bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A paradigm for regulation at the effector interface with RNA-binding proteins

Kriti Shah[1,2,8], Shiyang He[1,2,8], David J. Turner[3,8], Joshua Corbo[3,4], Khadija Rebbani[3], Joseph M. Bateman[5], Sihem Cheloufi[1,2,6], Cátia Igreja[7], Eugene Valkov[3,*], Jernej Murn[1,2,9,*]

Author Affiliations

[1]Department of Biochemistry, University of California, Riverside, 3401 Watkins Drive, Boyce Hall, Riverside, CA 92521, U.S.A.
[2]Center for RNA Biology and Medicine, 900 University Ave, Riverside, CA 92521, U.S.A.
[3]RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, U.S.A.
[4]Department of Molecular and Cellular Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, U.S.A.
[5]Maurice Wohl Clinical Neuroscience Institute, King's College London, 5 Cutcombe Road, London, SE5 9RX, U.K.
[6]Stem Cell Center, University of California, Riverside, 900 University Ave, Riverside, CA 92521, U.S.A.
[7]Department for Integrative Evolutionary Biology, Max-Planck-Ring 9, D-72076 Tübingen, Germany.
[8]These authors contributed equally
[9]Lead contact
*Correspondence: eugene.valkov@nih.gov, jernej.murn@ucr.edu

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUMMARY

RNA-binding proteins (RBPs) are key regulators of gene expression, but how RBPs convey regulatory instructions to the core effectors of RNA processing is unclear. We address this problem by dissecting the effector interface responsible for regulation by a developmentally essential RBP. We show that the sequence-specific RBP Unkempt, in addition to its RNA-binding domain, contains an extensive intrinsically disordered region (IDR) and a dimerization domain, all of which distinctly contribute to Unkempt's function. Within IDR, we identify an array of linear motifs that support direct contact with two central effectors, the CCR4-NOT complex and PABPC. The spatial arrangement, dual-purpose motifs, and obligate homodimerization allow for high-avidity effector interactions that are mandatory for the accuracy of RNA targeting, translational repression, and control of cell morphology by Unkempt. Our study establishes the molecular assembly and functional principles of an RBP–effector interface, with general implications for the evolution and function of RBP-operated regulatory networks.

Keywords: RNA-binding protein, gene expression, RNA processing, short linear motif (SLiM), intrinsically disordered region (IDR), Unkempt, CCR4-NOT, PABPC, translation, mRNA stability

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

RNA processing is executed by a diverse set of effector proteins and protein complexes that specialize in facilitating a particular molecular event during the lifetime of an RNA. Although several effectors may act on RNA with little specificity, effector activities are often regulated to affect certain transcripts or their parts more than others[1,2]. Specificity in RNA processing is essential for a range of key cellular functions, including cellular differentiation, timely responses to immune signaling, or synaptic plasticity[3-5]. Critical to regulated RNA processing are RNA-binding proteins (RBPs) that can, by interacting with both the RNA and the effector, divert effector activity towards specific, RBP-targeted transcripts[1,2,6].

One major unsolved problem associated with regulated RNA processing is a fragmented understanding of RBP–effector interactions. In particular, it is unclear how interactions of different RBPs with a given effector can direct different modes of RNA processing. For instance, regulatory RBPs that bind to the same relative positions on RNA can, through recruitment of spliceosomal components, either promote or antagonize the maturation of a functional spliceosome, resulting in opposing effects on RNA splicing[7]. Likewise, RBP-dependent recruitment of the CCR4-NOT effector complex can favor either decay or translational repression of targeted transcripts to serve different cellular functions[8-12]. It is unknown whether RBP–effector interactions also contribute to the RNA recognition by the RBP, which could, on its own, influence both the specificity and the mode of RNA processing[13,14].

An emerging theme is that RBP–effector interactions are often mediated by a short linear motif (SLiM) embedded within RBP's intrinsically disordered region (IDR), which supports direct contact with a structured domain of an effector protein[2]. A few essential residues in these SLiMs provide specificity in transient interactions[15]. However, functional studies of SLiMs and the derivation of generalizable principles are complicated by the evolutionary plasticity of SLiMs at the sequence level.

Here, we study a developmentally essential RBP–effector interface via its control of RNA processing. We use as a paradigm the sequence-specific RBP Unkempt (UNK), taking advantage of its distinct molecular features and a clear cellular phenotype that we utilize as sensitive functional readouts[16,17]. This includes UNK's strict requirement for a specific RNA binding motif, its potent transcriptional and translational activities, as well as its unique capacity to induce a bipolar cellular morphology, an activity that is required during early neurogenesis and that can be recapitulated in non-neuronal cells (Figures 1A and 1B)[16-18]. We identified numerous RBP–effector contacts maintained via IDR-embedded SLiMs and arranged via RBP dimerization. Interactions of UNK with each of its key effectors, CCR4-NOT and poly(A)-binding protein (PABPC), substantially contribute to the recognition of UNK's RNA-binding motif, with PABPC additionally playing a dominant role in positioning UNK on mRNA and with CCR4-NOT mediating target-specific translational repression. Our findings define an RBP–effector interface and elucidate its central role in specifying the regulatory function of an RBP.

RESULTS

Unkempt's intrinsically disordered region is a hub of regulatory activity

UNK is a cytoplasmic, translationally active RBP with a critical role in the development of the nervous system[16,19-22]. Prior to its identification as an RBP, however, we discovered that UNK was transcriptionally active in a dual luciferase reporter assay (Figures 1C, 1D, S1A). We mapped UNK's transcriptional activity to an extended and conserved IDR (Figures 1E, 1F, and Table S1). Deleting the entire IDR or its portions silenced transcriptional activation in this assay

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 1F). Moreover, we found that transcriptionally more potent UNK mutants induced stronger morphological transformation of cells, suggesting that IDR is required for UNK function (Figures 1A, 1B, 1G, and 1H).

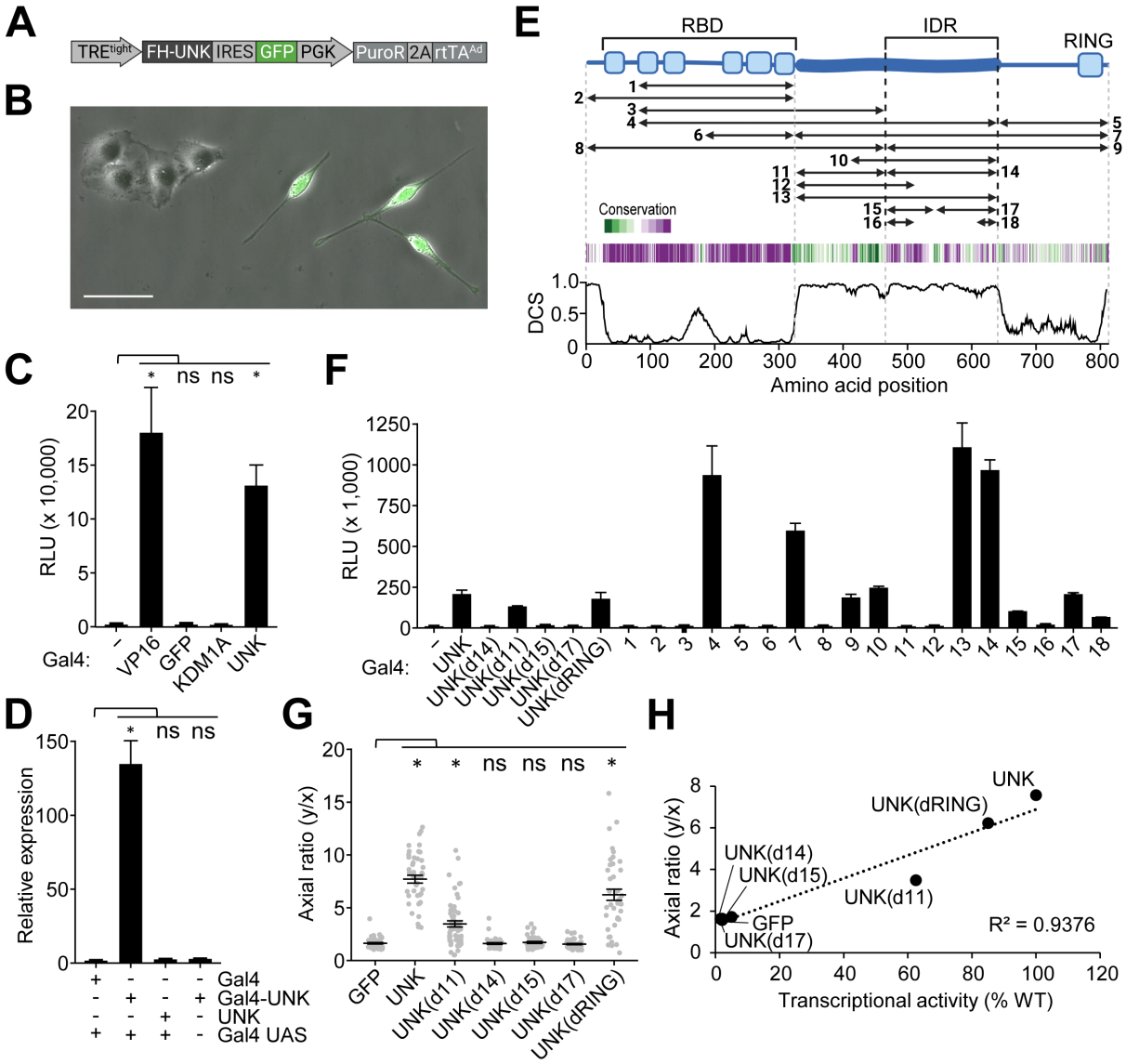


Figure 1. Morphogenetic and transcriptional activity of Unkempt maps to its IDR.

(A) Inducible system for UNK-driven cell morphogenesis. Shown is an all-in-one variant of the previously reported system used in this study (Methods)[16,17].

(B) HeLa cells are incubated with doxycycline (Dox) for 48 hours after which the morphology of GFP-expressing cells is evaluated (Methods). Scale bar, 50 µm.

(C) Transcriptional activity of UNK measured by a dual luciferase reporter assay at 24 hours after transfection of HeLa cells with constructs for expression of the indicated, Gal4 DNA-binding domain (Gal4)-tagged proteins. VP16 transcriptional activator served as a positive control. (–), Gal4 alone; RLU, relative luminescence units (n = 6). See also Figure S1A.

(D) Relative quantification by qPCR of firefly luciferase mRNA levels in cells, as in (C). Gal4 UAS, Gal4 upstream activating sequences (n = 3). See also Figure S1A.

(E) Domain map of UNK (blue) indicating its RNA-binding domain (RBD), intrinsically disordered region (IDR), and a RING finger domain. The studied segments of UNK are indicated (Table S1). Amino acid conservation

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(green, least conserved; purple, most conserved position) and disorder confidence profile of UNK are shown below the map (Methods). DCS, disorder confidence score.
(F) As in (C), transcriptional activities of Gal4-tagged truncation mutants, indicated in (E), and deletion mutants of UNK (n = 3). See also Table S1.
(G) As in (A) and (B), morphologies of cells co-expressing the indicated UNK mutant and GFP were quantified by calculating their axial ratios (y/x; Methods)[16] (n = 50 GFP-expressing cells per cell line).
(H) Correlation of cell morphologies shown in (G) with transcriptional activities of the corresponding Gal4-tagged UNK mutants shown in (F).
Data in (C), (D), (F), and (G) are presented as mean ± SD. Statistical significance was determined using Student's t-test with *$p < 0.0001$; ns, not significant.

In a further screen of IDR mutants, we defined two shortest active regions, minN and minC (Figures 2A-2B and Table S1). Strikingly, substituting L522 in minN or two residues in minC (W622 and F625) to alanines completely silenced the activities of either region (green arrows in Figure 2A). A combined mutant, UNK_{3M}, encompassing L522A/W622A/F625A, silenced not only IDR but also the full-length UNK protein (Figures 2C and 2D). Notably, as seen with the deletion mutants (Figure 1H), we observed positive correlation between the cell-polarizing activities of UNK residue mutants and their transcriptional activities, with UNK_{3M} as the minimal mutant that failed to elicit cellular polarization (Figures 2E and 2F).

Because UNK is nearly exclusively cytoplasmic, its transcriptional activity was unexpected (Figures S1B and S1C)[16,19,20]. To dissect a potential role of UNK in transcription, we performed ChIP-seq analyses of endogenous or ectopic UNK, RNA-guided recruitment of dCas9 fusions with UNK or its IDR to loci of endogenous genes to induce their transcription, and mass spectrometry analyses of affinity-purified nuclear protein complexes of UNK to identify any chromatin-associated interactors. None of these analyses suggested a transcriptional activity for UNK (data not shown), although we cannot rule out its biological relevance. We further pursued the function of IDR due to its strict requirement for the morphogenetic activity of UNK (Figures 1G, 1H, 2E, and 2F).

CCR4-NOT and PABPC are critical effectors of Unkempt

Proximity-dependent biotinylation (BioID) analysis identified > 100 UNK interactors in cells (Figure 3A)[23]. To assess their impact on UNK function, we compared compositions of complexes formed by the wild-type (WT) UNK (UNK_{WT}) or the inactive UNK_{3M} by mass spectrometry (Figure 3B). We observed a major difference in the association with the CCR4-NOT complex subunits; whereas all CCR4-NOT subunits were readily detected in the UNK_{WT} complex, they were absent in the UNK_{3M} complex (Figures 3B and S2A). This was confirmed by a co-IP/western analysis that further pointed to the contribution of each of the three mutated residues to the interaction between UNK and CCR4-NOT (Figure 3C). Thus, the 3M mutation that renders UNK morphogenetically inactive also specifically disrupts its association with the CCR4-NOT complex.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

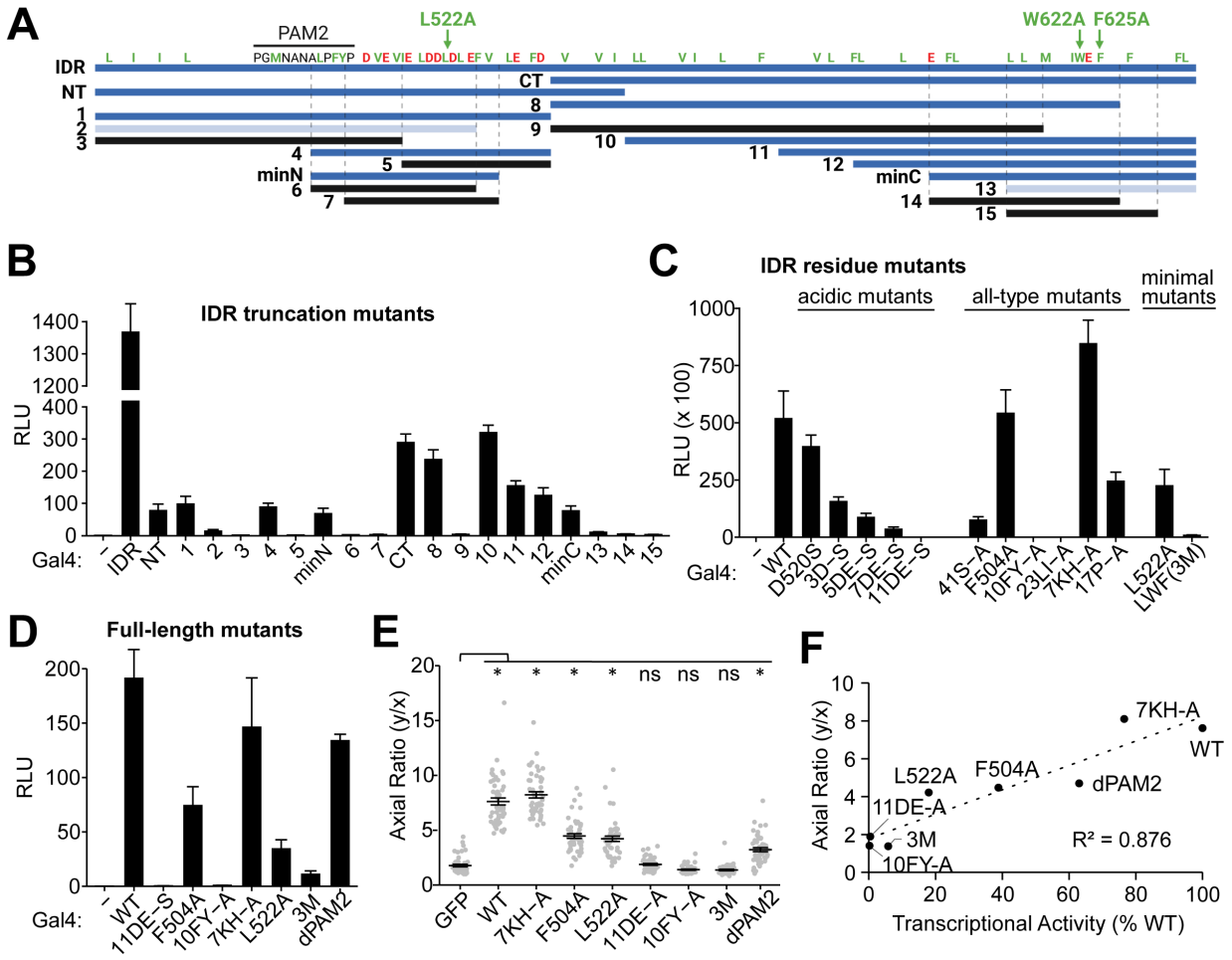


Figure 2. Identification of key activity-linked features in the IDR of Unkempt.

(A) Key truncation mutants of IDR (top blue line) that led to the identification of the smallest transcriptionally active regions, minN and minC. Highly active fragments are in blue, weakly active in light blue, and silent fragments are in black. Positions of hydrophobic residues (I, L, V, M, W, F) and Y are shown in green and acidic residues (D, E) are in red. Green arrows indicate mutations that silence minN (L522A) or minC (W622A and F625A combined). PAM2, the predicted PABPC-binding motif. See also Table S1 and Figure S2B.

(B) Dual luciferase reporter assay of Gal4 DNA-binding domain (Gal4)-fusions with IDR or its truncation mutants shown in (A) (n = 3). (–), Gal4 alone; RLU, relative luminescence units.

(C) Contribution of different residues to the transcriptional activity of Gal4-tagged IDR. The indicated mutants were analyzed as in (B). 'All-type mutants' have all residues of the indicated type mutated to alanines (n = 3). See also Table S1.

(D) Contribution of the indicated mutations to the transcriptional activity of the Gal4-tagged full-length UNK, analyzed as in (B) (n = 3). See also Table S1.

(E) Morphologies of cells inducibly co-expressing the indicated full-length UNK mutant and GFP at 48 h of incubation with Dox (n = 50 GFP-expressing cells per cell line).

(F) Correlation of cell morphologies shown in (E) with transcriptional activities shown in (D) for the indicated mutants.

Data in (B)-(E) are presented as mean ± SD. Statistical significance was determined using Student's t test with *p < 0.0001; ns, not significant.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

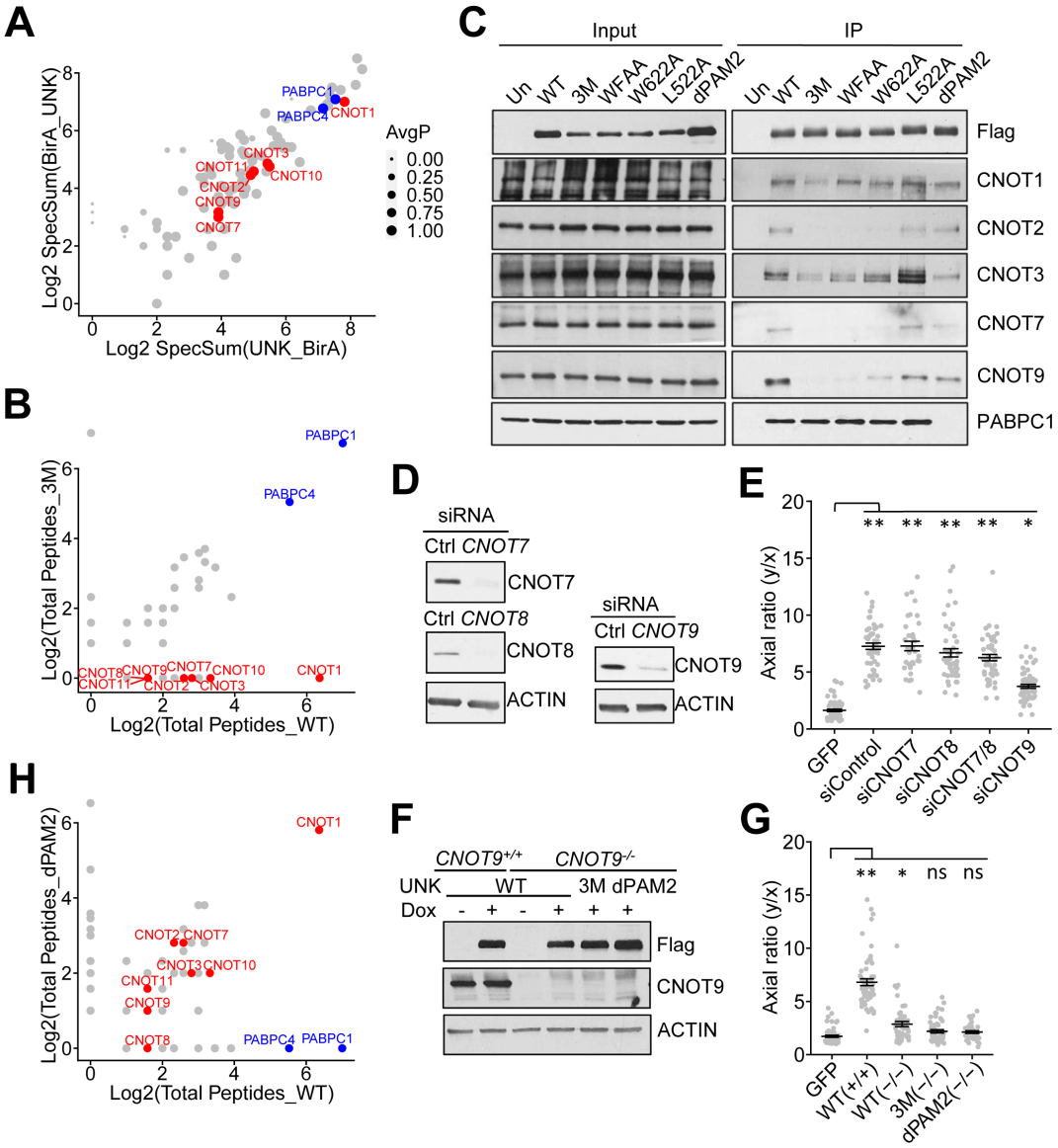


Figure 3. The CCR4-NOT complex is a critical effector of Unkempt.

