mRNA 3′UTRs chaperone intrinsically disordered regions to control protein activity

Yang Luo[1], Yaofeng Zhong[1,2], Sudipto Basu[1], and Christine Mayr[1,2†]

[1]Cancer Biology and Genetics Program, Sloan Kettering Institute, New York, NY 10065, USA

[2]Biochemistry, Cell, and Molecular Biology (BCMB) Allied Program, Weill Cornell Graduate School of Medical Sciences, New York, NY 10065, USA

[†]Correspondence: mayrc@mskcc.org

Abstract

More than 2,700 human mRNA 3′UTRs have hundreds of highly conserved (HC) nucleotides, but their biological roles are unclear. Here, we show that mRNAs with HC 3′UTRs mostly encode proteins with long intrinsically disordered regions (IDRs), including MYC, UTX, and JMJD3. These proteins are only fully active when translated from mRNA templates that include their 3′UTRs, raising the possibility of functional interactions between 3′UTRs and IDRs. Rather than affecting protein abundance or localization, we find that HC 3′UTRs control transcriptional or histone demethylase activity through co-translationally determined protein oligomerization states that are kinetically stable. 3′UTR-dependent changes in protein folding require mRNA-IDR interactions, suggesting that mRNAs act as IDR chaperones. These mRNAs are multivalent, a biophysical RNA feature that enables their translation in network-like condensates, which provide favorable folding environments for proteins with long IDRs. These data indicate that the coding sequence is insufficient for the biogenesis of biologically active conformations of IDR-containing proteins and that RNA can catalyze protein folding.

Introduction

Messenger RNAs (mRNAs) are the templates for protein synthesis which makes them central components of gene expression regulation. Each mRNA contains a coding sequence (CDS), 5′ untranslated region (UTR), and 3′UTR. Transcriptome-wide mRNA 3′ end sequencing revealed that overall, CDS and 3′UTRs both contribute nearly 50% of mRNA sequence[1,2]. Whereas the CDS is translated into the protein sequence, the biological roles of 3′UTRs are less clear.

Individual 3′UTRs have been shown to regulate mRNA stability, mRNA localization, and protein localization[3-7]. Moreover, 3′UTRs play important roles in cytoplasmic organization[8-11]. The view on cytoplasmic organization has substantially evolved within the past few years by the discovery that the cytoplasm is compartmentalized not only by membrane-enclosed organelles, but also by cytoplasmic condensates. Importantly, several research groups – by using diverse experimental approaches – found that mRNAs do not localize randomly within the cytoplasm but rather segregate into membrane-associated, condensate-associated, and unbound cytosolic mRNAs[10,12-17]. In a few cases, the 3′UTR-determined location of protein translation has been shown to control the functions of the newly synthesized proteins[7,8,18]. However, it is currently unclear, whether translation of an mRNA in cytoplasmic condensates versus the cytosol has broader biological consequences.

Most protein functions are not achieved alone and roughly two thirds of human enzymes can form multimers[19]. Cells use multiple strategies to establish interactions between correct partners, ranging from complex assembly during protein synthesis to post-translational

pathways that require covalent modifications or assembly factors[20-24]. Recently, a co-translational pathway that promotes formation of kinetically stable homodimers was discovered, where two neighboring nascent chains translated from the same mRNA on neighboring ribosomes co-fold. Ribosome profiling suggests that more than 3,000 proteins have structural features that make them susceptible for co-translational homodimerization[22,23]. Interestingly, many of the resulting homodimers evolved to be kinetically stable, which is necessary for their downstream functions, as quality control mechanisms probe complex shape and stability to eliminate unstable co-translationally formed dimers[24,25]. How co-translational pathways are regulated is currently unclear.

Intrinsically disordered regions (IDRs) are protein segments with a lower fraction of hydrophobic amino acids than folded domains[26]. IDRs are strongly enriched in regulatory proteins, including transcription factors and chromatin regulators; thus, it is of interest to understand how IDRs influence protein function[27]. However, the sequence-structure-function relationships remain poorly characterized, as IDRs do not adopt a single stable conformation in isolation but sample a wide array of distinct conformational states[28,29]. Nevertheless, two thirds of IDRs adopt defined secondary structures when bound to their partners, which can be other proteins or RNA[28,30-34]. The latter pathway could be widespread, supported by the recent discovery of prevalent RNA-IDR interactions in cells[35,36]. In contrast, most folded proteins, except canonical RNA-binding proteins, do not bind to RNA with high affinity[35]. The biological significance of widespread RNA-IDR interactions in cells is largely unclear.

Here, we show that inclusion of the highly conserved (HC) 3′UTRs in mRNA templates encoding MYC, UTX, and JMJD3 was required for full transcriptional or histone demethylase activity of the encoded proteins in the nucleus. The HC 3′UTRs control activity of IDR-containing proteins by changing the co-translationally determined protein oligomerization states. mRNAs with HC 3′UTRs are multivalent, which allows their translation in cytoplasmic network-like condensates, suggesting that these condensates act as folding environments for a large group of IDR-containing proteins. For HC 3′UTRs to alter protein oligomerization states requires interaction between multivalent 3′UTRs and IDRs, suggesting that multivalent mRNAs act as chaperones for long IDRs to promote protein conformations that are biologically active. Taken together, we show that protein activity of transcriptional regulators can be determined co-translationally, which requires IDR chaperone activity provided by HC 3′UTRs in the mRNA template.

Results

mRNAs with HC 3′UTRs encode proteins with long IDRs

We previously performed transcriptome-wide mapping of 3′UTRs and observed that 3′UTRs make up nearly 50% of human mRNA sequence (Fig. S1A)[1,2]. Moreover, we noticed that many mRNAs have 3′UTR sequences that are markedly conserved across vertebrates, including human and zebrafish (Fig. 1A). To systematically identify HC human 3′UTRs, we used phylogenetic *P* values (PhyloP scores), obtained from sequence conservation analysis across 100 vertebrate species, including birds, reptiles and fish[37,38]. This analysis revealed that 43% of 3′UTRs do not have any conserved nucleotides; however, 2,714 3′UTRs have more than 150 conserved nucleotides (Fig. 1B, Table S1). We define the latter group as HC 3′UTRs. We observed that mRNAs with non-conserved 3′UTRs mostly encode small, well-folded proteins, whereas mRNAs with HC 3′UTRs encode proteins with long IDRs (Fig. 1C, S1B). IDRs were determined using IUPred2A[39]. We obtained similar results when using AlphaFold predicted local distance difference test (pLDDT) scores for IDR identification (Fig. S1C)[40]. Gene ontology analysis showed that proteins encoded by mRNAs with HC 3′UTRs are substantially enriched in

chromatin regulators and transcription factors (Fig. 1D)[41], consistent with previous observations that these factors are overrepresented among proteins with long IDRs[27,42].

Including 3′UTRs in mRNA templates increases protein activity of transcriptional regulators

To investigate the biological roles of HC 3′UTRs, we transiently expressed the transcription factor MYC and two histone demethylases, UTX and JMJD3, encoded by the *KDM6A* and *KDM6B* genes, from cDNA expression vectors that either contained or lacked their corresponding full-length 3′UTRs within the vector construct (Fig. 1E). We add the suffix -NU (no 3′UTR) to denote vectors that only contain the CDS and added -U to indicate vectors that contain both the CDS and corresponding 3′UTR. The three 3′UTRs contained between 151 and 956 conserved nucleotides (Fig. 1A, 1F) and the three proteins have large regions with low confidence scores in structural predictions by AlphaFold, indicative of IDRs (Fig. 1G). We observed that protein abundance (Fig. 1H) and subcellular localization (Fig. S1D, S1E) of the transcriptional regulators were unaffected when they were expressed in the presence or absence of their 3′UTRs.

We then investigated activity of the proteins expressed from U or NU constructs. We expressed *MYC-U* or *MYC-NU* in U2OS cells, where endogenous MYC is low, and performed RNA-seq (Fig. 1I). Compared to the GFP-only control, the vector containing *MYC-U* induced or repressed nearly twice as many genes as the *MYC-NU* vector (Fig. 1I, Table S2). This indicates that inclusion of the 3′UTR in the mRNA template increases MYC protein activity in the nucleus in a manner that is independent of protein abundance. Similar results were obtained for the H3K27me3 demethylase UTX/KDM6A. UTX cooperates with transcription and signaling factors to control gene expression programs[43,44]. As UTX is expressed in MCF7 cells, we knocked down endogenous *KDM6A* and re-expressed HA-tagged UTX protein, followed by RNA-seq analysis (Fig. 1J, Fig. S1E, S1F). To assess 3′UTR-dependent UTX protein activity, we investigated the genes that significantly increased or decreased upon *KDM6A* knockdown and where re-expression of *KDM6A-NU* or *KDM6A-U* reversed at least partially the knockdown effect (Fig. 1J, Table S3). We detected nearly twice as many genes that reversed the knockdown effect when UTX protein was re-expressed from a vector that included the *KDM6A* 3′UTR compared to expression of *KDM6A-NU* (Fig. 1J). These results indicate that inclusion of the *KDM6A* 3′UTR in the mRNA template increases UTX activity in the nucleus in an abundance-independent manner. Taken together, these data indicate that for these two IDR-containing proteins with HC 3′UTRs, despite their protein abundance being similar with or without their 3′UTR, their functional activity was increased by including their HC 3′UTR. These data indicated an unexpected functional role for protein activity regulation by vectors with the HC 3′UTR.

Generation of a fully active JMJD3 histone demethylase requires the *KDM6B* 3′UTR

Given these results, we set out to measure 3′UTR-dependent activity more directly. JMJD3/KDM6B is an H3K27me3 demethylase with very low expression in HeLa cells[45,46]. We ectopically expressed *HA-KDM6B-U* or *-NU* and measured H3K27me3 fluorescence intensity in each cell using immunofluorescence (IF) staining. A reduction in H3K27me3 signal correlates with JMJD3 demethylase activity. In parallel, we quantified protein abundance of transfected HA-tagged JMJD3 using HA signal intensity (Fig. 1K): HA intensity was similar between cells expressing *KDM6B-U* or *-NU* (Fig. 1L). Cells transfected with the CDS-only construct showed a modest reduction in methylation level with a median decrease of less than 10%. By contrast, cells transfected with the vector containing the *KDM6B* 3′UTR showed a substantial reduction of

H3K27me3 staining, with a median decrease of 50% compared to untransfected cells (Fig. 1M). This comparison represents a five-fold difference in median demethylation efficiency, indicating that the JMJD3 demethylase is more active when translated from the *KDM6B-U* construct. These findings indicate that the CDS alone is insufficient for the biogenesis of a fully active JMJD3 demethylase.

JMJD3 IDR deletion makes the 3′UTR dispensable for high demethylase activity

Given these results, we investigated the molecular mechanism by which the *KDM6B* 3′UTR impacts JMJD3 protein activity. JMJD3 contains a long N-terminal IDR of over 1,000 residues and a structured C-terminus (Fig. 1G). Crystal structures of the C-terminus revealed a JmjC domain that folds into a characteristic β-barrel structure[47]. To date, nearly all studies on JMJD3 activity have used a truncated construct that excludes the long N-terminal IDR and the *KDM6B* 3′UTR[45,48-50]. We call the IDR-deleted protein IDRdel (Fig. 2A, 2B).

Repeating the IF assay with the IDRdel constructs showed that expression of only the structured enzymatic domain generates an active JMJD3 demethylase, resulting in a median reduction of 40-50% in H3K27me3 staining, which is comparable to the full-length protein encoded by the mRNA with its HC 3′UTR. Importantly, biogenesis of a functional protein that only contains the structured domain does not require the *KDM6B* 3′UTR in the construct (Fig. 2C-E). This experiment demonstrates a functional link between the JMJD3 IDR domain of the protein and the *KDM6B* 3′UTR in the mRNA: when both components are present or when both components are absent, a fully active protein is generated. This finding indicates that the JMJD3 IDR contains an autoinhibitory activity whose effect is overcome by the 3′UTR in the mRNA template.

IDR-dependent JMJD3 localization to chromatin is not regulated by the *KDM6B* 3′UTR

Chromatin modifiers are typically regulated by altered localization to genomic targets or by conformational changes that tune catalytic activity[51,52]. As 3′UTRs in mRNA templates can control the localization of the encoded proteins[4,5,7], we investigated whether the *KDM6B* 3′UTR directs JMJD3 protein to specific gene loci. We expressed our various HA-tagged KDM6B constructs (Fig. 1E, 2B) and performed Chromatin immunoprecipitation and sequencing (ChIP-seq). The HA signal indicating JMJD3 localization to chromatin revealed reduced binding to specific gene loci with the IDRdel protein (JMJD3 IDR deletion) compared to the full-length protein (Fig. 2F, 2G). The *KDM6B* 3′UTR, however, had no effect on subnuclear localization of full-length JMJD3 (Fig. 2F, 2G), thus the extent of protein localization to chromatin is regulated by the IDR, but not by the *KDM6B* 3′UTR.

In parallel, we performed H3K27me3 ChIP-seq to investigate the effects of IDR and mRNA 3′UTR on genome-wide demethylase activity of JMJD3 (Fig. 2H). The ChIP-seq results confirmed the previous IF results, as we detected full JMJD3 demethylase activity either when the IDR was deleted or when the full-length protein was expressed from a vector that included the *KDM6B* 3′UTR (Fig. 2I). In contrast, expressing the full-length protein from a vector lacking the 3′UTR resulted in a 50% reduction in demethylase activity (Fig. 2I). This finding indicates that the IDR domain inhibits the activity of the demethylase, if the *KDM6B* 3′UTR is lacking from the mRNA template. Taken together, these results demonstrate that the JMJD3 IDR has at least two independent functions: one is the regulation of protein localization to chromatin; the other is autoinhibition of JMJD3 demethylase activity. Our findings indicate that only one of them—namely IDR-dependent enzymatic activity—is regulated by the *KDM6B* 3′UTR.

The *KDM6B* 3′UTR increases the fraction of active JMJD3 protein molecules in cells

To obtain deeper insights into 3′UTR-dependent demethylase activity regulation, we plotted the IF data at single cell resolution to uncover 3′UTR-dependent differences in the relationship between JMJD3 protein abundance and activity. In cells expressing *KDM6B-NU*, demethylase activity mostly correlated with JMJD3 expression levels, as the largest fraction of cells with active demethylase had high protein abundance (Fig. 3A, 3B). For cells expressing *KDM6B-U*, JMJD3 reached similar maximum protein expression and again, most cells with high protein abundance showed high demethylase activity. Strikingly, however, for cells expressing *KDM6B-U*, there was on average a three-fold increase in the fraction of cells with low JMJD3 protein levels and low H3K27me3 staining, indicative of an active demethylase (Fig. 3A, 3B). These results suggest that the presence of the *KDM6B* 3′UTR increases the fraction of active JMJD3 molecules at low protein levels, pointing to a regulatory step during co-translational protein folding.

The *KDM6B* 3′UTR in the mRNA template suppresses JMJD3 oligomerization

These results raised the possibility that the *KDM6B* 3′UTR regulates IDR-dependent JMJD3 protein conformations. To obtain evidence for 3′UTR-dependent protein conformations, we needed a protein conformation assay that could be performed in cells and used on proteins with long IDRs. Chemical crosslinking-western blot analysis fulfils these criteria. After expressing *HA-KDM6B-U* or *-NU*, we thus performed Disuccinimidyl Glutarate (DSG) crosslinking in live cells, followed by western blot analysis using anti-HA (Fig. 3C).

Crosslinking of JMJD3 protein translated from the CDS-only vector produced smeared high-molecular weight products, indicative of multiple oligomeric states driven by the long IDR. The absence of a discrete ladder of oligomeric products suggests that multiple sites along the IDR are subject to crosslinking, giving different shapes with different electrophoretic mobilities. By contrast, for JMJD3 expressed from the vector that contained the *KDM6B* 3′UTR, we observed more monomeric protein (Fig. 3D). Importantly, the 3′UTR did not influence oligomerization of JMJD3 lacking the IDR: IDRdel was mostly monomeric, and the oligomerization pattern was not altered in a 3′UTR-dependent manner (Fig. 3E). These results demonstrate that the *KDM6B* 3′UTR in the mRNA template decreases JMJD3 oligomerization in cells in a manner dependent on the IDR domain of the protein.

The β-hairpin region in the IDR acts as an autoinhibitory JMJD3 dimerization domain

Given these data, we pursued identification of an oligomerization domain within the JMJD3 IDR. According to AlphaFold3, a region in the IDR predicted to adopt a β-hairpin (HP) conformation binds to the β-barrel of the structured enzymatic domain of another JMJD3 molecule through domain swapping and β-sheet addition (Fig. 3F, Fig. S2A)[53,54]. Intriguingly, within the JMJD3 IDR, the HP region has the highest sequence conservation and is conserved across vertebrates, including zebrafish (Fig. 3G). We tested the importance of the HP region for protein oligomerization using DSG crosslinking. As expression of *HA-KDM6B-NU* predominantly resulted in oligomerization, we deleted the HP region (HPdel) in the *NU* construct and examined the resulting oligomerization. The HPdel construct resulted in an increased fraction of monomeric JMJD3 protein (Fig. 3H). This experiment thus identified the β-HP region in the IDR as a major dimerization domain of JMJD3.

Combining the findings of JMJD3 oligomerization and demethylation indicates that protein demethylase activity negatively correlates with JMJD3 oligomerization. Monomeric JMJD3 protein appears to be active and can be generated through IDR deletion or endogenous 3′UTR inclusion during expression of full-length protein (Fig. 1M, 2E, 3D, 3E). Based on this model, HPdel should have high demethylase activity in cells, because deletion of the HP domain increases monomeric JMJD3 (Fig. 3H). We confirmed this prediction, as *HPdel-NU* generates a protein with high demethylase activity—as high as observed upon deletion of the JMJD3 IDR (Fig. 3I-K, S2B). These data indicate that the β-HP region is a dimerization and autoinhibitory domain within the IDR.

mRNA-IDR interaction is necessary for full JMJD3 histone demethylase activity in cells

Our data indicate that the IDR increases JMJD3 oligomerization when translated from a CDS-only construct. We next asked, how does the 3′UTR suppress JMJD3 oligomerization and inactivation? As the 3′UTR is an RNA region, we searched for an RNA-binding domain (RBD) within the IDR. Just upstream of the β-HP, we observed a region with the highest positive net charge within the entire protein, suggestive of a potential RBD (Fig. 4A). We expressed this RBD region as an MBP-tagged recombinant protein and tested its affinity for the Cy5-labeled *KDM6B* 3′UTR by using a fluorescence polarization assay (Fig. S3A, S3B). This domain bound the *KDM6B* 3′UTR with a Kd of 107 nM at physiological salt concentration, which is comparable to the RNA binding affinity of canonical RNA-binding proteins (Fig. 4B)[36].

To then test whether the RBD in the IDR mediates the effects of the *KDM6B* 3′UTR on protein oligomerization and demethylase activity, we deleted the RBD (RBDdel). Indeed, the *RBDdel-U* generated less monomeric protein and strongly decreased demethylase activity in cells (Fig. 4C-E, S3C). These results indicate that the mRNA-IDR interaction is necessary to release the IDR from the autoinhibited state. This is further supported by our prior observation that the full-length protein containing the RBD is largely inactive when expressed from a CDS-only vector (see Fig. 1M). Together, these results support that the JMJD3 oligomerization status controls JMJD3 demethylase activity, and that IDR-dependent oligomerization is suppressed by mRNA-IDR interaction, which appears to be the crucial regulatory event enabled by the 3′UTR.

mRNA-IDR binding decreases co-translational nascent chain dimerization

As our data indicated that the oligomeric state of JMJD3 protein determines its activity and that the regulation occurs through the *KDM6B* 3′UTR, these findings suggested that the protein oligomeric state is determined co-translationally. To address this, we established a ‘co-translational disome assay’ (Fig. 5A). Dimerization of nascent chains from neighboring ribosomes form pairs of ribosomes, called disomes. Disomes are connected in cell lysates even upon RNA digestion, whereas non-associating nascent chains localize to monosomes[22,55,56]. Monosomes and disomes can be extracted from cells by sucrose gradient centrifugation. As disome formation is a read-out for co-translational homodimerization, we expressed *HA-KDM6B-U* or *-NU*, purified monosomes and disomes, and labeled nascent chains in each fraction with puromycin. Immunoprecipitation with an HA antibody followed by blotting for puromycin quantified the amount of HA-tagged JMJD3 nascent chains in the monosome and disome fractions (Fig. 5A). When expressing JMJD3 from a vector without the *KDM6B* 3′UTR (FL-NU), there were significantly more JMJD3 nascent chains in disomes compared with monosomes (Fig. 5B). By contrast, inclusion of the 3′UTR in the vector (FL-U) shifted the majority of JMJD3 nascent chains to the monosome fraction, indicative of decreased JMJD3

homodimerization (Fig. 5B, Fig. S3D). These results provide evidence that the 3′UTR in the mRNA template controls the extent of co-translational protein homodimerization of JMJD3.

To define co-translational functions of the characterized IDR domains, we also examined the impact of deletion of the HP or the RBD on co-translational JMJD3 homodimerization. Deletion of the HP decreased protein oligomerization, whereas deletion of the RBD increased JMJD3 oligomerization (Fig. 5B). These results demonstrate that JMJD3 nascent chains use the β-HP domain to co-translationally dimerize in the absence of the 3′UTR, whereas mRNA-IDR binding through the RBD suppresses IDR-mediated JMJD3 nascent chain dimerization when expressed from an mRNA that includes the 3′UTR (Fig. 5B).

