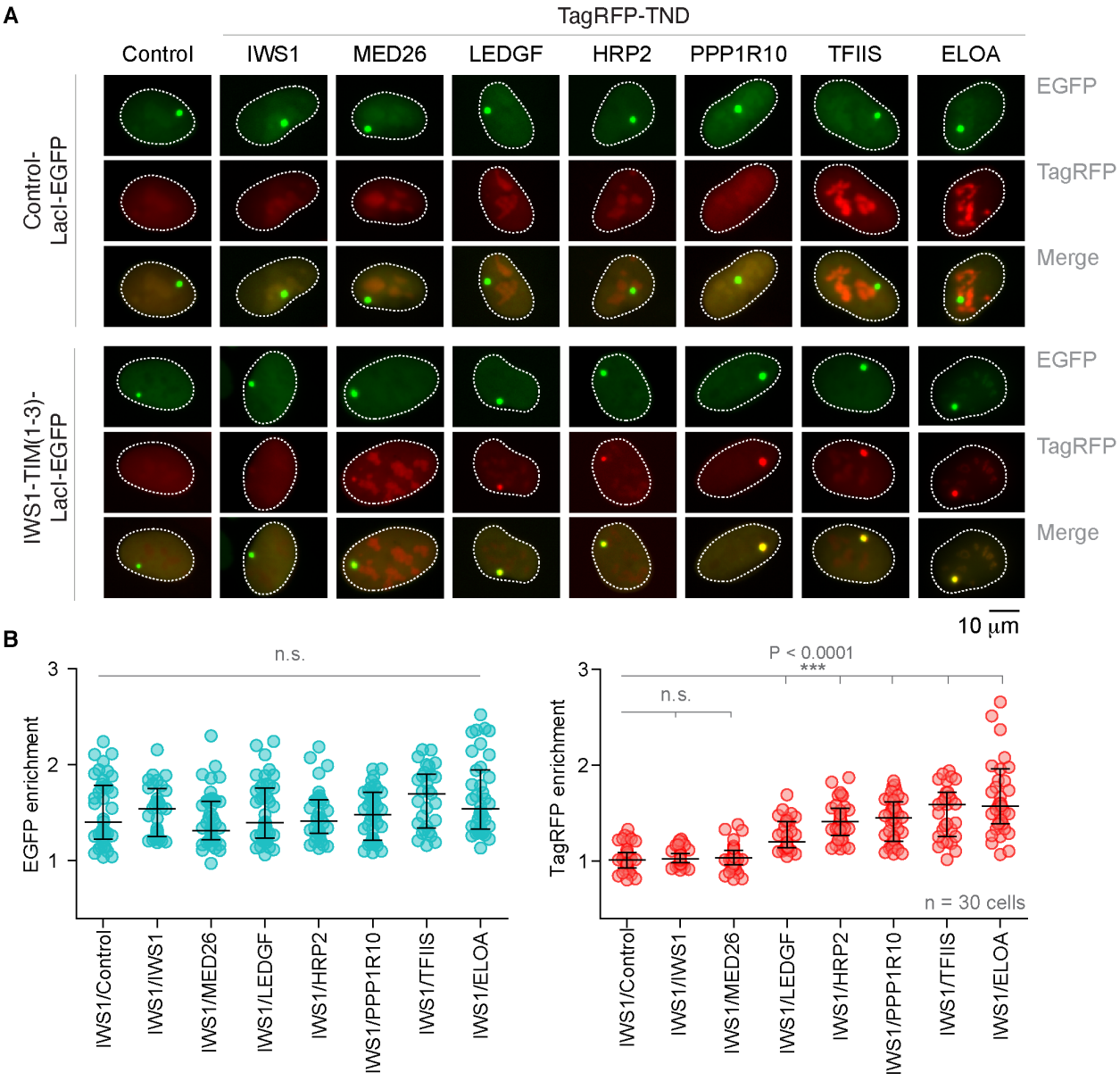

Fig. S11: IWS1 TIMs are sufficient for targeting of LEDGF, HRP2, PPP1R10, TFIIS and ELOA TNDs, but have no detectable interaction with IWS1 and MED26 TNDs in living cells. This figure is an extension and quantification of the data presented in Fig. 3B. (**A**) Representative images of U2OS 2-6-3 cells expressing TIM-EGFP-LacI control, or IWS1-TIM(1-3)-EGFP-LacI bait and library of TagRFP-NLS-TNDs or RFP-NLS control. (**B**) Quantification of EGFP and TagRFP enrichment in fluorescent foci in U2OS 2-6-3 cells (n = 30) from panel (A); *** $P < 0.0001$.

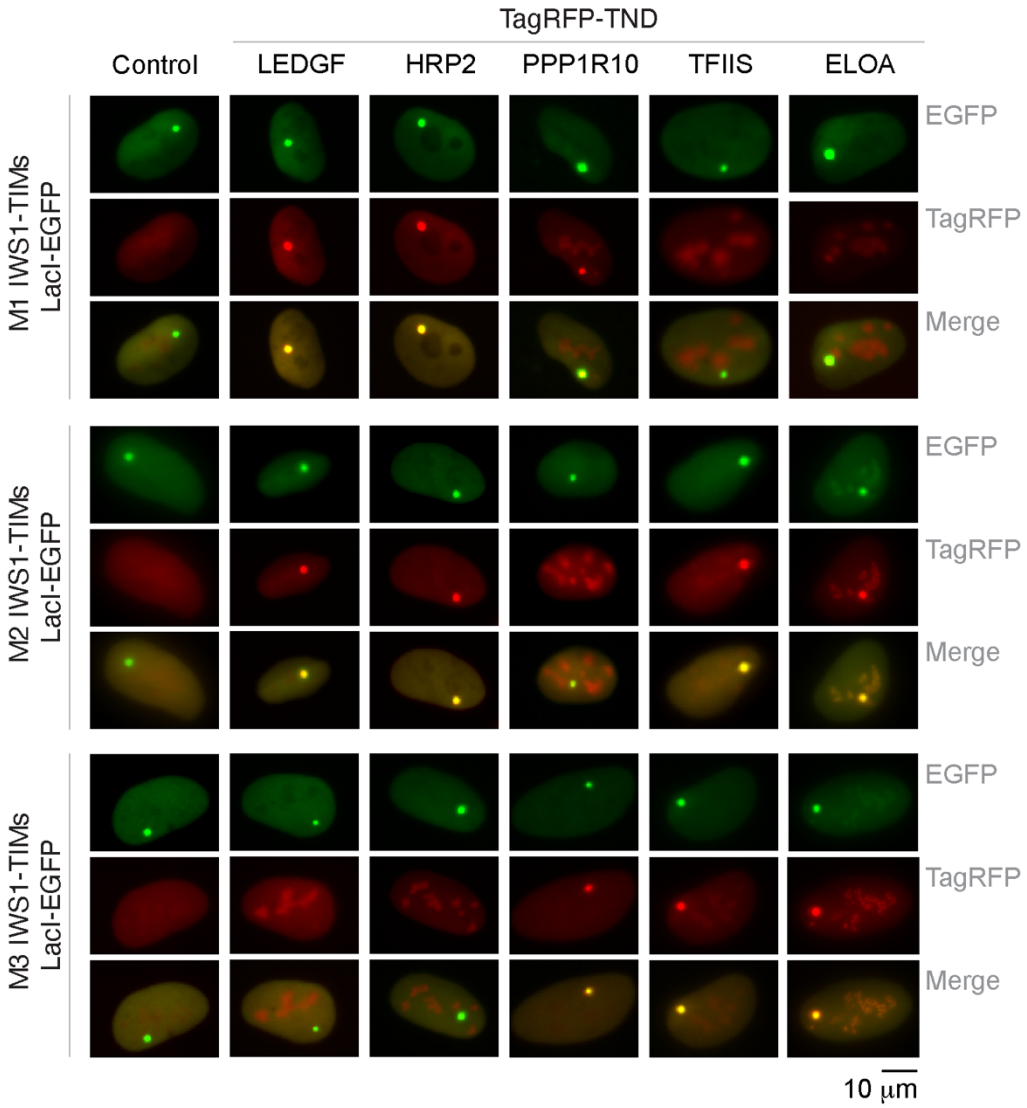

Fig. S12: Mutations in IWS1 TIMs selectively abolish interaction with cognate TNDs. Representative images of U2OS 2-6-3 cells expressing IWS1-TIM-EGFP-LacI L458A/F459A (M1), I476A/F477A (M2) or F488A/F491A (M3) mutant as bait and TagRFP-NLS-TNDs of LEDGF, HRP2, PPP1R10, TFIIS or ELOA TNDs as prey, or TagRFP-NLS control.