(A) Result of a BioID analysis showing the comparison of average spectral counts of peptides derived from each protein identified as a high-confidence interactor of UNK with an abortive biotin ligase (BirA) fused to either N-terminus (BirA_UNK) or C-terminus (UNK_BirA) of UNK. The diameter of each interactor is proportional to the average probability of interaction (AvgP). Red and blue circles highlight subunits of the CCR4-NOT complex and PABPC1/4 proteins, respectively. Reanalysis of data from Youn et al., 2018[23].

(B) Results of mass spectrometry analyses comparing total peptide counts detected in protein complexes of UNK_{WT} and UNK_{3M} (n = 2).

(C) Co-IP of endogenous CCR4-NOT subunits and PABPC1 with UNK from lysates of HeLa cells inducibly expressing the indicated Flag-HA-tagged UNK mutants. Precipitated proteins were detected by western blot analysis (n = 3). WFAA, UNK with W622A and F625A mutations.

(D) Knockdown of the indicated CCR4-NOT subunits at 48 hours after transfection of their targeting siRNAs in cells inducibly co-expressing UNK_{WT} and GFP cells (n = 3).

(E) As in (D), morphologies of siRNA-transfected cells were quantified at 48 hours of incubation with Dox (n = 50 GFP-expressing cells per cell line).

(F) Western analysis of WT (*CNOT9*$^{+/+}$) or CNOT9 knockout cells (CNOT9$^{-/-}$) inducibly co-expressing Flag-HA-tagged UNK_{WT}, UNK_{3M}, or UNK_{dPAM2} and GFP at 48 hours of induction with Dox (n = 2).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(G) As in (F), morphologies of cells were quantified at 48 hours of incubation with Dox (n = 50 GFP-expressing cells per cell line).
(H) Results of mass spectrometry analyses comparing total peptide counts detected in protein complexes of UNK_{WT} and UNK_{dPAM2} (n = 2).
Data in (E) and (G) are presented as mean ± SD. Statistical significance was determined using Student's t-test with $**p \leq 2.37 \times 10^{-19}$ and $*p = 1.46 \times 10^{-17}$ in (E) and with $**p = 1.01 \times 10^{-23}$ and $*p = 6.99 \times 10^{-5}$ in (G). ns, not significant.

To validate these results genetically, we tested the morphogenetic capacity of UNK_{WT} in HeLa cells following siRNA-mediated knockdown (KD) of individual CCR4-NOT subunits. Interestingly, suppressing the deadenylase activity of CCR4-NOT via a simultaneous KD of CNOT7 and CNOT8 did not impact cell polarization, whereas KD of CNOT9 substantially impaired the capacity of UNK_{WT} to transform cellular morphology (Figures 3D and 3E)[24]. We confirmed the requirement for CNOT9 in *CNOT9*-null cells that were nearly fully resistant to the morphogenetic activity of UNK_{WT} (Figures 3F and 3G). Furthermore, the absence of CNOT9 substantially reduced the interaction of UNK with other CCR4-NOT subunits (Figure S2C), implicating CNOT9 as the principal binding site for UNK on CCR4-NOT.

The 3M mutation did not affect interactions with two of the strongest binding partners of UNK, the cytoplasmic poly(A)-binding proteins PABPC1 and PABPC4 (collectively termed PABPC; Figures 3A-3C and S2A)[2,25]. In the UNK protein sequence, we identified a putative SLiM, known as a PAM2 motif, found in diverse proteins that bind to the MLLE domain of PABPC (Figures 2A and S2B)[26,27]. Deletion of the UNK's PAM2-like motif, rendering the UNK_{dPAM2} mutant, released from the UNK complex PABPC but not CCR4-NOT, as assessed by mass spectrometry and confirmed by co-IP/western analysis (Figures 3C, 3H, and S2A). Notably, UNK_{dPAM2} exerted weaker transcriptional and morphogenetic activities than UNK_{WT} (Figures 2D-2F). Together, these findings identify CCR4-NOT and PABPC as key effectors of UNK and further suggest that the role of CCR4-NOT is independent of its deadenylase activity.

Unkempt interacts directly with multiple CCR4-NOT complex subunits

To investigate the directness of interactions between UNK and CCR4-NOT, we carried out *in vitro* pull-down assays with recombinant full-length UNK (UNK_{FULL}) that was immobilized on beads and incubated with CCR4-NOT subcomplexes (modules) reconstituted from purified recombinant proteins (Figure 4A)[28]. UNK_{FULL} bound specifically to the NOT9 and NOT modules but not the catalytic or NOT10/11 modules (Figure 4B). Both the NOT9 and NOT modules were also directly bound by the functionally essential segment, UNK_{IDR} (Figures 1E and 4C). However, UNK_{IDR} with the 3M mutation only inefficiently pulled down the NOT module and did not bind the NOT9 module (Figure 4C). A mutational analysis of the individual 3M residues further pointed to a major role for W622 and minor contributions of L522 and F625 in supporting direct contact with the CCR4-NOT modules (Figure 4C).

We then used AlphaFold employing the rigorous approach proposed by Conti and co-workers to generate structure predictions of UNK_{IDR} in complex with the NOT9 module (Figures 4D-4G)[29-31]. Interestingly, these predictions suggested that a segment of UNK_{IDR} folds into a helix (residues 507-537; henceforth SLiM 1) that binds across the CNOT9 concave surface (Figures 4D and 4E), which serves as a protein-protein interaction site for several other factors[32], whereas the sole tryptophan residue within UNK_{IDR}, W622, inserts into either of the defined tryptophan (W)-binding pockets on the convex surface of CNOT9 (Figures 4F and 4G). We denote W622 and its neighboring residues in contact with CCR4-NOT (residues 617-625) as SLiM 2.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

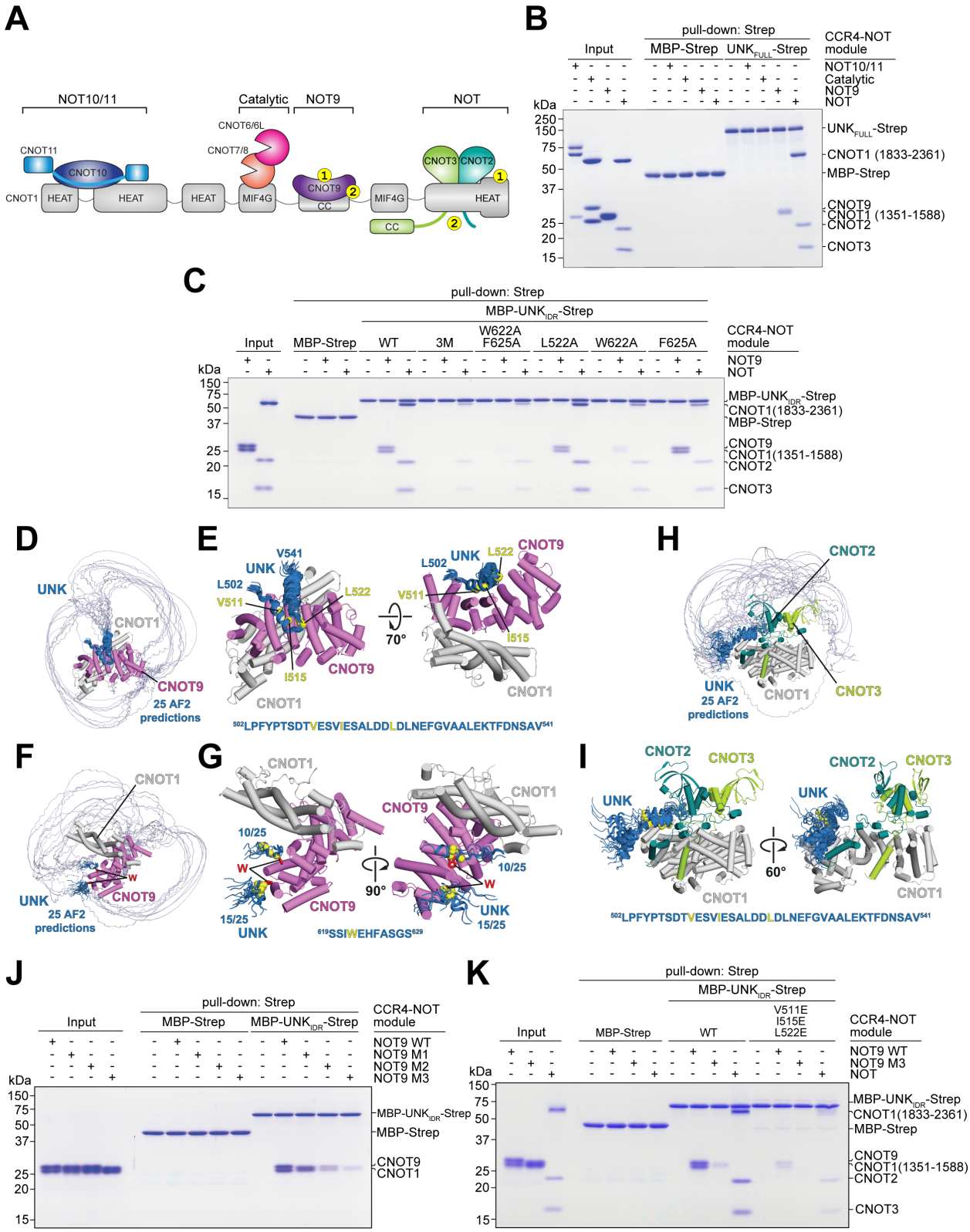


Figure 4. Definition of interactions between Unkempt and CCR4-NOT.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Schematic representation of the CCR4-NOT complex indicating its modules. Yellow circles indicate points of contact with UNK_{IDR} identified in this study. The numbering of contacts indicates either SLiM 1 or SLiM 2 binding sites, 1 or 2, respectively.
(B) Coomassie-stained polyacrylamide gel of *in vitro* pull-down assays with recombinant UNK_{FULL} tagged with the StrepII (Strep) affinity tag upon incubation with four different CCR4-NOT modules, as indicated (n = 3).
(C) As in (B) but with mutants of UNK_{IDR} constructs fused to MBP and Strep after incubation with the NOT9 or the NOT module (n = 3).
(D) Twenty-five AlphaFold predictions of interfaces of UNK_{IDR} interacting with the NOT9 module. The predictions are aligned on the CNOT9/CNOT1 heterodimer[33]. The region of UNK_{IDR} where the predictions converged is in dark blue. See also Methods.
(E) The converged region of UNK_{IDR} bound on the concave surface of CNOT9/CNOT1. The C-alpha atoms of the key interacting residues are shown as yellow spheres. The sequence of the converged region is shown with the key interacting residues highlighted in yellow.
(F) The same 25 predictions as in (D) but oriented to show the tryptophan (W)-binding pockets of CNOT9. The crystal structure of CNOT9 with two W residues (red sticks) bound[33] was superposed on the AlphaFold prediction.
(G) Two clusters comprising all 25 predictions of the converged region of UNK_{IDR} close to the W-binding pockets of CNOT9. The C-alpha atoms of the key interacting W residues are shown as yellow spheres. The sequence of the converged region is shown with W622 highlighted in yellow.
(H) Twenty-five AlphaFold predictions of interfaces of UNK_{IDR} interacting with the NOT module. The predictions are aligned on the CNOT1/CNOT2/CNOT3 heterotrimer[34]. The region of UNK_{IDR} where predictions converged is in dark blue. See also Methods.
(I) The converged region of UNK_{IDR} bound on the surface of CNOT1. The C-alpha atoms of the key interacting residues are shown as yellow spheres. The sequence of the converged region is shown with the key interacting residues highlighted in yellow.
(J) As in (B), but showing pull-down of WT or M1-M3 mutants of the NOT9 module by MBP-UNK_{IDR}-Strep. Residues in CNOT9 mutated to alanines in M1 (Y203 and R244) line the W-pocket 1, and those mutated in M2 (R205 and H208) line the W-pocket 2[33]. All four residues (Y203, R205, H208, and R244) were mutated in M3 (n = 2).
(K) As in (B), but showing pull-down of the WT or M3 NOT9 module or the NOT module by WT MBP-UNK_{IDR}-Strep or its mutant with three key residues in SLiM 1 (V511, I515, L522) substituted with glutamic acid.

Intriguingly, AlphaFold predicted that SLiM 1 may also mediate an interaction between UNK_{IDR} and the NOT module (Figures 4H and 4I). Specifically, SLiM 1 was predicted again to fold into a helix and bind to a conserved hydrophobic pocket on the surface of the CNOT1 C-terminal domain (Figures 4H and 4I). Although these predictions did not point to a clear SLiM 2 binding site on the NOT module, the pull-down assays showing that mutating either W622 or F625 reduces the binding of UNK_{IDR} with the NOT module suggested such interaction (Figure 4C).

We first tested whether UNK_{IDR} interacts with NOT9 W-pocket mutants to validate the predicted interfaces. The double W-pocket mutant (NOT9 M3) showed the most impaired interaction with UNK_{IDR} (Figure 4J). Next, we substituted three hydrophobic residues (V511, I515, and L522) within SLiM 1 to glutamates; this mutant UNK_{IDR} less efficiently recruited either the NOT9 or NOT module and was completely unable to interact with the NOT9 double W-pocket mutant (Figure 4K). These results support a multivalent mode of UNK_{IDR} interaction with CNOT9 and support the observation that UNK uses the same motifs to bind NOT9 and NOT modules of the CCR4-NOT complex.

Unkempt binds its effectors as a dimer stabilized by a conserved coiled coil

UNK contains a region with a distinct heptad repeat pattern of a coiled-coil motif (residues 643-767; Figures S3A). To investigate the possible structural role of this motif, we used AlphaFold to generate structure predictions for UNK_{FULL} and a C-terminal fragment, residues 637-810, termed UNK_{C}. Both predictions revealed two parallel coiled coils stabilizing a putative dimer (Figures

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5A and S3B). To see whether UNK indeed dimerized in solution, we measured the molecular weight of purified UNK_{FULL} and UNK_C by mass photometry and confirmed both as exclusive dimers (Figure 5B).

To validate the dimerization interface, we substituted hydrophobic residues in *d* positions of the coiled-coil for glutamates, generating UNK_{E8} and UNK_{E6}, with the latter having substitutions only in the more extended coiled-coil motif (Figure 5A). These substitutions placed negatively charged residues opposite each other in the coiled coil, leading to electrostatic repulsion and destabilization of the interface. UNK_{E6} was a mixed species of monomers and dimers, while UNK_{E8} was an exclusive monomer, suggesting that both coiled-coil motifs are essential for dimer stability (Figure 5C).

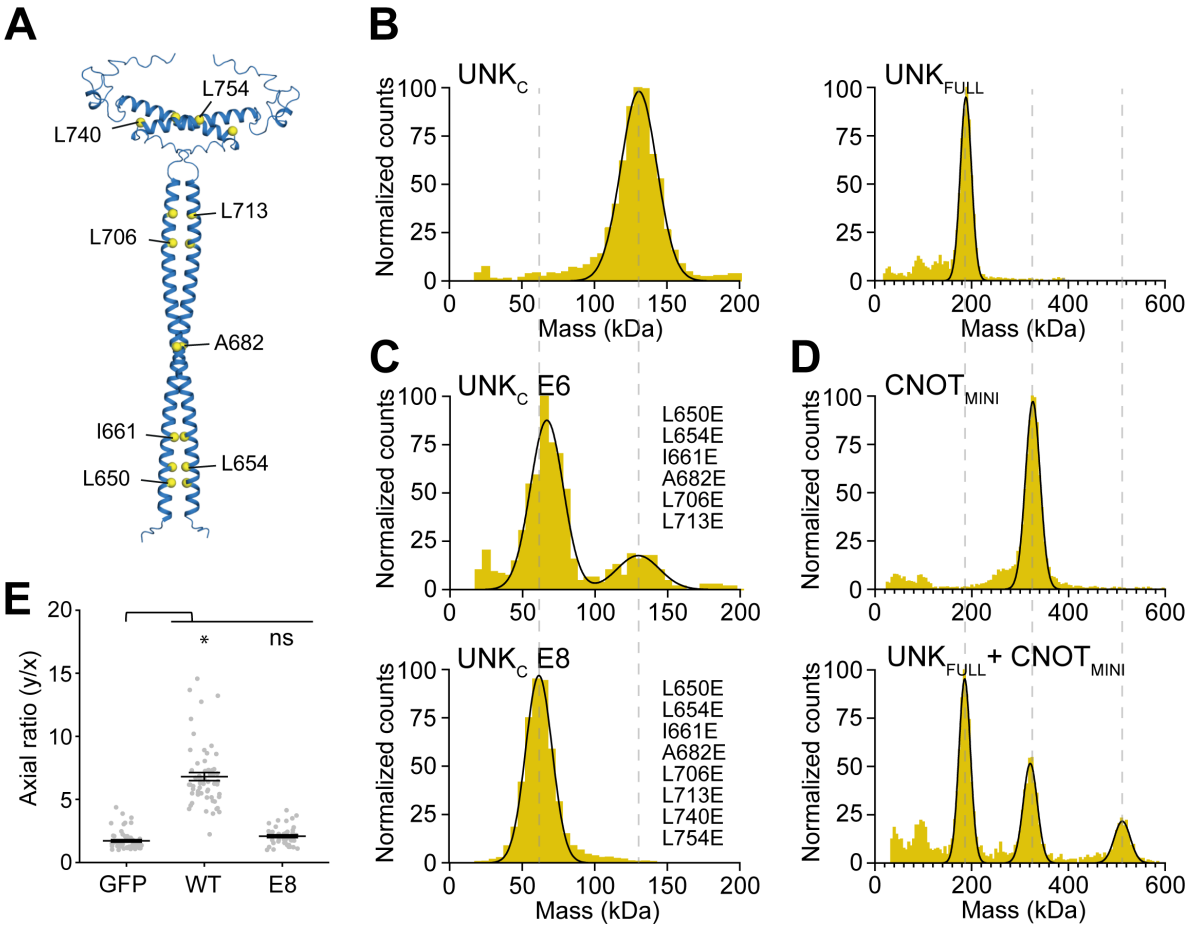


Figure 5. A critical role for dimerization of Unkempt.

(A) A representative AlphaFold prediction of the coiled coil homodimer formed by two copies of the C-terminal segment of UNK (UNK_C; residues 637-810). All 25 predictions were identical, bar some minor variations in the C-terminal loop regions. The C-alpha atoms of the key hydrophobic residues stabilizing the coiled coil are shown as yellow spheres. Predictions of the homodimer full-length UNK protein are shown in Figure S3B. See also Figures S3A and S3E.

(B)-(D) Mass photometry analyses of Strep-tagged UNK_C or UNK_{FULL} (B), UNK_C E6 or UNK_C E8 (C), $CNOT_{MINI}$ or in complex with UNK_{FULL} (D) (n = 2). Mutations in UNK_C E6 and UNK_C E8 are listed in (C). The calculated molecular weights are 62.5 kDa for UNK_C monomer, 125 kDa for UNK_C dimer, 90.5 kDa for UNK_{FULL} monomer, and 181 kDa for UNK_{FULL} dimer. The observed mean molecular weights are 126 kDa for UNK_C, 62 kDa and 125

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

kDa for UNK_C E6, 58 kDa for UNK_C E8, 189 kDa for UNK_{FULL}, 326 kDa for $CNOT_{MINI}$, and 186 kDa, 321 kDa, and 512 kDa for the mixture of UNK_{FULL} and $CNOT_{MINI}$. See also Figures S3C and S3D.
(E) Morphologies of cells inducibly co-expressing UNK_{WT} or UNK_{E8} and GFP compared to GFP-only expressing cells at 48 hours of incubation with Dox (n = 50 GFP-expressing cells per cell line). Data are presented as mean ± SD. Statistical significance was determined using Student's t-test with *$p = 6.27 \times 10^{-15}$; ns, not significant.