3′UTR-IDR binding switches alternative JMJD3 homodimer conformations

Given these results on how the domains impact nascent chain dimerization, we pursued molecular mechanisms by which mRNA-IDR binding influences protein-protein interactions between the JMJD3 IDR and its structured enzymatic domain. To address this, we performed RNA binding and protein dimerization assays using recombinant protein domains (Fig. S3E). Fluorescence polarization assays showed that the HP region alone only weakly bound to RNA, whereas fusing the RBD to the HP, called RHP, increased RNA affinity by 500-fold (Fig. S3F-H). In contrast, the structured enzymatic domain (ED) showed no RNA binding at the tested concentrations (Fig. S3I).

We then investigated dimer interfaces using the thiol-thiol crosslinker bismaleimidoethane (BMOE), which covalently links sulfhydryl groups. We used native cysteines in the ED and mutated all three naturally occurring cysteines in the HP region while introducing a single cysteine in the vicinity of the β-sheet (Fig. 5C). The ED alone did not form crosslinked product and addition of RNA (*KDM6B* 3′UTR) had no effect, consistent with the lack of RNA binding (Fig. 5D, lanes 4 and 6, S3J). In the absence of RNA, mixing of ED and RHP generated equal amounts of RHP:RHP and RHP:ED dimers (Fig. 5D, lane 7, S3J). Addition of RNA to the mixture generated a 90:10 dimer ratio, by exclusively promoting homotypic dimerization (RHP:RHP dimer) without affecting the heterotypic interaction (RHP:ED dimer; Fig. 5D, lane 8, S3J). In contrast, repeating the assay with the HP region lacking the RBD abrogated the RNA-mediated increase in HP:HP dimerization (Fig. 5E, S3K), indicating that 3′UTR-IDR interaction near the β-HP region is necessary for the switch from heterotypic to homotypic interaction of the recombinant proteins.

Combining the in vitro data with the results from cells, we propose that the *KDM6B* 3′UTR promotes the generation of an active demethylase in two ways: it increases the fraction of monomeric JMJD3 (Fig. 3D) and, when dimers do form, it biases the formation of an active homodimer by promoting IDR-IDR interactions (Fig. 5D). In the absence of the *KDM6B* 3′UTR, co-translational dimers/oligomers are predominantly formed (Fig. 3D). Approximately half of the dimers generate an active histone demethylase through IDR-IDR interactions of two molecules, whereas our data suggest that the other half of dimers generate an inactive enzyme through IDR-ED interaction. The high fraction of inactive homodimers formed in the absence of the *KDM6B* 3′UTR decreases the overall enzymatic activity measured in cells (Fig. 5F). Moreover, our results suggest that long IDRs participate in protein folding (Fig. 3D, 3H, 4C). However, we want to emphasize that our results do not indicate that the entire IDRs are folded. Instead, we suggest that local IDR regions, such as the JMJD3 β-HP region are involved in protein folding, thus generating kinetically stable complexes that function in the nucleus, whereas long intervening IDR regions remain intrinsically disordered, a feature that allows the IDR-containing JMJD3 protein to localize to specific genomic regions, which is not affected by the presence of the *KDM6B* 3′UTR (Fig. 2G, 5F).

mRNAs with HC 3′UTRs are multivalent and enriched in network-like condensates

Our findings on 3′UTR-dependent changes in JMJD3 protein oligomerization raised the possibility that other mRNA transcripts with HC 3′UTR may have similar functions encoded within the 3′UTR sequence. Thus, to generalize this finding, we set out to identify the RNA features of HC 3′UTRs that are necessary to regulate co-translational oligomerization states of IDR-containing proteins. We noted that mRNAs containing HC 3′UTRs are substantially longer and have high values for mRNA valency (Fig. 6A, 6B). How do we define mRNA valency[9]?

We previously found that mRNA biophysical features dictate whether condensates adopt a droplet- or a network-like morphology—the latter being characteristic of cytoplasmic network-like condensates, such as TIS granules or the FXR1 network[8-11]. Using in vitro phase separation assays with multivalent RNA-binding proteins, we had observed that addition of short 3′UTRs did not change droplet morphology, but long 3′UTRs transformed droplets into network-like condensates[9]. Importantly, when we used 3′UTRs normalized for length- and GC-content, we identified the number of RNA-RNA interaction sites (or RNA valency) as the crucial parameter for determining the switch in condensate morphology from droplet- to network-like[9].

RNA-RNA interactions can occur through palindromic or low complexity sequences, but these types of RNA binding sites differ in affinity and accessibility[9,57-59]. We used our previous dataset to identify sequence features that could serve as surrogate markers for RNA-RNA interactions, which would allow us to predict multivalent RNAs from sequence[9]. Whereas the sum of G and C homopolymers (with at least five consecutive nucleotides) did not correlate, the sum of A and U homopolymers in 3′UTRs correlated strongly with the formation of in vitro condensate networks (Fig. S4A, S4B). This relationship was confirmed for mRNAs in cells, where again the sum of G/C homopolymers did not correlate, but mRNA length or the sum of A/U homopolymers correlated strongly with mRNA enrichment in cytoplasmic network-like condensates[10,11]. Based on these findings, we suggest that the sum of A/U homopolymers in an RNA can serve as a surrogate marker for RNA valency (Fig. 6C, 6D, S4C, Table S1).

mRNA valency correlates with mRNA length; however, even after comparing groups of mRNAs with similar length, mRNAs with HC 3′UTRs have significantly higher valencies than mRNAs with non-conserved 3′UTRs (Fig. S4D). Taken together, these findings indicate that HC 3′UTRs are strongly multivalent, which correlates with their mRNA enrichment in cytoplasmic condensates (Fig. 6A-D, S4E, S4F).

Multivalent 3′UTRs enable mRNA localization to network-like condensates

mRNA valency correlates with subcytoplasmic mRNA localization: whereas mRNAs classified to be enriched in cytoplasmic network-like condensates are multivalent, mRNAs classified as cytosolic have low valencies (Fig. 6D). However, mRNA valency is not equally distributed along the mRNA, and the majority is contributed by the 3′UTR (Fig. 6E). Based on these findings, we predicted that exclusion of multivalent 3′UTRs, such as the 3′UTRs of *MYC*, *KDM6A*, and *KDM6B* from cDNA expression vectors, will change the mRNA localization from network-like condensates to the cytosol (Fig. 6F). To test this model, we performed a previously established ‘digitonin RNA-FISH assay’ that distinguishes between endogenous cytosolic and condensate-enriched mRNAs that encode non-membrane proteins[10]. As digitonin permeabilizes the plasma membrane but leaves internal membranes intact[60], unbound cytosolic mRNAs are removed, but condensate-enriched mRNAs are largely retained in the cytoplasm[10].

This assay is also suitable for mRNAs derived from cDNA expression vectors. RNA valency of 3′UTR-containing *MYC*, *KDM6A*, and *KDM6B* mRNAs is at least two-fold higher compared to their 3′UTR-deleted counterparts (Fig. 6G). Digitonin treatment of transfected cells, followed by RNA-FISH showed that 3′UTR-containing mRNAs are largely retained in the cytoplasm, whereas deletion of their multivalent 3′UTRs substantially decreases their cytoplasmic mRNA retention (Fig. 6G-I, Fig. S5A). These results show that inclusion of multivalent HC 3′UTRs in cDNA expression vectors allows mRNA localization to cytoplasmic condensates, whereas 3′UTR deletion promotes their localization to the cytosol.

Multivalent 3′UTRs are required for the biogenesis of an active JMJD3 histone demethylase

Expression of JMJD3 from a vector that includes the *KDM6B* 3′UTR promotes the biogenesis of an active JMJD3 histone demethylase (Fig. 1M). To investigate whether this 3′UTR contains special elements for protein activity regulation or whether 3′UTR multivalency and subsequent mRNA translation in cytoplasmic network-like condensates is sufficient to generate an active JMJD3 protein, we performed 3′UTR swapping experiments. We selected two multivalent and two low-valency 3′UTRs, fused them to the KDM6B CDS, and assessed whether these 3′UTRs impact *KDM6B* mRNA localization to condensates or the cytosol (Fig. 6J). Digitonin RNA-FISH confirmed that multivalent 3′UTRs were sufficient for enrichment of the chimeric mRNAs in cytoplasmic condensates, whereas low-valency 3′UTRs predominantly localized the chimeric mRNAs to the cytosol (Fig. 6K, 6L, Fig. S5B).

Importantly, we observed that 3′UTR multivalency also correlated with JMJD3 histone demethylase activity. Expression of JMJD3 protein from vectors that contained multivalent 3′UTRs resulted in high demethylase activity, similar to that observed with the endogenous *KDM6B* 3′UTR. In contrast, JMJD3 protein produced from vectors containing low-valency 3′UTRs resulted in significantly lower JMJD3 activity, comparable to the activity observed with the CDS-only construct (Fig. 6M, 6N, S5C, S5D). These results indicate that biophysical features of the encoding mRNA – such as RNA valency – control JMJD3 demethylase activity in the nucleus. Taken together with the results on the molecular mechanism by which the *KDM6B* 3′UTR controls demethylase activity, we propose that multivalent mRNAs act as IDR chaperones that co-translationally control protein oligomerization states during protein folding, and thus impact the level of functional protein.

Proteins encoded by low-valency mRNAs fold in a 3′UTR-independent manner, whereas mRNAs with HC 3′UTRs modulate folding of IDR-containing proteins

Although mRNAs with HC 3′UTRs mostly encode proteins with long IDRs (Fig. 1C), mRNAs with non-conserved 3′UTRs also encode proteins with IDRs, but these IDRs tend to be shorter (Fig. 7A). To determine if HC 3′UTRs broadly impact co-translationally determined protein oligomerization states, we examined additional candidates. Our findings so far indicated that IDR-containing proteins encoded by multivalent mRNAs with HC 3′UTRs are prone to undergo co-translational oligomerization, which results in kinetically stable protein complexes. Importantly, these oligomerization states are altered by the presence of multivalent mRNAs in the folding environment, which is provided by network-like condensates. Network-like condensates are strongly enriched in multivalent mRNAs, allowing them to co-translationally interact with IDRs, thus acting as IDR chaperones (Fig. 7B). This model further predicts that proteins encoded by low-valency mRNAs, which are predominantly translated outside of network-like condensates, fold in a 3′UTR-independent manner (Fig. 6F, 7B).

To test these predictions, we chose candidate proteins to examine based on mRNA valency and 3′UTR conservation status. 3′UTR-dependent changes in protein oligomerization states are expected for mRNAs with HC 3′UTRs and a minimum mRNA valency of eight, whereas 3′UTR-independent protein oligomerization is expected for low-valency mRNAs with non-conserved 3′UTRs (Table S4). We expressed protein candidates from cDNA vectors that either contained or lacked the corresponding 3′UTRs and determined 3′UTR-dependent protein oligomerization states using in-cell chemical crosslinking followed by western blot analysis. For 6/6 proteins encoded by low-valency mRNAs, we confirmed 3′UTR-independent protein oligomerization states (Fig. 7C, 7D, S6A-D). This was the case for all tested proteins encoded by low valency mRNAs regardless of whether the proteins contained structured domains (Fig. 7C, S6A, S6B) or structured domains and IDRs longer than 300 aa (Fig. 7D, S6C, S6D).

Moreover, we validated 3′UTR-dependent protein oligomerization states in 3/3 IDR-containing proteins encoded by multivalent mRNAs and HC 3′UTRs (Fig. 7E, 7F, S6E). The presence of the *MYC* 3′UTR in the encoding mRNA increased monomeric MYC protein (Fig. S6E), whereas expression of UTX and FOXP2 from mRNA templates containing their 3′UTRs changes the observed pattern of protein oligomeric states (Fig. 7E, 7F). These findings indicate that multivalent 3′UTRs change the co-translationally determined protein conformations.

Discussion

mRNAs are known as the templates for protein synthesis. Whereas the CDS is sufficient for proper folding of small and well-folded proteins[61], here we show that proteins with long IDRs use multivalent mRNAs as chaperones to adopt biologically active conformations during protein biosynthesis. Our findings indicate that RNA can catalyze protein folding and suggest that mRNA biophysical features, such as RNA valency, which is predominantly contributed by HC 3′UTRs, make mRNAs capable of chaperoning IDRs. We show that protein activity of transcriptional regulators can be determined co-translationally, as local regions in long IDRs participate in protein folding, leading to the formation of kinetically stable protein oligomerization states.

HC 3′UTRs act as chaperones for IDRs during co-translational protein folding

The importance of 3′UTRs in mRNA templates that encode proteins with long IDRs is suggested by their extensive sequence conservation (Fig. 1A-C). HC 3′UTRs often are as conserved as their corresponding CDS, highlighting a critical role in protein biogenesis. Proteins with long IDRs are substantially enriched in transcription factors and chromatin regulators (Fig. 1D)[27,42,52]. Importantly, for all three transcriptional regulators tested, we found that the mRNA 3′UTRs are necessary for the generation of fully functional proteins through the control of IDR-dependent protein oligomerization states. This raises the possibility that these findings are a general principle for these types of encoded proteins. Moreover, we could show that it is not the specific sequence, but the valency of the 3′UTRs that matters, at least for the tested situations.

Although, oligomerization is widely used to regulate protein activity[62-64], the novelty here is that the activity of IDR-containing proteins is determined co-translationally through protein folding. The surface area over volume ratio is greater for IDRs than for globular proteins and therefore, IDRs have a greater effect on solvent entropy[65]. A potential model for the participation of long IDRs in protein folding is that IDR regions with high local hydrophobicity may interact with structured domains to be shielded from the solvent and for the protein to attain a thermodynamically favorable state[66]. The absence of IDR chaperones during protein synthesis may thermodynamically trap IDR regions in non-functional conformations, thus decreasing the

overall activity of the proteins. The presence of HC 3′UTRs, however, enables co-translational mRNA-IDR interactions to modulate the oligomeric states and to promote the formation biologically active protein conformations (Fig. 5F, 7B).

Co-translational homodimerization of proteins with long IDRs

Our findings show a new mode of protein activity regulation by protein oligomerization, which is suppressed by the presence of multivalent mRNAs in the folding environment. Compared to the well-understood post-translational oligomerization pathway[62-64], where proteins associate and disassociate after completion of protein synthesis, this newly discovered assembly mode occurs during protein synthesis and involves long IDRs.

Co-translational protein homodimerization is well recognized. It is a widespread mode of protein dimerization and has been observed for ~3,000 proteins[22,23]. In the studied cases, specific structural domains at the N-termini of proteins co-translationally bind to each other, which results in mutual stabilization. Co-translational homodimerization allows co-folding, thus leading to intertwining of subunits and the generation of stable homodimers[25]. For example, BTB domain homodimerization in E3 ligases is necessary for their specificity. To ensure the biogenesis of stable homodimers, specific quality control mechanisms have evolved that probe shape and complex stability to eliminate co-translationally formed dimers that are not entirely stable[24,25]. Taken together, the co-translational formation of stable homodimers is widespread and functionally relevant. Whereas previous studies focused on the structural determinants of co-translationally formed homodimers[23], we expand these studies to show that long IDRs use multivalent mRNA-IDR interactions during protein synthesis to adopt kinetically stable protein states, where local IDR regions appear stabilized by interaction with structured domains.

The role of multivalency in mRNAs acting as IDR chaperones

The molecular mechanism by which mRNAs act as IDR chaperones is currently unknown, but previously we found that multivalent mRNAs often have a high conformational heterogeneity, as determined by a high RNA ensemble diversity[9]. This feature could enable dynamic mRNA-IDR interactions, leading to mutual stabilization. This idea is supported by previous structural studies on RNA-IDR interactions, which found that usually both RNA and peptides are disordered in their unbound states and become structured upon binding[31,32,67,68]. For example, RNA-IDR interactions can induce the formation of turns in peptides, which is often the initiating event for β-HP formation[32,67,69]. RNA-IDR interactions can also induce IDR sequences to adopt β-HP structures[31,70], demonstrating that RNA-IDR interactions can change IDR conformations.

Although the previously reported RNA-IDR interactions occur post-translationally[31,32,35,36,67,68,71], our data suggest that co-translational mRNA-IDR interactions are widespread, allowing mRNAs to act as chaperones for long IDRs. It is likely that mRNAs cooperate with protein chaperones during protein folding, where protein chaperones recognize surface-exposed hydrophobic residues and RNAs scaffold long disordered regions, taking advantage of the much larger sizes of RNA molecules. More detailed understanding of the molecular details by which mRNAs function as chaperones for IDRs promises to provide novel opportunities for targeting the functions of IDRs.

mRNA multivalency enables translation in cytoplasmic condensates (Fig. 6), which seem to provide favorable protein folding environments for many IDR-containing proteins. Cytoplasmic condensates are enriched in various RNA-binding proteins[10,11], which may facilitate dynamic mRNA-protein interactions, thus increasing accessibility of multivalent mRNAs. As large

numbers of proteins seem to be simultaneously translated in cytoplasmic condensates[8,10], the condensate environment may further promote stabilization of long IDRs through heteromeric, co-translational interactions[4,5]. However, the IDR features that determine whether an IDR participates in co-translational homo- or hetero-oligomerization are currently unknown. Taken together, our findings support a biologically important role of cytoplasmic network-like condensates for protein folding of IDR-containing proteins, where multivalent mRNAs are crucial components of the local folding environment.

Role of the long JMJD3 IDR

Long IDRs are particularly enriched in transcriptional regulators[42]. Through IDR deletion experiments, we observed that the JMJD3 IDR is necessary for proper localization of the protein to specific genomic loci (Fig. 2G). This finding is consistent with previous reports that long IDRs with hundreds of residues are important for transcriptional regulator localization to specific genomic or chromatin regions[52].

The molecular mechanism by which long IDRs mediate their functions are currently not well understood, largely due to the difficulty in studying them[52]. Our results suggest that long IDRs participate in protein folding (Fig. 3D, 7E, 7F, S6E). However, we emphasize that our results do not indicate that the entire IDRs are folded. Instead, we suggest that local IDR regions, such as the JMJD3 β-HP region, are involved in protein folding, whereas other regions may remain intrinsically disordered (Fig. 7B). This model is supported by the observation that 3′UTR-regulated JMJD3 conformations did not alter chromatin localization of JMJD3, and therefore, are thought to still have accessible short linear motifs that form transient, dynamic, and multivalent interactions, which seem to be important for IDR-mediated chromatin localization (Fig. 2G)[52].

Moreover, IDRs are known to adopt different conformational states[28,29], and this IDR feature is preserved during protein folding as our in vitro results indicate at least two different modes of co-translational homodimerization with IDR-IDR interaction or IDR-ED interaction as seen in Fig. 5D and 5F.

3′UTR conservation identifies mRNA-chaperoned proteins

Our analysis shows that not all proteins require the presence of mRNA-mediated chaperone function. For example, fully folded proteins, including E3 ligase adaptors, metabolic enzymes, and proteins involved in translational control, including DHPS, RACK1, and EIF3I use co-translational homodimerization through their structured domains[22]. These proteins form constitutive dimers whose conformations are not influenced by their mRNA 3′UTRs, which are usually entirely non-conserved (Fig. 7C, S6A, S6B). Similarly, the conformations of proteins with long IDRs that are encoded by mRNAs with non-conserved 3′UTRs, including SSBP4, SSRP1, and SF3B4, are also not regulated by their 3′UTRs (Fig. 7D, S6C, S6D). Only IDR-containing proteins that are encoded by multivalent mRNAs with HC 3′UTRs change their protein oligomeric states in a 3′UTR-dependent manner (Fig. 7E, 7F, S6E). As currently the features of IDRs that participate in co-translational folding are unknown, our best way to predict proteins that require chaperone mRNAs is through 3′UTR conservation analysis.

Implications for ectopic gene expression studies

Our work has widespread and important implications for the study of proteins with large IDRs using vectors. Ectopic gene expression is a widely used tool to study protein functions in cells,

transgenic animals, and through recombinant proteins. In nearly all cases, viral or cDNA expression vectors are used that drive expression of the CDS only. This has mostly historical reasons as vector technology was developed in the 1970s, when it was unknown that approximately 50% of mRNA sequence is contributed by 3′UTRs[1,2,72]. Using CDS-only vectors for expression of IDR-containing proteins may generate only partially active proteins due to the lack of IDR chaperones. Moreover, a recombinantly generated protein is already in a 'post-translational state'. The lack of IDR chaperones during protein synthesis may thermodynamically trap these proteins in non-functional conformations, thus potentially generate misleading results.

Our study suggests that the currently accepted models on protein-protein interactions and protein complex assembly are too simplistic, as they are mostly based on experimental approaches that are blind towards co-translationally formed kinetically stable complexes. As 40% of the human proteome contains IDRs, new experimental approaches need to be established that include 3′UTRs in mRNAs and IDRs in proteins. Applying these new approaches will likely change our current view by which proteins find each other in the crowded cellular environment.

Limitations of the study

Although we mapped the JMJD3 oligomerization domain to the β-HP region of the IDR, we currently have no insights into the molecular details of the interaction. As we focused on the IDR, we did not map the interacting residues in the JMJD3 structured domain important for demethylase activity.

Although, we observe 3′UTR-dependent changes in protein oligomeric states for several proteins (Fig. 3D, 7E, 7F, S6E), the crosslinking western blot does not allow us to distinguish between homo- and hetero-oligomerization of the investigated proteins.