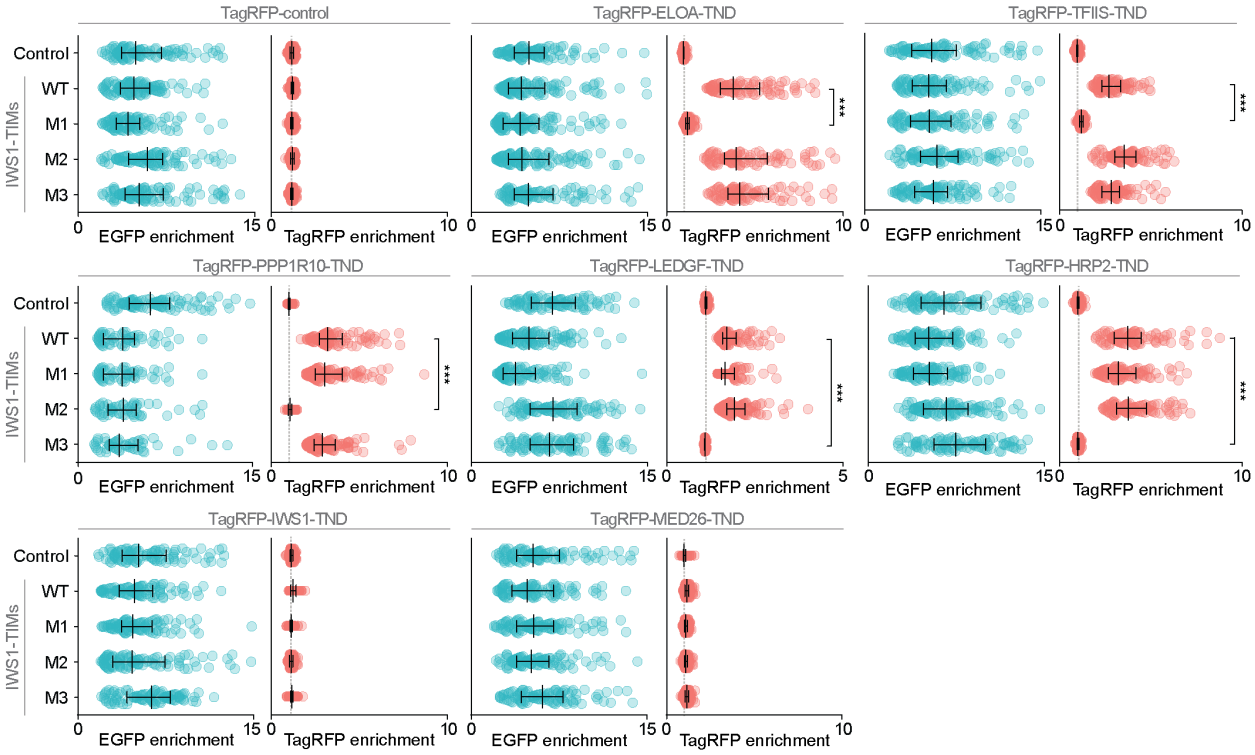

Fig. S13: Quantification of altered interactions of mutant TIMs with cognate TNDs. Quantification of EGFP and TagRFP enrichment in fluorescent foci in U2OS 2-6-3 cells (n = 100, except TagRFP-LEDGF-TND:IWS1-TIMs-M1-EGFP-LacI condition, with n = 88) IWS1-TIM-EGFP-LacI wild-type (WT) or L458A/F459A (M1), I476A/F477A (M2), F488A/F491A (M3) mutants and TagRFP-NLS-TND library as a prey or TagRFP-NLS control. For all panels: *** $P < 0.001$. Representative images for each of these conditions are presented in Figs. 3, S11, and S12.

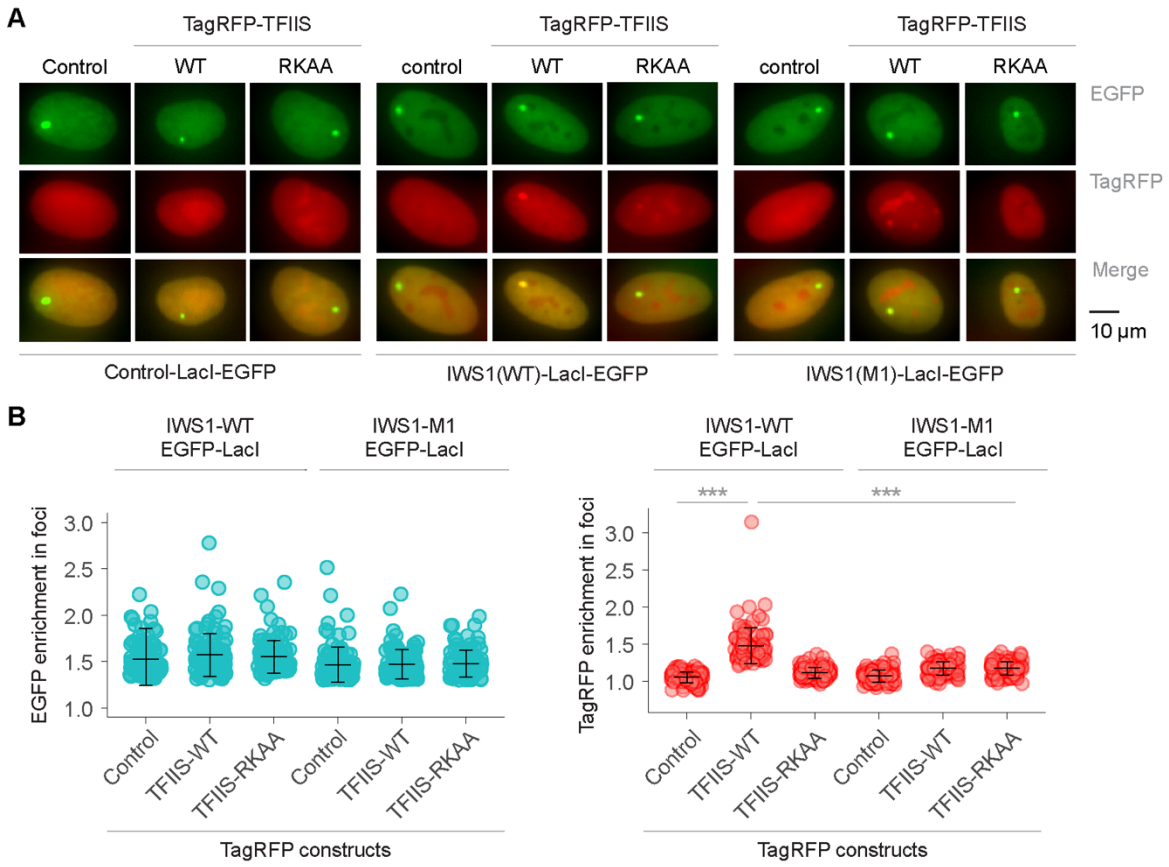

Fig. S14: Interaction between IWS1-TIM and TFIIS-TND is necessary to induce co-localization in living cells. (**A**) Representative images of U2OS 2-6-3 cells expressing EGFP-LacI control, IWS1(full length)-EGFP-LacI wild-type (WT) or L458A/F459A (M1) mutant as a bait and TagRFP-TFIIS (full length)-WT or R54A, K55A (RKAA) as a prey or TagRFP-NLS control. (**B**) Quantification of EGFP and TagRFP enrichment in fluorescent foci in U2OS 2-6-3 cells (n = 100) of the same conditions presented in panel (A). RK-AA mutation of the TFIIS TND resulted in 86% reduction of the signal compared to WT and when combined with IWS1 TIM1 mutation, enrichment was indistinguishable from background (Student's t-test P = 0.590, n = 100 each condition).