UNK_{FULL} efficiently pulled down abridged recombinant CCR4-NOT subcomplexes containing both the NOT9 and NOT modules, either the four-subunit CNOT1/2/3/9 or six-subunit $CNOT_{MINI}$ complex, consistent with direct, stable binding (Figures S3C, S3D, and data not shown)[28]. To determine the stoichiometry of binding, we measured the mass of a reconstituted complex of UNK_{FULL} with $CNOT_{MINI}$, revealing that two copies bind one $CNOT_{MINI}$ (Figure 5D). Although UNK_{FULL} also pulled down recombinant PABPC1, we could not determine the stoichiometry, suggesting that UNK does not bind PABPC1 as stably as CCR4-NOT (Figure S3D). However, including CNOT1/2/3/9 in the binding reaction had no apparent effect on the pull-down of PABPC1 (Figure S3D), suggesting that PABPC and CCR4-NOT may interact with UNK independently.

Given the capacity of both SLiM 1 and SLiM 2 to bind different CCR4-NOT subunits, dimerization may enhance the stability of the UNK–CCR4-NOT interface through avidity effects. We asked whether dimerization might be important for UNK's cellular function. Strikingly, the monomeric UNK_{E8} failed to alter cell morphology analogous to the UNK_{3M} phenotype (Figures 2E and 5E). Thus, the IDR-embedded SLiMs and homodimerization are essential for UNK's morphogenetic activity. Notably, the predicted conservation of UNK's propensity to dimerize (Figure S3E) and form SLiM-mediated contacts with CCR4-NOT (Figures S3F), suggest evolutionary constraints that may support the observed conservation of the morphogenetic activity of UNK[17].

Effector interactions regulate RNA sequence recognition by Unkempt

UNK's consensus RNA recognition sequence is specified by its two CCCH-type zinc finger clusters and consists of a UAG motif upstream of a U/A-rich trimer[16,17]. However, as often observed for sequence-specific RBPs, less than a quarter of the predicted mRNA binding sites are occupied by UNK in cells and the majority of the observed binding sites do not contain the consensus recognition sequence[16,35]. This led us to ask whether UNK–effector interface may function as an auxiliary determinant of RNA binding by UNK.

To test this, we performed crosslinking and immunoprecipitation using an improved protocol (iCLIP2) for UNK_{WT}, UNK_{dPAM2}, and UNK_{3M}[36]. Analysis of the UNK_{WT} dataset revealed several thousand mRNA targets, a substantial increase over the initially annotated pool of UNK-bound messages (Figures S4A-S4D and Table S2)[16]. In line with the earlier study, we found that UNK binding sites distributed broadly over the coding regions and 3'UTRs of mRNAs and were enriched in UAG and U/A-rich motifs just up- and downstream of the binding peak, respectively (Figures 6A, S4E, and S4F). Curiously, a similar analysis of UNK_{dPAM2} and UNK_{3M} revealed an altered RNA-binding pattern with a weaker enrichment of the critical UAG motif and with changes in the position-specific representation of several U/A-rich motifs (Figures 6A, 6B, and Table S2).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

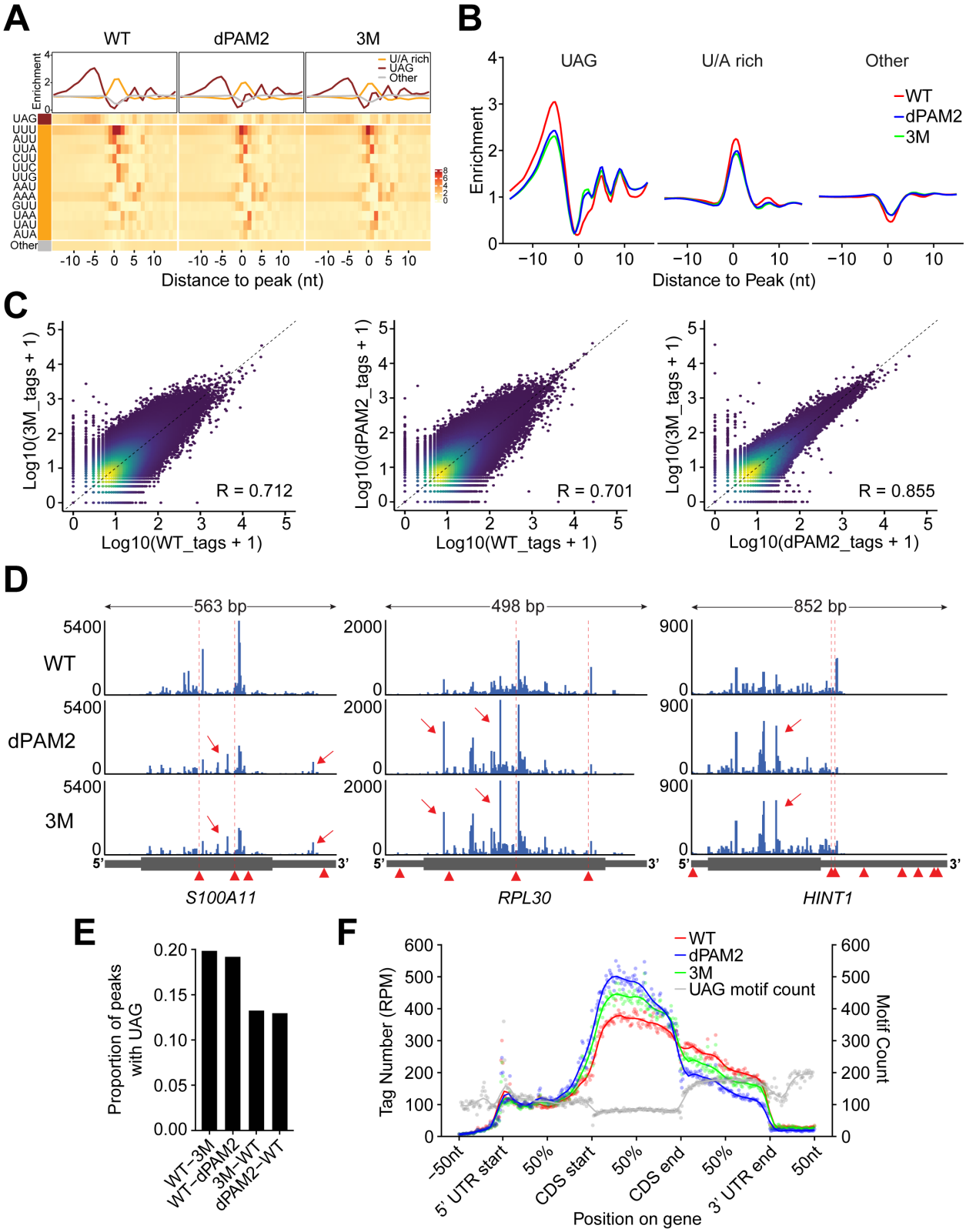


Figure 6. Impact of effector interactions on RNA binding by Unkempt.
(A), (B) Altered RNA sequence recognition by UNK mutants.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Heatmaps illustrate positional frequencies of the 64 possible trimers within UNK-binding sites between 15 nts upstream and downstream of the binding-site maxima for UNK_{WT}, UNK_{dPAM2}, and UNK_{3M}. Plots above the heatmaps profile the mean enrichment of different sets of trimers, considering the upstream UAG trimer (brown), the downstream U/A-rich trimers (orange), and all other trimers (gray; Methods).
(B) Overlays of the mean enrichment profiles for each set of trimers shown in (A).
(C) Density scatterplots comparing crosslink events per peak pair-wise among the combined iCLIP replicates (n = 4) for UNK_{WT}, UNK_{dPAM2}, and UNK_{3M}.
(D) Normalized UNK_{WT}, UNK_{dPAM2}, and UNK_{3M} iCLIP coverage tracks for three UNK target transcripts showing mutant-specific reduced binding strength at UAG-containing binding sites and the emergence of UAG-less satellite peaks. Arrowheads indicate positions of all UAG motifs; positions of UAG motifs immediately upstream of UNK_{WT} binding sites are highlighted by dashed vertical lines. Arrows point to UAG-less satellite peaks.
(E) Transcriptome-wide occurrence of the UAG trimer within 15 nts upstream of WT-specific (WT-3M and WT-dPAM2) or mutant-specific peaks (3M–WT and dPAM2-WT).
(F) Proportional metatranscript analysis of iCLIP data showing the positional frequency of crosslink events for UNK_{WT}, UNK_{dPAM2}, and UNK_{3M} on different segments of mRNA. Data points represent normalized crosslink events summarized over every percent of a given mRNA segment. Gray data points show total counts of the UAG trimer. Note the 5'-shift of RNA binding into coding sequences (CDS) by UNK_{3M} and especially by UNK_{dPAM2} despite the relative depletion of UAG in this mRNA segment.

Although both mutants retained the broad mRNA-targeting potential of UNK_{WT} (Figure S4G and Table S2), a consideration of individual RNA-binding events pointed to clear differences between either mutant and UNK_{WT} (Figure 6C). Interestingly, these differences were less apparent in a mutant-to-mutant comparison, much like the relative similarity in the consensus sequence recognition by UNK_{dPAM2} and UNK_{3M} (Figures 6A-6C). Inspection of individual target transcripts revealed weakened targeting of UAG-containing sites by the mutants compared to UNK_{WT} with concomitant emergence of UAG-less 'satellite' peaks (Figures 6D and 6E). Taken together, effector interactions distinctly contribute to the accuracy of RNA sequence recognition by UNK in cells.

PABPC controls the distribution of Unkempt on mRNA

Both CCR4-NOT and PABPC are thought to locate largely at the 3' ends of mRNAs, although the precise positions of the mammalian CCR4-NOT have not been determined[25,37-39]. We asked how effector localization may affect the distribution of UNK on mRNAs. Strikingly, the deletion of PAM2 SLiM caused a profound reduction in UNK binding to 3'UTRs and increased targeting of the coding sequences (Fig. 6F). In contrast, the 3M mutation led to a more moderate but still substantial upstream repositioning of UNK (Fig. 6F). This suggests that UNK position on mRNAs is controlled through association with effectors, with PABPC exerting a stronger influence compared to CCR4-NOT.

PABPC has a low nanomolar affinity for poly(A) RNA and is found largely at or very near the poly(A) tails[38,40]. However, factors such as PAIP2 or TNRC6 were reported to displace PABPC from mRNAs[41,42]. To determine whether UNK may function similarly, we first inspected the bulk interactions of PABPC1 with poly(A) tails (Figure S4H). UNK expression showed no effect on the pattern of ~27-nt footprints of PABPC on poly(A) tails in partially digested RNA co-precipitated with PABPC1 following *in vivo* UV crosslinking (Figure S4H), indicating a generally intact binding of PABPC1 to poly(A) tails.

To obtain a sequence-specific view of PABPC1 binding, we carried out iCLIP of endogenous PABPC1 and focused on its unique binding sites known to cluster around polyadenylation signals[38]. As with the gross analysis of poly(A) tails (Figure S4H), we observed no overt changes in the binding pattern of PABPC1 upon expression of UNK (Figures S4I-S4K

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and Table S3). However, UNK_{3M} or UNK_{dPAM2} showed weaker enrichment in the vicinity of PABPC1 binding sites than UNK_{WT}, consistent with the reduced presence of the UNK mutants on 3'UTRs (Figures 6F and S4K). These results support a model where PABPC strongly influences the distribution of UNK on mRNAs but not vice versa.

The Unkempt–effector interface indirectly regulates steady-state mRNA levels

UNK is a translational repressor that has little effect on transcript stability[16]. As CCR4-NOT and PABPC are principal factors affecting mRNA translation and stability, we asked whether the interactions of these effectors with UNK mediate its regulatory input.

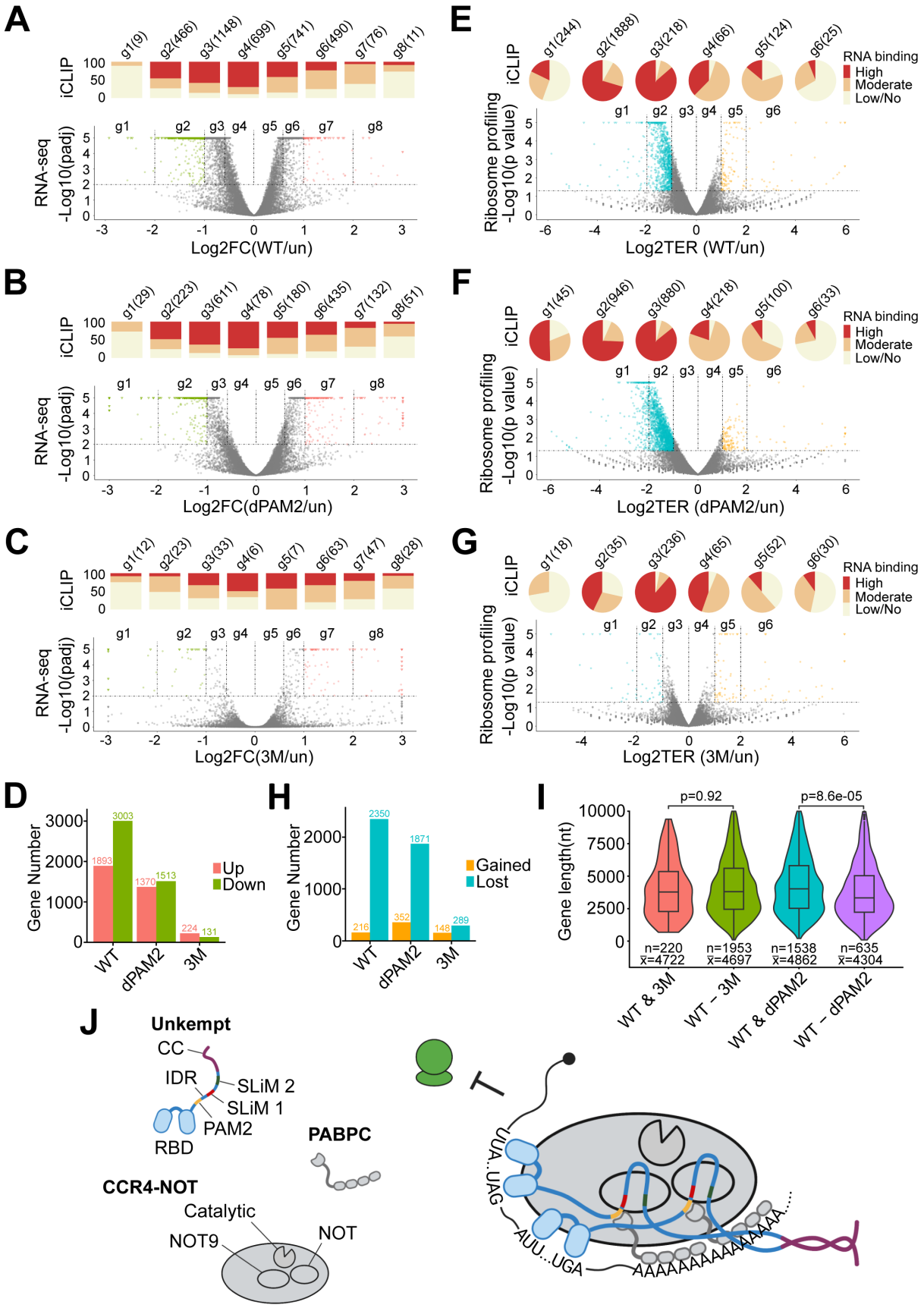
We first determined the imprint of UNK_{WT}, UNK_{dPAM2}, and UNK_{3M} on the cellular transcriptome 24 hours post-induction of expression (Table S4). UNK_{WT} perturbed steady-state mRNA levels of many transcripts with some bias towards downregulation (Figure 7A). A correlative analysis of iCLIP data indicated relatively weak binding of the highly regulated transcripts by UNK_{WT} and stronger targeting of transcripts that showed little regulation, again with a moderate preference for downregulated messages (Figure 7A). Similar trends were also noted upon induction of UNK mutants, however, with UNK_{dPAM2} affecting only about 60% and UNK_{3M} less than 10% of the number of transcripts regulated by UNK_{WT} (Figures 7B-7D). Thus, the largely indirect effect of UNK on steady-state mRNA levels relies heavily on its interactions with the CCR4-NOT complex and less on PABPC.

We then asked if UNK influences the shortening of mRNA poly(A) tails, a process known as deadenylation and in which PABPC and CCR4-NOT both play principal roles[24,25]. Using direct RNA sequencing, we derived mRNA poly(A) tail length estimates in different conditions of UNK expression in cells. Irrespective of UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} expression, we observed a length distribution consistent with relatively short tails of highly expressed mRNAs, a conserved feature of eukaryotic cells (Figures S5A, S5B, and Table S5)[43-45]. We found no correlation between the strength of mRNA targeting by UNK and mRNA poly(A) tail length, regardless of whether UNK_{WT} or its mutants were expressed (Figure S5C). We conclude that UNK does not substantially impact the metabolism of mRNA poly(A) tails in cells and that the effects on steady-state mRNA levels are mediated indirectly by its effector interface (Figures 7A-7D, S5D).

SLiM-mediated contacts with CCR4-NOT are critical conduits of translational control

To test whether the interactions with CCR4-NOT and PABPC mediate the translational regulation by UNK, we conducted ribosome profiling experiments to evaluate the impact of the effector interface on translational efficiencies of mRNAs while also considering the strength of mRNA targeting by UNK. The expression of UNK_{WT} resulted in a striking reduction in ribosome occupancy for the large majority of all significantly regulated mRNAs (91.6% or 2,350 mRNAs; Figure 7E and Table S6). Notably, most of these transcripts were highly bound by UNK, whereas the few with gains in ribosome occupancy were not (Figures S5E and 7E), pointing to a strong and direct repressive effect of UNK on translation. A separate analysis that only considered transcripts with no changes in expression indicated comparable, if not greater, bias towards translational silencing (Figure S6A).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7. Post-transcriptional regulation at the Unkempt–effector interface.
(A)-(D) Impact of UNK-effector interactions on steady-state mRNA levels.
(A)-(C) Volcano plots showing differential mRNA abundances between uninduced cells (un) and cells expressing UNK_{WT} (A), UNK_{dPAM2} (B), or UNK_{3M} (C) (RNA-seq data; n = 3). Significantly regulated transcripts (p adj. < 0.01) are binned into eight groups (g1-g8) according to the strength and sense of their regulation, with transcripts showing twofold or larger changes in abundance highlighted in color. Bar charts above the volcano plots indicate proportions of mRNAs bound highly, moderately, or lowly/not bound in each transcript group (iCLIP data). The total numbers of transcripts in each group are indicated. FC, fold change.
(D) Total numbers of up- or downregulated transcripts for each comparison shown in (A)-(C) (p adj. < 0.01).
(E)-(H) Direct translational repression mediated by UNK-effector interactions.
(E)-(G) Akin to (A)-(C), the volcano plots summarize ribosome profiling analyses showing differential ribosome occupancies of transcripts between uninduced cells and cells expressing UNK_{WT} (E), UNK_{dPAM2} (F), or UNK_{3M} (G) (n = 2). Significantly regulated transcripts (p < 0.05) are binned into six groups (g1-g6) according to the strength and sense of their regulation, with transcripts showing twofold or larger changes in ribosome occupancy highlighted in color. Pie charts above the volcano plots indicate RNA-binding information (iCLIP data) and transcript numbers, as in (A)-(C). TER, translational efficiency ratio. See also Figure S6A.
(H) Total numbers of transcripts with gained or lost ribosome occupancy for each comparison shown in (E)-(G) (p < 0.05).
(I) Loss of contact with PABPC leads to preferential translational derepression of shorter transcripts. Violin plots show the distribution of mRNA lengths in groups of UNK-targeted transcripts that are translationally significantly repressed (p < 0.05) by UNK_{WT} and UNK_{3M} (WT & 3M) or by UNK_{WT} and UNK_{dPAM2} (WT & dPAM2), as well as those that are uniquely repressed by UNK_{WT} but not UNK_{3M} (WT-3M) or by UNK_{WT} but not UNK_{dPAM2} (WT-dPAM2). The numbers of transcripts in each group (n) and their average length (x) are indicated. Statistical significance was calculated using Student's t-test.
(J) Model for regulation at the UNK-effector interface. Multivalent, high-avidity interactions between UNK and its effectors, CCR4-NOT and PABPC, are enabled by IDR-embedded SLiMs and the C-terminal coiled coil (CC) domain of UNK. See also Figure S7.

Similar analyses for UNK_{dPAM2} and UNK_{3M} revealed that CCR4-NOT and, to a lesser extent, PABPC are critical mediators of UNK-driven translational control (Figures 7E-7H). Specifically, removing the interaction with PABPC reduced the number of significantly repressed transcripts by about 20% (Figures 7F and 7H), whereas disrupting binding to CCR4-NOT essentially eliminated UNK-mediated translational repression (Figures 7G and 7H). Markedly, we could pinpoint the critical interface on the CCR4-NOT effector as the removal of the CNOT9 subunit, which weakened the interaction with UNK (Figure S2C), rendered UNK_{WT} incompetent for target repression (Figures S6B-S6G). Thus, the UNK–CCR4-NOT nexus is a critical conduit of translational repression for a large fraction of the cellular mRNA pool.