Only multivalent mRNAs act as IDR chaperones and these mRNAs are strongly enriched in cytoplasmic network-like condensates (Fig. 6N, 7E, 7F). However, it is currently not entirely clear if translation in cytoplasmic network-like condensates is required for IDR chaperone function. As several network-like condensates enrich multivalent mRNAs, it is impossible to simultaneously disrupt all network-like condensates in cells. Therefore, it is possible that a single multivalent 3′UTR could be sufficient to provide mRNA chaperone activity for the encoded protein.

Acknowledgements

This work was funded by a grant from the Pershing Square Foundation, William Ackman, and Neri Oxman, a grant from the G. Harold and Leila Y. Mathers Foundation, the NIH Director's Pioneer Award (DP1GM123454) and the NIH grant R35GM144046 to C.M. as well as the MSK Core Grant (P30 CA008748). We thank Mervin Fansler for early work on 3′UTR conservation and help with RNA-seq analysis, Ellen Horste for help with RNA-FISH, Chiara Evans for help with the co-translational disome assay, and Jiaxiang Ren for providing code for the analysis of immunofluorescence images. We thank Pierre-Jacques Hamard and Richard Koche from the Epigenetics Research Innovation Lab at MSKCC for performing and analyzing the ChIP-seq experiments. The ChIP-seq samples were sequenced at the MSKCC Integrated Genomics Operation core facility. We thank all members of the Mayr lab, especially Ting Cai and Ming-Chung Wu, as well as Thomas R. Cech (University of Colorado Boulder), Nancy M. Bonini (University of Pennsylvania), and Emmanuel Levy (University of Geneva) for critical reading of

the manuscript and helpful discussions. Author contributions: Y.L. performed all experiments, except in vitro RNA and protein production and the RNA binding and thiol-thiol crosslinking assays, which were performed and analyzed by Y.Z. S.B. performed all genome-wide analyses. Y.L. and C.M. conceived the project, designed the experiments, and wrote the manuscript with input from all authors.

Declaration of interest

C.M. is a member of the *Cell* advisory board.

References

1. Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev *27*, 2380-2396.
2. Fansler, M.M., Mitschka, S., and Mayr, C. (2024). Quantifying 3′UTR length from scRNA-seq data reveals changes independent of gene expression. Nature communications *15*, 4050.
3. Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673-684.
4. Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363-367.
5. Lee, S.H., and Mayr, C. (2019). Gain of Additional BIRC3 Protein Functions through 3'-UTR-Mediated Protein Complex Formation. Mol Cell *74*, 701-712 e709.
6. Mitschka, S., and Mayr, C. (2022). Context-specific regulation and function of mRNA alternative polyadenylation. Nature reviews. Molecular cell biology *23*, 779-796.
7. Gasparski, A.N., Moissoglu, K., Pallikkuth, S., Meydan, S., Guydosh, N.R., and Mili, S. (2023). mRNA location and translation rate determine protein targeting to dual destinations. Mol Cell *83*, 2726-2738.e2729.
8. Ma, W., and Mayr, C. (2018). A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions. Cell *175*, 1492-1506 e1419.
9. Ma, W., Zhen, G., Xie, W., and Mayr, C. (2021). In vivo reconstitution finds multivalent RNA-RNA interactions as drivers of mesh-like condensates. eLife *10*.
10. Horste, E.L., Fansler, M.M., Cai, T., Chen, X., Mitschka, S., Zhen, G., Lee, F.C.Y., Ule, J., and Mayr, C. (2023). Subcytoplasmic location of translation controls protein output. Mol Cell *83*, 4509-4523.e4511.
11. Chen, X., Fansler, M.M., Janjoš, U., Ule, J., and Mayr, C. (2024). The FXR1 network acts as a signaling scaffold for actomyosin remodeling. Cell *187*, 5048-5063.e5025.
12. Reid, D.W., and Nicchitta, C.V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J Biol Chem *287*, 5518-5527.

13. Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. Cell *178*, 473-490.e426.

14. Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc Natl Acad Sci U S A *116*, 19490-19499.

15. Villanueva, E., Smith, T., Pizzinga, M., Elzek, M., Queiroz, R.M.L., Harvey, R.F., Breckels, L.M., Crook, O.M., Monti, M., Dezi, V., Willis, A.E., and Lilley, K.S. (2024). System-wide analysis of RNA and protein subcellular localization dynamics. Nat Methods *21*, 60-71.

16. Keber, F.C., Nguyen, T., Mariossi, A., Brangwynne, C.P., and Wühr, M. (2024). Evidence for widespread cytoplasmic structuring into mesoscale condensates. Nature cell biology *26*, 346-352.

17. Boraas, L.C., Hu, M., Martino, P., Thornton, L., Vejnar, C.E., Zhen, G., Zeng, L., Parker, D.M., Cox, A.L., Giraldez, A.J., et al. (2025). G3BP1 ribonucleoprotein complexes regulate focal adhesion protein mobility and cell migration. Cell reports *44*, 115237.

18. Gasparski, A.N., Mason, D.E., Moissoglu, K., and Mili, S. (2022). Regulation and outcomes of localized RNA translation. Wiley interdisciplinary reviews. RNA *13*, e1721.

19. Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., and Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res *49*, D498-d508.

20. Duncan, C.D., and Mata, J. (2011). Widespread cotranslational formation of protein complexes. PLoS Genet *7*, e1002398.

21. Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., and Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. Nature *561*, 268-272.

22. Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., Auburger, J.J., Tans, S., Bukau, B., and Kramer, G. (2021). Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. Science *371*, 57-64.

23. Mallik, S., Venezian, J., Lobov, A., Heidenreich, M., Garcia-Seisdedos, H., Yeates, T.O., Shiber, A., and Levy, E.D. (2025). Structural determinants of co-translational protein complex assembly. Cell *188*, 764-777.e722.

24. Padovani, C., Jevtić, P., and Rapé, M. (2022). Quality control of protein complex composition. Mol Cell *82*, 1439-1450.

25. Mena, E.L., Jevtić, P., Greber, B.J., Gee, C.L., Lew, B.G., Akopian, D., Nogales, E., Kuriyan, J., and Rape, M. (2020). Structural basis for dimerization quality control. Nature *586*, 452-456.

26. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. Chemical reviews *114*, 6589-6631.

27. Cermakova, K., and Hodges, H.C. (2023). Interaction modules that impart specificity to disordered protein. Trends in biochemical sciences *48*, 477-490.

28. Davey, N.E. (2019). The functional importance of structure in unstructured protein regions. Current opinion in structural biology *56*, 155-163.

29. Holehouse, A.S., and Kragelund, B.B. (2024). The molecular basis for cellular function of intrinsically disordered protein regions. Nature reviews. Molecular cell biology *25*, 187-211.

30. Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. Current opinion in structural biology *19*, 31-38.

31. Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat Struct Mol Biol *18*, 680-687.

32. Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R.B., and Patel, D.J. (2011). Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. Nat Struct Mol Biol *18*, 796-804.

33. Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews. Molecular cell biology *16*, 18-29.

34. Zeke, A., Schád, É., Horváth, T., Abukhairan, R., Szabó, B., and Tantos, A. (2022). Deep structural insights into RNA-binding disordered protein regions. Wiley interdisciplinary reviews. RNA, e1714.

35. Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell *63*, 696-710.

36. Oksuz, O., Henninger, J.E., Warneford-Thomson, R., Zheng, M.M., Erb, H., Vancura, A., Overholt, K.J., Hawken, S.W., Banani, S.F., Lauman, R., et al. (2023). Transcription factors interact with RNA to regulate genes. Mol Cell *83*, 2449-2463.e2413.

37. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

38. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res *20*, 110-121.

39. Meszaros, B., Erdos, G., and Dosztanyi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res *46*, W329-W337.

40. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

41. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res *50*, W216-w221.

42. Minezaki, Y., Homma, K., Kinjo, A.R., and Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. Journal of molecular biology *359*, 1137-1149.

43. Dhar, S.S., Lee, S.H., Chen, K., Zhu, G., Oh, W., Allton, K., Gafni, O., Kim, Y.Z., Tomoiga, A.S., Barton, M.C., et al. (2016). An essential role for UTX in resolution and activation of bivalent promoters. Nucleic Acids Res *44*, 3659-3674.

44. Revia, S., Seretny, A., Wendler, L., Banito, A., Eckert, C., Breuer, K., Mayakonda, A., Lutsik, P., Evert, M., Ribback, S., et al. (2022). Histone H3K27 demethylase KDM6A is an epigenetic gatekeeper of mTORC1 signalling in cancer. Gut *71*, 1613-1628.

45. Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731-734.

46. Xiang, Y., Zhu, Z., Han, G., Lin, H., Xu, L., and Chen, C.D. (2007). JMJD3 is a histone H3K27 demethylase. Cell research *17*, 850-857.

47. Kruidenier, L., Chung, C.W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., et al. (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature *488*, 404-408.

48. De Santa, F., Totaro, M.G., Prosperini, E., Notarbartolo, S., Testa, G., and Natoli, G. (2007). The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell *130*, 1083-1094.

49. Miller, S.A., Mohn, S.E., and Weinmann, A.S. (2010). Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. Mol Cell *40*, 594-605.

50. Yang, M., Wang, C., Zhou, M., Bao, L., Wang, Y., Kumar, A., Xing, C., Luo, W., and Wang, Y. (2022). KDM6B promotes PARthanatos via suppression of O6-methylguanine DNA methyltransferase repair and sustained checkpoint response. Nucleic Acids Res *50*, 6313-6331.

51. Kartikasari, A.E., Zhou, J.X., Kanji, M.S., Chan, D.N., Sinha, A., Grapin-Botton, A., Magnuson, M.A., Lowry, W.E., and Bhushan, A. (2013). The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. Embo j *32*, 1393-1408.

52. Jonas, F., Navon, Y., and Barkai, N. (2025). Intrinsically disordered regions as facilitators of the transcription factor target search. Nature reviews. Genetics *26*, 424-435.

53. Remaut, H., and Waksman, G. (2006). Protein-protein interaction through beta-strand addition. Trends in biochemical sciences *31*, 436-444.

54. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493-500.

55. Zhao, T., Chen, Y.M., Li, Y., Wang, J., Chen, S., Gao, N., and Qian, W. (2021). Disome-seq reveals widespread ribosome collisions that promote cotranslational protein folding. Genome biology *22*, 16.

56. Aviner, R., Li, K.H., Frydman, J., and Andino, R. (2021). Cotranslational prolyl hydroxylation is essential for flavivirus biogenesis. Nature *596*, 558-564.

57. Bevilacqua, P.C., Williams, A.M., Chou, H.L., and Assmann, S.M. (2022). RNA multimerization as an organizing force for liquid-liquid phase separation. Rna *28*, 16-26.

58. Bose, M., Rankovic, B., Mahamid, J., and Ephrussi, A. (2024). An architectural role of specific RNA-RNA interactions in oskar granules. Nature cell biology *26*, 1934-1942.

59. Setti, A., Bini, G., Maiorca, V., Pellegrini, F., Proietti, G., Miltiadis-Vrachnos, D., Armaos, A., Martone, J., Monti, M., and Ruocco, G. (2025). Decoding RNA-RNA Interactions: The Role of Low-Complexity Repeats and a Deep Learning Framework for Sequence-Based Prediction. bioRxiv, 2025.2002. 2016.638500.

60. Lerner, R.S., Seiser, R.M., Zheng, T., Lager, P.J., Reedy, M.C., Keene, J.D., and Nicchitta, C.V. (2003). Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes. Rna *9*, 1123-1137.

61. Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223-230.

62. Marianayagam, N.J., Sunde, M., and Matthews, J.M. (2004). The power of two: protein dimerization in biology. Trends in biochemical sciences *29*, 618-625.

63. Selwood, T., and Jaffe, E.K. (2012). Dynamic dissociating homo-oligomers and the control of protein function. Arch Biochem Biophys *519*, 131-143.

64. Schweke, H., Pacesa, M., Levin, T., Goverde, C.A., Kumar, P., Duhoo, Y., Dornfeld, L.J., Dubreuil, B., Georgeon, S., Ovchinnikov, S., et al. (2024). An atlas of protein homo-oligomerization across domains of life. Cell *187*, 999-1010.e1015.

65. Watson, J.L., Seinkmane, E., Styles, C.T., Mihut, A., Krüger, L.K., McNally, K.E., Planelles-Herrero, V.J., Dudek, M., McCall, P.M., Barbiero, S., et al. (2023). Macromolecular condensation buffers intracellular water potential. Nature *623*, 842-852.

66. Dyson, H.J., Wright, P.E., and Scheraga, H.A. (2006). The role of hydrophobic interactions in initiation and propagation of protein folding. Proc Natl Acad Sci U S A *103*, 13057-13061.

67. Puglisi, J.D., Chen, L., Blanchard, S., and Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. Science *270*, 1200-1203.

68. Williamson, J.R. (2000). Induced fit in RNA-protein recognition. Nature structural biology *7*, 834-837.

69. Dyson, H.J., Sayre, J.R., Merutka, G., Shin, H.C., Lerner, R.A., and Wright, P.E. (1992). Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. Journal of molecular biology *226*, 819-835.

70. Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol *18*, 688-692.

71. Horos, R., Büscher, M., Kleinendorst, R., Alleaume, A.M., Tarafder, A.K., Schwarzl, T., Dziuba, D., Tischer, C., Zielonka, E.M., Adak, A., et al. (2019). The Small Non-coding Vault RNA1-1 Acts as a Riboregulator of Autophagy. Cell *176*, 1054-1067.e1012.

72. Nora, L.C., Westmann, C.A., Martins-Santana, L., Alves, L.F., Monteiro, L.M.O., Guazzaroni, M.E., and Silva-Rocha, R. (2019). The art of vector engineering: towards the construction of next-generation genetic tools. Microb Biotechnol *12*, 125-147.

73. Afroz, T., Hock, E.M., Ernst, P., Foglieni, C., Jambeau, M., Gilhespy, L.A.B., Laferriere, F., Maniecka, Z., Plückthun, A., Mittl, P., et al. (2017). Functional and dynamic

polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. Nature communications *8*, 45.

74. Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry, A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., et al. (2022). A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature *604*, 310-315.

75. Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O., and Pappu, R.V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophysical journal *112*, 16-21.

76. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein science : a publication of the Protein Society *32*, e4792.

77. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods *14*, 417-419.

78. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

79. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.

Figure legends

Figure 1. HC 3′UTRs regulate activity of transcription and chromatin regulators independently of protein abundance.

A. Sequence conservation of 100 vertebrate genomes, obtained from the UCSC PhyloP100way track is shown for the genomic context depicting the last two exons of the *KDM6A* gene.

B. Classification of human genes based on the number of conserved nucleotides (nt) in human 3′UTRs obtained from PhyloP. NC, non-conserved, 0 conserved nt (N = 7947), IM, intermediate, 4-150 nt (N = 7824), HC, highly conserved >150 nt (N = 2714). Kruskal-Wallis test, P = 0.

C. IDR length distributions of proteins encoded by mRNAs shown in (B). IDR if IUPreD2A score ≥ 0.33. aa, amino acid. Mann-Whitney test: P = 0.

D. Gene ontology analysis of mRNAs with HC 3′UTRs. The top six functional gene classes and Benjamini Hochberg-adjusted P values are shown.

E. cDNA vector constructs used for protein activity assays in cells, showing the tag, CDS (grey) and full-length 3′UTR (blue).

F. As in (A) but shown for *MYC* and *KDM6B*.

G. Prediction of structured domains (pLDDT ≥ 80, brown) and IDRs (pLDDT < 80, grey) using AlphaFold2 pLDDT scores. Names of annotated protein domains are given.

H. Western blot analysis after transfection of constructs shown in (E) into U2OS, MCF7, and HeLa cells. Actin, GAPDH, and Tubulin were used as loading controls.

I. Genes induced (red) or repressed (blue) upon transfection of GFP-MYC-U or GFP-MYC-NU compared to GFP control in U2OS cells shown as heatmap after clustering analysis. Shown are genes with log2FC > 0.3 and adjusted P value < 0.1. Differentially expressed genes were classified into genes with significant change only in NU vs GFP (NU only; N = 185), significant change only in U vs GFP (U only, N = 753), and significant change in NU vs GFP and U vs GFP (both, N = 686). TPM, transcripts per million.

J. As in (I) but using MCF7 cells with shRNA-mediated knockdown of endogenous *KDM6A,* which were subsequently transfected with HA-KDM6A-NU or HA-KDM6A-U. Shown are genes significantly up- or downregulated by *KDM6A* knockdown relative to control. Subsequent clustering analysis was performed separately on the two groups showing genes, whose expression was at least partially rescued by expression of HA-KDM6A-NU or HA-KDM6A-U. NU only, N = 77; U only, N = 173; both, N = 151.

K. Immunofluorescence assay to assess 3′UTR-dependent JMJD3 activity in HeLa cells after transfection of HA-KDM6B-NU or HA-KDM6B-U. HeLa cells have low levels of endogenous JMJD3. Nuclei were stained by DAPI. Anti-HA staining (green) shows transfected JMJD3 protein, and a reduction in anti-H3K27me3 staining (magenta) is used to infer JMJD3 demethylase activity. White arrowheads mark transfected cells. Representative images are shown.

L. Quantification of transfected JMJD3 abundance as HA intensity values from (K). Shown are boxplots depicting median, 25th and 75th percentiles and 5-95% confidence intervals from N = 3 independent experiments. Mann-Whitney test: ns, not significant.

M. 3′UTR-dependent JMJD3 demethylase activity was measured as a reduction in nuclear H3K27me3 intensity in transfected cells compared to non-transfected cells from the experiment shown in (K). The nuclear H3K27me3 intensity in each cell was normalized to the median

intensity of non-transfected cells (HA < 500 units). Shown as in (L) from *N* = 3 independent experiments. Mann-Whitney test, ****, *P* < 0.0001.

Figure 2. The JMJD3 IDR mediates 3′UTR-dependent activity regulation.

A. Protein models for full-length (FL) and IDR-deleted (IDRdel) JMJD3. The structured JmjC domain is indicated in brown, whereas the IDR is shown in grey.

B. JMJD3 IDR deletion constructs are shown as in Fig. 1E. The SV40 nuclear localization signal (NLS) was added.

C. As in Fig. 1K, but 3′UTR-dependent activity was measured for the IDRdel protein using constructs from (B).

D. As in Fig. 1L, but IDRdel protein abundance was measured as HA intensity from *N* = 3 independent experiments shown in (C). Mann-Whitney test: ns, not significant.

E. As in Fig. 1M, but 3′UTR-dependent demethylase activity was measured as reduction in H3K27me3 for IDRdel-transfected cells. Normalized H3K27me3 intensity values obtained from *N* = 3 independent experiments, as shown in (C). Mann-Whitney test: ****, *P* < 0.0001; ns, not significant.

F. Genome browser view of genomic context of *LONRF2* for anti-HA ChIP-seq signal (depth-normalized read count of one of the replicates) is shown. ChIP-seq was performed after transfection of the indicated KDM6B constructs, shown in Fig. 1E and 2B. No HA indicates negative control.

G. Metaplot analysis of experiment from (F) using all anti-HA ChIP-seq peaks with *P* < 0.001.

H. As in (F), but spike-in normalized anti-H3K27me3 ChIP-seq signal (depth-normalized read count of one of the replicates) is shown.

I. Metaplot analysis of experiment from (H) using all anti-H3K27me3 ChIP-seq peaks with *P* < 0.001 that showed at least a 50% signal reduction in KDM6B-U transfected samples compared to the control.

Figure 3. The *KDM6B* 3′UTR increases the proportion of active JMJD3 protein by regulating its oligomerization state.

A. Scatterplot showing transfected JMJD3 abundance (as HA intensity) and resulting H3K27me3 intensity (as negative measure of JMJD3 demethylase activity) in individual cells after transfection of HA-KDM6B-NU (left panel) or HA-KDM6B-U (right panel) in HeLa cells. Cells with HA intensity < 500 are non-transfected (neg). The cells below the horizontal red dotted line are defined as having active demethylase. The vertical red dotted line separates cells with low and high JMJD3 abundance.

B. Quantification of the fraction of cells expressing active JMJD3 in populations with low versus high JMJD3 expression levels, obtained from experiments in Fig. 1M and 3A. Data represent results from four independent replicates. Unpaired t-test: *, *P* < 0.05; ****, *P* < 0.0001.

C. Workflow for DSG chemical crosslinking in cells, followed by western blot.

D. Western blot analysis of whole-cell lysates from HeLa cells transfected with HA-KDM6B-NU or HA-KDM6B-U constructs before and after incubation with DSG. Blue dot indicates the monomer. Tubulin was used as loading control. Protein model shown as in Fig. 2A.

E. As in (D) but after transfection of IDRdel constructs. Homodimer is indicated by same-colored dots, whereas different colors indicate heterodimer formation.

F. AlphaFold3 prediction of full-length JMJD3 homodimer but shown are only the interacting domains (aa 1086-1510), color-coded as in the protein model shown above. Purple, HP region; brown, JmjC domain. (A) and (B) indicate two different JMJD3 molecules.

G. Top: 100-way PhyloP scores of KDM6B CDS. Bottom amino acid alignment of HP region across selected species. Dark grey, perfect conservation; light grey, conservative substitution.

H. As in (D) but after transfection of indicated full-length and HPdel constructs lacking aa 1088-1128.

I. As in Fig. 1K, but JMJD3 abundance and demethylase activity assay of full-length and HPdel is shown.

J. As in Fig. 1L, but JMJD3 protein abundance of full-length and HPdel was measured as HA intensity from $N = 3$ independent experiments shown in (I). Mann-Whitney test: ns, not significant.