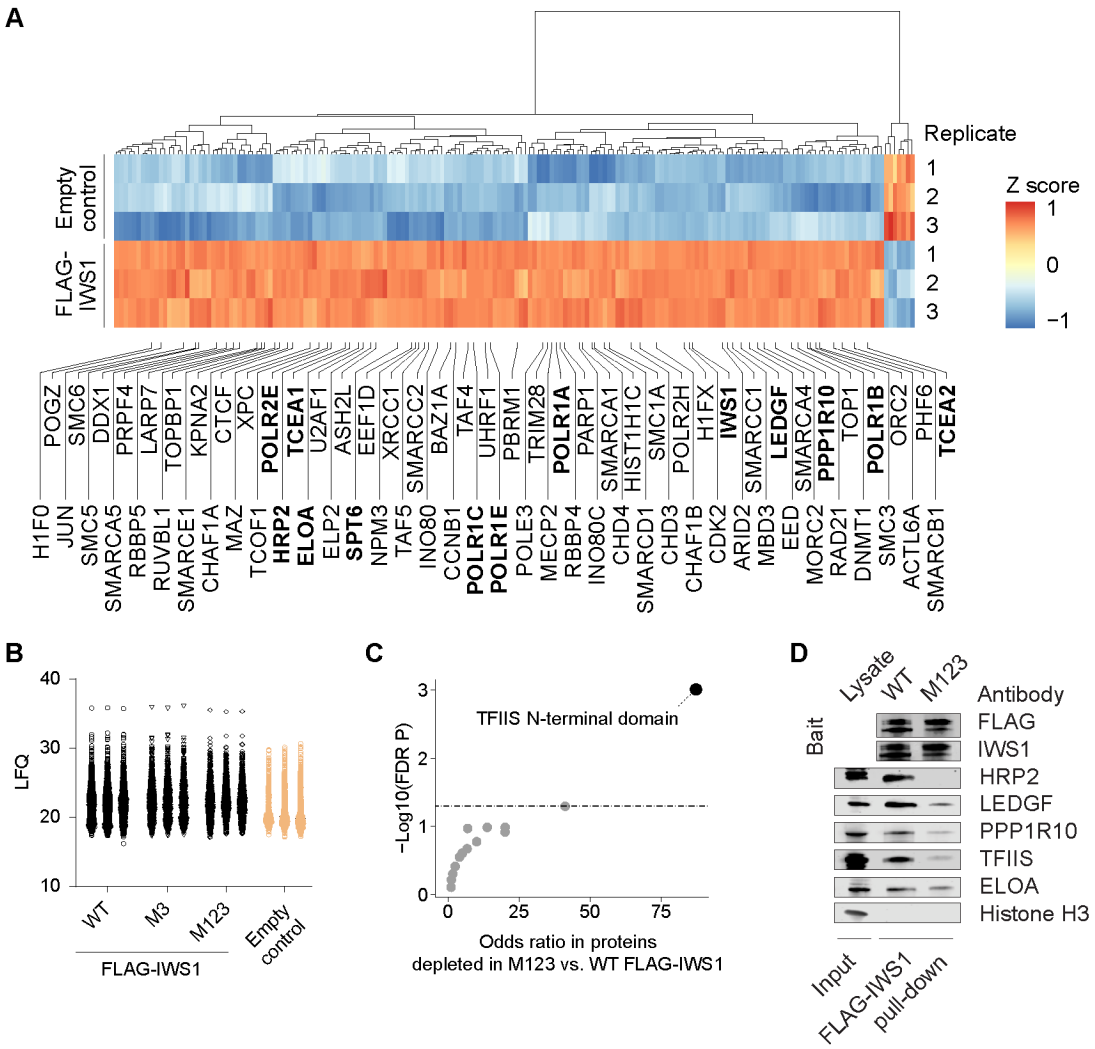

Fig. S15: Mutation of IWS1-TIMs results in loss of TND-containing factors from the IWS1 interactome. (A) Heatmap of the WT FLAG-IWS1 interactome as detected by mass spectrometry. **(B)** Comparison of overall protein abundance across experimental conditions detected by mass spectrometry. LFQ, label-free quantification intensity. **(C)** Enrichment of structural domains in proteins depleted from M123 FLAG-IWS1 interactome. **(D)** Western blot analysis of WT and M123 FLAG-IWS1 pull-down.

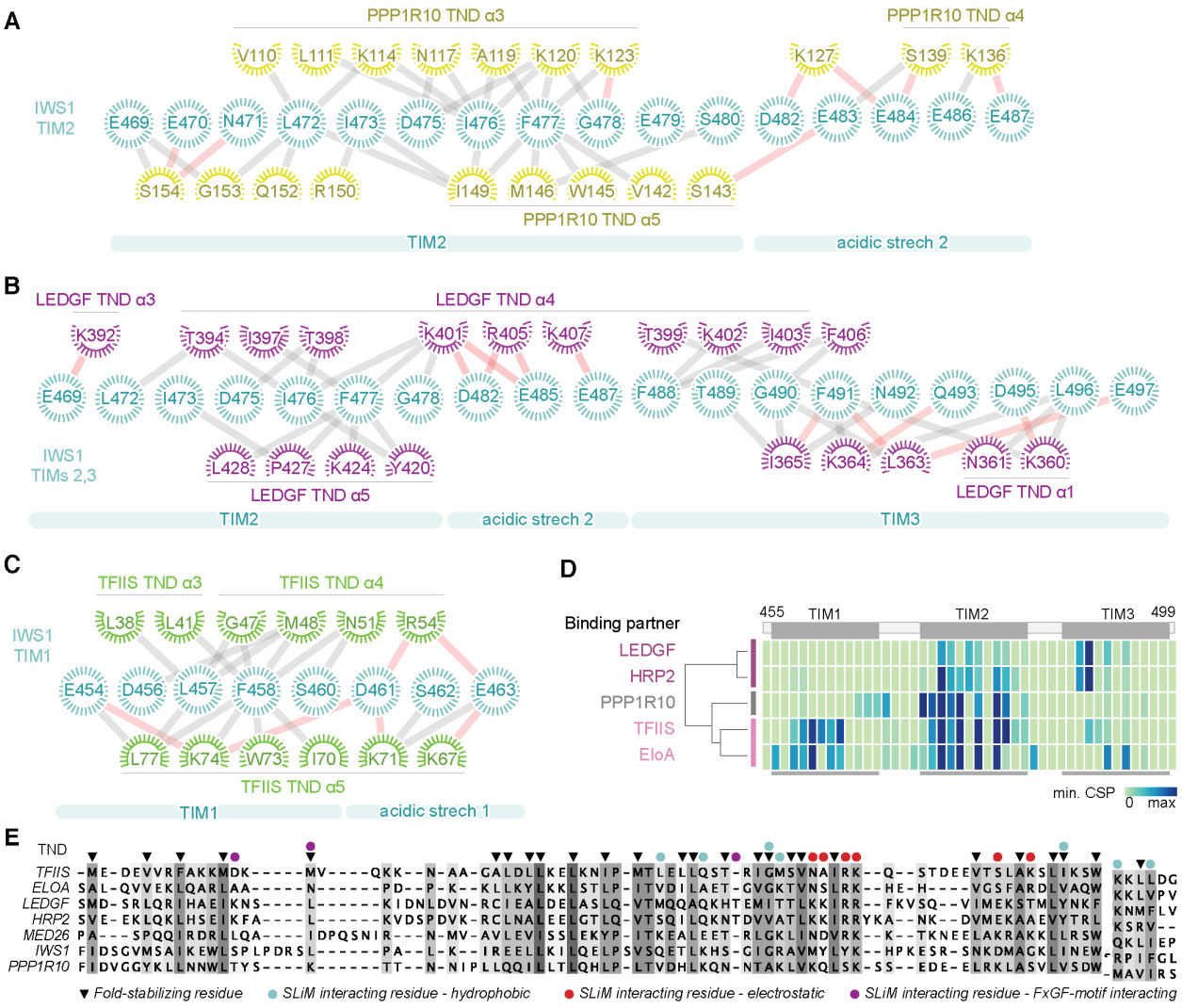

Fig. S16: Detailed structural analysis of contacts maintained by TND-TIM surfaces in protein complexes. Contact map of (**A**) PPP1R10-IWS1 (**B**) LEDGF-IWS1 and (**C**) TFIIS-IWS1 interfaces. Charged interactions are highlighted in red, uncharged contacts are depicted in gray. (**D**) The minimal chemical shift perturbations (CSPs) in IWS1 backbone amide signals (aa 455-499) in the presence of different TNDs. (**E**) Structure-guided multiple sequence alignment of TND sequences.