We also considered that the ability of the UNK–PABPC nexus to repress translation may be limited by the distance from the PABPC binding sites on poly(A) tails and 3'UTRs. To test this, we considered the lengths of UNK targets whose translational silencing depends on UNK maintaining contacts with either PABPC or CCR4-NOT (Figures 7E-7G). Transcripts silenced by UNK_{WT} but not by UNK_{dPAM2} were significantly shorter than those with sustained repression ($p = 8.6e^{-05}$), whereas the large population of mRNAs derepressed due to weakened contacts between UNK_{3M} and CCR4-NOT showed no such bias (Figure 7I). These findings indicate differential functional requirements for UNK–effector interactions in the context of a translationally repressive RNP.

It is of interest to note that the direct effect of the studied perturbations on translation is closely matched by changes in the mRNA levels as well as the morphogenetic potential of UNK (Figures 2E, 7A-7D, S6B-S6E). The principal implication is that much of UNK's cellular activity is coupled to its SLiM-mediated regulation of protein translation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In conclusion, the identified critical disordered segments of UNK (Figures 1 and 2) and their conserved interactions with its identified effectors (Figures 3, 4, and 5) show a clear relationship with the RNA-binding capacity and regulation (Figures 6 and 7), with implications for fundamental understanding of structural and functional principles of RBP–effector interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION

How do RBPs interact with their effectors to instruct different types of RNA processing? To address this question, we introduce an integrative approach combining *in vitro* and *in vivo* biochemistry, structure prediction, and multiple levels of systems analyses to study the RBP–effector interface, taking as a paradigm an essential RBP with distinct general molecular and cellular activities, Unkempt. This strategy allowed us to identify critical effectors of UNK, define the interactions constituting the interface, and evaluate the functional contribution of these interactions to UNK's activity.

UNK forms multivalent interactions with its effectors, the CCR4-NOT complex and PABPC. This includes an extensive IDR that contains three effector-binding SLiMs, two of which, SLiM 1 and SLiM 2, specifically interact with CCR4-NOT, and one, a PAM2-like SLiM, that interacts with PABPC, as well as a C-terminal coiled coil domain that induces UNK to homodimerize. Each of the CCR4-NOT–binding SLiMs can contact the NOT9 and the NOT module of the CCR4-NOT complex. This suggests an assembly of a functionally competent RNP that is held together by multivalent, SLiM-mediated RBP-effector interactions (Figures 7J and S7A-S7C). In the modeled RNP, two RNA-binding domains are brought together and can bind to the same or different molecules of mRNA, and this multivalency may facilitate the assembly of a larger RNP (Figure S7C).

We identify CCR4-NOT as the principal effector of UNK, with PABPC in a supportive regulatory role. Disrupting the UNK–CCR4-NOT interface essentially eliminates the signature activity of UNK at multiple scales in cells, including its effects on protein translation, perturbation of the transcriptome, and cell morphogenesis. Converting UNK to a monomeric state results in a complete loss of its morphogenetic activity. Although interrupting the UNK–PABPC contacts results in a weaker reduction of UNK's activity, the effect is again consistent across scales and thus in line with the central regulatory role of the UNK–effector interface.

For many RBPs, low proportions of observed versus expected RNA-binding events suggest that additional determinants of RNA binding must exist[35,46,47]. Effector interactions have traditionally been viewed as serving a recruiting role and have not been thought to feed back onto RNA binding by the recruiting RBP[2]. Unexpectedly, we find that interactions with CCR4-NOT or PABPC exert a substantial, two-prong auxiliary effect on RNA binding by UNK. First, both effectors assist UNK with the specificity of RNA sequence recognition; this is seen globally with compromised effector contacts leading to reduced binding of the critical UAG and the adjacent U/A-rich motifs, as well as locally by the emergence of numerous UAG-less satellite peaks (Figures 6A-6E). The remarkably similar defects in RNA binding by UNK_{dPAM2} or UNK_{3M} suggest that PABPC and CCR4-NOT may both stabilize UNK on mRNA (Figure S7B).

Independent of RNA sequence recognition, PABPC and, to a lesser extent, CCR4-NOT appear to control the distribution of UNK on mRNAs by facilitating its binding to 3'UTRs on targets. This is supported by a transcriptome-wide repositioning of UNK mutants to coding sequences (Figure 6F), which may be assisted by UNK's ability to bind endogenously paused ribosomes[16]; this could also explain the relative paucity of UNK in 5'UTRs of its targets. Along with the largely unaltered RNA binding by PABPC, the observed repositioning suggests that the 3'-anchored PABPC secures UNK to 3'UTRs rather than it being recruited by UNK, similar to how PABPC promotes the association of miRISC with mRNAs or the positioning of Makorin 1 RBP upstream of premature poly(A) tails[14,48]. In contrast to PABPC, the more subtle 5'–shift of UNK_{3M} and suppression of mRNA targets of all lengths (versus the regulatory bias of PABPC towards shorter mRNAs; Figure 7I) point to recruitment of CCR4-NOT to UNK binding sites on

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mRNAs. We also note that UNK_{dPAM2} or UNK_{3M} target similar but not identical sets of transcripts compared to UNK_{WT} (Figure S4G), implying that effector contribution to RNA binding may itself play a regulatory role.

The subunits of CCR4-NOT with which UNK interacts via SLiMs are also known to be targeted by divergent SLiMs of other RBPs and non-RBPs, indicating an independent but convergent evolution of IDR-embedded SLiMs of UNK[2,9,33,49-53]. However, unlike the multivalent interface between the UNK homodimer and CCR4-NOT described here, only one or at most two contacts have been experimentally validated between any other protein and CCR4-NOT, although more numerous contacts are suspected to exist[9,39]. Given the generally weak affinity of SLiM-mediated interactions, the large surface of the CCR4-NOT complex, and the commonly reported recruitment of CCR4-NOT by RBPs, it is plausible that additional, yet unidentified contacts with CCR4-NOT facilitate regulation by RBPs.

The existence of multi-purpose SLiMs suggests an economical evolutionary adaptation serving to eliminate a need to maintain a separate SLiM for each effector contact. This would permit combinatorial flexibility of RBP–effector interactions or facilitate their synchronous regulation, e.g., via post-translational modifications (Figure S7A). Remarkably, structural predictions based on evolutionary data suggest that the key protein features participating in the formation of UNK–CCR4-NOT contacts already existed in the earliest known UNK ortholog that emerged more than 500 million years ago and was specific for the UAG motif with some level of cell-morphogenetic activity (Figures S3E and S3F)[17,54].

Past studies of RBP- or miRNA-mediated gene silencing commonly relied on tethering assays and reporter transcripts to investigate the translational repression by the CCR4-NOT complex that is decoupled from its impact on mRNA stability[9,55-58]. We comprehensively show on a transcriptome-wide scale that mRNA regulation via CCR4-NOT in cells is not necessarily accompanied by deadenylation and mRNA decay. In the case of UNK, we speculate that deadenylation may be inhibited by the extensive interactions of IDR with the NOT9 and NOT modules, which are known to directly stimulate deadenylation by CCR4-NOT[28,34].

Limitations of the study

Several aspects of our study merit further consideration. For instance, we find that effector interactions are required for the accuracy of RNA sequence recognition by UNK, but cannot explain why this is so, why both PABPC and CCR4-NOT show a similar requirement, and why some but not other UNK binding sites require effector contacts. Rationalizing these observations will provide new fundamental knowledge about determinants of RNA-binding site selection and functional organization of RNPs. Towards this goal, it will be interesting to determine the *in vivo* positions of CCR4-NOT on mRNAs and its repositioning by UNK, evaluate whether or not both effectors associate with UNK at its mRNA-binding sites, and carry out structural studies of UNK RNPs, especially those reconstituted *in vitro* from purified components. The striking reliance on multivalent interactions between UNK and the multisubunit CCR4-NOT complex, multiplied by a high number of RNA binding sites that generally track with efficient translational repression by UNK, could conceivably lead to molecular-scale condensation. Finally, we speculate that the reporter plasmid-linked transcriptional activity of the principally cytoplasmic UNK may at least in part rely on its recruitment of CCR4-NOT, which is known to participate in diverse gene regulatory processes, including control of gene transcription[32,59]. Although there is currently no evidence for such activity of native UNK, we leave open a possibility that its regulatory repertoire encompasses processes in addition to the control of translation. While beyond the

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

scope of this study, links between the molecular and cellular biology of UNK, including its control of cell morphogenesis, present intriguing avenues to explore in future studies.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

The authors thank M. Oakes, K. Jepsen, Y. Kanomata, S.-A. Chung, R. Tomaino, Y. Wu, and C. Cai for assistance; F. Lan for the pDEST-pcDNA3-Gal4 plasmid; N. Ingolia for the pNTK576-pET28a-His6x-KmHnt3 plasmid; A.B. Shyu for the anti-CNOT7 antibody; the UCI Genomics Research & Technology Hub, the UCSD IGM Genomics Center, the Taplin Mass Spectrometry Facility at Harvard Medical School, and the Biophysics Resource in the Center for Structural Biology, Center for Cancer Research, NCI at Frederick. This work was supported in part by the U.S. National Institutes of Health grant R01 GM144693 (to J.M.). D.J.T., J.C., K.R., and E.V. are supported by the Intramural Research Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health.

AUTHOR CONTRIBUTIONS

K.S., S.H., D.J.T., E.V., and J.M. conceptualized the study and designed experiments and data analyses. K.S., D.J.T., K.R., J.C., and J.M. performed all experiments. S.H. and E.V. performed bioinformatic analyses. J.M.B., S.C., and C.I. contributed key reagents and provided guidance. E.V. and J.M. acquired funding and supervised this work. J.M. wrote the manuscript with assistance from D.J.T. and E.V. and input from all other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Doxycycline-inducible cell lines

Human cell lines, including SH-SY5Y, HeLa, and 293T, were maintained in DMEM medium supplemented with 10% FBS and penicillin/streptomycin at 37ºC and 5% CO_2. Cells were authenticated by ATCC using STR profiling and were regularly tested for mycoplasma contamination during experimentation.

For evaluation of the morphogenetic activity of UNK, Dox-inducible HeLa cells were created via infection with an all-in-one lentivirus expressing a puromycin resistance gene (PuroR), advanced reverse tetracycline-controlled transactivator protein ($rtTA^{Ad}$), and a TREtight-driven transcript encoding GFP alone or GFP and either Flag-HA-tagged UNK_{WT} or UNK mutants (pLIX-IRES-GFP; see Figure 1A and Plasmid constructs). Dox-inducible HeLa cells used in all other experiments were generated analogously, using a similar all-in-one lentivirus that did not express GFP (pLIX-403; Addgene_41395), and were made monoclonal via single-cell sorting to ensure comparable inducible expression of UNK in cells within a population and among populations expressing Flag-HA-tagged UNK_{WT}, UNK_{dPAM2}, or UNK_{3M}. To induce transgene expression, puromycin-resistant cells were treated with doxycycline (Millipore Sigma) at 1 ug/ml. CRISPR/Cas9-mediated knockout of CNOT9 was achieved via transduction with a lentivirus for expression of gRNA, Cas9, and a blasticidin resistance gene (lentiCRISPR v2-Blast; Addgene_83480). Successful CRISPR/Cas9-mediated genome editing was monitored in single-cell clones by PCR and sequencing of the genomic locus. All lentiviral particles were produced in 293T cells by co-transfection of a lentiviral expression vector, the lentiviral packaging vector pCMV delta R8.2 (Addgene_12263) and the pMD2.G vector (Addgene_12259) with polyethylenimine (Polysciences, 23966-100; pH7.0). Growth medium was exchanged 16 hours post transfection. After 2 days, virus-containing supernatant was filtered through a 0.45 µm syringe filter and used for transduction.

Plasmid constructs

Plasmids for transient expression of Gal4-tagged UNK mutants (Figures 1 and 2) were created using the Gateway cloning strategy where UNK mutants in the pENTR/D-TOPO vector backbone (Thermo Fisher Scientific, K240020) were transferred in an LR reaction using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific, 11791020) into the pDEST-pcDNA3-Gal4 vector (generated by insertion of the ccdB cassette into the pcDNA3-Gal4 construct[60]; gift from Fei Lan) following manufacturer's protocol. The entry clones were generated by first inserting the full-length mouse UNK_{WT} that was amplified by PCR from the pTtight-UNK-IGPP vector[16] using the Gateway BP Clonase II Enzyme Mix (Thermo Fisher Scientific, 11789020) into the pENTR/D-TOPO vector, yielding pENTR-UNK. Full-length UNK deletion mutants and residue mutants F504A, L522A, 3M, WFAA, and W622A (Table S1) were prepared by mutating the pENTR-UNK vector. Specifically, the deletion mutants were generated by PCR with oligos flanking the deleted regions and amplification of the entire plasmid. The resulting reactions were treated with DpnI and transformed into One Shot TOP10 *E. coli* (Thermo Fisher Scientific, C404003). The above residue mutants were created by site-directed mutagenesis following instructions provided in the QuickChange XL Site-Directed Mutagenesis Kit manual (Agilent Technologies). All other full-length UNK residue mutants, including 11DE-A, 10FY-A, 7KH-A, and E8 (Table S1), were created by replacing IDR_{WT} in pENTR-UNK with corresponding mutant IDRs synthesized as GeneArt Strings DNA fragments (Thermo Fisher

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Scientific). To enable the replacements, BamHI and BspEI restriction enzyme cut sites were introduced by silently mutating UNK sequences immediately 5' and 3' to the IDR, respectively, by site-directed mutagenesis. The replacements were performed by cutting the resulting entry vector with BamHI and BspEI to release IDR_{WT} and clone in either of the mutant IDRs amplified by PCR from the synthetic DNA fragments. All full-length UNK_{WT} and IDR_{WT} truncation mutants analyzed in Figures 1 and 2 (see Table S1), as well as IDR residue mutants F504A, L522A, 3M, 11DE-A, 10FY-A, and 7KH-A were created by PCR amplification of the corresponding UNK segments from the above entry vectors followed by their insertion into the EcoRI- and XbaI-cut pENTR vector. IDR residue mutants D520S and 3D-S were created by site-directed mutagenesis of pENTR-IDR. IDR residue mutants 5DE-S, 7DE-S, 41S-A, 23LI-A, and 17P-A were ordered as synthetic DNA fragments and cloned into pENTR as above.
For the expression of MBP-IDR-Strep proteins in bacteria (Figure 4), WT or mutant IDR were amplified from the above entry vectors or a synthetic DNA fragment encoding IDR with V511E/I515E/L522E substitutions such that two StrepII tags (GSGWSHPQFEKGSWSHPQFEK) were added in-frame straight after the C-terminal residue of IDR in each protein. The amplicons were then inserted individually in the pnYC-NvHM_M plasmid (Addgene_146932) between NdeI and MfeI sites. The same strategy was employed for cloning of UNK_C, UNK_C E6, and UNK_C E8 (Figure 5), which were amplified from pENTR-UNK (for UNK_C) or synthetic DNA fragments (for UNK_C E6 and UNK_C E8) for insertion in the pnYC-NvHM_M plasmid.

Plasmids for the expression in insect cells of the full-length UNK fused C-terminally to two StrepII tags (Figures 4, 5, and S3) were generated by insertion of the full-length UNK amplified from the pENTR-UNK vector by PCR, which also introduced two C-terminal StrepII tags, in the pLIB plasmid (Addgene_80610) between BamHI and SalI sites.

All-in-one lentiviral plasmids for Dox-inducible expression of UNK without GFP in HeLa cells were created by insertion of the Flag-HA-tagged full-length WT or mutant UNK amplified from the corresponding entry vector into the pLIX_403 plasmid (Addgene_41395) between NheI and AgeI sites. For Dox-inducible expression of GFP with or without UNK, the pLIX-IRES-GFP plasmid was first created by subcloning the IRES-GFP segment from pTREtight-IRES-GFP-PGK-Puro[16] into pLIX_403 between NheI and AgeI sites. WT or mutant Flag-HA-tagged UNK, amplified from the entry vectors above, were then cloned individually in pLIX-IRES-GFP between NheI and MluI sites.

To generate a plasmid for stable knockdown of the endogenous UNK in SH-SY5Y cells (Figure S1), an shRNA targeting human *UNK* gene was cloned in the pLKO.1 puro plasmid (Addgene_8453) between AgeI and EcoRI sites (Key resources table)[61]. We used the scramble shRNA plasmid (Addgene_1864) for the expression of non-targeting control shRNA (Figure S1). The *UNK*-targeting or non-targeting control guide sequences were introduced into the BsmBI-digested lentiCRISPR v2-Blast plasmid (Addgene_83480) as pairs of annealed oligos (Key resources table)[62].

Transfection of siRNAs

HeLa cells inducibly expressing GFP and Flag-HA-tagged UNK_{WT} were seeded in 6-well dishes and transfected 24 hours later at about 40% confluence using the *Trans*IT-X2 Dynamic Delivery System (Mirus, MIR6003) with a pool of siRNAs targeting CNOT7 (Horizon, L-012897-00-0005), CNOT8 (Horizon, L-018791-00-0005), CNOT9 (Horizon, L-019972-00-0005), or a non-targeting siRNA pool (Horizon, D-001206-13) at 50 nM. Cells were induced with Dox at 24

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hours after transfection and cell morphologies or the efficiency of knockdown were evaluated at 48 hours after induction.

Dual luciferase reporter assay

Dual luciferase reporter assays were performed by co-transfecting 400 ng of a Gal4-tagged UNK mutant-expressing plasmid, 200 ng of the pGL4.35[luc2P/9XGAL4UAS/Hygro] Vector (Promega, E1370), and 20 ng of the control pRL-TK Vector (Promega, E2241) into 293T cells using Lipofectamine 2000 Transfection Reagent (Themo Fisher Scientific, 11668019). Forty-eight hours after transfection, cells were harvested and processed using Dual-Luciferase Reporter Assay System (Promega, E1960) according to the manufacturer's instructions. The emitted luminescence was detected using SpectraMax L Luminescence Microplate Reader (Molecular Devices).

RT-qPCR analysis

Total RNA was extracted from samples equivalent to those used for the dual luciferase assays using TRIzol Reagent (Themo Fisher Scientific, 15596018) and Direct-zol RNA Miniprep (Zymo Research, R2050) according to the manufacturer's instructions. cDNA was prepared from equal amounts of RNA using PrimeScript RT Reagent Kit (Takara, RR037A) following manufacturer's instructions. qPCR was performed using PowerUp Sybr Green Master Mix (Thermo Fisher Scientific, A25742) and the oligonucleotide primers listed in the Key resource table to amplify the cDNA on the CFX Connect Real-Time PCR Detection System at the annealing temperature of 63 °C. Relative firefly luciferase mRNA levels were normalized to relative expression levels of the RPS18 gene that was used as an internal control.

Immunofluorescence

SH-SY5Y cells and HeLa cells ectopically expressing Flag-HA-tagged UNK_{WT} or UNK_{3M} were fixed in 4% paraformaldehyde for 10 min at room temperature, permeabilized in 0.5% Triton X-100, blocked in 5% goat serum, and probed with anti-UNK (Millipore Sigma, HPA023636) or anti-HA antibodies at 4°C for 24 hours. After an overnight incubation, the cells were probed with fluorochrome-conjugated secondary antibodies for 1 hour at room temperature and mounted using VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200-10). Images were taken with the LSM 880 confocal microscope (Zeiss).

Quantification of cell morphologies

Cell morphologies were quantified essentially as reported previously[16]. Briefly, after 48 hours of incubation with Dox, HeLa cells inducibly expression either GFP alone or GFP and WT or mutant UNK were imaged and the axes of GFP-positive cells were measured with Adobe Illustrator software (Adobe). The morphologies of at least 50 GFP-positive cells were quantified for each induced transgene by calculation of their axial ratios, y/x, where y is the length of the absolute longest cellular axis and x is the length of the longest axis perpendicular to the y axis.

SDS-PAGE and western blot analysis

Whole cell lysates and eluates from immunoprecipitations were run on 10% SDS-polyacrylamide gels and transferred to supported nitrocellulose membrane (Bio-Rad) by standard methods. Membranes were then blocked for 1 hour in 5% non-fat dry milk in 1 x TBS with 0.1% Tween-20 (TBST), rinsed, and incubated with primary antibody diluted in 3% BSA in TBST

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

overnight at 4 ºC. The following primary antibodies were used: anti-Flag (Millipore Sigma, F1804), anti-UNK (Millipore Sigma, HPA023636), anti-CNOT1 (Proteintech, 14276-1-AP), anti-CNOT2 (Cell Signaling Technology, 34214), anti-CNOT3 (Proteintech, 11135-1-AP), anti-CNOT9 (Fine Test, FNab07487), anti-PABPC (Abcam, ab21060), and anti-β-Actin-peroxidase (Millipore Sigma, A3854). Blots were washed in TBST, incubated with HRP-conjugated secondary antibodies in 5% milk in TBST for 1 hour (except for anti-β-Actin-peroxidase antibody), and washed again. HRP signal was detected by Western Lightning Plus chemiluminescent substrate (NEL103001EA).