K. As in Fig. 1M, but JMJD3 demethylase activity of full-length and HPdel was measured as reduction in H3K27me3. Normalized H3K27me3 intensity values obtained from $N = 3$ independent experiments shown in (I). Mann-Whitney test: ****, $P < 0.0001$; ns, not significant.

Figure 4. mRNA-IDR interaction suppresses oligomerization of JMJD3.

A. Net charge per residue (NCPR) of JMJD3 is shown. Protein model of JMJD3 showing a putative RNA-binding domain (RBD) in green (aa 973-1087), identified as a region with high positive net charge.

B. Fluorescence polarization assay to measure RNA-protein interactions between the MBP-tagged JMJD3 RBD from (A) and Cy5-labeled *KDM6B* 3′UTR in a solution containing 150 mM salt. Shown is the FP saturation binding curve from $N = 3$ independent experiments. The dissociation constant (K_d) is reported as ± error of fitting.

C. As in Fig. 3D but JMJD3 oligomerization was assessed after transfection of full-length or RBDdel constructs, which lacks aa 973-1087.

D. As in Fig. 1K, but JMJD3 abundance and demethylase activity assay of full-length and RBDdel is shown.

E. As in Fig. 1M, but JMJD3 demethylase activity of full-length and RBDdel was measured as reduction in H3K27me3. Normalized H3K27me3 intensity values obtained from $N = 3$ independent experiments, as shown in (D). Mann-Whitney test: ****, $P < 0.0001$; ns, not significant.

Figure 5. The *KDM6B* 3′UTR regulates co-translational homodimerization of JMJD3.

A. Workflow of co-translational disome assay. HeLa cells transfected with the indicated construct were lysed, treated with ribonuclease, and separated into monosome and disome fractions using sucrose gradient centrifugation. Nascent chains in each fraction were labeled with puromycin (green hexagon). HA-tagged target protein was immunoprecipitated, newly synthesized target was detected by immunoblotting with anti-puromycin, and results were normalized to RPLP0.

B. Abundance of HA-tagged nascent chains after transfection of the indicated constructs in disome over monosome fractions. Puromycylated HA-KDM6B band intensity in immunoprecipitated fractions was normalized to RPLP0 band intensity in the input to normalize for ribosome amount in each fraction. Shown are mean ±□SD from $N = 5$ independent experiments. Raw data are shown in Fig. S3D. Unpaired t-test: ns, not significant; *, $P < 0.05$; **, $P < 0.01$.

C. Constructs used for RNA-binding and thiol-thiol crosslinking experiments in vitro. All constructs contain an N-terminal 6xHis-tag (medium grey) and C-terminal Strep-tag (dark grey). All three naturally occurring cysteines in the RHP region were mutated, while a single cysteine was introduced at E1102 (asterisk) to enable crosslinking. The aa boundaries of the protein domains used are shown.

D. Immunoblots showing RNA-dependent protein-protein interactions between recombinant JMJD3 RHP and ED detected by BMOE-mediated thiol-thiol crosslinking. To resolve different dimer bands, a 6% Tris-Glycine gel was used, which does not capture the monomer bands. The same reactions were run on a separate 8% Tris-Glycine gel to resolve the monomer bands. An 8% gel showing monomers and dimers on one gel is presented in Fig. S3J.

E. As in (D) but using the HP instead of RHP region to detect RNA-dependent protein-protein interactions. The reaction was run on an 8% Tris-Glycine gel. The full gel is presented in Fig. S3K.

F. Model for 3′UTR-dependent regulation of JMJD3 demethylase activity via co-translationally determined protein conformations. Details, see text.

Figure 6. mRNAs with highly conserved 3′UTRs are multivalent, enabling mRNA translation in network-like condensates and 3′UTR-dependent protein activity regulation.

A. As in Fig. 1C, but mRNA length is shown. Mann-Whitney test, $P = 0$.

B. As in Fig. 1C, but mRNA valency is shown. Mann-Whitney, $P = 0$.

C. mRNA length distribution of mRNAs enriched in the cytosol ($N = 1328$) or in cytoplasmic condensates ($N = 3888$). Mann-Whitney test, $P = 1.6 \times 10^{-286}$.

D. As in (C) but mRNA valency is shown. Mann-Whitney test, $P = 1.7 \times 10^{-318}$.

E. Contribution of RNA valency in 3′UTRs to mRNAs. $N = 17091$ mRNAs with at least one A/U homopolymer are shown. Median = 0.77.

F. A schematic of a cell with cytoplasmic network-like condensate (in blue) is shown. Zoom-in shows enrichment of multivalent mRNAs in cytoplasmic network-like condensates. Network-like condensates are generated through RNA-RNA interactions of multivalent mRNAs. In contrast, low-valency mRNAs freely diffuse and are enriched in the cytosol.

G. GFP-tagged constructs used for RNA-FISH are depicted as in Fig. 1E. The white dots indicate A/U homopolymers and serve as surrogate marker for RNA valency.

H. RNA-FISH (white) of mRNAs from the indicated transfected constructs in HeLa cells before (-) and after (+) digitonin extraction. DAPI (blue) stains nuclei. Thin yellow dashed lines indicate cell boundaries, and thick dotted lines mark the nucleus. As digitonin only extracts cytoplasmic mRNAs, only cytoplasmic RNA-FISH signal was analyzed. Representative images are shown.

I. Quantification of (H). Fraction of digitonin-resistant mRNA in the cytoplasm shown as mean ± SEM from $N = 3$ independent experiments. The number of analyzed cells is shown. Mann-Whitney test, ****, $P < 0.0001$.

J. GFP- or HA-tagged constructs for 3′UTR-swapping experiments are shown as in (G). The KDM6B CDS was fused with two different multivalent 3′UTRs (blue) or with two different low-valency 3′UTRs (gray).

K. As in (H) but using the indicated GFP-tagged constructs from (J).

L. As in (I), but experiments from (K) were quantified. t test for independent samples, *, $P < 0.05$.

M. As in Fig. 1K but JMJD3 abundance and 3′UTR-dependent activity was measured for chimeric KDM6B constructs containing 3′UTRs that differ in RNA valency. The used constructs are shown in (J).

N. As in Fig. 1M, but 3′UTR-dependent JMJD3 demethylase activity was measured when using the KDM6B CDS fused to low-valency vs multivalent 3′UTRs (constructs shown in J). $N = 3$ independent experiments were performed. A reduction in H3K27me3 intensity indicates active demethylase. Mann-Whitney test, ****, $P < 0.0001$.

Figure 7. Inclusion of HC 3′UTRs during mRNA translation changes oligomerization states of proteins with long IDRs.

A. Proteins are grouped according to their IDR length (determined using IUPred2A score ≥ 0.33). The fraction and number of proteins encoded by multivalent mRNAs with HC 3′UTRs (MVMHC) or low-valency mRNAs with non-conserved 3′UTRs (LVMNC) is shown for the four groups.

B. Proposed model by which multivalent mRNAs with HC 3′UTRs co-translationally control protein oligomerization states of their encoded IDR-containing proteins, whereas low-valency mRNAs with non-conserved 3′UTRs are translated in the cytosol and have no influence on protein folding.

C-F. For each candidate to be validated, shown are mRNA multivalency (indicated by white dots), 3′UTR conservation (indicated by the color of the 3′UTR with NC in grey and HC in blue), and protein domains (annotated structured domains from uniprot are indicated in brown). Protein oligomerization patterns in cells after transfection of the indicated HA-tagged constructs were analyzed by DSG cross-linking and western blot analysis. Distinct oligomeric states are indicated by Roman numerals with increasing shades of gray indicating higher order of protein assemblies. The stacked bar plot shows the fraction of each oligomeric species for each condition as quantified from the western blot.

C. DSG crosslinking western blot of HA-tagged DHPS shows 3′UTR-independent protein oligomerization pattern. Actin was used as loading control.

D. As in (C) but shown is HA-SSBP4.

E. As in (C) but shown is HA-KDM6A, showing a protein oligomerization pattern that is altered by inclusion of the 3′UTR in the mRNA template. Tubulin was used as loading control.

F. As in (E) but shown is HA-FOXP2, showing a protein oligomerization pattern that is altered by inclusion of the 3′UTR in the mRNA template.

Detailed methods

Cell lines

HeLa (human cervical cancer line, female origin) was previously described[8]. MCF7 (human breast cancer, female origin) and U2OS (human osteosarcoma line, female origin) were previously described[3]. All cells were maintained at 37°C with 5% CO2 injection in Dulbecco's Modified Eagle Medium (DMEM) containing 4.5 g/l glucose, 10% heat inactivated fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin.

Vector constructs

GFP-tagged constructs. For GFP-tagged constructs, pcDNA3.1-puro mGFP (monomeric GFP, A207K), which was previously reported[8], was used. To generate GFP-MYC, the *MYC* promoter (2780 bp; hg38, chr8:127733806-127736585) was PCR-amplified from HeLa genomic DNA and inserted between *BglII* and *NheI* sites. To generate GFP-MYC-NU, the *MYC* coding sequence (1365 bp) was PCR-amplified from HeLa cDNA and inserted between *BsrGI* and *BamHI* sites. For GFP-MYC-U, the *MYC 3′UTR* (473 bp) was amplified and inserted between *BamHI* and *EcoRI* of the NU construct.

To obtain GFP-KDM6A and GFP-KDM6B constructs for RNA-FISH, the coding sequences were PCR-amplified from HA-tagged constructs (see below) and inserted into the pcDNA3.1-puro mGFP vector between *BsrGI* and *NotI* or *KpnI* sites, respectively. The 3′UTRs were then inserted between *NotI* or *KpnI* and *XhoI* sites.

HA-tagged constructs. KDM6A (NM_001291416) and KDM6B (NM_001080424) expression constructs in the pCMV-HA backbone were obtained from Addgene (#24168 and #24167)[45] and used as NU constructs. For the U constructs, the full-length *KDM6A 3′UTR* (1368 bp) and full-length *KDM6B 3′UTR* (1274 bp) were each amplified from HeLa genomic DNA and inserted between *NotI* and *XhoI* restriction sites. For ChIP-seq, a plasmid containing NLS-KDM6Bs (amino acids 1027–1685)[45] was used as IDRdel. The NLS (PKKKRKV) was derived from SV40 large tumor antigen and was added to the N-terminus via PCR. For all other assays, IDRdel contains this NLS at the N-terminus and amino acids 1141-1685.

All other HA-tagged expression constructs were cloned into the pCMV-HA backbone using similar approaches and were used in the chemical crosslinking experiments. For NU constructs, coding sequences (CDS, lacking a start codon) of SSRP1 (2127 bp, NM_003146), SF3B4 (1272 bp, NM_005850), SSBP4 (1155 bp, NM_032637), EIF3I (975 bp, NM_003757), DHPS (1107 bp, NM_001930) and RACK1 (951 bp, NM_006098) were PCR amplified from HeLa cDNA and inserted between *KpnI* and *NotI,* except for the FOXP2 coding sequence (2145 bp, NM_014491), which was PCR-amplified from HEK293 cDNA. The FOXP2-NU construct also includes a C-terminal FLAG-tag (DYKDDDK), added via PCR. For the FOXP2-U construct, the full-length *FOXP2 3′UTR* (3838 bp) was inserted between *NotI* and *XhoI* sites. For the U constructs of low-valency mRNAs, the coding sequences together with the 3′UTRs (SSRP1, 2546 bp; SF3B4, 1497 bp; SSBP4, 1471 bp; EIF3I, 1383 bp; DHPS, 1228 bp; RACK1, 1030 bp) were directly amplified from cDNA and inserted between *KpnI* and *NotI*.

For chimeric constructs containing the KDM6B CDS and various 3′UTRs, 3′UTR sequences (*IDH1 3′UTR*, 850 bp; *HSPA1B 3′UTR*, 379 bp; *PIN1 3′UTR*, 490 bp; *MAP2K2 3′UTR*, 277 bp) were amplified from HeLa genomic DNA and inserted between *NotI* and *XhoI* sites of the KDM6B-NU construct.

shRNA constructs. pLKO.1 puro was used to create shRNA constructs for stable knockdown cell lines. The sequence of the shRNA against KDM6A was previously reported[45]. DNA

oligonucleotides (Table S5) were annealed and inserted into pLKO.1 puro between *SgrAI* and *EcoRI*. As control, we used a scramble shRNA (Addgene #1864). To generate shRNA-resistant KDM6A expression constructs, eight synonymous mutations (CTATGAATCTCTAATCTT to tTAcGAgTCctTgATtcT) were introduced to the shRNA binding site of the original KDM6A plasmids using Q5 site-directed mutagenesis (NEB).

pET28a-MBP constructs. The pET28a vector containing a 6×His-MBP tag was previously reported[9] and used to clone all constructs for recombinant protein expression, except for the KDM6B enzymatic domain (ED) construct, which was cloned into the original pET28a backbone. For constructs used in thiol-thiol crosslinking, a C-terminal Strep-Tag II (SAWSHPQFEK) was included in the reverse primer used to amplify the inserts. The KDM6B RBD (aa 971-1020) and KDM6B ED (aa 1141-1643) were inserted between the *NheI* and *BamHI* sites.

The amino acid boundaries of the HP sequence are 1081-1178, whereas the amino acid boundaries of KDM6B RBD-HP (called RHP) are 971-1178. To generate the 6×His-MBP-RHP-Strep-Tag-II construct, the wild-type RHP sequence was PCR-amplified and all the cysteines (C972, C990, C1138) were mutated to serines using the NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621L) with forward and reverse primers carrying the mutations and E1102 was mutated to cysteine for thiol-thiol crosslinking. The 6×His-MBP-HP-Strep-Tag-II construct was generated through PCR-amplification using the mutated RHP construct as template and inserted between the *NheI* and *BamHI* sites.

All constructs were verified by sequencing and all primer sequences and insert lengths are listed in Table S5.

DNA templates for RNA in vitro transcription. All DNA templates were PCR-amplified and purified by PCR cleanup. As the full-length *KDM6B* 3′UTR could not be transcribed, the sequence without the first 110 bp (1176 bp) was amplified from the pCMV-HA-KDM6B-U construct and used as *KDM6B* 3′UTR for in vitro experiments. The 3′UTR of HSPA1B (379 bp) was PCR-amplified from pCMV-HA-KDM6B-HSPA1B-UTR. The T7 promoter (TAATACGACTCACTATAGGG) was incorporated into the forward primers used for generating these templates.

Plasmid transfection and viral transduction

Lipofectamine 3000 (Thermo Fisher, L3000015) was used to transfect pcDNA3-mGFP-MYC construct into U2OS cells. FuGENE HD (Promega, E2311) was used to transfect all pCMV-HA-derived plasmids and all pcDNA3-mGFP-KDM6B constructs into HeLa cells, except for HA-KDM6A, which was transfected into HeLa and MCF7 using Lipofectamine 3000. NU and U expression constructs for the same protein were transfected at the same molarity by default. For specific constructs requiring equalization of protein expression, the molarity was adjusted. Specifically, chimeric KDM6B constructs containing multivalent UTRs and U constructs of DHPS, SF3B4, SSRP1 and SSBP4 were transfected at a 1.3-fold higher molarity than their corresponding NU constructs. The FOXP2-U construct was transfected at a 1.5-fold higher molarity than its corresponding NU construct.

MCF7 cells stably expressing shKDM6B were generated as previously described[8]. To generate virus, HEK293T cells were transfected using Lipofectamine 3000 with pCMV-dR8.2, pCMV-VSV-G, and a pLKO.1-shRNA_KDM6A plasmid. Viral particles were harvested 48 h after transfection and filtered through a 0.45 μm filter. MCF7 cells grown in 6-well plates were transduced with 200 μl viral particles in the presence of 8 μg/ml polybrene. Puromycin (1 μg/ml)

was added 48 h post-transduction. Cells were maintained in puromycin for at least four more days. Knockdown efficiency was validated by western blotting.

Western blotting

Cells were washed twice with PBS and lysed directly with 2x Laemmli sample buffer (Thermo Scientific, J61337.AC). The lysate was transferred to a 1.5 ml Eppendorf tube and boiled at 95°C for 10 min, followed by sonication for three times in 30s on/off cycles using a Diagenode Bioruptor. Samples were subjected to SDS-PAGE on NuPAGE 4 -12% Bis-Tris gradient gel (Invitrogen). For KDM6B, FOXP2, and KDM6A chemical crosslinking experiments, NuPAGE 3 - 8% Tris-Acetate protein gels were used. For thiol-thiol crosslinking *in vitro*, 6% or 8% Tris-Glycine gels were used. Membranes were scanned with an Odyssey DLx Imaging system (LI-COR). The primary antibodies used were: anti-GAPDH (1:2000, Sigma-Aldrich, G8795), anti-α-Tubulin (1:1000, Sigma-Aldrich, T9026), anti-Actin (1:1000, Sigma-Aldrich, A2066), anti-GFP (1:2000, Abcam, ab13970), and anti-HA (1:2000, Covance, MMS-101P). The secondary antibodies used were: IRDye 680RD donkey anti-chicken (LI-COR, 926-68075), IRDye 680RD goat anti-mouse (LI-COR, 926-68070), and IRDye 800CW donkey anti-rabbit (LI-COR, 926-32213).

Chemical Crosslinking

Chemical crosslinking with DSG (Thermo Fisher) was performed as previous described[73], with minor modifications. Briefly, HeLa cells were transfected with the indicated constructs. After 24 hours, cells grown in 24-well plates were washed twice with 0.5 ml PBS, and DSG in 200 µl PBS was directly added to each well. The reaction was incubated at room temperature (RT) with shaking for 30 min and quenched by adding 50 mM Tris, pH 7.5, for 15 min at RT. After discarding the buffers, cells were lysed with 2x Laemmli buffer. The samples were sonicated 3-5 times in 30s on/off cycles using a Diagenode Bioruptor and boiled at 95°C for 10 min. The crosslinked proteins were then analyzed by SDS-PAGE followed by western blotting.

Immunofluorescence staining

HeLa cells were seeded on 4-well Millicell EZ Slides and transfected. After 24 hours, the transfected cells were briefly washed twice with PBS, fixed with 4% paraformaldehyde for 10 min at RT, washed three times with PBS, and permeabilized in 0.1% Triton X-100 in PBS for 7 min. Cells were washed three times with PBST (PBS with 0.1% Tween-20) and incubated in blocking buffer (3% BSA in PBST) for 1 hour at RT. The cells were stained with the following primary antibodies diluted in blocking buffer at 4°C overnight: mouse anti-HA (1:1000, Covance, MMS-101P) and rabbit anti-H3K27me3 (1:2000, Cell Signaling Technology, 9733). Cells were then washed 3x with PBST and incubated with secondary antibody in blocking buffer (goat anti-rabbit Alexa 647, 1:1000, Thermo Fisher A-21244; donkey anti-mouse Alexa 488, 1:1000, Thermo Fisher A-21202) for 1 hour at RT. Cells were washed three times with PBST, mounted with mounting media with DAPI, and a coverslip was added. Samples were then imaged using a Nikon CSU-W1 SoRa spinning disk microscope with a 63x/1.40 CFI Plan Apo oil immersion objective. All images were taken at the same laser intensity and exposure settings.

Digitonin RNA-FISH

RNA-FISH against GFP was performed before and after digitonin extraction of the cytosol, as previously described[8,10]. Briefly, HeLa cells were seeded on 4-well Millicell EZ Slide and transfected with GFP-fusion constructs. Fourteen to 16 hours after transfection, cells were incubated with 400 µl of digitonin solution (40 µg/ml digitonin, 150 mM NaCl, 20 mM HEPES pH 7.4, 0.2 mM EDTA, 2 mM DTT, 2 mM $MgCl_2$) for 10 min at 4°C. The digitonin solution was aspirated, and coverslips were rinsed once with 400 µl PBS. For untreated samples, cells were briefly rinsed twice with PBS and subjected to the same RNA-FISH procedure described below. Specifically, cells were fixed with 4% paraformaldehyde for 10 min at RT, followed by two 5-minute washes with PBS and permeabilized using 1 ml of 70% ethanol in each well for 8 hours at 4°C. The slides were washed once with 1 ml wash buffer (2x SSC, 10% formamide) for 5 min at RT. A 200 µl hybridization mix containing Stellaris FISH Probes for eGFP with Quasar 670 Dye (Biosearchtech, VSMF-1015-5) was prepared and added to each well as previously reported. Slides were incubated at 37°C overnight, followed by two 30-minute washes with 1 ml pre-warmed wash buffer at 37°C. After a brief wash with PBST, slides were mounted with ProLong™ Gold Antifade Mountant with DAPI DNA stain (Thermo Fisher). Images were captured using a Nikon CSU-W1 SoRa spinning disk microscope with a 63x/1.40 CFI Plan Apo oil immersion objective. Excitations were performed sequentially using 405, 488, and 633 nm laser wavelengths. Eleven z-stacks were taken at 0.5 µm increments.

Co-translational disome assay

Monosome and disome isolation. For isolation of monosomes and disomes, HeLa cells were plated on 15 cm dishes and transfected with HA-tagged constructs. Twenty-four hours after transfection, cells were washed with 10 ml 1x PBS supplemented with 100 µg/ml CHX (Sigma Aldrich) and 10 mM $MgCl_2$. The PBS solution was aspirated, and 100 µl of ice-cold 5x lysis buffer (250 mM HEPES, pH 7.0, 50 mM $MgCl_2$, 750 mM KCl, 5% NP-40, 50 mM DTT, 500 µg/mL CHX, 125 U/ml Turbo DNase (Invitrogen) and cOmplete EDTA-free protease inhibitor (Roche)) was added. The cells were scraped from the plate, and 1 unit/µl of RNase I (Ambion) was directly added to the lysate. The lysate was transferred to a 1.5 ml RNase-free Eppendorf tube and incubated on ice for 30 min.