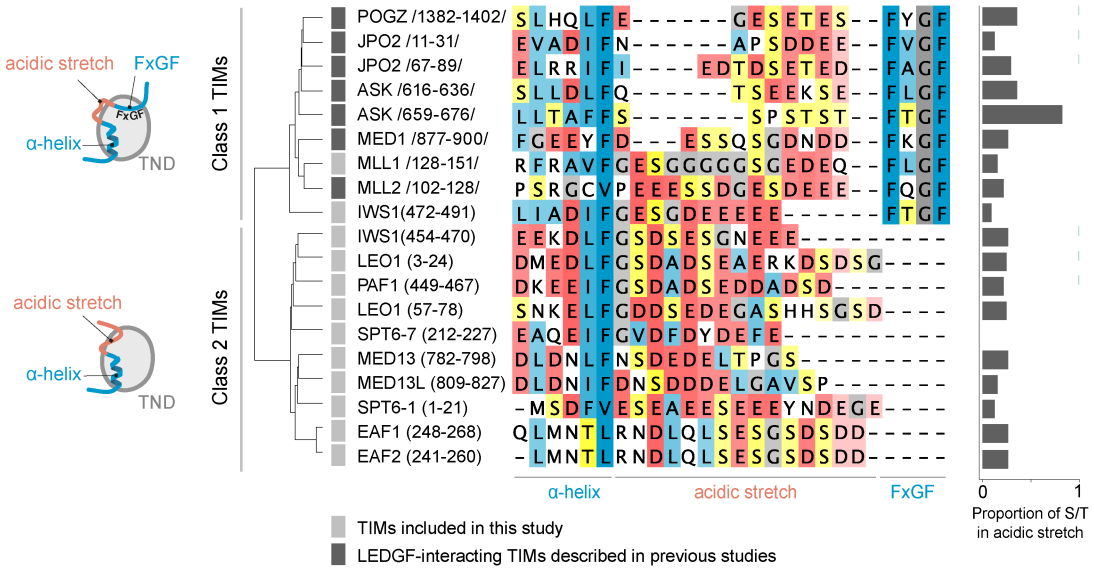

Fig. S17: Classification of TND-interacting motifs (TIMs). Two classes of TIMs can be found in elongation factors. Class 1 is characterized by a region with α-helical propensity followed by an acidic stretch and an FxGF motif. The FxGF portion of the motif is engaged while binding to HRP2 and LEDGF TND. Class 2 motifs contain only a region with α-helical propensity followed by an acidic stretch. The acidic stretches of TIMs that undergo phosphorylation may serve as a regulatory element, controlling the affinity for cognate TNDs by variable charge density.

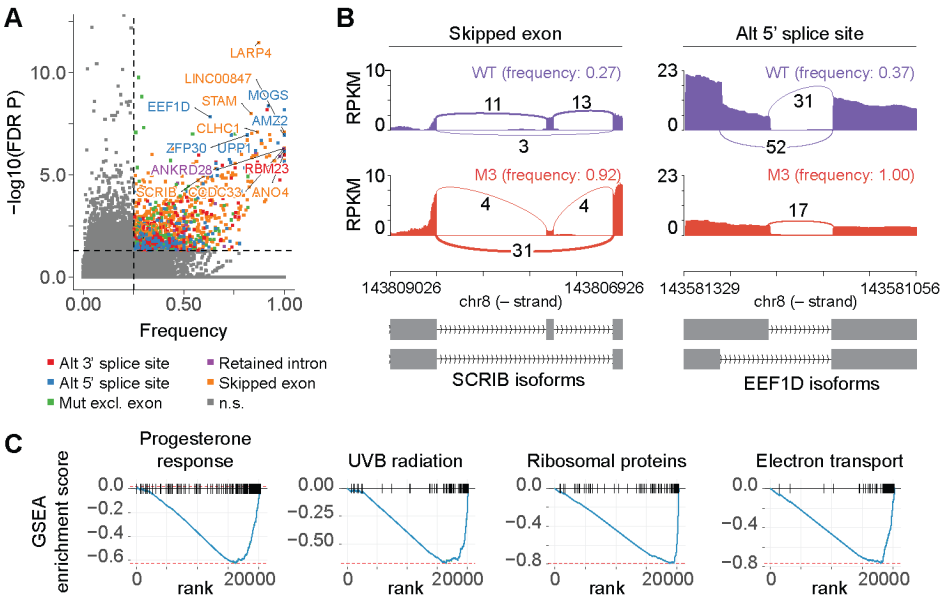

Fig. S18: Mutation of IWS1 TIM3 alters gene expression and induces differential splicing. (**A**) Multivariate Analysis of Transcript Splicing (rMATS) of RNA-seq data from Fig. 5. Genes subject to alternative mRNA splicing are color-coded based on the splicing event and the most significantly affected genes are labeled. (**B**) Browser tracks of example genes that are differentially spiced in the presence of M3 compared to WT IWS1. (**C**) Gene set enrichment analysis (GSEA) of RNA-seq data presented in Fig. 5. Gene sets with reduced expression upon IWS1 mutation are presented.