Co-immunoprecipitation from cell lysates

For co-IP experiments, HeLa cells inducibly expressing Flag-HA-tagged UNK were treated for 24 hours with Dox. Uninduced samples were processed in parallel. Cells were harvested, washed once with PBS, and lysed in whole-cell lysis buffer (20 mM HEPES-KOH pH 7.9, 10% glycerol, 300 mM KCl, 0.1% IGEPAL, 1 mM DTT) supplemented with cOmplete Protease Inhibitor Cocktail (Millipore Sigma, 11697498001) for 30 min at 4 ºC. Supernatants were cleared off debris by a 30-min centrifugation at 13,000 rpm at 4 ºC. The lysates were then mixed with an equal volume of no-salt lysis buffer (20 mM HEPES-KOH pH 7.9, 10% glycerol, 0.1% IGEPAL, 1 mM DTT) supplemented with cOmplete Protease Inhibitor Cocktail to lower the final salt concentration to 150 mM KCl (IP buffer), added to anti-Flag or normal mouse IgG antibody-conjugated Protein G Dynabeads (Invitrogen 10003D), and rotated for 2 hours at 4 ºC. To prepare antibody-conjugated magnetic beads, 50 µl of Protein G Dynabeads per experiment were washed with the IP buffer, resuspended in 100 µl IP buffer with 2 µg antibody, rotated at room temperature for 45 min, and washed twice with the IP buffer before being mixed with the cleared lysate. After the IP, the beads were washed thoroughly with the IP buffer and the bound proteins were eluted with 200 µg/ml Flag (DYKDDDDK) peptide (GenScript, RP10586) in thermomixer at 4 ºC, shaking at 1250 rpm for 1 hour. The eluates were analyzed by western blotting.

Protein complex purification

To purify protein complexes of UNK_{WT}, UNK_{dPAM2}, and UNK_{3M}, approximately 300 million HeLa cells per experiment were harvested at 24 hours of induction with Dox, flash-frozen in liquid nitrogen, and stored at -80 ºC until use. Cells were resuspended in buffer A (20 mM HEPES-KOH pH 7.9, 10% glycerol, 300 mM KCl, 0.1% IGEPAL, 1 mM DTT, cOmplete Protease Inhibitor Cocktail; 100 µl of buffer A was used per 1e6 cells) and rotated at 4 ºC for 30 min. The lysates were centrifuged at 13,000 rpm for 30 min at 4 ºC, supernatants were collected, and dialyzed in dialysis buffer (20 mM Tris-Cl, pH 7.3, 100 mM KCl, 0.2 mM EDTA, 20% Glycerol, 0.2 mM PMSF, 10 mM beta-merceptoethanol) for 1 hour at 4 ºC. The lysates were centrifuged at 13,000 rpm for 30 min at 4 ºC, then 250 µl anti-Flag M2 affinity gel (Millipore Sigma, A2220) was added and the mixture was rotated for 2 hours at 4 ºC. The affinity gel was then washed with TAP-wash buffer (50 mM Tris-Cl, 100 mM KCl, 5 mM $MgCl_2$, 10% Glycerol, 0.2 mM PMSF, 0.1% NP40). The bound proteins were eluted with Flag peptide (200 µg/ml; GenScript, RP10586) in thermomixer at 4 ºC, shaking at 1250 rpm for 1 hour. The eluate was mixed with anti-HA magnetic beads (Thermo Fisher Scientific, 88837) and rotated for 2 hours at 4 ºC. The beads were washed with TAP-wash buffer and proteins were eluted using HA peptide (200 µg/ml; GenScript, RP11735) by shaking the beads in thermomixer at 1400 rpm for 45 min at 30 ºC. The eluate was was TCA-precipitated and analyzed by mass spectrometry.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mass spectrometry

Fifty µl of 50mM ammonium bicarbonate with 10% acetonitrile were added to the dry tubes containing the TCA-precipitated protein and gently vortexed. Next, 10 µl (20 ng/µl) of modified sequencing-grade trypsin (Promega, V5111) was spiked into the solutions and the samples were incubated at 37 ºC overnight. Samples were acidified by spiking in 5 µl 20% formic acid solution and then desalted by a STAGE tip[63]. On the day of analysis, the samples were reconstituted in 10 µl of HPLC solvent A. A nano-scale reverse-phase HPLC capillary column was created by packing 2.6 µm C18 spherical silica beads into a fused silica capillary (100 µm inner diameter, 30 cm in length) with a flame-drawn tip[64]. After equilibrating the column, each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ Orbitrap Velos Elite ion-trap mass spectrometer (Thermo Fisher Scientific). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences and hence protein identities were determined by matching protein databases with the acquired fragmentation pattern by the software program Sequest (Thermo Fisher Scientific)[65]. All databases included a reversed version of all the sequences and the data was filtered at between 1% to 2% peptide false discovery rate.

Recombinant protein expression and purification

Full-length mouse UNK with two C-terminal StrepII tags was produced in Sf21 insect cells using the MultiBac baculovirus expression system as previously described (Plasmid constructs)[66,67]. In brief, DH10-EmBacY cells were transformed with pLIB-UNK, transposition onto the baculoviral genome was selected by blue-white screening, the bacmid DNA was isolated and transfected into Sf21 cells to generate baculovirus. Sf21 cells were grown to a density of 2×10^6 cells/ml at 27 °C in Sf900II medium (Thermo Fisher Scientific), infected with the V1 UNK stock of baculovirus, and harvested 48 hours after they stopped dividing. Cells were resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, pH 7.5) and lysed using a Branson Ultrasonics Sonifier SFX550. The lysate was cleared by centrifugation at 40,000 g for 1 hour at 4 °C and filtered through 0.45 µm syringe-driven filters (Millipore). The cleared and filtered lysate was loaded onto a 1 ml StrepTrap XT column (Cytiva). The bound protein was eluted in one step with binding buffer (50 mM HEPES, 200 mM NaCl, pH 7.5) supplemented with 50 mM biotin.
An MBP-tagged C-terminal fragment of UNK (residues 637-810) was produced in *E. coli* BL21 (DE3) Star cells (Thermo Fisher Scientific) in LB medium at 20 °C as a fusion protein carrying an N-terminal His6-MBP tag and two C-terminal StrepII tags (Plasmid constructs). Cells were resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 30 mM Imidazole, pH 7.5) and lysed using a Branson Ultrasonics Sonifier SFX550. The lysate was cleared by centrifugation at 40,000 g for 1 hour at 4 °C. The cleared lysate was loaded onto a 5 ml HisTrap column (Cytiva). The bound protein was eluted over a linear gradient with elution buffer (50 mM HEPES, 200 mM NaCl, 500 mM Imidazole, pH 7.5). The final step was size exclusion chromatography on a Superdex 200 26/600 column in a buffer containing 10 mM HEPES, 200 mM NaCl, 2 mM DTT, pH7.5. In addition, two mutated constructs UNK_{E6} (residues 637-810 with L650E, L654E, I661E, A682E, L706E, L713E substitutions) and UNK_{E8} (residues 637-810 with L650E, L654E,

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

I661E, A682E, L706E, L713E, L740E, L754E substitutions) were produced and purified in the same manner as the WT version.

To prepare the recombinant, thermostable 5'-deadenylase (Hnt3p protein from a thermophilic eukaryote *K. marxianus*) used in ribosome profiling experiments, BL21 (DE3) bacteria were transformed with the pNTK576-pET28a-His6x-KmHnt3 plasmid (gift from Nicholas Ingolia)[68]. Individual colonies were picked and 500 ml cultures were grown to an OD of 0.4. Liquid cultures were then induced with 500 µl of 1 M IPTG and transferred to a shaker at 16°C for 18 hours. Cultures were pelleted by centrifugation at 3,000 g for 20 min at 4 °C and the pellets were flash-frozen with liquid nitrogen. To lyse the cells, pellets were maintained on ice and resuspended in 15 ml lysis buffer (500 mM NaCl, 0.5% NP-40 (Igepal), 10 mM Imidazole, 20 mM HEPES, 10 mM bMe, pH7.5). Resuspended pellets were sonicated for a total of 90 seconds then centrifuged at 10,000 g for 20 min at 4 °C. The supernatant was collected and incubated with 1.5 ml of lysis buffer-equilibrated Ni-NTA beads for batch binding. The slurry was rotated for 1 hour at 4 °C and then spun at 3,000 rpm for 3 min. The supernatant was carefully removed and the Ni-NTA beads were washed with 10 ml high-salt wash buffer (1 M NaCl, 20 mM Imidazole, 10 mM bMe, 20 mM HEPES, pH 7.5) followed by low-salt wash buffer (10 mM NaCl, 20 mM Imidazole, 10 mM bMe, 20 mM HEPES, pH 7.5). Beads were incubated for 5 min with 1.5 ml elution buffer (10 mM NaCl, 250 mM Imidazole, 10 mM bMe, 20 mM HEPES, pH 7.5) and the elution fractions were collected. This process was repeated a total of three times. Glycerol was added to each collected fraction to a final concentration of 10% before flash-freezing in liquid nitrogen for long term storage.

StrepTactin pull-down assay

StrepII-tagged MBP, as well as StrepII-tagged and MBP-tagged UNK IDR (residues 467-640) WT and mutant constructs were produced in *E. coli* BL21 (DE3) Star cells (Thermo Fisher Scientific) grown in auto-induction medium overnight at 37 °C. Cells were resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, pH 7.5) and lysed using a Branson Ultrasonics Sonifier SFX550, the lysate was then cleared by centrifugation at 40,000 g for 1 hour at 4 °C. The cleared lysate or purified UNK was incubated with StrepTactin Sepharose resin (Cytiva, 28935599). After a 1-hour incubation beads were washed twice with 50 mM HEPES, 500 mM NaCl, pH 7.5, 0.03% Tween, once with 50 mM HEPES, 500 mM NaCl, pH 7.5, and once with binding buffer (50 mM HEPES, 200 mM NaCl, pH 7.5). Purified modules of the human CCR4-NOT complex, prepared as previously described[28], or purified PABPC1 were added to the bead-bound proteins. After a 1-hour incubation, beads were washed four times with binding buffer and proteins were eluted with 50 mM biotin in binding buffer. The eluted proteins were analyzed by SDS polyacrylamide gel electrophoresis followed by Coomassie blue staining.

AlphaFold-Multimer prediction methods

Predictions were generated with AlphaFold-Multimer[29,30] version 2.3.2 following a published approach[31] and using a computing cluster with these key settings:
--db_preset=full_dbs --max_template_date=2020-05-14 --models_to_relax=best --model_preset=multimer --num_multimer_predictions_per_model=5
The resulting predicted models were aligned in PyMOL v2.5.4 to assess prediction convergence, and this software was used to prepare all structural figures. For UNK, the IDR segment, the C-terminal region, or the full-length sequences were provided. For the CCR4-NOT subunits,

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequences corresponding to experimentally determined structures were used (PDB accession codes 4crv and 4c0d)[33,34].

Mass Photometry

Mass photometry of UNK, UNK_C, UNK_{E6}, UNK_{E8}, and $CNOT_{MINI}$ was performed using the Refeyn TwoMP mass photometry instrument in buffer containing 50mM HEPES, 200mM NaCl, pH7.5. Molecular weight calibrations were performed using two protein oligomer solutions, β-amylase (56, 112, and 224 kDa) and Thyroglobulin (670 kDa). The data acquisition was performed with AcquireMP (version 2023 R1.1) software and data analysis was performed with DiscoverMP (version 2023 R1.2) software.

Individual-nucleotide resolution UV-crosslinking and immunoprecipitation (iCLIP)

All iCLIP experiments were performed in replicates following the iCLIP2 protocol[36]. Briefly, monoclonal HeLa cells inducibly expressing Flag-HA-tagged UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} were grown in 10 cm plates and harvested at 85% confluence. Prior to harvest, the cells were treated with Dox for 24 hours or were left untreated. The cells were then washed with ice-cold PBS and irradiated with UV-light at 254 nm on ice. The irradiated cells were scraped, aliquoted into three 2-ml tubes and centrifuged at 5,000 g for 2 min at 4 °C. The supernatant was removed and the cell pellets were flash-frozen in liquid nitrogen and stored at -80 ºC until use. Immunoprecipitation of the crosslinked UNK-RNA or PABPC1-RNA complexes was carried out using anti-Flag antibody (Millipore Sigma, F1804) or anti-PABPC1 antibody (Abcam, ab21060). The complete iCLIP experiment, including deep sequencing of the prepared cDNA libraries, was repeated in four and two replicates for UNK and PABPC1 iCLIP libraries, respectively.

RNA-seq library preparation

Total RNA from aliquots of samples used for ribosome profiling experiments (see Ribosome profiling) of each uninduced cells or cells expressing Flag-HA-tagged UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} for 24 hours was extracted using Direct-zol RNA miniprep kit (Zymo Research, R2050) according to the manufacturer's instructions. Poly-A containing RNA was enriched from the total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, E7490S) and sequencing libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep with Sample Purification Beads (NEB E7765S). RNA-seq libraries for each sample type were prepared, sequenced, and analyzed in triplicates.

Ribosome profiling

Ribosome profiling experiments with uninduced cells or cells expressing Flag-HA-tagged UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} for 24 hours were carried out in duplicates essentially as described[69] and by following the cDNA library-making protocol as for the iCLIP experiments[36]. Briefly, cells were grown in 15 cm dishes and harvested at 70% confluence. Prior to harvest, cells were treated or not with Dox for 24 hours to induce the expression of the transgenes. Cells were then washed in ice-cold PBS, lysed for 10 min on ice in a lysis buffer, triturated by passing twice through a syringe fitted with a 26-gauge needle, and spun at 20,000 g for 10 min at 4 °C. Cell lysates were digested with RNase I for 45 min at room temperature followed by the addition of SUPERase•In RNase Inhibitor (Themo Fisher Scientific, AM2696). The lysates were underlaid with 1 M sucrose and spun in a 50.4 Ti rotor at 50,000 rpm for 2 hours at 4 °C. Pellets were resuspended in TRIzol reagent (Life Technologies, 15596018) and RNA was extracted

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

according to the manufacturer's instructions. The extracted RNA was size-selected by denaturing PAGE, retaining only fragments between 26 and 34 nts, and 3' end-dephosphorylated with T4 PNK for 30 min at 37 °C followed by ligation to a pre-adenylated linker (L3-App) as described for the iCLIP procedure[36]. Unligated 3' linker was removed by incubating the samples with the 5'-deadenylase KmHnt3 (see Recombinant protein expression and purification) and RecJ exonuclease (New England Biolabs, M0264S) for 45 min at 37 ºC. The 3L-App ligated RNA was purified with Oligo Clean & Concentrator (Zymo Research, D4060), reverse transcribed by SuperScript IV Reverse Transcriptase (Themo Fisher Scientific, 18090010), and converted to cDNA libraries for high-throughput sequencing, as described[36].

Poly(A) tail length analysis

Total RNA was extracted from uninduced cells or cells expressing Flag-HA-tagged UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} for 24 hours using Direct-zol RNA Miniprep (Zymo Research, R2050) according to the manufacturer's instructions. RNA quality was assessed using the 2100 Bioanalyzer (Agilent Technologies) with the RNA integrity number ranging from 8.4 to 10. Libraries for direct RNA sequencing were prepared from mRNA in duplicates using Library Kit SQK-RNA002 (Oxford Nanopore Technologies) and sequenced on the GridION or PromethION 2 Solo device (Oxford Nanopore Technologies) using FLO-MIN106D or FLO-PRO002 flow-cells, respectively. One flow-cell was used for each sample.

Quality control of cDNA libraries and High-throughput sequencing

iCLIP, RNA-seq, and ribosome profiling cDNA libraries were analyzed by non-denaturing PAGE and the 2100 Bioanalyzer (Agilent Technologies), quantified with the Qubit 2.0 Fluorometer (Life Technologies), pooled by library type, and sequenced using the HiSeq 2500, HiSeq 4000, or NovaSeq 6000 systems (all Illumina).

Computational analyses

Conservation and disorder score calculation

Amino acid conservation of UNK was calculated using https://consurf.tau.ac.il, applying default settings. The disorder confidence score was calculated using the DISOPRED3 algorithm with default settings on the full-length mouse UNK amino acid sequence on the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipre).

Analysis of iCLIP data

The iCLIP data were processed essentially as described previously[70]. Briefly, data was assessed with FastQC (v0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). High quality data was chosen with fastq_quality_filter from FASTX Toolkit (v0.0.13, http://hannonlab.cshl.edu/fastx_toolkit/), with these parameters: -Q 33 -q 10 -p 100. The indexed sequencing reads were demultiplexed with flexbar (v3.5.0)[71,72] and then mapped to UCSC hg38 genome with STAR genome aligner(v2.7.3a)[73] using these parameters: --outFilterMismatchNoverReadLmax 0.04 --outFilterMismatchNmax 999 --outFilterMultimapNmax 1 --alignEndsType Extend5pOfRead1 --sjdbGTFfile gencode.v35.annotation.gtf --sjdbOverhang 75 --outReadsUnmapped Fastx --outSJfilterReads Unique --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --runThreadN 8. PCR duplicates were removed with umi_tools (v1.0.1)[74]. PureCLIP (v1.3.1, parameter: -ld -nt 8)[75] was utilized to identify individual crosslink events and for calling of peaks, i.e., binding

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sites. A minimum of 20 crosslink events were required for each peak. All peaks called by PureCLIP were expanded to a 9-nt region and assigned to Gencode (v35) comprehensive gene annotation[76]. For assessing the genomic distribution of iCLIP crosslink nucleotides, we used the following hierarchy: ncRNA > CDS > 3'UTR > 5'UTR > intron > other > intergenic (Figure S4E). Peaks mapping to different isoforms of a gene were assigned to the gene.

The classification of gene-binding strength (high, moderate, or low/no; Figures 7 and S6) was based on the total tag number in all gene's peaks or on maximal tag number of individual, ~9-nt peaks. In particular, we classified as highly bound those genes that were either among the top 25% in their tag number within the called peaks or had a maximal peak height that ranked among the top 25% of all genes. We also classified as lowly or non-bound those genes whose both total tag number and maximal individual peak height ranked in the bottom 25% of genes in the respective categories. All other genes were classified as moderately bound.

The sequence composition at UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} binding sites was assessed as described previously (Figures 6A and 6B)[17]. First, we identified the position of the maximum within each binding site (i.e., the nucleotide with the highest number of crosslink events; the first was taken in case of multiple nucleotides with equal counts) and extracted an extended window of 51 nts on either side. We counted the frequency of all 64 possible trinucleotides (triplets) at each position across all binding sites, counting each triplet on the first of three nucleotides. To correct for different background levels, we further normalized the frequency profile of each triplet to its median frequency across the complete 103-nt window, generating enrichment scores.

To compare the spatial arrangement of different triplets, we performed unsupervised hierarchical clustering of the normalized triplet profiles in a 31-nt window around the binding site maxima of UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} (Figure 6A; the heatmap profiles the enrichment scores). The resulting dendrogram was split into subtrees to obtain three sets of triplets with similar spatial distribution: (1) UAG, (2) U/A-rich triplets (UUU, AUU, UUA, CUU, UUC, UUG, AAU, AAA, GUU, UAA, UAU, and AUA), and (3) all remaining triplets. Triplet frequencies in each set were combined into a summarized profile (Figure 6A, top).

To assess the prevalence of the UAG triplet in WT-specific and mutant-specific binding sites (Figure 6E), the ratio of respective peaks with UAG within 15 nts upstream of the binding site maxima versus all analyzed peaks was calculated. The occurrence of PABPC1 iCLIP peaks in the vicinity of UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} iCLIP peaks on mRNA (Figure S4K), the UNK binding sites on mRNA were slopped with bedtools to upstream and downstream for 20 nt[77]. The overlapping PABPC1 peaks were counted with the bedmap[78]. An intersection of more than 1 nt was considered as overlapping.

Ribosome profiling data analysis

After quality control with FastQC (v0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), sequencing reads were demultiplexed with flexbar (v3.5.0)[71,72] and mapped to the human rRNA with bowtie2 (v2.4.5)[79]. Hg38 rRNA sequences were retrieved from UCSC repeatmask database using table browser tool (https://genome.ucsc.edu/cgi-bin/hgTables). Reads not mapping to human rRNA were then mapped to hg38 lncRNA (Gencode v35; https://www.gencodegenes.org/human/) and the unmapped reads were aligned to hg38 protein-coding transcripts, keeping only uniquely mapping reads. Ribosome protected fragments (RPFs) in each gene (Gencode v35) were counted with samtools (v1.14)[80]. The differential ribosome occupancy was performed with DESeq2 package (v1.38.2)[81]. Because UNK targets the majority of all expressed transcripts, as indicated by the

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

iCLIP analysis, instead of using DESeq2-inherent normalization, we normalized RPFs to 311 highly expressed genes (RPKM > 10) that had no UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} binding sites. We normalized RPF counts for individual genes using the formula 40,000*R/N, where 40,000 is an arbitrary number that is close to the average RPF count for the 311 genes across all conditions, N is the total RPF count for the 311 genes, and R is the RPF count for an individual gene. We considered as differentially translated genes with a P value < 0.05 and fold change in ribosome occupancy > 2. The volcano plots in Figures 7 and S6 were drawn with ggplot2 package (ggplot2_3.4.0).