The cell lysates were gently triturated ten times through a 26 Guage needle and cleared by centrifugation for 10 min at 14,000 rpm at 4°C. Cleared lysates were loaded onto 10-25% sucrose gradients in sucrose buffer (50 mM HEPES, pH 7.0, 5 mM $MgCl_2$, 150 mM KCl, 100 µg/ml CHX, and EDTA-free protease inhibitor) and subjected to ultracentrifugation at 35,000 rpm in a TH-641 swinging bucket rotor (Thermo Fisher) for 3 hours at 4°C. Fourteen equal-volume fractions were collected using a Piston Gradient Fractionator (Biocomp) with continuous monitoring of RNA at UV260. Fractions 8-10 were pooled to obtain the monosome fraction and fractions 11-13 were pooled as the disome fraction.

Puromycylation and detection of puromycylated nascent proteins. To label and detect HA-tagged nascent chains in monosome and disome fractions, each fraction was incubated with 100 µM puromycin (Millipore) for 5 min at 37°C with shaking. 10% of puromycylated fractions were precipitated by TCA (TCI) and saved as input samples to adjust for total amount of ribosome loading. The remaining fractions were incubated with 20 µl of anti-HA magnetic beads (Thermo Fisher) for 2 hours at 4°C with constant tumbling. Beads were washed three times with 0.5 ml PBST and resuspended in 20 µl 2x Laemmli buffer. Immunoprecipitated puromycylated proteins were resolved on SDS-PAGE and probed using anti-puromycin antibody (1:1000, MilliporeSigma, MABE343) and anti-HA (1:2000, Covance, MMS-101P). To normalize for sample loading, total ribosome content in the input of each fraction was assessed by probing for RPLP0 using rabbit anti-RPLP0 (1:1000, Antibodies Online, ABIN310003). Band intensity of the

HA-tagged nascent chain (co-localized HA and puromycin signal) in each IP fraction was quantified by FIJI and normalized to the RPLP0 signal from the corresponding input. This yielded the normalized abundance of HA-tagged nascent chain in each fraction. The ratio of HA-tagged nascent chain in the disome over the monosome fraction was calculated and plotted.

Protein purification

All expression constructs were transformed into BL21 *E. coli* (NEB) for protein expression. With the exception of RHP, all proteins were purified using TALON resin under native conditions, followed by size exclusion chromatography. Specifically, a fresh colony was used to inoculate an overnight culture in 5 ml LB/kanamycin media and shaken vigorously at 250 rpm at 37°C. The overnight culture was diluted in 500 ml LB media containing 50 mg/l kanamycin and grown at 37°C until OD600 reached 0.6-0.8. Protein expression was induced by adding 0.5 mM IPTG, and the bacteria were cultured at 18°C overnight. Bacteria were harvested by centrifugation at 6,000 g for 15 min. The pellet was resuspended in 30 ml cold lysis buffer (50 mM sodium phosphate pH 7.4, 600 mM NaCl, 1 mM PMSF), supplemented with 10 mg lysozyme (Life Technologies), 1x Halt protease inhibitor cocktail (Thermo Fisher), and 1 mg DNase I (Roche). After incubation on ice for 30 min, bacteria were sonicated on ice for 3 min 30 seconds with on/off intervals of 1 and 3 s (Misonix S-4000). The lysate was cleared by centrifugation at 38,000 g at 4°C for 30 min.

5 ml TALON metal affinity resin (TaKaRa Bio) was equilibrated with five column volumes of native lysis buffer. The supernatant of cleared bacteria lysate was transferred to a new 50 ml Falcon tube and incubated with TALON resin with rotation at 4°C for 2 hours. The slurry was transferred into a gravity column and washed with 10 column volumes of native wash buffer (50 mM sodium phosphate pH 7.4, 600 mM NaCl, 10 mM imidazole). The protein was eluted with five column volumes of elution buffer (50 mM sodium phosphate pH 7.4, 600 mM NaCl, 200 mM imidazole).

The TALON-eluted protein was buffer-exchanged into protein storage buffer (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 1 mM TCEP, 5% glycerol, and 0.01% Tween-20) using centrifugal filters with a 30 kDa cutoff (Pall) and concentrated to 1 ml. The concentrated protein was spun down at 20,000 g for 5 min at 4°C, and the supernatant was loaded onto a Superdex 200 Increase 10/300 column using the AKTA Purifier system (Cytiva). Peak fractions were collected, aliquoted, and snap-frozen in liquid nitrogen for storage at -80°C.

As the 6×His-MBP-StrepII-tagged RHP protein degrades on the size exclusion column, a two-step purification was performed. RHP protein was first purified using TALON resins under native conditions as described above. The eluted protein was concentrated to 10 ml using a 30 kDa cutoff centrifugal filter. The concentrated protein was passed through 2.5 ml of pre-equilibrated Strep-Tactin® 4Flow® high-capacity resin (IBA Lifesciences) twice. The protein-bound beads were then washed with 5 column volumes of PBS and eluted with 5 column volumes of PBS supplemented with 10 mM desthiobiotin (Sigma Aldrich). The eluted RHP protein was buffer-exchanged into the protein storage buffer by using a 10 kDa cutoff centrifugal filter, aliquoted, and snap-frozen in liquid nitrogen for storage at -80°C. Protein concentrations were determined by BCA assay (Thermo Fisher). The integrity of the purified protein was examined by SDS-PAGE and Coomassie blue staining.

RNA *in vitro* transcription

All RNAs were prepared by *in vitro* transcription using the T7 MEGAscript kit (Invitrogen). 250 ng of PCR-amplified DNA template was used for each 20 µl reaction. For Cy5 labeling, Cy5-UTP (Enzo life sciences, ENZ-42506) was added at a ratio of 1:300 labeled UTP:unlabeled UTP. The reaction mix was prepared following the manufacturer's protocol and incubated at 30°C for 6 hours. The products were then treated with Turbo DNase provided in the kit and incubated at 37°C for 15 min. The integrity and sizes of the RNAs were verified using agarose gel electrophoresis. All RNAs were precipitated with LiCl overnight at -20°C. The RNAs were pelleted by centrifugation at 20,000 g for 20 min at 4°C, washed with 1 ml of 70% ethanol, redissolved in nuclease-free water, aliquoted, and stored at -80 °C. RNA concentration was measured by NanoDrop.

Fluorescence polarization assay to measure RNA affinity

MBP-tagged protein domains were serially diluted using dilution buffer (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.01% Tween-20, and 400 U/ml SUPERase·In™ RNase Inhibitor (ThermoFisher)). Then, Cy5-labeled RNA (2.4 ng/µl) was added to the diluted proteins for a final reaction volume of 20 µL. All experiments were performed in triplicate. Reactions were incubated in a flat-bottom black 384 well-plate (Revvity) on a rocker at RT for 1 hour. Then, fluorescence polarization was measured using an Infinite M1000 Pro (Tecan) with the following parameters in i-control: excitation wavelength: 635 nm; emission wavelength: 665 nm; excitation/emission bandwidth: 5 nm; gain: 100 (manual); number of flashes: 30; settle time: 5 ms; Z-position: 24836 µm (manual); G-Factor: 1.1 (manual). Each sample was measured three-times, and the mean values were used for binding affinity calculations. Binding curves were fitted to the fluorescence polarization data using the nonlinear regression “One site – Total” model in Prism 10.

Thiol-thiol crosslinking

KDM6B RHP (5 µM) and ED (5 µM) proteins were mixed with Cy5-labeled RNA (6 ng/µl) on ice in an assay buffer containing 25 mM Tris-Cl pH 7.4, 150 mM NaCl, and 0.01% Tween-20. BMOE (Thermo Fisher) was added to a final concentration of 5 µM in a total reaction volume of 20 µl. After 1 hour of incubation at 25°C, the reactions were quenched with 10 µl of 6x Laemmli SDS sample buffer. The samples were boiled at 95°C for 7 min and loaded onto 6% and 8% Tris-Glycine gels to resolve the dimer or monomer bands, respectively. Western blot was then performed, and proteins were probed with Streptavidin-680 (LI-COR, 1:3000 dilution).

RNA-seq

To identify 3′UTR-dependent MYC targets, U2OS cells were transfected with GFP-MYC constructs and FACS-sorted 24 hours post-transfection to isolate GFP-positive cells. RNA was extracted from GFP-positive cells using TRIzol according to the manufacturer’s protocol and submitted to Genewiz for standard, 150 bp paired-end RNA sequencing (53-67 million reads were obtained per sample).

To identify 3′UTR-dependent KDM6A targets, endogenous *KDM6A* was knocked down with shRNAs and HA-tagged shRNA-resistant KDM6A constructs were transfected into the *KDM6A* knockdown cells. At 24 hours post-transfection, expression levels of total KDM6A and HA-tagged constructs were validated by western blotting using rabbit anti-UTX (1:2000, Bethyl, A302-374A) and mouse anti-HA (1:2000, Covance, MMS-101P) antibodies, respectively. RNA was then extracted using TRIzol according to the manufacturer’s protocol and submitted to

Genewiz for standard, 150 bp paired-end RNA sequencing (19-30 million reads per sample were obtained).

ChIP-seq

To analyze genome-wide 3′UTR-dependent JMJD3/KDM6B protein localization and activity, HeLa cells were co-transfected with HA-tagged KDM6B constructs and pcDNA3-mGFP at a 10:1 ratio. Twenty-four hours after transfection, cells were trypsinized, washed twice with ice-cold PBS, and stained with LIVE/DEAD fixable far-red stain (Thermo Fisher) following the manufacturer's protocol. The cells were washed twice with ice-cold PBS and fixed with 1% paraformaldehyde for 10 min at room temperature on a nutator. Paraformaldehyde was quenched with 0.125 M glycine and mixed for 5 min at RT. The cells were centrifuged at 500 g for 5 min at 4°C, washed twice with ice-cold PBS supplemented with 1% FBS and EDTA-free protease inhibitor, and sorted by FACS to collect GFP-positive cells.

ChIP was performed by the MSKCC Epigenetics Research Innovation Lab. Cells were resuspended in SDS buffer (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.5% SDS, 1x protease inhibitor cocktail from Roche). The resulting nuclei were spun down, resuspended in the immunoprecipitation buffer at 1 ml per 0.5 million cells (SDS buffer and Triton Dilution buffer (100 mM NaCl, 100 mM Tris HCl pH 8.0, 5 mM EDTA, 5% Triton X-100) mixed in 2:1 ratio with the addition of 1x protease inhibitor cocktail (Millipore Sigma, #11836170001) and processed on a Covaris E220 Focused-ultrasonicator to achieve an average fragment length of 200-300 bps with the following parameters: PIP=140, Duty Factor=5, CBP/Burst per sec=200, Time = 20 min. Chromatin concentrations were estimated using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific #23227) according to the manufacturer's instructions. The immunoprecipitation reactions were set up in 500 µl of the immunoprecipitation buffer in Protein LoBind tubes (Eppendorf, #22431081) and pre-cleared with 50 µl of Protein G Dynabeads (ThermoFisher Scientific #10004D) for two hours at 4°C. After pre-clearing, the samples were transferred into new Protein LoBind tubes and incubated overnight at 4°C with 5 µg of HA antibody (CST, #3724) and H3K27me3 antibody (CST, #9733). For normalization purposes, 2 µl of KMet SNAP-ChIP Panel (Epicypher, #191100) were added to each reaction. The next day, 50 µl of BSA-blocked Protein G Dynabeads were added to the reactions and incubated for 2 hours at 4°C. The beads were then washed twice with low-salt washing buffer (150 mM NaCl, 1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris HCl pH 8.0), twice with high-salt washing buffer (500 mM NaCl, 1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris HCl pH 8.0), twice with LiCL wash buffer (250 mM LiCl, 10 mM Tris HCl pH 8.0, 1 mM EDTA, 1% Na-Deoxycholate, 1% IGEPAL CA-630) and once with TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA). Samples were then reverse-crosslinked overnight in the elution buffer (1% SDS, 0.1 M NaHCO3) and purified using the ChIP DNA Clean & Concentrator kit (Zymo Research, #D5205) following the manufacturer's instructions. After quantification of the recovered DNA fragments, libraries were prepared using the ThruPLEX®DNA-Seq kit (R400676, Takara) following the manufacturer's instructions, purified with SPRIselect magnetic beads (B23318, Beckman Coulter), and quantified using a Qubit Flex fluorometer (ThermoFisher Scientific) and profiled with a TapeStation (Agilent). The libraries were sent to MSKCC Integrated Genomics Operation core facility for sequencing on an Illumina NovaSeq 6000 (aiming for 30-40 million 100 bp paired-end reads per library).

Data analysis

mRNA and protein sequence features

PhyloP sequence conservation in 3′UTRs. The human MANE gene annotations (hg38) were used for all analyses[74]. The genome-wide phylogenetic *P* value (phyloP conservation score) was obtained from the PHAST package for multiple alignments of 99 vertebrate genomes with the human genome in a form of a bigwig file. The genome-wide phyloP score was mapped to the 3′UTR boundaries of the human GTF (genome-build: GRCh38.p14, genome-build-accession: GCA_000001405.29), and the phyloP conservation values for 3′UTRs were obtained. To determine the number of conserved nucleotides in human 3′UTRs, we used a minimum phyloP score ≥ 2 and counted the nucleotides if at least four consecutive nucleotides had phyloP scores above this cutoff. The sum of conserved nucleotides for each 3′UTR is listed in Table S1.

RNA valency. To identify RNA sequence features that are predictive of RNA-RNA interactions and the formation of network-like condensates, homopolymer stretches (≥ 5 consecutive nucleotides) were counted in a given RNA and the sum of G and C, as well as the sum of A and U homopolymers were determined and listed in Table S1. If a homopolymer contained multiples of five consecutive nucleotides, each stretch of five was counted as an individual homopolymer. The initial analysis was performed on the dataset obtained from Ma et al., eLife (2021)[9], which contains experimentally-verified 3′UTRs that induce network-like or droplet-like condensates, respectively. RNA valency for human 5′UTR, CDS, 3′UTR, and mRNA were calculated based on the MANE gene annotations[74].

mRNAs enriched in cytosol or cytoplasmic network-like condensates. The list of cytosol-enriched mRNAs was obtained from Horste et al., Mol Cell (2023), Table S1, V60=4[10]. mRNAs enriched in cytoplasmic network-like condensates were obtained from Horste et al., Mol Cell (2023), Table S1, V60=1 (enriched in TIS granules), V60=3 (enriched in the TIA-ER region)[10] and Chen et al., Cell (2024), Table S1, FXR1 mRNA targets[11]. mRNAs (*N* = 142) that were classified as cytosol-enriched in Horste et al., Mol Cell (2023) and as FXR1 network-enriched in Chen et al., Cell (2024) were considered here as condensate-enriched (see Table S1).

Identification of IDRs. The coding sequences obtained from MANE gene annotations were translated into amino acids and IUPred2A scores were determined[39]. All amino acids with IUPred2A score ≥ 0.33 were defined as IDRs and their number of amino acids are listed in Table S1. These values were used throughout the paper.

pLDDT (predicted Local Distance Difference Test) scores were obtained from AlphaFold and amino acids with pLDDT scores < 0.8 were considered unstructured. These values were used in Fig. 1G and S1C.

Net charge per residue (NCPR). The net charge per residue for KDM6B was calculated with localCIDER[75] using a step size of 50.

CDS PhyloP. Nucleotide scores for the KDM6B CDS were plotted with a sliding window of 50.

Gene ontology analysis

Gene ontology analysis was performed using DAVID[41]. Bonferroni-adjusted *P* values are reported.

Protein structure prediction

JMJD3 dimer structures were generated using the AlphaFold3 online server with default parameters. For each full-length protein, five structures were generated, and the highest-ranked structure was selected and rendered using UCSF ChimeraX (v1.9)[76].

RNA-seq analysis

Raw sequencing reads were quantified using Salmon (v1.10.1)[77]. The GENCODE release v44 and v47 basic annotations were used as the references for MYC RNA-seq and UTX RNA-seq, respectively. Transcript-level data was imported using tximport (v1.22.0), and differential gene expression analysis was performed using DESeq2 (v1.38.3 or Galaxy Version 2.11.40.8+galaxy0)[78]. Only protein-coding genes defined by the GENCODE annotation were included in the following analysis.

To identify genes regulated by MYC-NU or MYC-U, each condition was compared against the GFP-only control. We focused on mRNAs expressed in the GFP sample with mean TPM (transcripts per million) > 3. Differentially expressed genes have a |log2FC| > 0.3 and an adjusted *P* value < 0.1 in at least one of the tested conditions. Hierarchical clustering was performed on differentially expressed genes identified in either condition. The clustering was based on the z-score of the mean TPM value for each mRNA obtained from replicates and utilized the “ward.D” agglomeration method in R. 3′UTR-dependent groups were then defined using the obtained clusters and are presented in Table S2. The heatmap was visualized using the pheatmap (v1.0.13) package in R.

For the UTX RNA-seq analysis, UTX-regulated genes were first identified by comparing samples expressing shRNA targeting *KDM6A* to those expressing a scrambled control shRNA. Transcripts with a *P* value < 0.05 according to DESeq2 were considered significantly regulated. Based on the direction of change, these were classified as either “UTX-activated” (downregulated upon *KDM6A* knockdown) or “UTX-repressed” (upregulated upon *KDM6A* knockdown).

To identify genes rescued by expression of either the KDM6A-NU or KDM6A-U construct, we first filtered for robustly expressed genes, requiring a mean TPM > 7 in either the shCtrl or shKDM6A samples. We next selected genes that exhibited partial rescue, defined as those whose expression in KDM6A-NU or KDM6A-U transfected samples shifted back towards the levels observed in the shCtrl samples. mRNAs were divided into “UTX-activated” and “UTX-repressed” categories and clustered separately. Hierarchical clustering was performed on the z-scored mean TPM values using the “average” linkage method in R. 3′UTR-dependent groups were then defined using the clustering results and presented in Table S3.

ChIP-seq analysis

Raw sequencing reads were trimmed and filtered for quality and library adapters using version 0.4.5 of TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore), with a quality setting of 15 and running cutadapt (v1.15) and FastQC (v0.11.5). Reads were aligned to human assembly hg38 with version 2.3.4.1 of bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) and were deduplicated using MarkDuplicates in version 2.16.0 of Picard Tools. To ascertain enriched regions, MACS2 (https://github.com/taoliu/MACS) was used with a p-value setting of 0.001 and run against a matched control for each condition. A peak atlas was created by combining the superset of all peaks using the ‘merge’ function in the BEDTools suite v2.29.2. Read density profiles were created using deepTools ‘bamCoverage’ v3.3.0, normalized to 10 million uniquely mapped reads

and with read pileups extended to 200 bp. For histone modification data, ChIP signal was scaled to the inverse of the normalized counts to reads aligning to H3K27me3-modified synthetic nucleosome spike-in barcodes. H3K27me3 enrichment was classified as two-fold enrichment of spike in signal in comparison with control. Peak-gene associations were created by assigning all intragenic peaks to that gene, while intergenic peaks were assigned using linear genomic distance to transcription start site. Composite and tornado plots were created using deepTools v3.3.0 by running computeMatrix and plotHeatmap on normalized bigwigs with average signal sampled in 25 bp windows and flanking region defined by the surrounding 2 kb for HA signal and 10 kb for H3K27me3. Motif signatures were obtained using Homer v4.5 (http://homer.ucsd.edu).

Immunofluorescence staining analysis

To analyze JMJD3/KDM6B protein abundance and protein activity after transfection of HA-tagged KDM6B constructs at the single-cell level, we loaded the obtained images using the nd2 (v0.10.0) package and processed them with the scikit-image package (v0.22.0). We first generated nuclear masks from DAPI channel grayscale images. Specifically, images were denoised using a Gaussian filter (sigma = 4, truncate = 4). Then, we created binary masks using each image's mean intensity as a threshold. The masks were further refined through binary closing/opening, hole filling and removal of border-touching objects. Next, nuclei were segmented using the watershed algorithm, seeded by local maxima. Lastly, fluorescence signals from the H3K27me3 and HA channels were quantified using the refined masks.

The median fluorescence H3K27me3 signal obtained from untransfected cells (HA < 500 units) of each sample were used for normalization.

Digitonin RNA-FISH analysis

To determine the fraction of mRNA transcripts retained in the cytoplasm after digitonin treatment, we calculated the ratio of mean fluorescence intensity in the cytoplasm of digitonin-treated cells over untreated cells. Specifically, we used FIJI[79] to generate masks for the whole cell and nucleus using the maximum projection of fluorescence signals from RNA-FISH and DAPI channels, respectively. Next, maximum projection of RNA-FISH signals were quantified using the whole cell and nucleus masks. Mean cytoplasmic signals were calculated by subtracting total nuclear signals from total whole cell signals, divided by the cytoplasmic area. At least six fields of view were quantified for each sample from each replicate. For each experiment, mean cytoplasmic RNA-FISH signal in each digitonin-treated cell was divided by the median of the mean cytoplasmic RNA-FISH signal from all untreated cells. At least three independent experiments per construct were performed.

Further statistical analysis

Statistical parameters are reported in the figures and figure legends, including the definitions and exact values of *N* and experimental measures (mean ± SD or SEM or boxplots depicting median, 25th and 75th percentile (boxes) and 5% and 95% confidence intervals (error bars)). Transcriptome-wide feature comparisons were performed using a Kruskal-Wallis test or two-sided Mann-Whitney test in SPSS. For transcriptomic analyses, statistical significance is indicated by asterisks: **, $1 \times 10^{-5} < P < 1 \times 10^{-20}$; ****, $1 \times 10^{-80} < P < 0$.