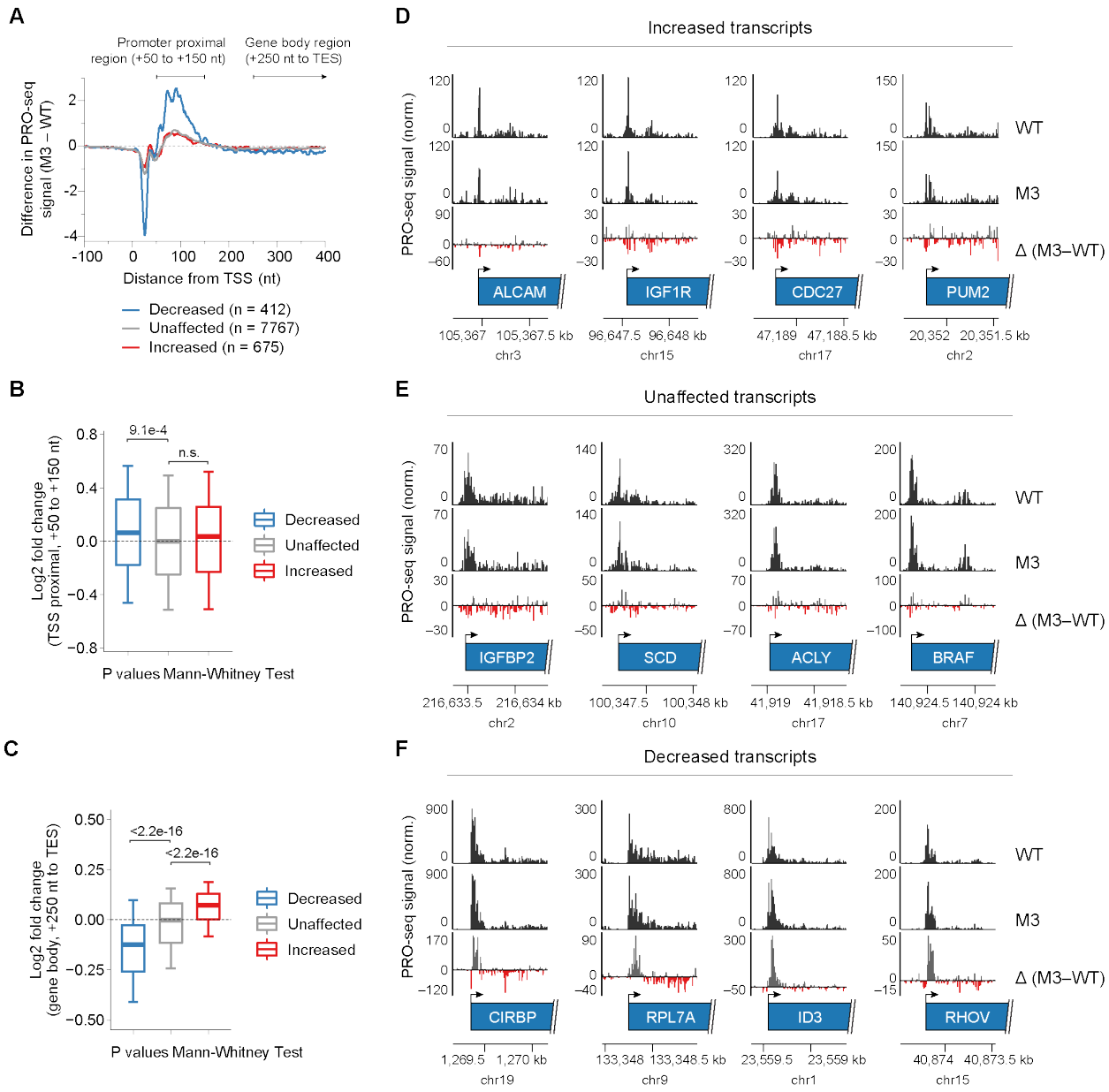

Fig. S19: Mutation of a single IWS1 TIM induces genome-wide defects in transcription elongation. (**A**) The difference in PRO-seq signal near promoters calculated with single-nucleotide resolution. (**B**) Log2 fold change in PRO-seq signal between M3 and WT IWS1 at each gene class at the TSS proximal pause site and (**C**) gene body. (**D**) Example PRO-seq browser tracks at increased, (**E**) unaffected, and (**F**) decreased genes. The difference in mean PRO-seq signal between WT and M3 conditions is presented below individual browser tracks. Abbreviations: TSS, transcription start site; TES, transcription end site.

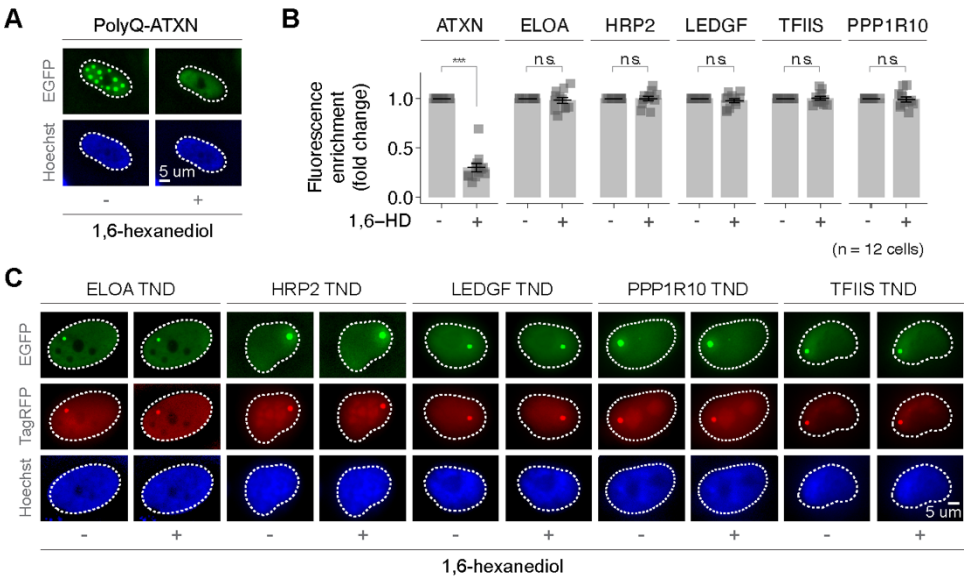

Fig. S20: 1,6-hexanediol does not dissolve binary TND-TIM interactions. (**A**) Positive control; representative images of U2OS 2-6-3 cells expressing EGFP-polyQ-ATXN in the presence or absence of 5% 1,6-hexanediol (*76*). (**B**) Quantification of panel (A) and signal intensity in fluorescent foci of U2OS 2-6-3 cells before and after treatment with 5% 1,6-hexanediol (1,6-HD). (**C**) Images corresponding to the data presented in panel (B). Representative images of U2OS 2-6-3 cells expressing WT IWS1-TIM-EGFP-LacI as bait with a panel of TagRFP-NLS-TNDs as prey in the presence or absence of 5% 1,6-hexanediol.

Table S1: Summary of GO terms analyzed in Figure 1 to calculate domain enrichment in different nuclear processes

Name	GO term #	Biological process defined by GO term
Initiation	GO:0006367	transcription initiation from RNA polymerase II promoter
Elongation	GO:0006368	transcription elongation from RNA polymerase II promoter
Termination	GO:0006369	termination of RNA polymerase II transcription
RNAP1	GO:0006360	transcription by RNA polymerase I
RNAP2	GO:0006366	transcription by RNA polymerase II
RNAP3	GO:0006383	transcription by RNA polymerase III
Transcription	GO:0006351	transcription, DNA-templated
Translation	GO:0006412	translation
RNA processing	GO:0006396	RNA processing
mRNA export	GO:0006406	mRNA export from nucleus
Protein maturation	GO:0051604	protein maturation
Gene expression	GO:0010467	gene expression
DNA damage	GO:0006974	cellular response to DNA damage stimulus
Nuclear transport	GO:0051169	nuclear transport
Cell cycle	GO:0007049	cell cycle
Cell death	GO:0008219	cell death
Chromatin organization	GO:0051276	chromosome organization
Differentiation	GO:0030154	cell differentiation
Homeostasis	GO:0042592	homeostatic process
Signaling	GO:0023052	signaling

Table S2: Roles of TND- and TIM-containing factors in transcription elongation

Protein	UniProtKB	Role in elongation	Citation
IWS1	Q96ST2	defines the composition of the RNAP2 elongation complex	Yoh et al. Genes Dev. 2007;21(2):160-74.; Yoh et al. Genes Dev. 2008;22(24):3422-34.
HRP2	Q7Z4V5	replaces the FACT role in overcoming the nucleosome-induced barrier to transcription elongation in differentiated cells; interacts with various proteins known to be involved in transcription elongation	Gao et al. Biochem Biophys Res Commun. 2015;458(4):849-55.; LeRoy et al. Sci Adv. 2019;5(10):eaay3068
LEDGF	O75475	replaces the FACT role in overcoming the nucleosome-induced barrier to transcription elongation in differentiated cells	LeRoy et al. Sci Adv. 2019;5(10):eaay3068
PPP1R10	Q96QC0	negatively regulates RNAP2 elongation rate; inhibits the assembly of core components of the transcription elongation factor P-TEFb	Cortazar et al. Mol Cell. 2019;76(6):896-908.e4.; Kapoor et al. Biochem J. 2015;470(3):293-302.
TFIIS	P23193	stimulates RNAP2 cleavage activity; rescues arrested and backtracked elongation complexes	Reines et al. J Biol Chem. 1992;267(6):3795-800.; Cheunget al. Nature. 2011;471(7337):249-53.
TCEA2	Q15560	stimulates RNAP2 cleavage activity; rescues arrested and backtracked elongation complexes; TFIIS isoform (testis-specific)	Reines et al. J Biol Chem. 1992;267(6):3795-800.; Cheunget al. Nature. 2011;471(7337):249-53.
TCEA3	O75764	stimulates RNAP2 cleavage activity; rescues arrested and backtracked elongation complexes; TFIIS isoform (highly expressed in embryonic stem cells)	Reines et al. J Biol Chem. 1992;267(6):3795-800.; Cheunget al. Nature. 2011;471(7337):249-53.
TCEANC2	Q96MN5	TFIIS homolog	
MED26	O95402	recruits Super Elongation Complex members to RNAP2 to facilitate elongation; component of the Mediator complex	Takahashi et al. Cell. 2011;146(1):92-104.; Conaway et al. Trends Biochem Sci. 2005;30(5):250-5.
TCEANC	Q8N8B7	TFIIS homolog	
ELOA	Q14241	suppresses RNAP2 transient pausing; subunit of the transcription factor B (SIII) complex	Aso et al. Science. 1995;269(5229):1439-43.
ELOA2	Q8IYF1	suppresses RNAP2 transient pausing; subunit of the transcription factor B (SIII) complex	Aso et al. Science. 1995;269(5229):1439-43.
TCEB3CL	Q3SY89	suppresses RNAP2 transient pausing; subunit of the transcription factor B (SIII) complex	Aso et al. Science. 1995;269(5229):1439-43.
TCEB3L2	Q8NG57	suppresses RNAP2 transient pausing; subunit of the transcription factor B (SIII) complex	Aso et al. Science. 1995;269(5229):1439-43.
TCEB3CL2	A6NLF2	suppresses RNAP2 transient pausing; subunit of the transcription factor B (SIII) complex	Aso et al. Science. 1995;269(5229):1439-43.
SPT6	Q7KZ85	stimulates RNAP2 transcription elongation	Endoh et al. Mol Cell Biol. 2004;24(8):3324-36.