RNA-seq data analysis

The RNA-seq data were mapped to the human genome (hg38) using STAR (v2.7.3a). Read counts in each gene (Gencode v35) were calculated with featureCounts tools (v2.0.0) from the Rsubread package[82]. Differentially expressed genes (DEGs) were identified with DESeq2 (v1.38.2)[81]. Similar to the ribosome profiling data analysis, reads uniquely mapping to protein-coding genes were normalized to the 311 highly expressed, UNK-unbound genes. DEGs were chosen based on the adjusted P value threshold of 0.05 and fold change in expression > 2. RPKM values were calculated with cufflinks (v2.2.1)[83].

Poly(A) tail-seq data analysis

Raw reads from the poly(A) tail-seq libraries were base-called with guppy (v6.3.7, https://community.nanoporetech.com/downloads). Passing reads were mapped to Gencode v35 transcripts using minimap2[84,85]. Poly(A) tail length for each read was estimated using the Nanopolish (https://github.com/jts/nanopolish) and only length estimates with the QC tag reported as PASS were considered in subsequent analyses. Read numbers for genes with different transcript isoforms were combined. Reads per million (RPM) values were calculated and correlated with transcript abundance (RPKM; RNA-seq data; Figure S5A) and average tail length (Figure S5B). Tail length distributions were determined for different gene/transcript groups based on the numbers UNK binding sites (BS): No BS, 1-2 BS, 3-10 BS, 11-30 BS, and 31 or more BS.

QUANTIFICATION AND STATISTICAL ANALYSIS

Information on statistical tests used to determine significance of specific differences is described in figure legends.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-UNK antibody produced in rabbit	Millipore Sigma	Cat#HPA023636; RRID:AB_10601774
Monoclonal ANTI-FLAG M2 antibody produced in mouse	Millipore Sigma	Cat#F1804; RRID:AB_262044
CNOT1 Polyclonal antibody	Proteintech	Cat#14276-1-AP; RRID:AB_10888627
CNOT2 (D8Z8P) Rabbit mAb	Cell Signaling Technology	Cat#34214; RRID:AB_2799049

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CNOT3 Polyclonal antibody	Proteintech	Cat#11135-1-AP; RRID:AB_2229682
Rabbit polyclonal anti-CNOT7 antibody	Yamashita et al.[86]	N/A
Anti-RQCD1 antibody	Fine Test	Cat#FNab07487
Anti-PABP antibody	Abcam	Cat#ab21060; RRID:AB_777008
Purified anti-HA.11 Epitope Tag Antibody	BioLegend	Cat#901502; RRID:AB_2565007
Normal mouse IgG	Santa Cruz Biotechnology	Cat#sc-2025; RRID:AB_737182
Anti-β-Actin−Peroxidase antibody, Mouse monoclonal	Millipore Sigma	Cat#A3854; RRID:AB_262011
Anti-Mouse IgG (Fc specific)–Peroxidase antibody produced in goat	Millipore Sigma	Cat#AP124P; RRID:AB_90456
Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat	Millipore Sigma	Cat#AP307P; RRID:AB_92641
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat#11029; RRID:AB_2534088
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat#A-11034; RRID:AB_2576217
Bacterial and virus strains		
BL21 (DE3) Competent *E. coli*	New England Biolabs	Cat#C2527I
One Shot BL21 Star (DE3) Chemically Competent *E. coli*	Thermo Fisher Scientific	Cat#C601003
One Shot TOP10 Chemically Competent *E. coli*	Thermo Fisher Scientific	Cat#C404003
NEB Stable Competent *E. coli* (High Efficiency)	New England Biolabs	Cat#C3040H
Chemicals, peptides, and recombinant proteins		
Dynabeads Protein G for Immunoprecipitation	Thermo Fisher Scientific	Cat#10004D
ANTI-FLAG M2 Affinity Gel	Millipore Sigma	Cat#A2220; RRID:AB_10063035
Pierce Anti-HA Magnetic Beads	Thermo Fisher Scientific	Cat#88837
DMEM, high glucose, pyruvate	Themo Fisher Scientific	Cat#11995065
Fetal Bovine Serum	Themo Fisher Scientific	Cat#10437
Doxycycline hyclate	Millipore Sigma	Cat#D9891
Penicillin-Streptomycin (10,000 U/mL)	Themo Fisher Scientific	Cat#1514022
PfuTurbo DNA Polymerase	Agilent Technologies	Cat#600250
Deoxynucleotide (dNTP) Solution Mix	New England Biolabs	Cat#N0447S
T4 DNA Polymerase	New England Biolabs	Cat#M0203S
T4 Polynucleotide Kinase	New England Biolabs	Cat#M0201L
T4 DNA Ligase	New England Biolabs	Cat#M0202L
DpnI	New England Biolabs	Cat#R0176S
VECTASHIELD Antifade Mounting Medium with DAPI	Vector Laboratories	Cat#H-1200-10
Gateway BP Clonase II Enzyme mix	Themo Fisher Scientific	Cat#11789020

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gateway LR Clonase II Enzyme mix	Themo Fisher Scientific	Cat#11791020
TRIzol Reagent	Themo Fisher Scientific	Cat#15596018
Polyethylenimine, Linear, MW 25000, Transfection Grade	Polysciences	Cat#23966-100
DYKDDDDK Peptide	GenScript	Cat#RP10586
HA Peptide	GenScript	Cat#RP11735
Supported Nitrocellulose Membrane	BioRad	Cat#1620097
TURBO DNase (2 U/µL)	Themo Fisher Scientific	Cat#AM2238
Ambion RNase I, cloned, 100 U/µL	Themo Fisher Scientific	Cat#AM2295
SUPERase•In RNase Inhibitor (20 U/µL)	Themo Fisher Scientific	Cat#AM2696
SYBR Gold Nucleic Acid Gel Stain (10,000X Concentrate in DMSO)	Themo Fisher Scientific	Cat#S11494
Novex TBE-Urea Sample Buffer (2X)	Themo Fisher Scientific	Cat#LC6876
Novex™ TBE-Urea Gels, 15%	Themo Fisher Scientific	Cat#EC68852BOX
T4 RNA Ligase 2, truncated K227Q	New England Biolabs	Cat#M0351S
$RecJ_f$	New England Biolabs	Cat#M0264S
ProNex Size-Selective Purification System	Promega	Cat#NG2001
Phusion High-Fidelity PCR Master Mix with HF Buffer	New England Biolabs	Cat#M0531L
Dynabeads MyOne Silane	Themo Fisher Scientific	Cat#37002D
Ni-NTA Agarose	Genesee Scientific	Cat#20-512
cOmplete Protease Inhibitor Cocktail	Millipore Sigma	Cat#11697498001
Paraformaldehyde, 16% w/v aq. soln., methanol free	Themo Fisher Scientific	Cat#043368.9M
Sodium deoxycholate	Millipore Sigma	Cat#D6750
Sf-900 II SFM	Thermo Fisher Scientific	Cat#10902088
StrepTactin Sepharose High Performance resin for Strep - tag II	Cytiva	Cat#28935599
PrimeScript RT Reagent Kit (Perfect Real Time)	Takara	Cat#RR037A
SuperScript IV Reverse Transcriptase	Themo Fisher Scientific	Cat#18090010
*Trans*IT-X2 Dynamic Delivery System	Mirus	Cat#MIR6003
Lipofectamine 2000 Transfection Reagent	Themo Fisher Scientific	Cat#11668019
Western Lightning Plus, Chemiluminescent Substrate	PerkinElmer	Cat#NEL103001EA
Critical commercial assays		
MycoAlert Mycoplasma Detection Kit	Lonza	Cat#LT07-318
Dual-Luciferase Reporter Assay System	Promega	Cat#E1960
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	Cat#E7490S
NEBNext Ultra II Directional RNA Library Prep with Sample Purification Beads	New England Biolabs	Cat#E7765S
Oligo Clean & Concentrator	Zymo Research	Cat#D4060

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Direct-zol RNA Miniprep	Zymo Research	Cat#R2050
PowerUp SYBR Green Master Mix for qPCR	Themo Fisher Scientific	Cat#A25742
Qubit dsDNA Quantification Assay Kits, High Sensitivity	Themo Fisher Scientific	Cat#Q32851
Deposited data		
iCLIP, RNA-seq, ribosome profiling, and poly(A)Tail-seq data	This paper	Sequencing data deposited at GEO under accession number GSE240571
Experimental models: Cell lines		
Human: 293T cells	ATCC	CRL-3216
Human: HeLa cells	ATCC	CCL-2
Human: SH-SY5Y cells	ATCC	CRL-2266
Spodoptera frugiperda: Sf21 insect cells	Thermo Fisher Scientific	Cat#11497013
Oligonucleotides		
ON-TARGETplus Human CNOT7 (29883) siRNA - SMARTpool	Horizon	Cat#L-012897-00-0005
ON-TARGETplus Human CNOT8 (9337) siRNA - SMARTpool	Horizon	Cat#L-018791-00-0005
ON-TARGETplus Human CNOT9 (9125) siRNA - SMARTpool	Horizon	Cat#L-019972-00-0005
siGENOME Non-Targeting siRNA Pool #1	Horizon	Cat#D-001206-13
Primer for qPCR: FireflyLuciferase_fw GAGCTATTCTTGCGCAGCTT	Murn et al.[16]	N/A
Primer for qPCR: FireflyLuciferase_rev CCTCACCTACCTCCTTGCTG	Murn et al.[16]	N/A
Primer for qPCR: RPS18_fw GATGGGCGGCGGAAAATAG	Murn et al.[16]	N/A
Primer for qPCR: RPS18_rev GCGTGGATTCTGCATAATGGT	Murn et al.[16]	N/A
Oligo for CNOT9-targeting guide: gCNOT9-top CACCgcccatgctgtggcattcatt	Sgromo et al.[50]	N/A
Oligo for CNOT9-targeting guide: gCNOT9-bottom AAACaatgaatgccacagcatgggC	Sgromo et al.[50]	N/A
shRNA targeting sequence in human *UNK*: CAGGTACCACCTTCGTTACTA	This paper	N/A
Recombinant DNA		
pENTR/D-TOPO	Thermo Fisher Scientific	Cat#K240020
pDEST-pcDNA3-Gal4	Gift from Fei Lan[60]	N/A
pGL4.35[*luc2P*/9X*GAL4*UAS/Hygro] Vector	Promega	Cat#E1370
pRL-TK Vector	Promega	Cat#E2241
pCMV delta R8.2	N/A	RRID:Addgene_12263
pMD2.G	N/A	RRID:Addgene_12259
pLIX-403	N/A	RRID:Addgene_41395
pLIX-IRES-GFP	This paper	N/A

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pTREtight-IRES-GFP-PGK-Puro	Murn et al.[16]	N/A
pNTK576-pET28a-His6x-KmHnt3	Riepe et al.[68]	N/A
pLIB	Weissmann et al.[66]	RRID:Addgene_80610
pnYC-NvHM_M	Diebold et al.[87]	RRID:Addgene_146932
lentiCRISPR v2-Blast	N/A	RRID:Addgene_83480
pLKO.1 puro	Stewart et al.[61]	RRID:Addgene_8453
scramble shRNA	Sarbassov et al.[88]	RRID:Addgene_1864
Software and algorithms		
nanopolish (version 0.14.0)	Oxford Nanopore Technology	https://github.com/jts/nanopolish
guppy (version 6.3.7-GPU)	Oxford Nanopore Technology	https://community.nanoporetech.com/downloads
minimap2 (version 2.26-r1175)	Li[85]	https://github.com/lh3/minimap2/releases
tailfindr (version 1.4)	Krause et al.[89]	https://github.com/adnaniazi/tailfindr
samtools 1.14	Danecek et al.[80]	http://github.com/samtools/
featureCounts v2.0.0	Liao et al.[90]	https://subread.sourceforge.net/
cufflinks v2.2.1	Trapnell et al.[91]	http://cole-trapnell-lab.github.io/cufflinks/install/
DESeq2 v2_1.40.1	Love et al.[81]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ggplot2 v3.4.2	N/A	https://www.rdocumentation.org/packages/ggplot2/versions/3.4.2
bowtie2 (version 2.4.5)	Langmead & Salzberg[79]	https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.4.5/
bedtools (version 2.29.2)	Quinlan & Hall[77]	https://bedtools.readthedocs.io/en/latest/content/installation.html
UMI-tools (version 0.5.5)	Smith et al.[74]	https://github.com/CGATOxford/UMI-tools
PureCLIP (version 1.3.0)	Krakau et al.[75]	https://github.com/skrakau/PureCLIP
kentUtils (version v365)	N/A	https://github.com/ENCODE-DCC/kentUtils

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FASTX-Toolkit (version 0.0.14)	N/A	http://hannonlab.cshl.edu/fastx_toolkit/
seqtk (version 1.3)	N/A	https://github.com/lh3/seqtk
GenomicFeatures (version 1.50.3)	Lawrence et al.[92]	https://www.bioconductor.org/packages/release/bioc/html/GenomicFeatures.html
rtracklayer (version 1.58.0)	Lawrence et al.[93]	https://www.bioconductor.org/packages/release/bioc/html/rtracklayer.html
ChIPseeker (version 1.34.1)	Yu et al.[94]	https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html
AnnotationDbi (version 1.60.0)	N/A	https://www.bioconductor.org/packages/release/bioc/html/AnnotationDbi.html
org.Hs.eg.db (version 3.16.0)	N/A	https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html
STAR-2.7.3a	Dobin et al.[73]	https://github.com/alexdobin/STAR/releases/tag/2.7.3a
Other		
Code used to analyze data and generate figures	This paper	https://github.com/Shiyang-He/Unkempt-Project-data-analysis/

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTAL FIGURES

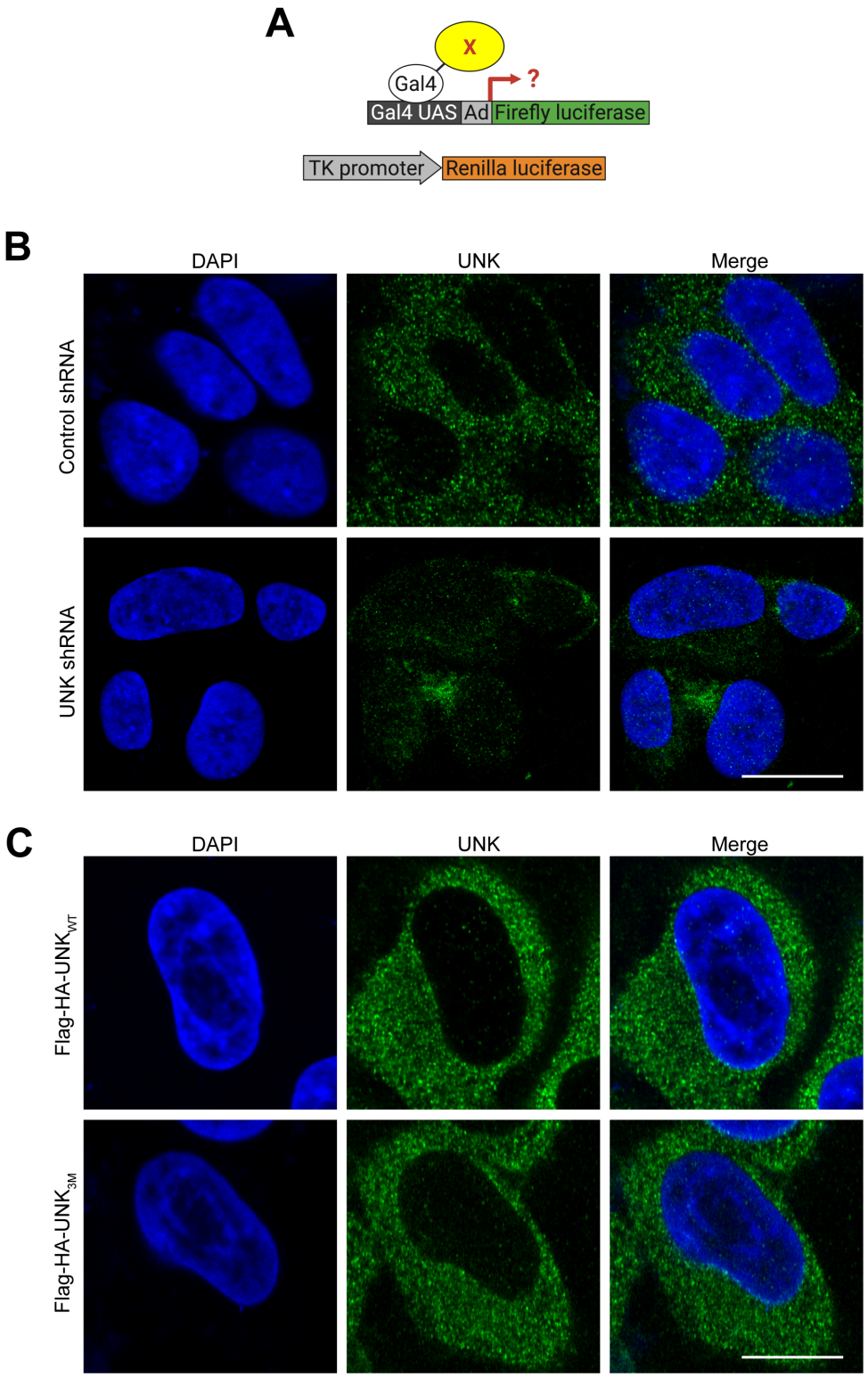


Figure S1. Transcriptional activity and intracellular localization of Unkempt, related to Figures 1 and 2.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(A) Principle of a dual luciferase reporter assay used to detect transcriptional activity of UNK. A tested protein (X) is tagged with the Gal4 DNA-binding domain for recruitment via Gal4 upstream activating sequences (Gal4 UAS) to the adenovirus major late promoter (Ad) driving minimal expression of the firefly luciferase reporter gene. The thymidine kinase (TK) promoter-driven Renilla luciferase serves as an internal control (Methods).
(B) Endogenous UNK detected by immunofluorescence using UNK-specific antibody in SH-SY5Y human neuroblastoma cells stably expressing either shRNA targeting luciferase (Control shRNA, top) or UNK (bottom). Scale bar, 10 μm.
(C) Ectopic Flag-HA-tagged UNK_{WT} or UNK_{3M} (D) detected by immunofluorescence using an HA-specific antibody in inducible HeLa cells treated with doxycycline for 24 h. Scale bar, 10 μm.
DAPI was used to visualize nuclei. Confocal images shown in (B) and (C) are representative of $n \geq 3$ experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A

Gene Symbol	MW (kDa)	Total peptide count			
		uninduced	WT	3M	dPAM2
CNOT1	266.77	0	80	0	54
CNOT10	82.26	0	9	0	3
CNOT3	81.82	0	6	0	3
CNOT7	32.72	0	5	0	6
CNOT2	59.7	0	4	0	6
CNOT11	55.18	0	2	0	2
CNOT8	33.52	0	2	0	0
CNOT9	33.61	0	2	0	1
PABPC1	70.63	4	127	105	0
PABPC4	70.74	0	45	32	0

B

```
LIG_PAM2_1: ..[LFP][NS][PIVTAFL].A..(([FY].[PYLF])|(W..)).

                          495        506
                           |          |
     H.sapiens  UNKEMPT    PGMNANALPFYP
     M.musculus Unkempt    PGMNANALPFYP
     H.sapiens  PAIP1      SKLSVNAPEFYP
     H.sapiens  PAIP2      SNLNPNAKEFVP
                           . :. **  * *
```

C

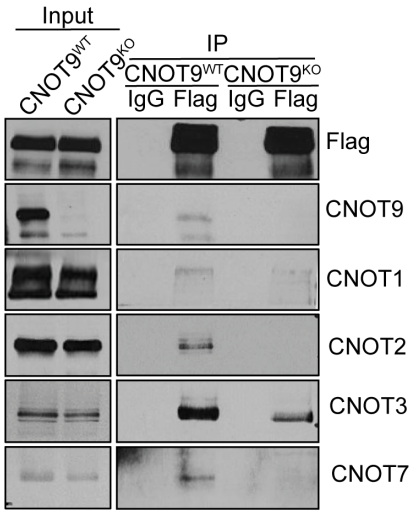


Figure S2. Analyses of interactions between Unkempt and its effectors, related to Figure 3.
(A) Total peptide counts detected by mass spectrometry analysis of tandem affinity-purified protein complexes prepared from uninduced (mock) or UNK_{WT}, UNK_{3M}, or UNK_{dPAM2} - expressing cells (Methods).
(B) Identification of the PAM2-like motif in UNK. A search for either of the two PAM2 motifs annotated in the ELM database (LIG_PAM2_1 and LIG_PAM2_2; http://elm.eu.org/) identified a 12-residue PAM2-like peptide sequence that deviates from the canonical LIG_PAM2_1 motif (regular expression shown on top) at one position, highlighted in red. PAM2 motifs of PABP-interacting proteins 1 and 2 (PAIP1 and PAIP2) are shown for comparison[41,95].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(C) Co-IP of endogenous CCR4-NOT subunits with Flag-HA-tagged UNK_{WT} from *CNOT9*^{WT} or *CNOT9*^{KO} HeLa cells. Precipitated proteins were detected by western blot analysis using the indicated antibodies (n = 2).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