For experimental validations, *P* values were calculated using GraphPad Prism. Statistical analysis used unpaired, two-sided t-test for bulk analysis, and two-sided Mann-Whitney test for comparison of values obtained at single-cell levels. To compare mRNA localization of low-valency and multivalent mRNAs, t-test for independent samples was performed. Statistical significance is indicated by asterisks with the following *P* value ranges: *, $P < 0.05$, **, $P < 0.01$, ***, $P < 0.001$, ****, $P < 0.0001$.

Data and code availability

PhyloP scores, RNA valency, mRNA length, and IDR length are reported in Table S1. The processed RNA-seq data for MYC and UTX/KDM6A are reported in Tables S2 and S3 and their raw data are accessible through GEO accession number GSE301135. The JMJD3/KDM6B HA and H3K27me3 ChIP seq data are accessible through GEO accession number GSE301136.

Supplementary Tables

Table S1. mRNA and protein features of human protein coding genes, related to Figure 1, 6 and 7.

Table S2. RNA-seq analysis of 3′UTR-regulated MYC activity, related to Figure 1I.

Table S3. RNA-seq analysis of 3′UTR-regulated UTX transcriptional targets, related to Figure 1J.

Table S4. mRNA and protein features of the ten candidates experimentally tested for 3′UTR-dependent protein oligomerization states, related to Figure 7.

Table S5. Sequences of oligonucleotides used, related to Key Resources Table.

Supplementary Figure legends

Figure S1. mRNAs with HC 3′UTRs encode large proteins.

A. Proportion of nucleotides assigned to 5′UTRs, CDS, and 3′UTRs in human mRNAs (N = 19556)[2].

B. Protein length distributions based on the number of conserved nucleotides (nt) in human 3′UTRs obtained from PhyloP. Non-conserved (NC), 0 conserved nt (N = 7364); highly conserved (HC) >150 conserved nt (N = 2714). Mann-Whitney test, $P = 4.5 \times 10^{-176}$.

C. As in (B) but shown is IDR length distribution when classified using pLDDT scores from AlphaFold2. IDRs are defined as pLDDT < 80. Mann-Whitney test, $P = 5.5 \times 10^{-248}$.

D. Live-cell fluorescence microscopy of GFP-MYC-NU or GFP-MYC-U in U2OS cells. Representative image is shown. The yellow dotted line indicates the cell outline. MYC protein localizes to the nucleus in a 3′UTR-independent manner.

E. As in (D), but anti-HA immunofluorescence staining of transfected HA-KDM6A-NU or HA-KDM6A-U in MCF7 cells.

F. Western blot of endogenous UTX/KDM6A levels in MCF7 cells stably expressing either a control shRNA or an shRNA targeting KDM6A. GAPDH was used as loading control.

Figure S2. The β-HP region in the JMJD3 IDR acts as an autoinhibitory dimerization domain.

A. As in Fig. 3F, but a different orientation is shown to visualize predicted β-sheet addition between the β-HP in the IDR of one JMJD3 molecule with the β-barrel of the JmjC domain of another JMJD3 molecule.

B. JMJD3 demethylase activity as measured by reduction of fluorescence intensity of anti-H3K27me3 from experiment in Fig. 2E and Fig. 3K are shown. HPdel results in similar JMJD3 demethylase activity as IDRdel. Mann-Whitney test, ns, not significant.

Figure S3. The JMJD3 IDR binds RNA in vitro and regulates co-translational protein dimerization.

A. Coomassie-stained SDS–PAGE showing recombinant 6xHis-MBP-tagged JMJD3 RBD.

B. Agarose gel showing in vitro transcribed Cy5-labeled *KDM6B* (1164 nt) and *HSPA1B* 3′UTRs (378 nt). The first 110 nt of the *KDM6B* 3′UTR are GC-rich and were removed to allow transcription.

C. As in Fig. 1L, but JMJD3 protein abundance of full-length and RBDdel was measured as HA intensity from $N = 3$ independent experiments as shown in Fig. 4D. Mann-Whitney test: ns, not significant.

D. Representative immunoblots of the co-translational disome assay. Shown is the abundance of newly synthesized nascent chains derived from various HA-tagged KDM6B constructs in either the monosome (M) or disome (D) fractions obtained by sucrose gradient centrifugation after their transfection into HeLa cells. As input, RPLP0 in each M or D fraction was blotted and used for normalization purposes. Puro, puromycin. Black arrows indicate the nascent chains that overlap with the full-length HA-KDM6B bands and were used for quantification. The blue arrow indicates a lower molecular weight, HA-tagged nascent chain that displayed a similar trend as the full-length band.

E. Coomassie-stained SDS–PAGE showing recombinant MBP-tagged RHP, MBP-tagged HP, and ED from JMJD3. All proteins contain a N-terminal 6xHis and a C-terminal Strep tag II.

F-I. Fluorescence polarization assay to measure RNA-protein interactions. Top: As RNA, the Cy5-labeled *KDM6B* 3′UTR (1164 nt) was used. Bottom: As RNA, the *HSPA1B* 3′UTR (378 nt) was used. Fluorescence polarization values obtained from $N = 3$ independent experiments are shown. The values were fitted to a sigmoidal binding curve and the K_d was calculated in Prism using the one-site binding model. The dissociation constant (K_d) is reported as ± error of fitting.

F. MBP alone does not bind to the RNA.

G. As in (F), but as protein the MBP-tagged RHP was used.

H. As in (F), but as protein the MBP-tagged HP was used.

I. As in (F), but as protein the ED (aa 1141-1643) was used.

J. The uncropped blot showing RNA-dependent protein-protein interactions generated through thiol-thiol crosslinking between recombinant JMJD3 RHP and ED by BMOE-mediated thiol-thiol crosslinking. The same reactions were resolved on either an 8% (left) or 6% (right) Tris-Glycine gel to allow clear visualization of individual monomer and dimer bands, respectively.

K. As in (J) but showing the crosslinking experiment between JMJD3 HP and ED.

Figure S4. HC 3′UTRs are multivalent.

A. Number of G/C homopolymers in 3′UTRs that are associated with droplet-like condensates ($N = 19$) or that induced formation of network-like condensates ($N = 27$). Data were obtained from reference[9]. Mann-Whitney test, $P = 0.47$ (ns).

B. As in (A), but the number of A/U homopolymers was determined. Mann-Whitney test, $P < 0.001$.

C. Number of G/C homopolymers in mRNAs enriched in the cytosol (*N* = 1328) versus cytoplasmic network-like condensates (*N* = 3888). Mann-Whitney test, *P* = 0.70 (ns).

D. RNA valency correlates with mRNA length, but for each mRNA length bin, HC 3′UTRs have significantly higher 3′UTR valency values than NC 3′UTRs. Mann-Whitney test, $P < 10^{-6}$.

E. 3′UTR length distribution of human mRNAs separated by NC (*N* = 7947) vs HC 3′UTRs (*N* = 2714). Mann-Whitney test, *P* = 0.

F. As in (E) but RNA valency in 3′UTRs is shown. Mann-Whitney test, *P* = 0.

Figure S5. Heterologous multivalent 3′UTRs promote full JMJD3 demethylase activity.

A. RNA-FISH (white) of mRNAs from the indicated transfected constructs in HeLa cells before (-) and after (+) digitonin extraction. Thin yellow dashed lines indicate cell boundaries, and thick dotted lines mark the nucleus. Representative images are shown. DAPI (blue) stains nuclei.

B. As in (A), but RNA-FISH was performed on *KDM6B* mRNA with swapped 3′UTRs.

C. Immunofluorescence assay to assess 3′UTR-dependent JMJD3 activity in HeLa cells after transfection of the indicated constructs. Nuclei were stained by DAPI. Anti-HA staining (green) shows transfected JMJD3 protein, and a reduction in anti-H3K27me3 staining (magenta) is used to infer JMJD3 demethylase activity. White arrowheads mark transfected cells. Representative images are shown.

D. Quantification of transfected JMJD3 abundance as HA intensity values from (C) and from Fig. 5M. Shown are boxplots generated from *N* = 3 independent experiments. Mann-Whitney test: ns, not significant; **, *P* <0.01; ****, *P* < 0.0001.

Figure S6. Low-valency mRNAs with non-conserved 3′UTRs do not influence the oligomerization states of their encoded proteins.

A-E. For each candidate to be validated, shown are mRNA multivalency (indicated by white dots), 3′UTR conservation (indicated by the color of the 3′UTR with NC in grey and HC in blue), and protein domains (annotated structured domains from uniprot are indicated in brown). Protein oligomerization patterns in cells after transfection of the indicated HA-tagged constructs were analyzed by DSG cross-linking and western blot analysis. Distinct oligomeric states are indicated by Roman numerals with increasing shades of gray indicating higher order of protein assemblies.

A. DSG crosslinking western blot of HA-tagged EIF3I shows 3′UTR-independent protein oligomerization pattern. Actin was used as loading control.

B. As in (A) but shown is HA-RACK1.

C. As in (A) but shown is HA-SSRP1. Tubulin was used as loading control.

D. As in (A) but shown is HA-SF3B4.

E. As in (A) but shown is HA-MYC, showing a protein oligomerization pattern that is altered by inclusion of the 3′UTR in the mRNA template. GAPDH was used as loading control.

Figure 1

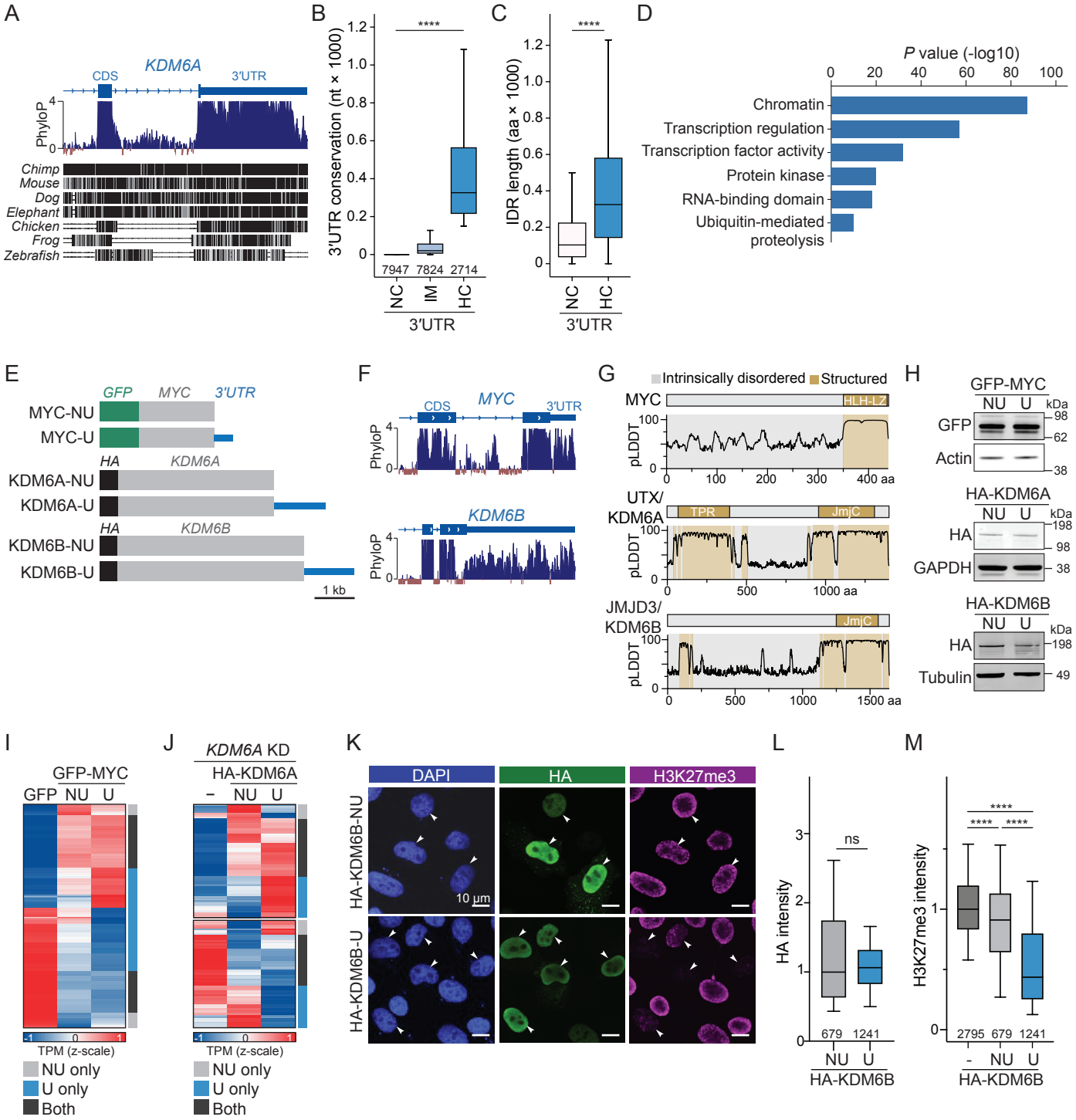

A
KDM6A
CDS
3'UTR
PhyloP
Chimp
Mouse
Dog
Elephant
Chicken
Frog
Zebrafish
B
3'UTR conservation (nt × 1000)
****
7947 7824 2714
NC IM HC
3'UTR
C
IDR length (aa × 1000)
****
NC HC
3'UTR
D
P value (-log10)
Chromatin
Transcription regulation
Transcription factor activity
Protein kinase
RNA-binding domain
Ubiquitin-mediated proteolysis
E
GFP MYC 3'UTR
MYC-NU
MYC-U
HA KDM6A
KDM6A-NU
KDM6A-U
HA KDM6B
KDM6B-NU
KDM6B-U
1 kb
F
CDS MYC 3'UTR
PhyloP
KDM6B
G
Intrinsically disordered
Structured
MYC
HLH-LZ
pLDDT
400 aa
UTX/ KDM6A
TPR
JmjC
1000 aa
JMJD3/ KDM6B
1500 aa
H
GFP-MYC
NU U
kDa
GFP
Actin
HA-KDM6A
HA
GAPDH
HA-KDM6B
Tubulin
I
GFP-MYC
GFP NU U
TPM (z-scale)
NU only
U only
Both
J
KDM6A KD
HA-KDM6A
– NU U
TPM (z-scale)
NU only
U only
Both
K
DAPI
HA
H3K27me3
HA-KDM6B-NU
HA-KDM6B-U
10 µm
L
HA intensity
ns
679 1241
NU U
HA-KDM6B
M
H3K27me3 intensity
****
2795 679 1241
- NU U
HA-KDM6B

Figure 2

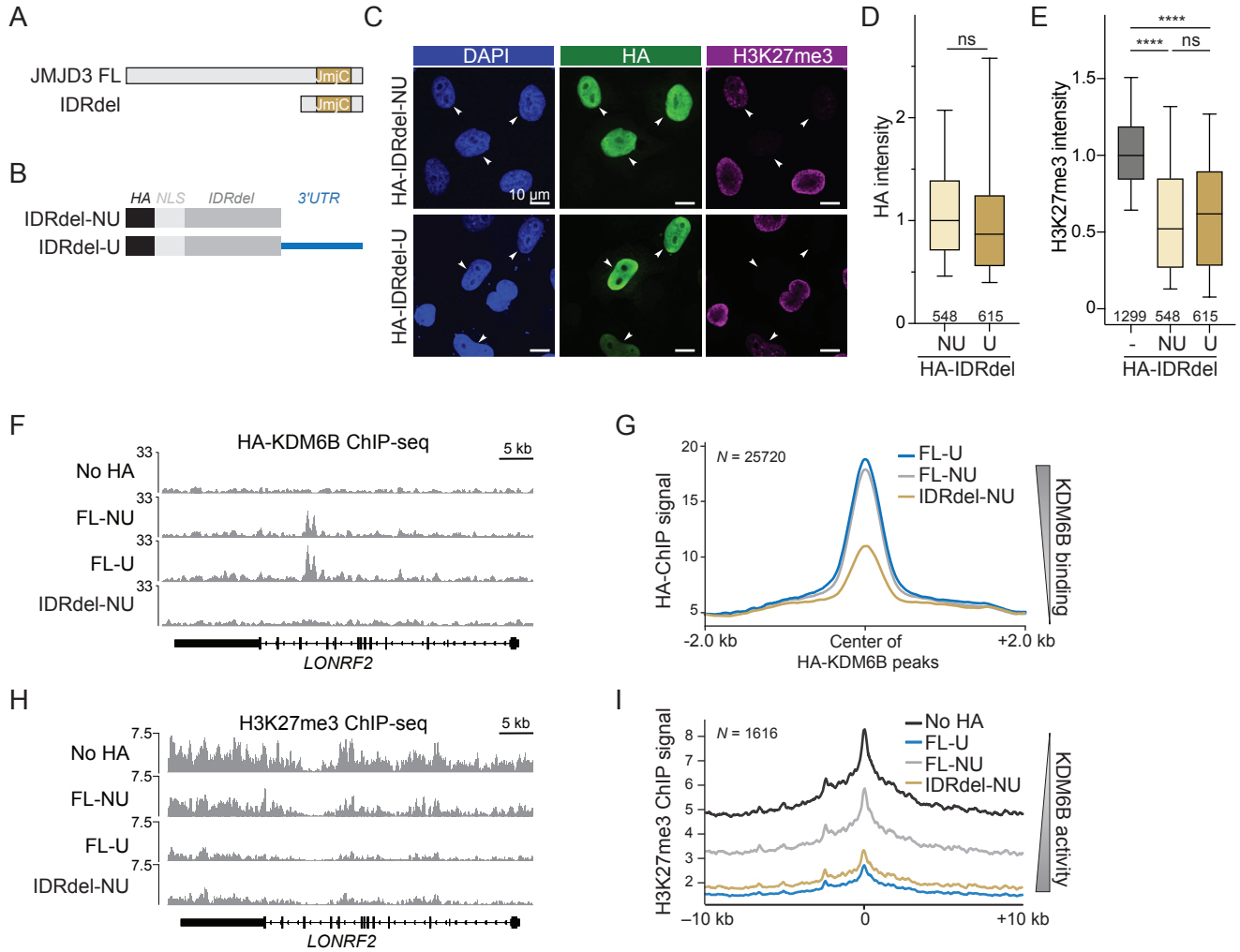

A
JMJD3 FL
JmjC
IDRdel
JmjC
B
HA NLS IDRdel 3'UTR
IDRdel-NU
IDRdel-U
C
DAPI
HA
H3K27me3
HA-IDRdel-NU
HA-IDRdel-U
10 µm
D
ns
HA intensity
548 615
NU U
HA-IDRdel
E
****
****
ns
H3K27me3 intensity
1299 548 615
- NU U
HA-IDRdel
F
HA-KDM6B ChIP-seq
5 kb
No HA
FL-NU
FL-U
IDRdel-NU
LONRF2
G
N = 25720
FL-U
FL-NU
IDRdel-NU
HA-ChIP signal
KDM6B binding
-2.0 kb
Center of
HA-KDM6B peaks
+2.0 kb
H
H3K27me3 ChIP-seq
5 kb
No HA
FL-NU
FL-U
IDRdel-NU
LONRF2
I
N = 1616
No HA
FL-U
FL-NU
IDRdel-NU
H3K27me3 ChIP signal
KDM6B activity
–10 kb
0
+10 kb