MED13	Q9UHV7	component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNAP2 dependent genes	Weryet et al. EMBO J. 2004;23(21):4232-42.; Conaway et al. Trends Biochem Sci. 2005;30(5):250-5.; Baek et al. JBC (2006) P15172-15181.
MED13L	Q71F56	component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNAP2 dependent genes	Weryet et al. EMBO J. 2004;23(21):4232-42.; Conaway et al. Trends Biochem Sci. 2005;30(5):250-5.; Baek et al. JBC (2006) P15172-15181.
PAF1	Q8N7H5	regulation of promoter-proximal pausing by RNAP2; component of the PAF1 complex	Chen et al. Cell. 2015;162(5):1003-1015; Yu et al. Science. 2015 Dec 11;350(6266):1383-6.; Chen et al. Genes Dev. 2009;23(23):2765-77; Hou et al. PNAS 2019; 116(29) 14583-14592
EAF1	Q96JC9	transcriptional transactivator of ELL and ELL2 elongation activities	Kong et al. Proc Natl Acad Sci U S A. 2005;102(29):10094-8.
EAF2	Q96CJ1	transcriptional transactivator of ELL and ELL2 elongation activities	Kong et al. Proc Natl Acad Sci U S A. 2005;102(29):10094-8.
LEO1	Q8WVC0	regulation of promoter-proximal RNAP2 pausing; component of the PAF1 complex	Chen et al. Cell. 2015;162(5):1003-1015; Yu et al. Science. 2015 Dec 11;350(6266):1383-6.; Chen et al. Genes Dev. 2009;23(23):2765-77.
MLL1	Q03164	MLL associates with RNAP2 at actively transcribed genes. Loss of MLL results in RNAP2 re-distribution	Milne et al. Proc Natl Acad Sci U S A. 2005;02(41):14765-70.

Table S3: Statistics for the final water-refined sets of structures

	TFIIS	ELOA	LEDGF	IWS1	PPP1R10	MED26	TFIIS-IWS1
Non-redundant constrains							
NOE restraints	2043	1957	1950	2907	3088	1819	2885
Intra-residue (i = j)	484	443	593	754	784	500	588
Sequential (| i - j | = 1)	407	468	392	689	771	391	603
Medium-range (1 < | i - j | < 5)	501	524	484	660	743	445	741
Long-range (| i - j | ≥ 5)	651	522	481	804	790	483	953
Torsion angles	138	128	154	220	246	144	162
Total number of restricting restraints	2181	2085	2104	3127	3334	1963	3047
Total restricting restraints per restrained residue	29.7	25.7	24.2	20.9	23.3	21.2	29.7
Residual constr. violations							
Distance viol. per structure							
0.1 – 0.2 Å	4.8	4.3	1.6	10.5	6.9	11.7	23.3
0.2 – 0.5 Å	1.1	1.6	0.4	1.8	2.0	6.4	8.1
> 0.5 Å	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r.m.s. of distance violation per constraint	0.01 Å	0.02 Å	0.01 Å	0.02 Å	0.01 Å	0.03 Å	0.02 Å
Maximum distance violation	0.45 Å	0.47 Å	0.49 Å	0.47 Å	0.50 Å	0.50 Å	0.50 Å
Dihedral angle viol. per structure							
1 – 10 °	1.2	2.1	1.7	3.3	2.3	2.3	3.5
> 10 °	0.0	0.0	0.0	0.0	0.0	0.0	0.0
r.m.s. of dihedral violations per constraint	0.36°	0.52°	0.40°	0.42°	0.36°	0.38°	0.49°
Maximum dihedral angle viol.	5.1°	5.1°	5.0°	5.0°	5.0°	4.9°	5.1°
Ramachandran plot							
Most favored regions	96.4 %	95.3 %	91.2 %	95.2 %	97.5 %	92.8 %	96.4 %
Additionally allowed regions	3.6 %	4.7 %	8.8 %	4.6 %	2.5 %	7.2 %	3.6 %
Generously allowed regions	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %
Disallowed regions	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
r.m.s.d. to the mean structure	*all/order*[1]	*all/order*[1]	*all/order*[1]	*all/order*[1]	*all/order*[1]	*all/order*[1]	*all/order*[1]
All backbone atoms	0.8/0.2 Å	3.2/0.2 Å	1.5/0.3 Å	2.6/0.3 Å	2.3/0.5 Å	1.1/0.3 Å	2.9/0.2 Å
All heavy atoms	1.1/0.7 Å	3.2/0.7 Å	1.7/0.8 Å	2.7/0.8 Å	2.3/0.8 Å	1.4/0.9 Å	3.0/0.7 Å
PDB entry	6ZUY	6ZUZ	6ZV0	6ZV1	6ZV2	6ZV3	6ZV4
BMRB accession code	34535	34536	34537	34538	34539	34540	34541

[1]Residues with sum of phi and psi order parameters > 1.8

Data S1. (separate file)

Raw data for domain enrichment analysis presented in Figure 1; GO terms used for this analysis are presented in Table S1.

Data S2. (separate file)

Raw data for proteome-wide search for putative human TND-interacting motifs (TIMs) using SLiMSearch4. Summary of these results is presented in Fig. S3.

Data S3. (separate file)

Raw data for mass spectrometry analysis of IWS1 interactome presented in Fig. 3 and Fig S15.

References and Notes

1. L. J. Core, J. J. Waterfall, J. T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (80-.).* (2008), doi:10.1126/science.1162228.

2. C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, C. Bustamante, Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (80-.).* (2009), doi:10.1126/science.1172926.

3. A. Mayer, J. Di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos, L. S. Churchman, Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* (2015), doi:10.1016/j.cell.2015.03.010.

4. T. E. Miller, B. B. Liau, L. C. Wallace, A. R. Morton, Q. Xie, D. Dixit, D. C. Factor, L. J. Y. Kim, J. J. Morrow, Q. Wu, S. C. Mack, C. G. Hubert, S. M. Gillespie, W. A. Flavahan, T. Hoffmann, R. Thummalapalli, M. T. Hemann, P. J. Paddison, C. M. Horbinski, J. Zuber, P. C. Scacheri, B. E. Bernstein, P. J. Tesar, J. N. Rich, Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. *Nature* (2017), doi:10.1038/nature23000.

5. C. Lin, E. R. Smith, H. Takahashi, K. C. Lai, S. Martin-Brown, L. Florens, M. P. Washburn, J. W. Conaway, R. C. Conaway, A. Shilatifard, AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. *Mol. Cell* (2010), doi:10.1016/j.molcel.2010.01.026.