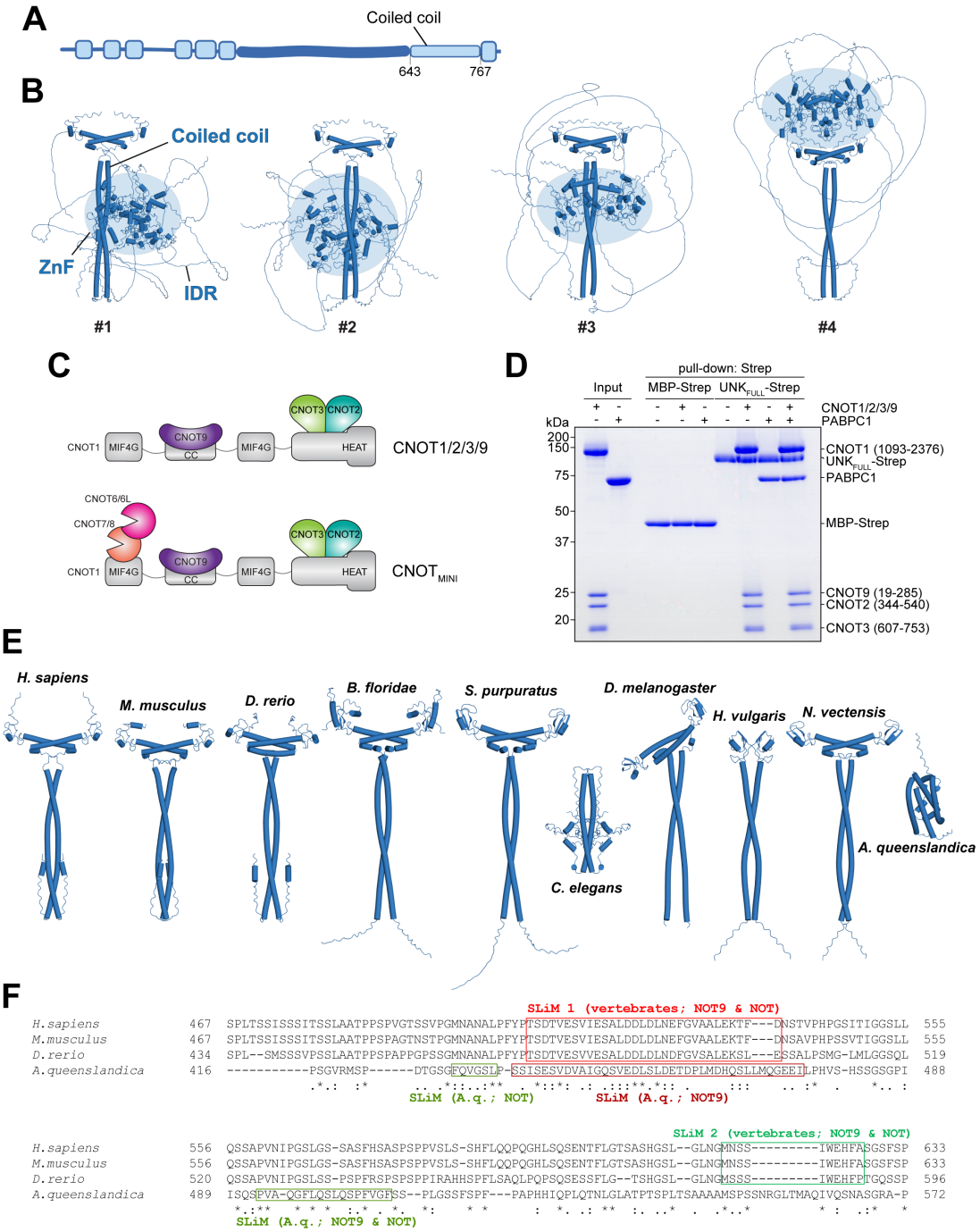
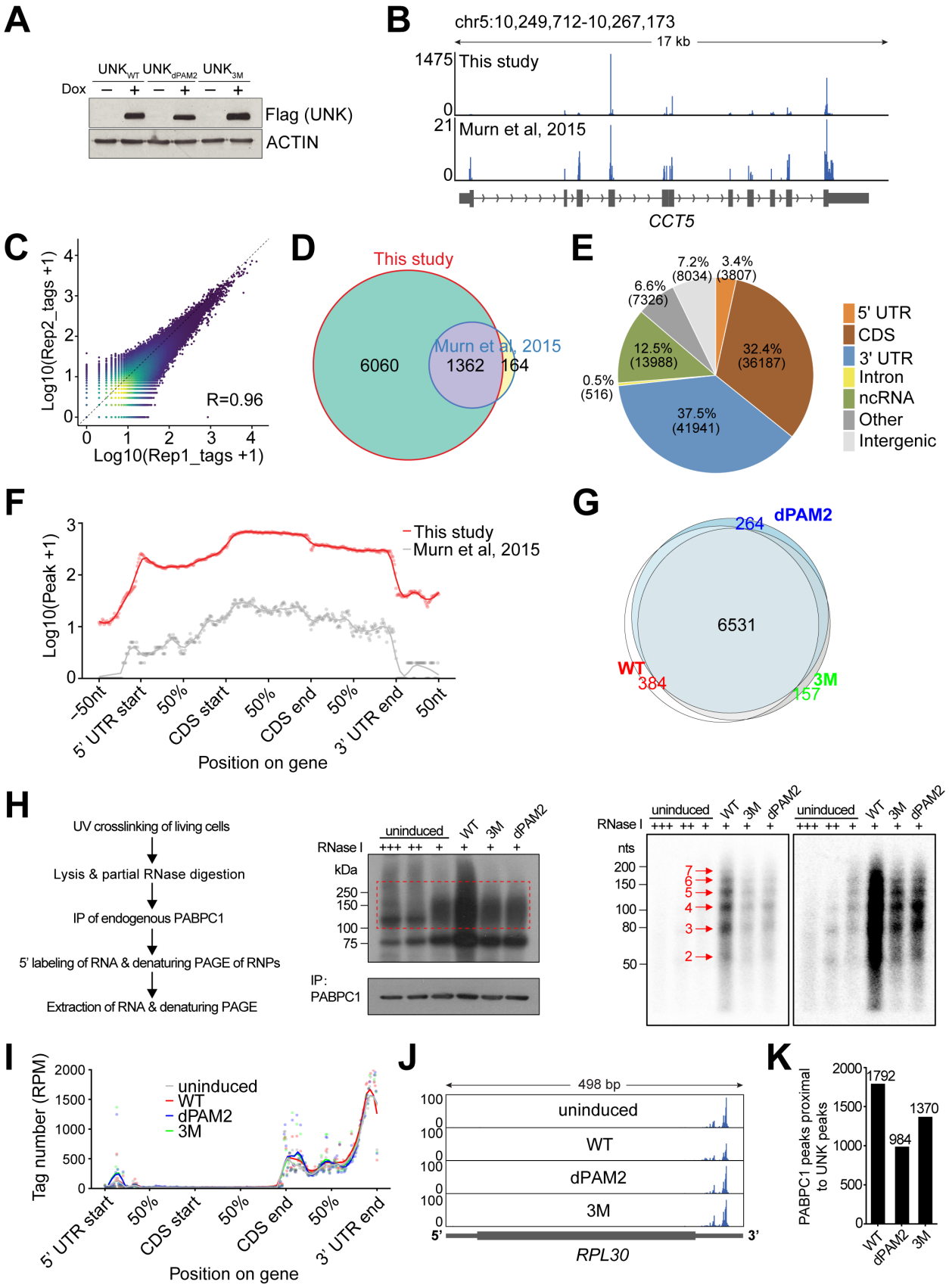


Figure S3. Structural and functional insights into the Unkempt-effector interface, related to Figures 4 and 5.
(A) Map of UNK protein indicating the location of the predicted coiled coil.
(B) Four AlphaFold predictions of the homodimer full-length mouse UNK protein showing the central coiled coil stabilizing the dimer with the positioning of the zinc fingers (ZnF, blue sphere) and the unstructured IDR being entirely stochastic with respect to the coil.
(C) Schematic representation of the reconstituted CNOT1/2/3/9 (top) and CCR4-NOT$_{MINI}$ (bottom) complexes. The rationale for using the larger, six-subunit CCR4-NOT$_{MINI}$ complex instead of CNOT1/2/3/9 for the mass photometry experiment shown in Figure 5D was to avoid signal overlap (the molecular weight of CNOT1/2/3/9 is similar to UNK dimer).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(D) Coomassie-stained polyacrylamide gel of an in vitro pull-down assay with MBP-Strep alone or recombinant UNK_{FULL}-Strep after incubation with PABPC1 and/or CNOT1/2/3/9, as indicated (n = 3).
(E) A representative AlphaFold prediction of the UNK coiled coil homodimer for each of the shown UNK orthologs (*Homo sapiens*, human; *Mus musculus*, house mouse; *Danio rerio*, zebrafish; *Branchiostoma floridae*, Florida lancelet; *Stronglycentrotus purpuratus*, sea urchin; *Caenorhabditis elegans*, roundworm; *Drosophila melanogaster*, fruit fly; *Hydra vulgaris*, fresh-water polyp; *Nematostella vectensis*, starlet sea anemone; *Amphimedon queenslandica*, sponge). See also Figure 5A.
(F) Sequence alignment of IDRs of the indicated UNK orthologs. Locations of SLiMs in contact with CCR4-NOT, as predicted by AlphaFold, are highlighted with red (matching SLiM 1 in mouse UNK) and green (matching SLiM 2 in mouse UNK) rectangles separately for vertebrates and *A. queenslandica* (*A.q.*), which harbors the evolutionarily most distant known UNK ortholog[17,54]. "NOT9" or "NOT" indicates that the SLiM is predicted to only contact one specific region in the respective module and "NOT9 & NOT" indicates repeated SLiM use.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S4. RNA binding by UNK_{WT}, UNK_{3M}, UNK_{dPAM2}, and PABPC1, related to Figure 6.

(A) Western blot analysis of HeLa cells inducibly expressing Flag-HA-tagged UNK_{WT}, UNK_{3M}, or UNK_{dPAM2} at 24 hours of treatment with Dox (n = 4).

(B) Genome browser view of the *CCT5* gene locus visualizing reproducibility of UNK_{WT} crosslink events between iCLIP replicates of this (top) and a previous (bottom) study[16].

(C) Density scatterplot comparing crosslink events per peak between replicates 1 and 2 of the UNK_{WT} iCLIP experiment. UNK_{3M} and UNK_{dPAM2} showed similar levels of reproducibility among iCLIP replicates (data now shown).

(D) Comparison of mRNA targets of UNK_{WT} in HeLa cells identified by iCLIP in this study and those reported previously[16].

(E) Distribution of UNK_{WT} iCLIP peaks among different RNA biotypes and mRNA segments determined in this study.

(F) Metatranscript analyses of the current (red) and previously reported (gray) iCLIP datasets showing the positional frequency of UNK_{WT} binding sites on mRNAs[16].

(G) Comparison of mRNA targets of UNK_{WT}, UNK_{3M}, and UNK_{dPAM2} identified by iCLIP. n = 4 iCLIP replicates were performed for each condition. See also Table S2 and Methods.

(H) Analysis of bulk interactions of PABPC1 with poly(A) tails. Labeled, partially RNase-digested PABPC1-bound RNA was extracted from cells expressing or not UNK_{WT}, UNK_{3M}, or UNK_{dPAM2} as outlined in the protocol on the left. Middle panels show autoradiogram of labeled PABPC1 RNP complexes (top) and the corresponding western blot of immunoprecipitated PABPC1 (bottom). Red dashed frame demarcates the area from which RNA was extracted. Two autoradiograms on the right are different exposures of the extracted RNA after resolution by denaturing PAGE. Red arrows point to footprints of 27-nt oligomers (red numbers) protected from RNase, matching the pattern of serial binding of PABPC1 to poly(A) tails[24,41]. Note that the periodic footprint pattern of PABPC1 is not perturbed by expression of UNK but is sensitive to the concentration of RNase I (plus signs) (n = 4).

(I)-(K) iCLIP analysis of PABPC1-RNA interactions.

(I) Proportional metatranscript analysis of iCLIP data showing the positional frequency of crosslink events for PABPC1 on different segments of mRNA in cells expressing UNK_{WT}, UNK_{3M}, or UNK_{dPAM2}, or in uninduced cells. Data points represent normalized crosslink events summarized over every percent of a given mRNA segment. CDS, coding sequence.

(J) Mapping PABPC1 iCLIP data onto UNK target transcripts. iCLIP data from four replicates were summed, normalized, and mapped to *RPL30* mRNA.

(K) Occurrence of PABPC1 iCLIP peaks in the vicinity (+/- 20 nts) of UNK_{WT}, UNK_{dPAM2}, or UNK_{3M} iCLIP peaks on mRNA. Note the reduction in the number of observed neighboring peaks upon disrupting UNK–PABPC interactions (UNK_{dPAM2}) and a weaker similar effect upon disrupting UNK–CCR4-NOT interactions (UNK_{3M}). See also Figure 6F.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

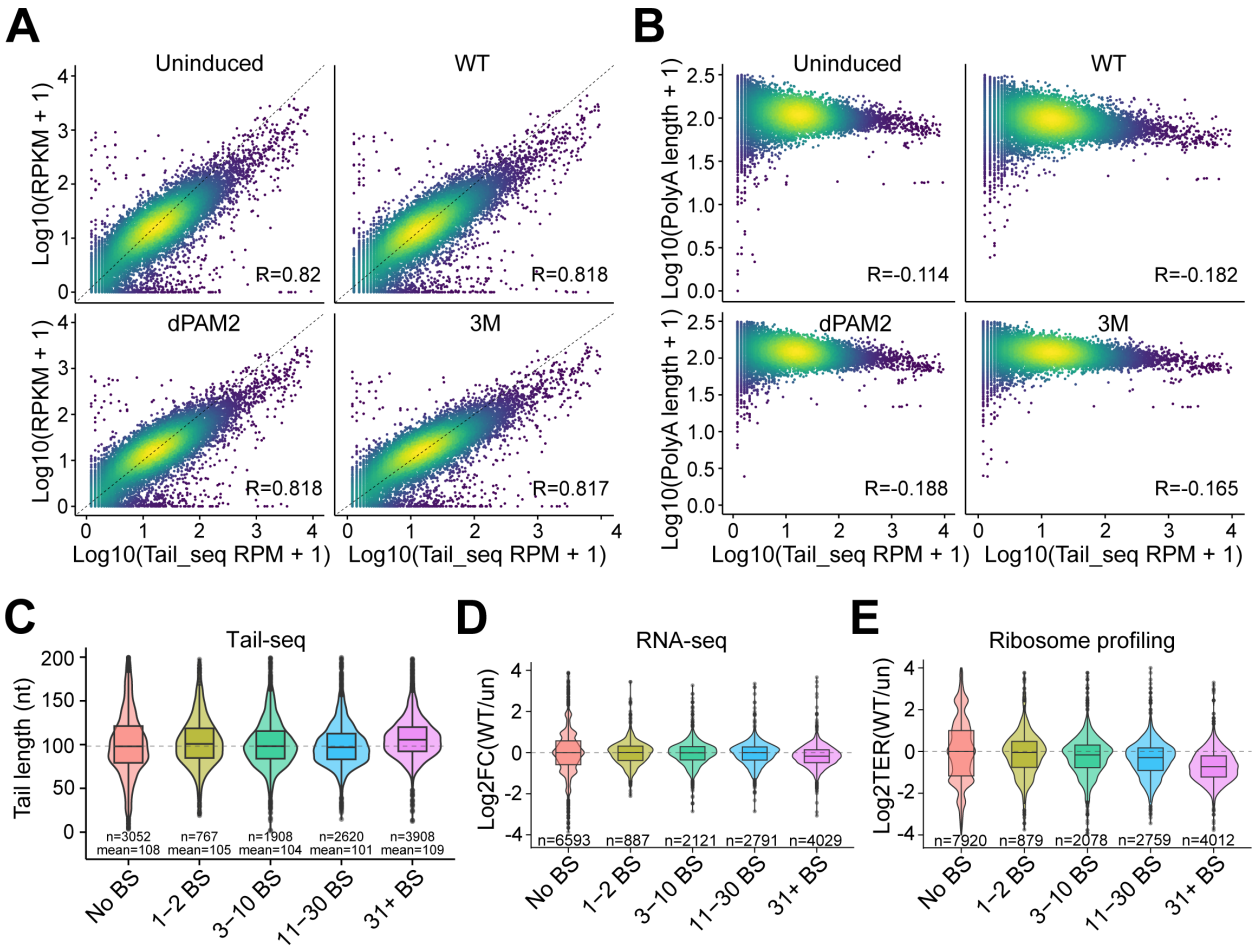


Figure S5. Impact of Unkempt on mRNA poly(A) tail length. Related to Figure 7.
(A) Scatterplots comparing log10-scaled transcript abundance determined by mRNA-seq (RPKM + 1) and Tail-seq (RPM + 1) analyses of each indicated sample, showing high correlation (n = 2).
(B) Scatterplots comparing log10-scaled transcript abundance (RPM + 1) and poly(A) tail length (value + 1), both determined by Tail-seq. Note the slight and expected tendency of highly expressed mRNAs to have relatively short poly(A) tails[43-45] (n = 2).
(C) Comparison of average poly(A) tail lengths of transcripts binned into groups according to the number of UNK binding sites (BS) per transcript. Data are shown for UNK_{WT}-expressing cells. Similar, minimal changes were observed for uninduced cells and cells expressing UNK_{3M} or UNK_{dPAM2} (data now shown). Numbers of transcripts (n) and mean lengths of their poly(A) tails are indicated for each group of transcripts (n = 2).
(D) RNA-seq results for transcripts binned into groups as in (C), showing changes in steady-state mRNA levels between UNK_{WT}-expressing and uninduced cells (un; n = 3). See also Figures 7A and 7D.
(E) Ribosome profiling data for transcripts binned into groups as in (C), showing changes in ribosome occupancy between UNK_{WT}-expressing and uninduced cells (un; n = 2). See also Figures 7E and 7H.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

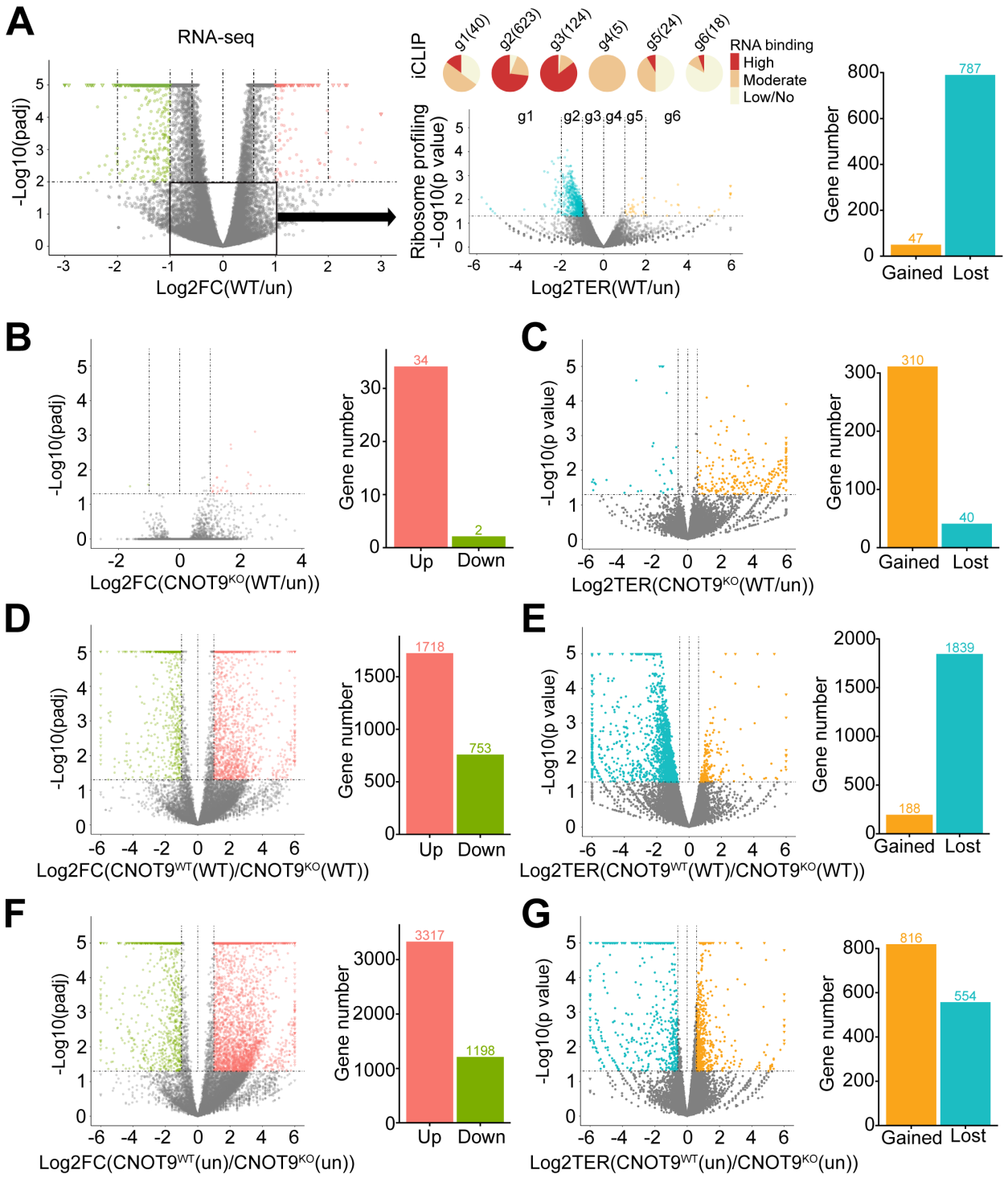


Figure S6. Reduced translational efficiency mediated by Unkempt. Related to Figure 7.
(A) Translational repression of mRNAs that show little change in abundance upon expression of UNK. mRNAs with small (< 2-fold) and insignificant (adjusted p > 0.01) changes in steady-state levels comparing UNK_{WT}-expressing to uninduced cells (group of mRNAs indicated by a black frame in the left plot) were analyzed for changes in ribosome occupancy (middle plot). Pie charts above the middle plot show RNA-binding information and transcript numbers in