Figure 3

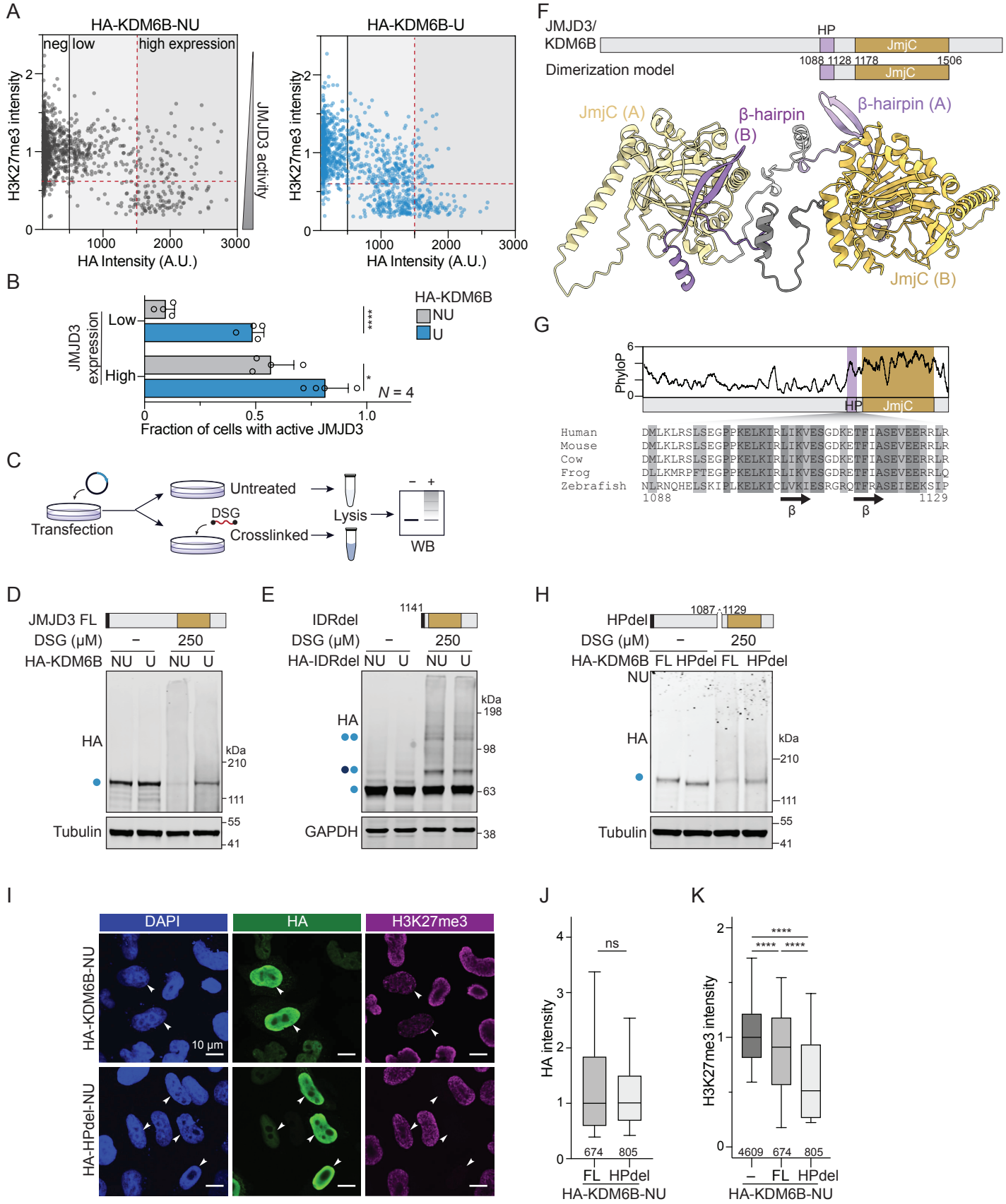
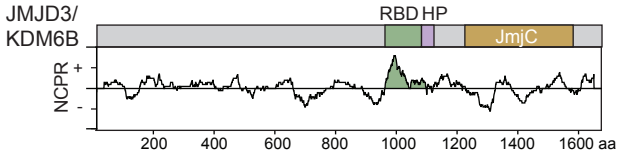
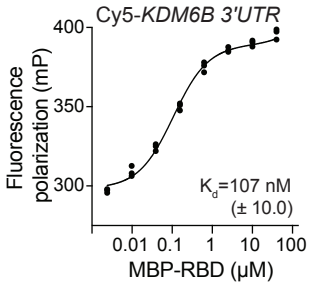

A
HA-KDM6B-NU
HA-KDM6B-U
neg
low
high expression
H3K27me3 intensity
HA Intensity (A.U.)
JMJD3 activity
B
JMJD3 expression
Low
High
HA-KDM6B
NU
U
N = 4
Fraction of cells with active JMJD3
C
Transfection
Untreated
DSG
Crosslinked
Lysis
WB
D
JMJD3 FL
DSG (μM)
HA-KDM6B
HA
Tubulin
kDa
E
IDRdel
HA-IDRdel
GAPDH
F
JMJD3/ KDM6B
HP
JmjC
Dimerization model
JmjC (A)
β-hairpin (B)
β-hairpin (A)
JmjC (B)
G
PhyloP
Human
Mouse
Cow
Frog
Zebrafish
H
HPdel
HA-KDM6B
I
DAPI
HA
H3K27me3
HA-KDM6B-NU
HA-HPdel-NU
10 μm
J
HA intensity
ns
FL
HPdel
K
H3K27me3 intensity

Figure 4

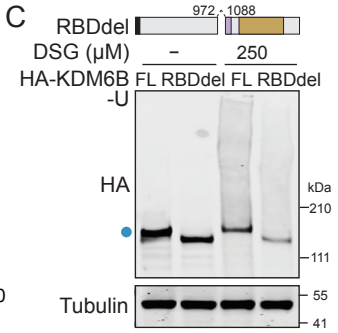
A



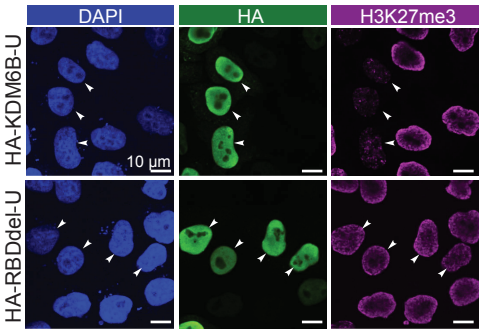
B



C



D



E

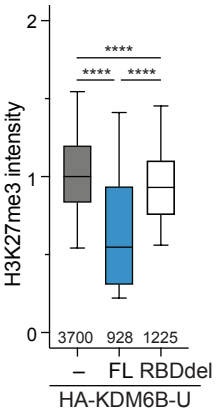


Figure 5

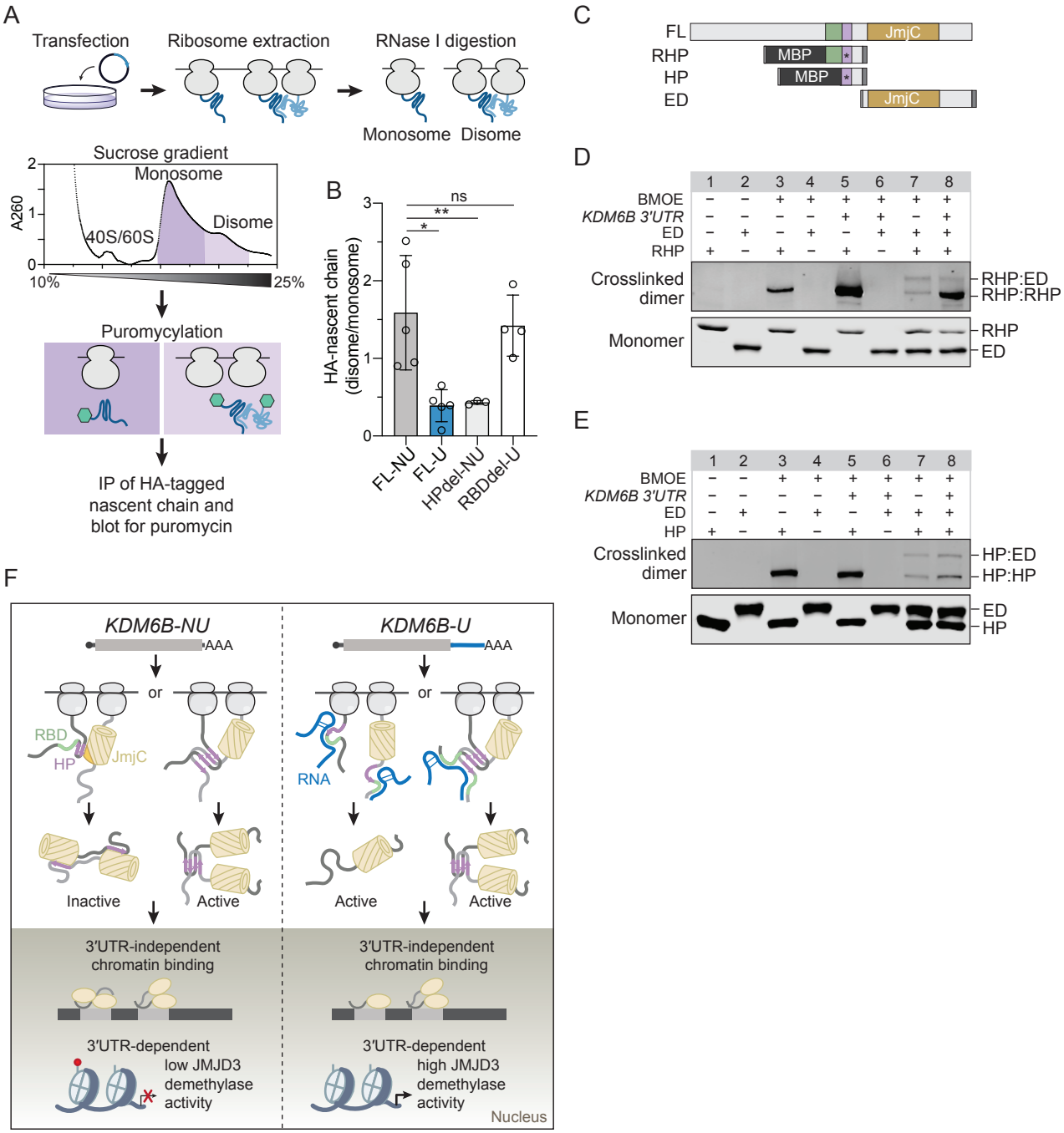

A
Transfection
Ribosome extraction
RNase I digestion
Monosome
Disome
Sucrose gradient
Monosome
Disome
40S/60S
A260
10%
25%
Puromycylation
IP of HA-tagged nascent chain and blot for puromycin
B
HA-nascent chain (disome/monosome)
ns
**
*
FL-NU
FL-U
HPdel-NU
RBDdel-U
C
FL
RHP
HP
ED
JmjC
MBP
D
BMOE
KDM6B 3'UTR
ED
RHP
Crosslinked dimer
Monomer
RHP:ED
RHP:RHP
RHP
ED
E
BMOE
KDM6B 3'UTR
ED
HP
Crosslinked dimer
Monomer
HP:ED
HP:HP
ED
HP
F
KDM6B-NU
KDM6B-U
AAA
or
RBD
HP
JmjC
RNA
Inactive
Active
3'UTR-independent chromatin binding
3'UTR-dependent
low JMJD3 demethylase activity
high JMJD3 demethylase activity
Nucleus

Figure 6

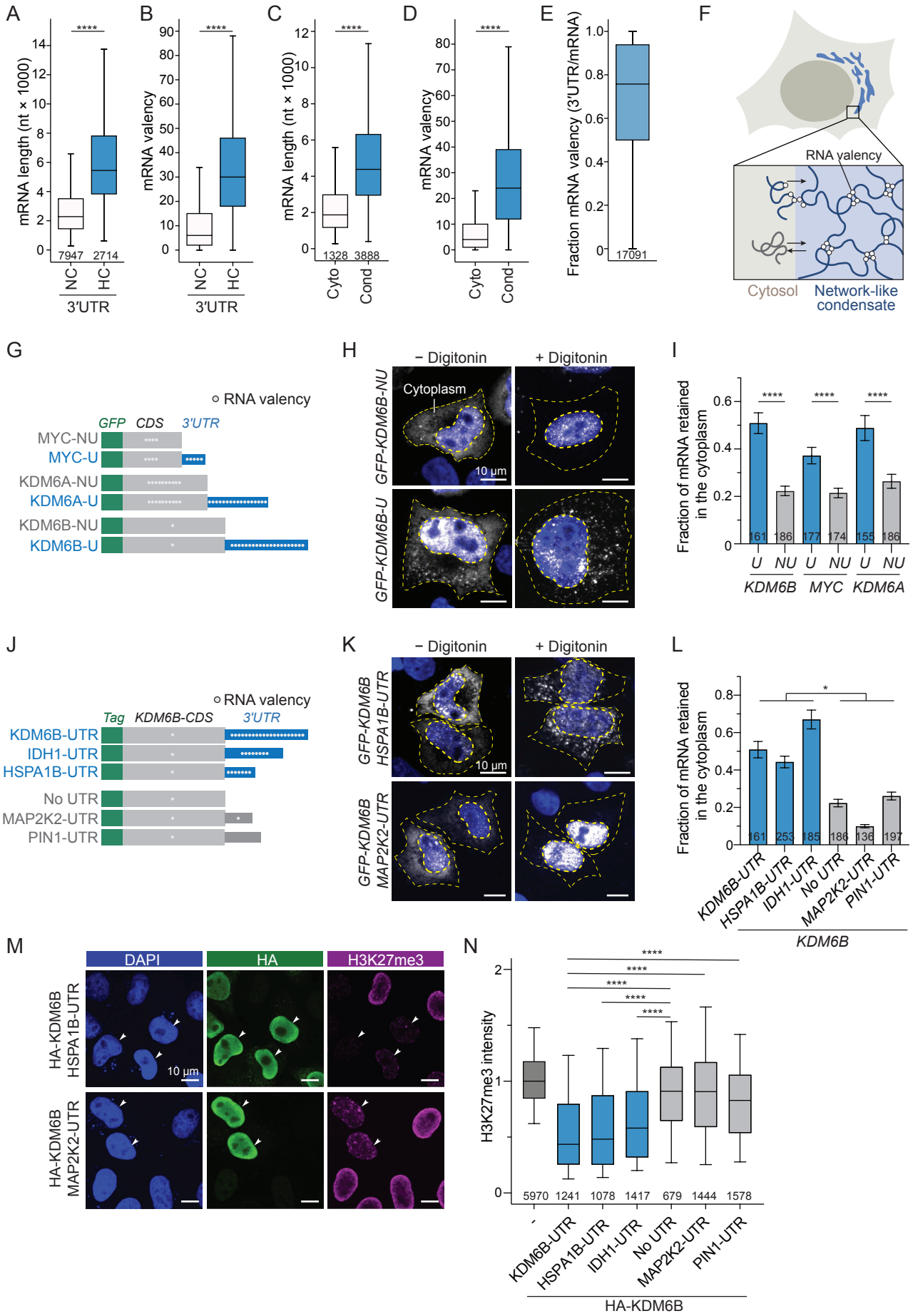

A
mRNA length (nt × 1000)
****
7947
2714
NC
HC
3'UTR
B
mRNA valency
****
NC
HC
3'UTR
C
mRNA length (nt × 1000)
****
1328
3888
Cyto
Cond
D
mRNA valency
****
Cyto
Cond
E
Fraction mRNA valency (3'UTR/mRNA)
17091
F
RNA valency
Cytosol
Network-like condensate
G
RNA valency
GFP
CDS
3'UTR
MYC-NU
MYC-U
KDM6A-NU
KDM6A-U
KDM6B-NU
KDM6B-U
H
− Digitonin
+ Digitonin
GFP-KDM6B-NU
GFP-KDM6B-U
Cytoplasm
10 µm
I
Fraction of mRNA retained in the cytoplasm
****
161
186
177
174
155
186
U
NU
KDM6B
MYC
KDM6A
J
RNA valency
Tag
KDM6B-CDS
3'UTR
KDM6B-UTR
IDH1-UTR
HSPA1B-UTR
No UTR
MAP2K2-UTR
PIN1-UTR
K
− Digitonin
+ Digitonin
GFP-KDM6B HSPA1B-UTR
GFP-KDM6B MAP2K2-UTR
10 µm
L
Fraction of mRNA retained in the cytoplasm
*
161
253
185
186
136
197
KDM6B-UTR
HSPA1B-UTR
IDH1-UTR
No UTR
MAP2K2-UTR
PIN1-UTR
KDM6B
M
DAPI
HA
H3K27me3
HA-KDM6B HSPA1B-UTR
HA-KDM6B MAP2K2-UTR
10 µm
N
H3K27me3 intensity
****
5970
1241
1078
1417
679
1444
1578
KDM6B-UTR
HSPA1B-UTR
IDH1-UTR
No UTR
MAP2K2-UTR
PIN1-UTR
HA-KDM6B

Figure 7

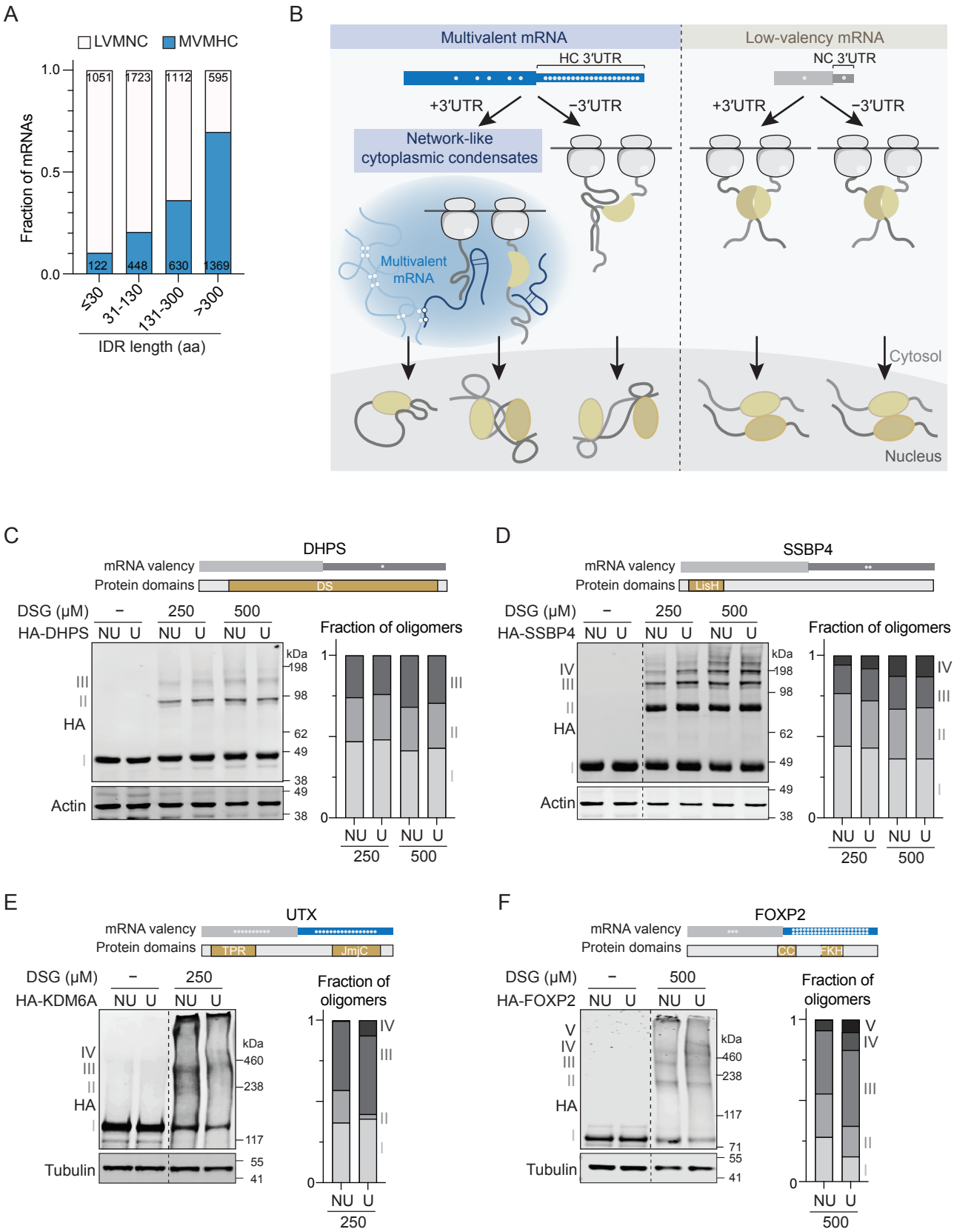

A
LVMNC
MVMHC
Fraction of mRNAs
1.0
0.5
0.0
1051
1723
1112
595
122
448
630
1369
≤30
31-130
131-300
>300
IDR length (aa)
B
Multivalent mRNA
Low-valency mRNA
HC 3′UTR
NC 3′UTR
+3′UTR
−3′UTR
+3′UTR
−3′UTR
Network-like cytoplasmic condensates
Multivalent mRNA
Cytosol
Nucleus
C
DHPS
mRNA valency
Protein domains
DS
DSG (μM)
−
250
500
HA-DHPS
NU U NU U NU U
kDa
198
98
62
49
38
49
38
III
II
HA
I
Actin
Fraction of oligomers
1
0
III
II
I
NU U NU U
250 500
D
SSBP4
mRNA valency
Protein domains
LisH
DSG (μM)
−
250
500
HA-SSBP4
NU U NU U NU U
IV
III
II
HA
I
Actin
kDa
198
98
62
49
49
38
Fraction of oligomers
1
0
IV
III
II
I
NU U NU U
250 500
E
UTX
mRNA valency
Protein domains
TPR
JmjC
DSG (μM)
−
250
HA-KDM6A
NU U NU U
IV
III
II
HA
I
Tubulin
kDa
460
238
117
55
41
Fraction of oligomers
1
0
IV
III
II
I
NU U
250
F
FOXP2
mRNA valency
Protein domains
CC
FKH
DSG (μM)
−
500
HA-FOXP2
NU U NU U
V
IV
III
II
HA
I
Tubulin
kDa
460
238
117
71
55
41
Fraction of oligomers
1
0
V
IV
III
II
I
NU U
500