6. Y. Yuva-Aydemir, S. Almeida, G. Krishnan, T. F. Gendron, F. B. Gao, Transcription elongation factor AFF2/FMR2 regulates expression of expanded GGGGCC repeat-containing C9ORF72 allele in ALS/FTD. *Nat. Commun.* (2019), doi:10.1038/s41467-019-13477-8.

7. S. A. Yukl, P. Kaiser, P. Kim, S. Telwatte, S. K. Joshi, M. Vu, H. Lampiris, J. K. Wong, HIV latency in isolated patient CD4+ T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci. Transl. Med.* (2018), doi:10.1126/scitranslmed.aap9927.

8. K. Adelman, M. A. Kennedy, S. Nechaev, D. A. Gilchrist, G. W. Muse, Y. Chinenov, I. Rogatsky, Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proc. Natl. Acad. Sci. U. S. A.* (2009), doi:10.1073/pnas.0910177106.

9. M. Fousteri, W. Vermeulen, A. A. van Zeeland, L. H. F. Mullenders, Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo. *Mol. Cell* (2006), doi:10.1016/j.molcel.2006.06.029.

10. X. Bai, J. Kim, Z. Yang, M. J. Jurynec, T. E. Akie, J. Lee, J. LeBlanc, A. Sessa, H. Jiang,

A. DiBiase, Y. Zhou, D. J. Grunwald, S. Lin, A. B. Cantor, S. H. Orkin, L. I. Zon, TIF1γ Controls Erythroid Cell Fate by Regulating Transcription Elongation. *Cell* (2010), doi:10.1016/j.cell.2010.05.028.

11. K. Izumi, R. Nakato, Z. Zhang, A. C. Edmondson, S. Noon, M. C. Dulik, R. Rajagopalan, C. P. Venditti, K. Gripp, J. Samanich, E. H. Zackai, M. A. Deardorff, D. Clark, J. L. Allen, D. Dorsett, Z. Misulovin, M. Komata, M. Bando, M. Kaur, Y. Katou, K. Shirahige, I. D. Krantz, Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. *Nat. Genet.* (2015), doi:10.1038/ng.3229.

12. B. M. Peterlin, D. H. Price, Controlling the Elongation Phase of Transcription with P-TEFb. *Mol. Cell* (2006), doi:10.1016/j.molcel.2006.06.014.

13. K. Glover-Cutter, S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher, D. L. Bentley, TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Mol. Cell. Biol.* (2009), doi:10.1128/mcb.00637-09.

14. M. A. Cortazar, R. M. Sheridan, B. Erickson, N. Fong, K. Glover-Cutter, K. Brannan, D. L. Bentley, Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Mol. Cell* (2019), doi:10.1016/j.molcel.2019.09.031.

15. E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, C. Bustamante, Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* (2007), doi:10.1038/nature05701.

16. T. Aso, W. S. Lane, J. W. Conaway, R. C. Conaway, Elongin (SIII): A multisubunit regulator of elongation by RNA polymerase II. *Science (80-.).* (1995), doi:10.1126/science.7660129.

17. Y. Liu, K. Zhou, N. Zhang, H. Wei, Y. Z. Tan, Z. Zhang, B. Carragher, C. S. Potter, S. D'Arcy, K. Luger, FACT caught in the act of manipulating the nucleosome. *Nature* (2020), doi:10.1038/s41586-019-1820-0.

18. A. Bortvin, F. Winston, Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science (80-.).* (1996), doi:10.1126/science.272.5267.1473.

19. G. LeRoy, O. Oksuz, N. Descostes, Y. Aoi, R. A. Ganai, H. O. Kara, J. R. Yu, C. H. Lee, J. Stafford, A. Shilatifard, D. Reinberg, LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. *Sci. Adv.* (2019), doi:10.1126/sciadv.aay3068.

20. Z. Luo, C. Lin, A. Shilatifard, The super elongation complex (SEC) family in transcriptional control. *Nat. Rev. Mol. Cell Biol.* (2012), doi:10.1038/nrm3417.

21. B. A. Gibson, Y. Zhang, H. Jiang, K. M. Hussey, J. H. Shrimp, H. Lin, F. Schwede, Y. Yu, W. L. Kraus, Chemical genetic discovery of PARP targets reveals a role for PARP-1

in transcription elongation. *Science (80-.).* (2016), doi:10.1126/science.aaf7865.

22. H. Takahashi, T. J. Parmely, S. Sato, C. Tomomori-Sato, C. A. S. Banks, S. E. Kong, H. Szutorisz, S. K. Swanson, S. Martin-Brown, M. P. Washburn, L. Florens, C. W. Seidel, C. Lin, E. R. Smith, A. Shilatifard, R. C. Conaway, J. W. Conaway, Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* (2011), doi:10.1016/j.cell.2011.06.005.

23. L. Baranello, D. Wojtowicz, K. Cui, B. N. Devaiah, H. J. Chung, K. Y. Chan-Salis, R. Guha, K. Wilson, X. Zhang, H. Zhang, J. Piotrowski, C. J. Thomas, D. S. Singer, B. F. Pugh, Y. Pommier, T. M. Przytycka, F. Kouzine, B. A. Lewis, K. Zhao, D. Levens, RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* (2016), doi:10.1016/j.cell.2016.02.036.

24. K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, N. E. Davey, Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* (2014), doi:10.1021/cr400585q.

25. B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).* (2018), doi:10.1126/science.aar3958.

26. N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, T. J. Gibson, Attributes of short linear motifs. *Mol. Biosyst.* (2012), doi:10.1039/c1mb05231d.

27. S. Sharma, K. Čermáková, J. De Rijck, J. Demeulemeester, M. Fábry, S. El Ashkar, S. Van Belle, M. Lepšík, P. Tesina, V. Duchoslav, P. Novák, M. Hubálek, P. Srb, F. Christ, P. Řezáčová, H. C. Hodges, Z. Debyser, V. Veverka, Affinity switching of the LEDGF/p75 IBD interactome is governed by kinase-dependent phosphorylation. *Proc. Natl. Acad. Sci.* (2018), doi:10.1073/pnas.1803909115.

28. B. A. Gibson, L. K. Doolittle, M. W. G. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, M. K. Rosen, Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* (2019), doi:10.1016/j.cell.2019.08.037.

29. V. Booth, C. M. Koth, A. M. Edwards, C. H. Arrowsmith, Structure of a conserved domain common to the transcription factors TFIIS, Elongin A, and CRSP70. *J. Biol. Chem.* (2000), doi:10.1074/jbc.M002595200.

30. P. Tesina, K. Cermáková, M. Horejší, K. Procházková, M. Fábry, S. Sharma, F. Christ, J. Demeulemeester, Z. Debyser, J. De Rijck, V. Veverka, P. Rezácová, Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. *Nat. Commun.* (2015), doi:10.1038/ncomms8968.

31. K. Čermaková, P. Tesina, J. Demeulemeester, S. El Ashkar, H. Méreau, J. Schwaller, P. Řezáčová, V. Veverka, J. De Rijck, Validation and structural characterization of the

LEDGF/p75-MLL interface as a new target for the treatment of MLL-dependent leukemia. *Cancer Res.* (2014), doi:10.1158/0008-5472.CAN-13-3602.

32. M. L. Diebold, M. Koch, E. Loeliger, V. Cura, F. Winston, J. Cavarelli, C. Romier, The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. *EMBO J.* (2010), doi:10.1038/emboj.2010.272.

33. S. M. McDonald, D. Close, H. Xin, T. Formosa, C. P. Hill, Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding. *Mol. Cell* (2010), doi:10.1016/j.molcel.2010.11.014.

34. S. M. Janicki, T. Tsukamoto, S. E. Salghetti, W. P. Tansey, R. Sachidanandam, K. V. Prasanth, T. Ried, Y. Shav-Tal, E. Bertrand, R. H. Singer, D. L. Spector, From silencing to gene expression: Real-time analysis in single cells. *Cell* (2004), doi:10.1016/S0092-8674(04)00171-0.