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

each group of significantly regulated transcripts (g1-g6), as in Figures 7E-7G. Bar chart on the right indicates total numbers of transcripts with gained or lost ribosome occupancy ($p < 0.05$) ($n = 2$). See also Figures 7A, 7E, and 7H. (B)-(G) Reliance of translational repression by UNK on CNOT9. Three pairs of samples are compared by RNA-seq and ribosome profiling to document this reliance as well as the effect of CNOT9 alone: (B, C) CNOT9 KO cells ($CNOT9^{KO}$) expressing or not UNK_{WT}, (D, E) uninduced (un) CNOT9 WT ($CNOT9^{WT}$) and $CNOT9^{KO}$ cells, and (F, G) CNOT9 WT and CNOT9 KO cells expressing UNK_{WT}. Significantly regulated transcripts with twofold or larger changes in abundance (RNA-seq) or ribosome occupancy (ribosome profiling) are highlighted in color. Bar charts in panels (C), (E), and (G) indicate total numbers of transcripts with gained or lost ribosome occupancy ($p < 0.05$). ($n = 3$ for RNA-seq; $n = 2$ for ribosome profiling).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

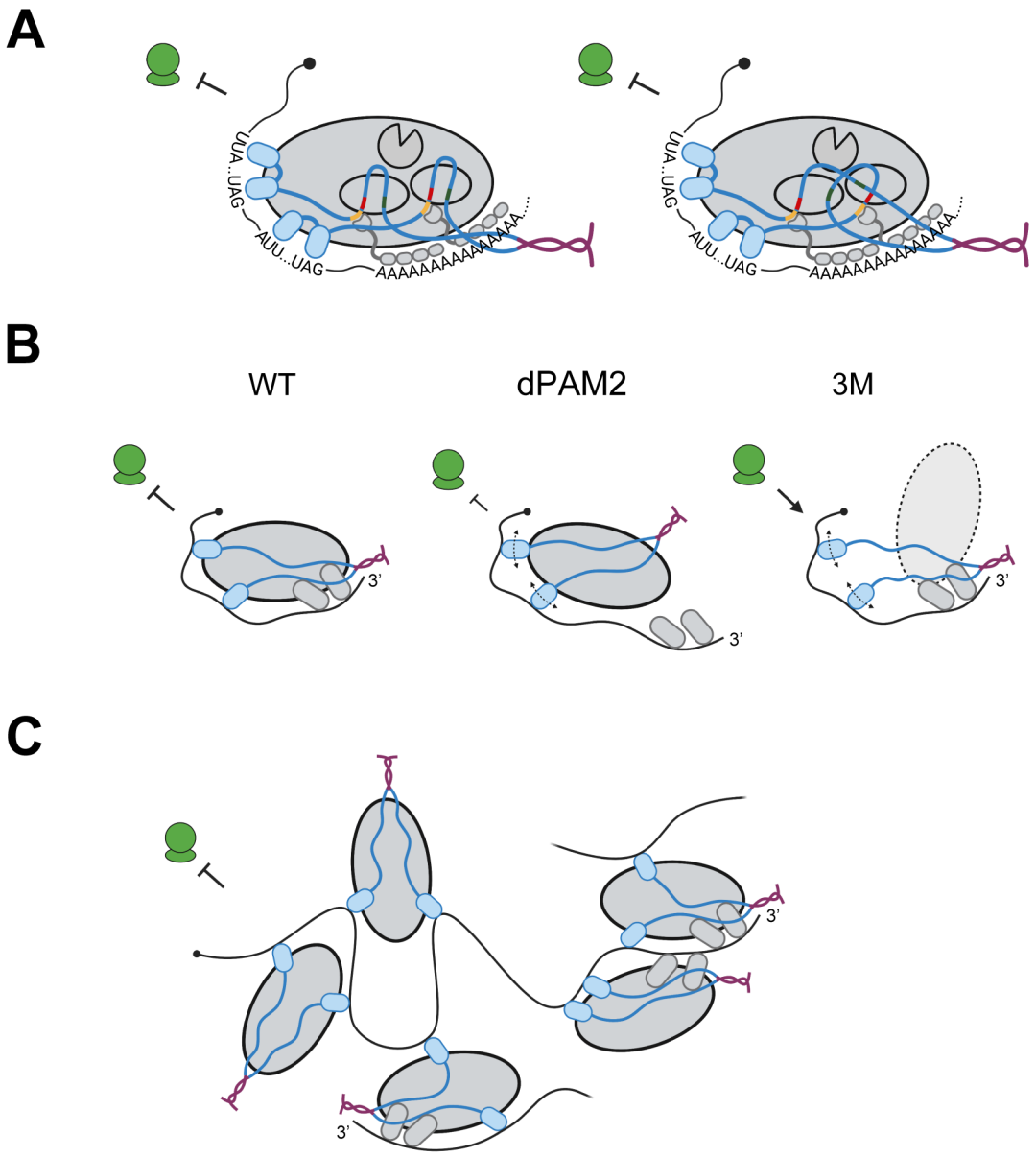


Figure S7. PABPC1-RNA interactions and models of Unkempt-effector interface. Related to Figure 7.
(A)-(C) Models of UNK-effector interface.
(A) Two possible modes of interaction between the IDR-embedded SLiMs of UNK and the NOT9 and NOT modules of CCR4-NOT. See also Figure 7J.
(B) Models illustrating how disrupting the interactions between UNK and PABPC (UNK_{dPAM2}) or CCR4-NOT (UNK_{3M}) might affect RNP organization and function. Curved dashed lines with arrowheads on each end indicate compromised RNA sequence recognition by UNK. Inhibitory arrows pointing to the ribosome (green shapes) or the arrow pointing to the RNP indicate translational repression or lack thereof, respectively.
(C) A larger RNP particle formed via association of multiple copies of the UNK dimer–CCR4-NOT–PABPC complex with mRNA.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTAL TABLES

Table S1. List of UNK mutants used in this study.

Table S2. iCLIP targets of UNK_{WT}, UNK_{3M}, and UNK_{dPAM2} in HeLa cells.

Table S3. PABPC1 iCLIP targets in uninduced HeLa cells and in HeLa cells expressing UNK_{WT}, UNK_{3M}, or UNK_{dPAM2}.

Table S4. RNA-seq analyses of uninduced HeLa cells and of HeLa cells expressing UNK_{WT}, UNK_{3M}, or UNK_{dPAM2}.

Table S5. Poly(A) tail length analyses of uninduced HeLa cells and of HeLa cells expressing UNK_{WT}, UNK_{3M}, or UNK_{dPAM2}.

Table S6. Ribosome profiling analyses of uninduced HeLa cells and of HeLa cells expressing UNK_{WT}, UNK_{3M}, or UNK_{dPAM2}.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1 Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**, 75-87, doi:10.1038/nrg2673 (2010).

2 He, S., Valkov, E., Cheloufi, S. & Murn, J. The nexus between RNA-binding proteins and their effectors. *Nat Rev Genet* **24**, 276-294, doi:10.1038/s41576-022-00550-0 (2023).

3 Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* **18**, 437-451, doi:10.1038/nrm.2017.27 (2017).

4 Carpenter, S., Ricci, E. P., Mercier, B. C., Moore, M. J. & Fitzgerald, K. A. Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* **14**, 361-376, doi:10.1038/nri3682 (2014).

5 Udagawa, T. *et al.* Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol Cell* **47**, 253-266, doi:10.1016/j.molcel.2012.05.016 (2012).

6 Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-543, doi:10.1038/nrg2111 (2007).

7 Erkelenz, S. *et al.* Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* **19**, 96-102, doi:10.1261/rna.037044.112 (2013).

8 Kadyrova, L. Y., Habara, Y., Lee, T. H. & Wharton, R. P. Translational control of maternal Cyclin B mRNA by Nanos in the Drosophila germline. *Development* **134**, 1519-1527, doi:10.1242/dev.002212 (2007).

9 Bhandari, D., Raisch, T., Weichenrieder, O., Jonas, S. & Izaurralde, E. Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes Dev* **28**, 888-901, doi:10.1101/gad.237289.113 (2014).

10 Duchaine, T. F. & Fabian, M. R. Mechanistic Insights into MicroRNA-Mediated Gene Silencing. *Cold Spring Harb Perspect Biol* **11**, doi:10.1101/cshperspect.a032771 (2019).

11 Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869-881, doi:10.1016/j.cell.2013.04.016 (2013).

12 Du, H. *et al.* YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat Commun* **7**, 12626, doi:10.1038/ncomms12626 (2016).

13 Witten, J. T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet* **27**, 89-97, doi:10.1016/j.tig.2010.12.001 (2011).

14 Hildebrandt, A. *et al.* The RNA-binding ubiquitin ligase MKRN1 functions in ribosome-associated quality control of poly(A) translation. *Genome Biol* **20**, 216, doi:10.1186/s13059-019-1814-0 (2019).

15 Van Roey, K. *et al.* Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* **114**, 6733-6778, doi:10.1021/cr400585q (2014).

16 Murn, J. *et al.* Control of a neuronal morphology program by an RNA-binding zinc finger protein, Unkempt. *Genes Dev* **29**, 501-512, doi:10.1101/gad.258483.115 (2015).

17 Murn, J., Teplova, M., Zarnack, K., Shi, Y. & Patel, D. J. Recognition of distinct RNA motifs by the clustered CCCH zinc fingers of neuronal protein Unkempt. *Nat Struct Mol Biol* **23**, 16-23, doi:10.1038/nsmb.3140 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18 Baskaran, P. *et al.* Phosphorylation of the novel mTOR substrate Unkempt regulates cellular morphogenesis. *J Biol Chem* **299**, 102788, doi:10.1016/j.jbc.2022.102788 (2023).

19 Avet-Rochex, A. *et al.* Unkempt is negatively regulated by mTOR and uncouples neuronal differentiation from growth control. *PLoS Genet* **10**, e1004624, doi:10.1371/journal.pgen.1004624 (2014).

20 Maierbrugger, K. T., Sousa-Nunes, R. & Bateman, J. M. The mTOR pathway component Unkempt regulates neural stem cell and neural progenitor cell cycle in the Drosophila central nervous system. *Dev Biol* **461**, 55-65, doi:10.1016/j.ydbio.2020.01.006 (2020).

21 Vinsland, E. *et al.* The zinc finger/RING domain protein Unkempt regulates cognitive flexibility. *Sci Rep* **11**, 16299, doi:10.1038/s41598-021-95286-y (2021).

22 Mohler, J. *et al.* The embryonically active gene, unkempt, of Drosophila encodes a Cys3His finger protein. *Genetics* **131**, 377-388, doi:10.1093/genetics/131.2.377 (1992).

23 Youn, J. Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol Cell* **69**, 517-532 e511, doi:10.1016/j.molcel.2017.12.020 (2018).

24 Yi, H. *et al.* PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol Cell* **70**, 1081-1088 e1085, doi:10.1016/j.molcel.2018.05.009 (2018).

25 Passmore, L. A. & Coller, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* **23**, 93-106, doi:10.1038/s41580-021-00417-y (2022).

26 Xie, J., Kozlov, G. & Gehring, K. The "tale" of poly(A) binding protein: the MLLE domain and PAM2-containing proteins. *Biochim Biophys Acta* **1839**, 1062-1068, doi:10.1016/j.bbagrm.2014.08.001 (2014).

27 Kumar, M. *et al.* The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res* **50**, D497-D508, doi:10.1093/nar/gkab975 (2022).

28 Raisch, T. *et al.* Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation. *Nat Commun* **10**, 3173, doi:10.1038/s41467-019-11094-z (2019).

29 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

30 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034, doi:10.1101/2021.10.04.463034 (2022).

31 Bonneau, F. *et al.* Nuclear mRNPs are compact particles packaged with a network of proteins promoting RNA-RNA interactions. *Genes Dev* **37**, 505-517, doi:10.1101/gad.350630.123 (2023).

32 Raisch, T. & Valkov, E. Regulation of the multisubunit CCR4-NOT deadenylase in the initiation of mRNA degradation. *Curr Opin Struct Biol* **77**, 102460, doi:10.1016/j.sbi.2022.102460 (2022).

33 Chen, Y. *et al.* A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol Cell* **54**, 737-750, doi:10.1016/j.molcel.2014.03.034 (2014).

34 Boland, A. *et al.* Structure and assembly of the NOT module of the human CCR4-NOT complex. *Nat Struct Mol Biol* **20**, 1289-1297, doi:10.1038/nsmb.2681 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35 Taliaferro, J. M. *et al.* RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Mol Cell* **64**, 294-306, doi:10.1016/j.molcel.2016.08.035 (2016).
36 Buchbender, A. *et al.* Improved library preparation with the new iCLIP2 protocol. *Methods* **178**, 33-48, doi:10.1016/j.ymeth.2019.10.003 (2020).
37 Sohrabi-Jahromi, S. *et al.* Transcriptome maps of general eukaryotic RNA degradation factors. *Elife* **8**, doi:10.7554/eLife.47040 (2019).
38 Kini, H. K., Silverman, I. M., Ji, X., Gregory, B. D. & Liebhaber, S. A. Cytoplasmic poly(A) binding protein-1 binds to genomically encoded sequences within mammalian mRNAs. *RNA* **22**, 61-74, doi:10.1261/rna.053447.115 (2016).
39 Webster, M. W., Stowell, J. A. & Passmore, L. A. RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* **8**, doi:10.7554/eLife.40670 (2019).
40 Gorlach, M., Burd, C. G. & Dreyfuss, G. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res* **211**, 400-407, doi:10.1006/excr.1994.1104 (1994).
41 Khaleghpour, K. *et al.* Translational repression by a novel partner of human poly(A) binding protein, Paip2. *Mol Cell* **7**, 205-216, doi:10.1016/s1097-2765(01)00168-x (2001).
42 Zekri, L., Kuzuoglu-Ozturk, D. & Izaurralde, E. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *EMBO J* **32**, 1052-1065, doi:10.1038/emboj.2013.44 (2013).
43 Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66-71, doi:10.1038/nature13007 (2014).
44 Lima, S. A. *et al.* Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* **24**, 1057-1063, doi:10.1038/nsmb.3499 (2017).
45 Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* **16**, 1297-1305, doi:10.1038/s41592-019-0617-2 (2019).
46 Li, X., Quon, G., Lipshitz, H. D. & Morris, Q. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* **16**, 1096-1107, doi:10.1261/rna.2017210 (2010).
47 Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-466, doi:10.1016/j.cell.2012.12.023 (2013).
48 Moretti, F., Kaiser, C., Zdanowicz-Specht, A. & Hentze, M. W. PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. *Nat Struct Mol Biol* **19**, 603-608, doi:10.1038/nsmb.2309 (2012).
49 Sgromo, A. *et al.* A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat Commun* **8**, 14307, doi:10.1038/ncomms14307 (2017).
50 Sgromo, A. *et al.* Drosophila Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-NOT complex to elicit repression of mRNA targets. *RNA* **24**, 381-395, doi:10.1261/rna.064584.117 (2018).
51 Poetz, F. *et al.* RNF219 attenuates global mRNA decay through inhibition of CCR4-NOT complex-mediated deadenylation. *Nat Commun* **12**, 7175, doi:10.1038/s41467-021-27471-6 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52 Keskeny, C. *et al.* A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4-NOT complex. *Genes Dev* **33**, 236-252, doi:10.1101/gad.320952.118 (2019).

53 Bulbrook, D. *et al.* Tryptophan-Mediated Interactions between Tristetraprolin and the CNOT9 Subunit Are Required for CCR4-NOT Deadenylase Complex Recruitment. *J Mol Biol* **430**, 722-736, doi:10.1016/j.jmb.2017.12.018 (2018).

54 Srivastava, M. *et al.* The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720-726, doi:10.1038/nature09201 (2010).

55 Lai, F., Zhou, Y., Luo, X., Fox, J. & King, M. L. Nanos1 functions as a translational repressor in the Xenopus germline. *Mech Dev* **128**, 153-163, doi:10.1016/j.mod.2010.12.001 (2011).

56 Hanet, A. *et al.* HELZ directly interacts with CCR4-NOT and causes decay of bound mRNAs. *Life Sci Alliance* **2**, doi:10.26508/lsa.201900405 (2019).

57 Cooke, A., Prigge, A. & Wickens, M. Translational repression by deadenylases. *J Biol Chem* **285**, 28506-28513, doi:10.1074/jbc.M110.150763 (2010).

58 Chekulaeva, M. *et al.* miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol* **18**, 1218-1226, doi:10.1038/nsmb.2166 (2011).

59 Collart, M. A. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **7**, 438-454, doi:10.1002/wrna.1332 (2016).

60 Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953, doi:10.1016/j.cell.2004.12.012 (2004).

61 Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493-501, doi:10.1261/rna.2192803 (2003).

62 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

63 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670, doi:10.1021/ac026117i (2003).

64 Peng, J. & Gygi, S. P. Proteomics: the move to mixtures. *J Mass Spectrom* **36**, 1083-1091, doi:10.1002/jms.229 (2001).

65 Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989, doi:10.1016/1044-0305(94)80016-2 (1994).

66 Weissmann, F. *et al.* biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc Natl Acad Sci U S A* **113**, E2564-2569, doi:10.1073/pnas.1604935113 (2016).

67 Weissmann, F. & Peters, J. M. Expressing Multi-subunit Complexes Using biGBac. *Methods Mol Biol* **1764**, 329-343, doi:10.1007/978-1-4939-7759-8_21 (2018).

68 Riepe, C. *et al.* Double stranded DNA breaks and genome editing trigger loss of ribosomal protein RPS27A. *FEBS J* **289**, 3101-3114, doi:10.1111/febs.16321 (2022).

69 McGlincy, N. J. & Ingolia, N. T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112-129, doi:10.1016/j.ymeth.2017.05.028 (2017).

70 Busch, A., Bruggemann, M., Ebersberger, S. & Zarnack, K. iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. *Methods* **178**, 49-62, doi:10.1016/j.ymeth.2019.11.008 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

71 Roehr, J. T., Dieterich, C. & Reinert, K. Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* **33**, 2941-2942, doi:10.1093/bioinformatics/btx330 (2017).
72 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* **1**, 895-905, doi:10.3390/biology1030895 (2012).
73 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
74 Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491-499, doi:10.1101/gr.209601.116 (2017).
75 Krakau, S., Richard, H. & Marsico, A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol* **18**, 240, doi:10.1186/s13059-017-1364-2 (2017).
76 Frankish, A. *et al.* Gencode 2021. *Nucleic Acids Res* **49**, D916-D923, doi:10.1093/nar/gkaa1087 (2021).
77 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
78 Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-1920, doi:10.1093/bioinformatics/bts277 (2012).
79 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
80 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
81 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
82 Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47, doi:10.1093/nar/gkz114 (2019).
83 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
84 Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572-4574, doi:10.1093/bioinformatics/btab705 (2021).
85 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
86 Yamashita, A. *et al.* Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol* **12**, 1054-1063, doi:10.1038/nsmb1016 (2005).
87 Diebold, M. L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* **175**, 178-188, doi:10.1016/j.jsb.2011.02.001 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.20.558714; this version posted September 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88 Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098-1101, doi:10.1126/science.1106148 (2005).

89 Krause, M. *et al.* tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA* **25**, 1229-1241, doi:10.1261/rna.071332.119 (2019).

90 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

91 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

92 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

93 Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841-1842, doi:10.1093/bioinformatics/btp328 (2009).

94 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

95 Craig, A. W., Haghighat, A., Yu, A. T. & Sonenberg, N. Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* **392**, 520-523, doi:10.1038/33198 (1998).