Figure S1

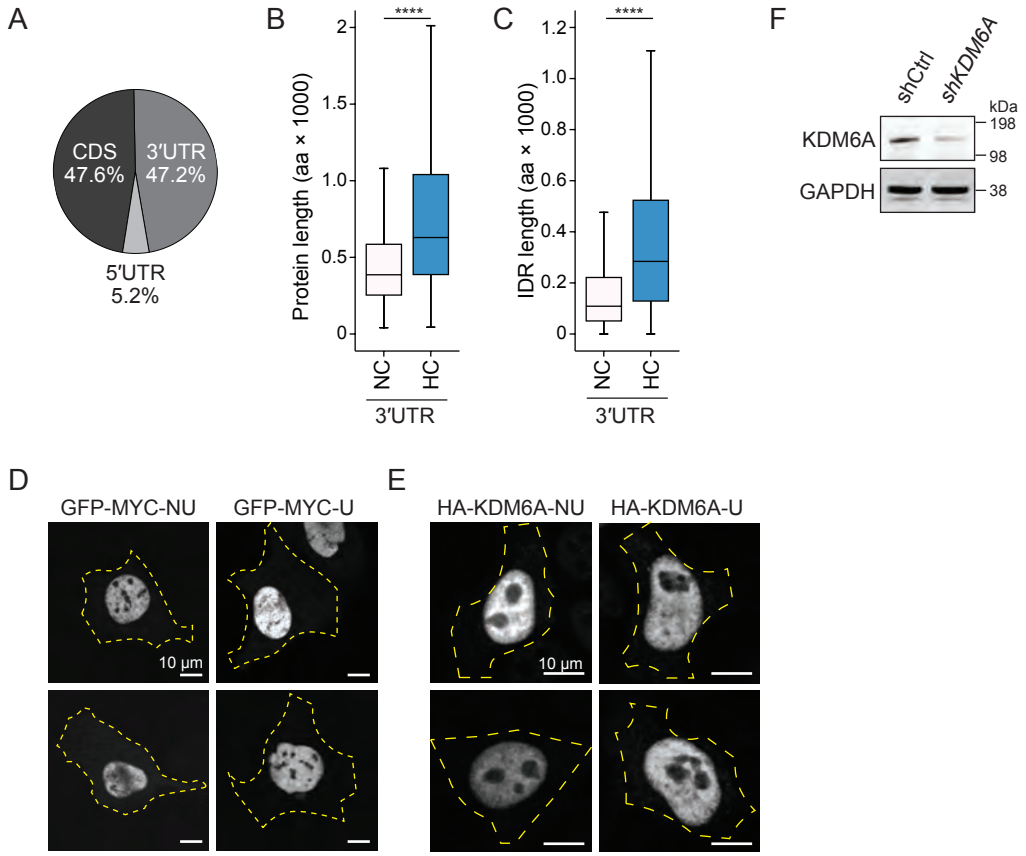

A
CDS
47.6%
3′UTR
47.2%
5′UTR
5.2%
B
****
Protein length (aa × 1000)
2
1.5
1.0
0.5
0
NC
HC
3′UTR
C
****
IDR length (aa × 1000)
1.2
1.0
0.8
0.6
0.4
0.2
0
NC
HC
3′UTR
F
shCtrl
shKDM6A
kDa
KDM6A
198
98
GAPDH
38
D
GFP-MYC-NU
GFP-MYC-U
10 μm
E
HA-KDM6A-NU
HA-KDM6A-U
10 μm

Figure S2

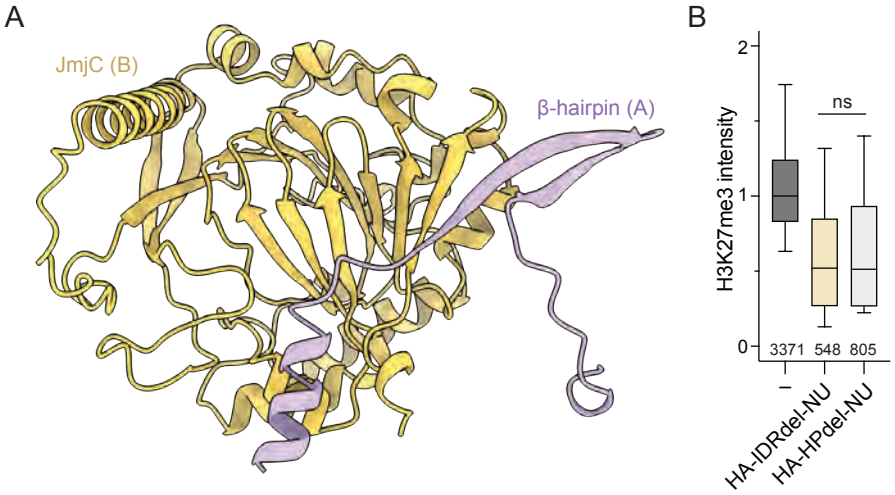

A
JmjC (B)
β-hairpin (A)
B
H3K27me3 intensity
2
1
0
ns
3371
548
805
–
HA-IDRdel-NU
HA-HPdel-NU

Figure S3

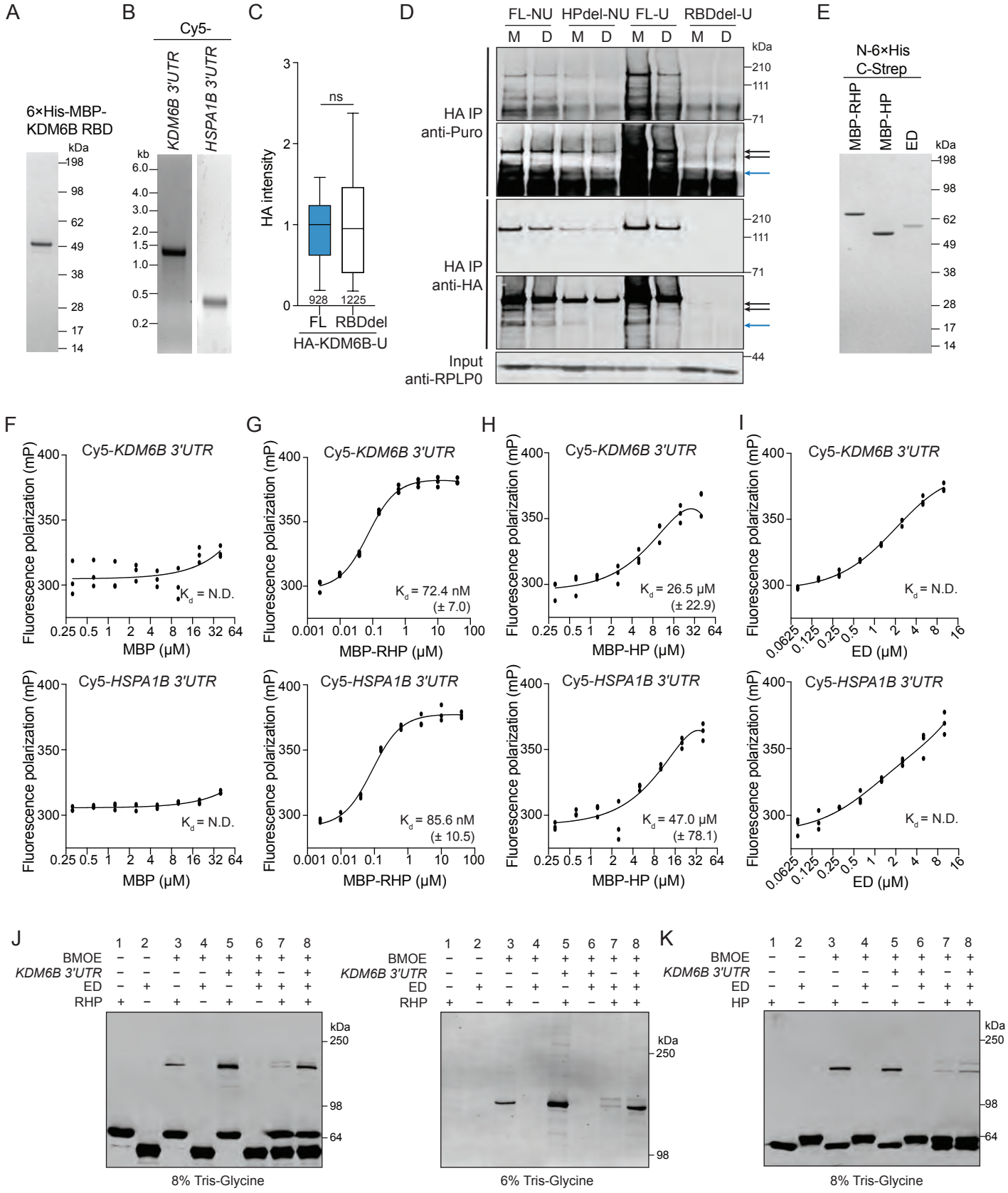

A
6×His-MBP-KDM6B RBD
kDa
198
98
62
49
38
28
17
14
B
Cy5-
KDM6B 3'UTR
HSPA1B 3'UTR
kb
6.0
4.0
3.0
2.0
1.5
1.0
0.5
0.2
C
HA intensity
ns
928
1225
FL
RBDdel
HA-KDM6B-U
D
FL-NU
HPdel-NU
FL-U
RBDdel-U
M D M D M D M D
HA IP anti-Puro
HA IP anti-HA
Input anti-RPLP0
kDa
210
111
71
44
E
N-6×His C-Strep
MBP-RHP
MBP-HP
ED
F
Cy5-KDM6B 3'UTR
Cy5-HSPA1B 3'UTR
Fluorescence polarization (mP)
MBP (μM)
Kd = N.D.
G
MBP-RHP (μM)
Kd = 72.4 nM (± 7.0)
Kd = 85.6 nM (± 10.5)
H
MBP-HP (μM)
Kd = 26.5 μM (± 22.9)
Kd = 47.0 μM (± 78.1)
I
ED (μM)
J
BMOE
KDM6B 3'UTR
ED
RHP
8% Tris-Glycine
6% Tris-Glycine
K
HP
8% Tris-Glycine

Figure S4

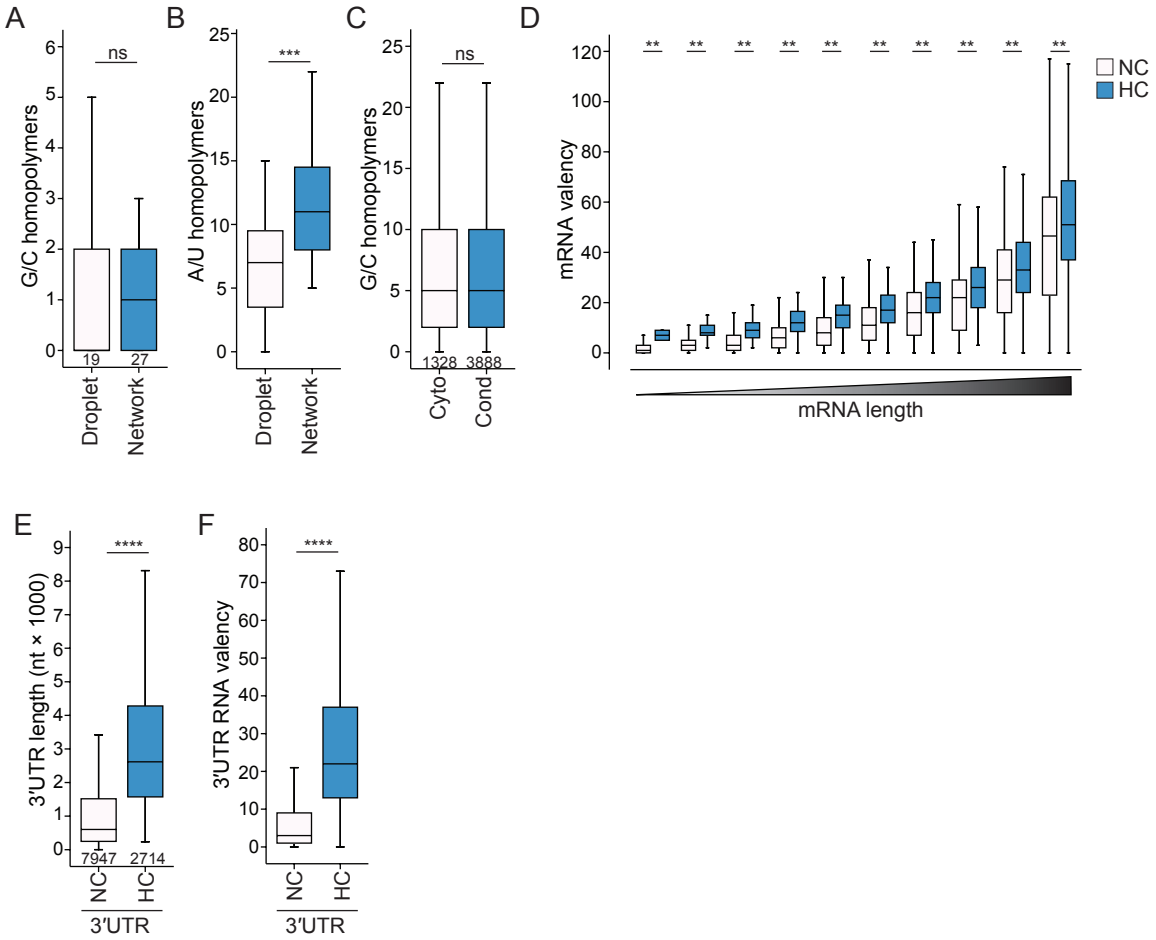

A
B
C
D
E
F
G/C homopolymers
A/U homopolymers
G/C homopolymers
mRNA valency
3'UTR length (nt × 1000)
3'UTR RNA valency
ns
***
ns
**
****
****
19
27
1328
3888
7947
2714
Droplet
Network
Cyto
Cond
NC
HC
3'UTR
mRNA length

Figure S5

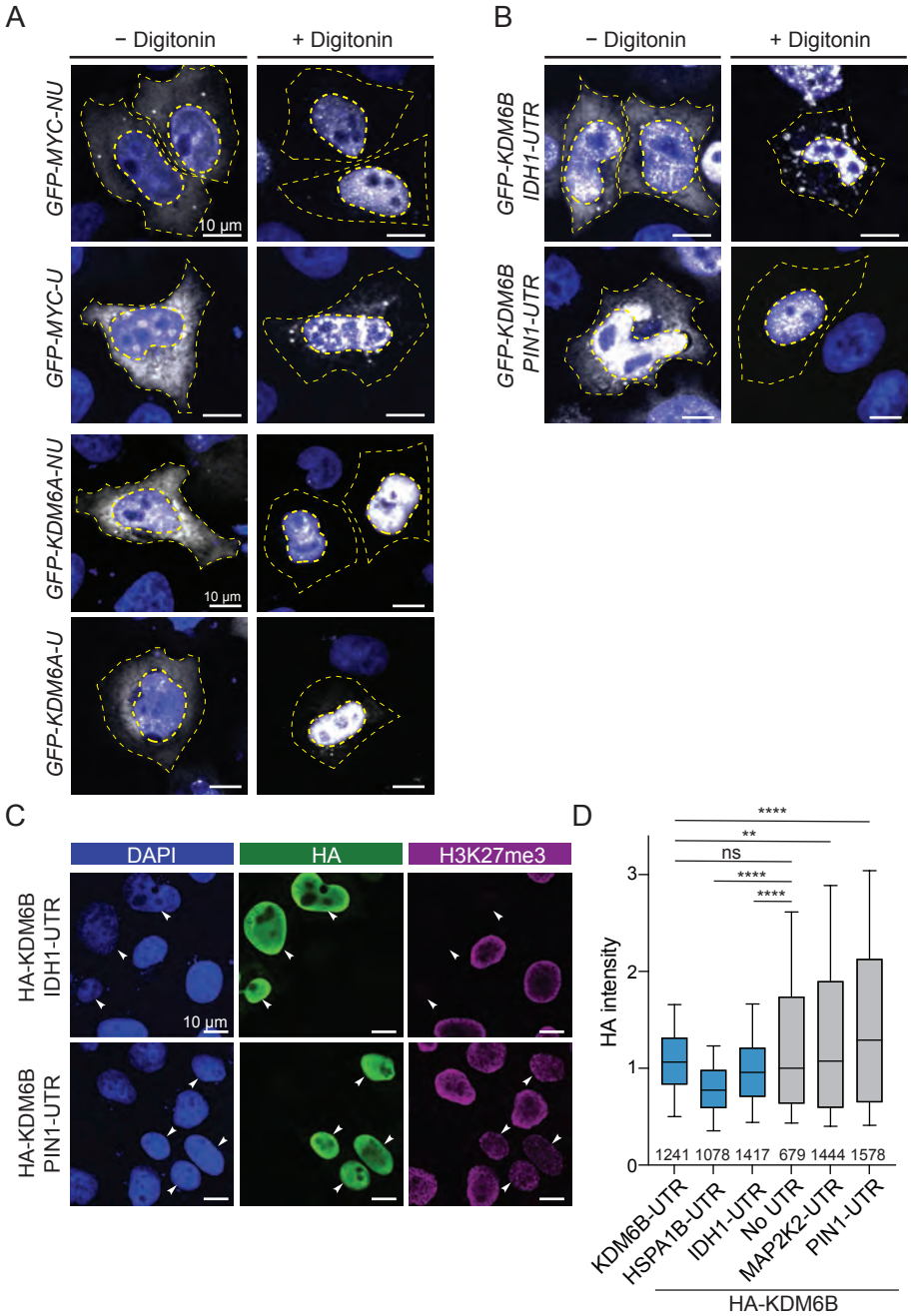
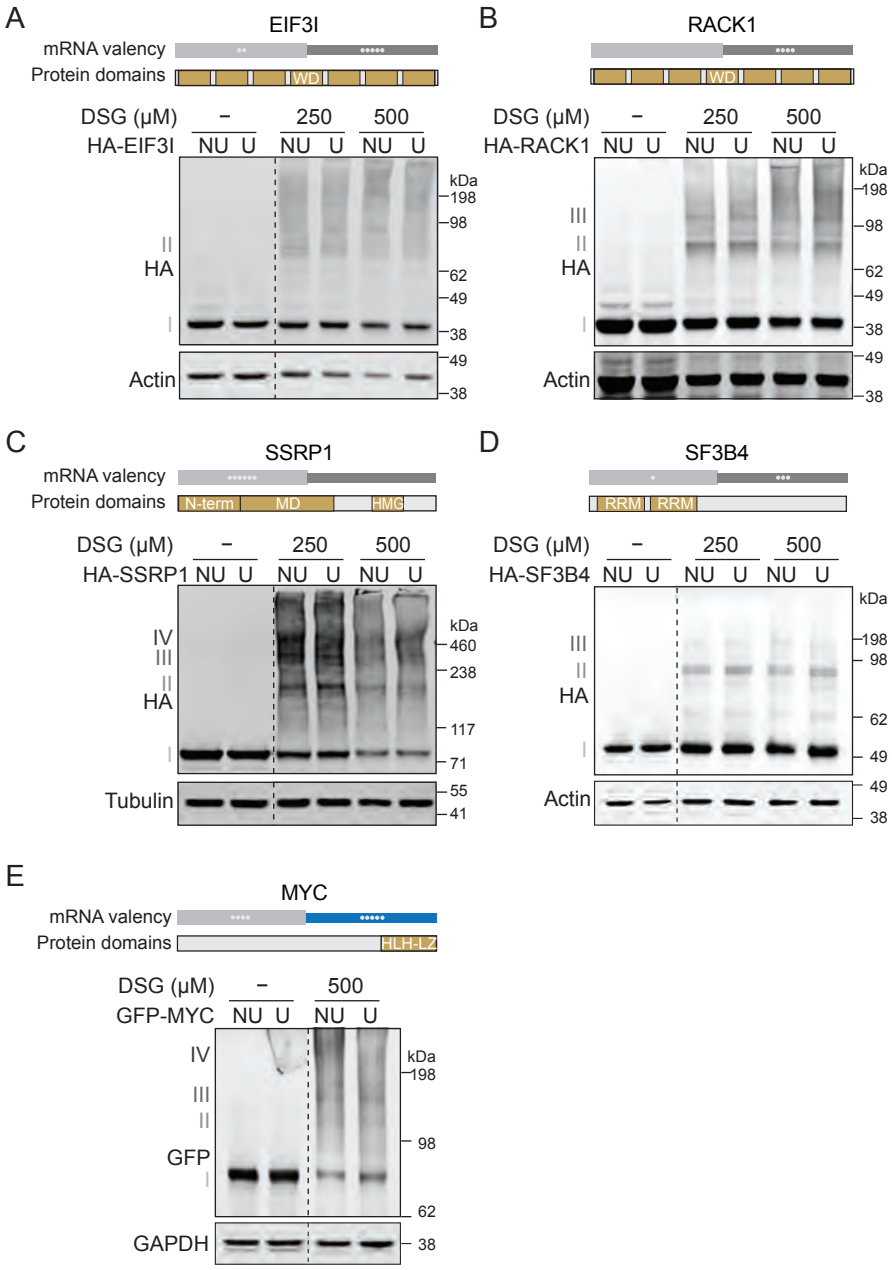

A
− Digitonin
+ Digitonin
GFP-MYC-NU
GFP-MYC-U
GFP-KDM6A-NU
GFP-KDM6A-U
10 µm
B
− Digitonin
+ Digitonin
GFP-KDM6B IDH1-UTR
GFP-KDM6B PIN1-UTR
C
DAPI
HA
H3K27me3
HA-KDM6B IDH1-UTR
HA-KDM6B PIN1-UTR
10 µm
D
HA intensity
0
1
2
3
****
**
ns
****
****
1241 1078 1417 679 1444 1578
KDM6B-UTR
HSPA1B-UTR
IDH1-UTR
No UTR
MAP2K2-UTR
PIN1-UTR
HA-KDM6B

Figure S6


A
EIF3I
mRNA valency
Protein domains
WD
DSG (μM) – 250 500
HA-EIF3I NU U NU U NU U
kDa
198
98
62
49
38
II
HA
I
Actin
49
38
B
RACK1
WD
DSG (μM) – 250 500
HA-RACK1 NU U NU U NU U
III
II
HA
I
Actin
C
SSRP1
mRNA valency
Protein domains
N-term
MD
HMG
DSG (μM) – 250 500
HA-SSRP1 NU U NU U NU U
IV
III
II
HA
I
Tubulin
460
238
117
71
55
41
D
SF3B4
RRM
RRM
DSG (μM) – 250 500
HA-SF3B4 NU U NU U NU U
III
II
HA
I
Actin
E
MYC
mRNA valency
Protein domains
HLH-LZ
DSG (μM) – 500
GFP-MYC NU U NU U
IV
III
II
GFP
I
GAPDH