35. N. I. Reim, J. Chuang, D. Jain, B. H. Alver, P. J. Park, F. Winston, The conserved elongation factor Spn1 is required for normal transcription, histone modifications, and splicing in Saccharomyces cerevisiae. *Nucleic Acids Res.* (2020), doi:10.1093/nar/gkaa745.

36. S. M. Yoh, H. Cho, L. Pickle, R. M. Evans, K. A. Jones, The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* (2007), doi:10.1101/gad.1503107.

37. R. Dronamraju, J. L. Kerschner, S. A. Peck, A. J. Hepperla, A. T. Adams, K. D. Hughes, S. Aslam, A. R. Yoblinski, I. J. Davis, A. L. Mosley, B. D. Strahl, Casein Kinase II Phosphorylation of Spt6 Enforces Transcriptional Fidelity by Maintaining Spn1-Spt6 Interaction. *Cell Rep.* (2018), doi:10.1016/j.celrep.2018.11.089.

38. D. B. Mahat, H. Kwak, G. T. Booth, I. H. Jonkers, C. G. Danko, R. K. Patel, C. T. Waters, K. Munson, L. J. Core, J. T. Lis, Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* (2016), doi:10.1038/nprot.2016.086.

39. L. Core, K. Adelman, Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes Dev.* (2019), doi:10.1101/gad.325142.119.

40. C. M. Weber, S. Ramachandran, S. Henikoff, Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase. *Mol. Cell* (2014), doi:10.1016/j.molcel.2014.02.014.

41. H. C. Hodges, Code: “hodgeslab/workflows”, Version 20210915. *Zenodo* (2021), doi:10.5281/zenodo.5511049.

42. Adelman lab, Code: “AdelmanLab/NIH_scripts”, Version 1.0. *Zenodo* (2021), doi:10.5281/zenodo.5519915.

43. Adelman lab, Code: “AdelmanLab/GetGeneAnnotation_GGA”, Version 1.0. *Zenodo* (2021), doi:10.5281/zenodo.5519928.

44. S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto, R. D. Finn, The Pfam protein families database in 2019. *Nucleic Acids Res.* (2019), doi:10.1093/nar/gky995.

45. A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, J. Zhang, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* (2017), doi:10.1093/nar/gkw1099.

46. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. Von Mering, STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015), doi:10.1093/nar/gku1003.

47. P. Cherepanov, Z. Y. J. Sun, S. Rahman, G. Maertens, G. Wagner, A. Engelman, Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat. Struct. Mol. Biol.* (2005), doi:10.1038/nsmb937.

48. I. Krystkowiak, N. E. Davey, SLiMSearch: A framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res.* (2017), doi:10.1093/nar/gkx238.

49. N. E. Davey, J. L. Cowan, D. C. Shields, T. J. Gibson, M. J. Coldwell, R. J. Edwards, SLiMPrints: Conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res.* (2012), doi:10.1093/nar/gks854.

50. B. Mészáros, G. Erdös, Z. Dosztányi, IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* (2018), doi:10.1093/nar/gky384.

51. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins*

Proteomics (2010), doi:10.1016/j.bbapap.2010.01.011.

52. J. Söding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* (2005), doi:10.1093/nar/gki408.

53. J. Chiu, P. E. March, R. Lee, D. Tillett, Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.* (2004), doi:10.1093/nar/gnh172.

54. P. S. Renshaw, V. Veverka, G. Kelly, T. A. Frenkiel, R. A. Williamson, S. V. Gordon, R. G. Hewinson, M. D. Carr, Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the Mycobacterium tuberculosis protein CFP-10 and ESAT-6. *J. Biomol. NMR* (2004), , doi:10.1023/B:JNMR.0000048852.40853.5c.

55. V. Veverka, G. Lennie, T. Crabbe, I. Bird, R. J. Taylor, M. D. Carr, NMR assignment of the mTOR domain responsible for rapamycin binding [3]. *J. Biomol. NMR* (2006), , doi:10.1007/s10858-005-4324-1.

56. V. Veverka, T. Crabbe, I. Bird, G. Lennie, F. W. Muskett, R. J. Taylor, M. D. Carr, Structural characterization of the interaction of mTOR with phosphatidic acid and a novel class of inhibitor: Compelling evidence for a central role of the FRB domain in small molecule-mediated regulation of mTOR. *Oncogene* (2008), doi:10.1038/sj.onc.1210693.

57. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* (2002), doi:10.1016/S0022-2836(02)00241-3.

58. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* (2009), doi:10.1007/s10858-009-9333-z.

59. E. Harjes, S. Harjes, S. Wohlgemuth, K. H. Müller, E. Krieger, C. Herrmann, P. Bayer, GTP-Ras Disrupts the Intramolecular Complex of C1 and RA Domains of Nore1. *Structure* (2006), doi:10.1016/j.str.2006.03.008.

60. R. Peruzzini, Z. Lens, A. Verger, F. Dewitte, E. Ferreira, J. L. Baert, V. Villeret, I. Landrieu, F. X. Cantrelle, 1H, 15N and 13C assignments of the N-terminal domain of the Mediator complex subunit MED26. *Biomol. NMR Assign.* (2016), doi:10.1007/s12104-016-9673-z.

61. Z. Lens, F. X. Cantrelle, R. Peruzzini, X. Hanoulle, F. Dewitte, E. Ferreira, J. L. Baert, D. Monté, M. Aumercier, V. Villeret, A. Verger, I. Landrieu, Solution Structure of the N-Terminal Domain of Mediator Subunit MED26 and Molecular Characterization of Its Interaction with EAF1 and TAF7. *J. Mol. Biol.* (2017), doi:10.1016/j.jmb.2017.09.001.

62. C. Dominguez, R. Boelens, A. M. J. J. Bonvin, HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* (2003), doi:10.1021/ja026939x.

63. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* (2012), doi:10.1158/2159-8290.CD-12-0095.

64. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* (2013), doi:10.1126/scisignal.2004088.

65. J. Erde, R. R. O. Loo, J. A. Loo, Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* (2014), doi:10.1021/pr4010019.

66. H. Langerová, B. Lubyová, A. Zábranský, M. Hubálek, K. Glendová, L. Aillot, J. Hodek, D. Strunin, V. Janovec, I. Hirsch, J. Weber, Hepatitis B Core Protein Is Post-Translationally Modified through K29-Linked Ubiquitination. *Cells* (2020), doi:10.3390/cells9122547.

67. S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* (2016), doi:10.1038/nmeth.3901.

68. H. C. Hodges, B. Z. Stanton, K. Cermakova, C.-Y. Y. Chang, E. L. Miller, J. G. Kirkland, W. L. Ku, V. Veverka, K. Zhao, G. R. Crabtree, Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72 (2018).

69. S. Shen, J. W. Park, Z. X. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou, Y. Xing, rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* (2014), doi:10.1073/pnas.1419161111.

70. S. Shen, J. W. Park, J. Huang, K. A. Dittmar, Z. X. Lu, Q. Zhou, R. P. Carstens, Y. Xing, MATS: A Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* (2012), doi:10.1093/nar/gkr1291.

71. J. W. Park, C. Tokheim, S. Shen, Y. Xing, Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. *Methods Mol. Biol.* (2013), doi:10.1007/978-1-62703-514-9_10.

72. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* (2009), doi:10.1186/gb-2009-10-3-r25.

73. A. Pohl, M. Beato, bwtool: A tool for bigWig files. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu056.

74. K. A. Reimer, C. A. Mimoso, K. Adelman, K. M. Neugebauer, Co-transcriptional splicing

regulates 3′ end cleavage during mammalian erythropoiesis. *Mol. Cell* (2021), doi:10.1016/j.molcel.2020.12.018.

75. L. Lopez-Delisle, L. Rabbani, J. Wolff, V. Bhardwaj, R. Backofen, B. Grüning, F. Ramírez, T. Manke, pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* (2021), doi:10.1093/bioinformatics/btaa692.

76. K. Cermakova, H. Courtney Hodges, Next-generation drugs and probes for chromatin biology: From targeted protein degradation to phase separation. *Molecules* (2018), doi:10.3390/molecules23081958.