Computational and Structural Biotechnology Journal 27 (2025) 2719–2729

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Short communication

Design and characterization of G-quadruplex RNA aptamers reveal RNA-binding by KDM5 lysine demethylases

Johanna Luige [a], Thomas Conrad [b], Alexandros Armaos [c], Annita Louloupi [d], Anna Vincent [a], David Meierhofer [d], Michael Gajhede [e], Gian Gaetano Tartaglia [c,f,*], Ulf Andersson Vang Ørom [a,**]

[a] Aarhus University, RNA Biology and Innovation, Aarhus, Denmark
[b] BIMSB, Max Delbrück Center, Berlin, Germany
[c] Italian institute of Technology, Genova, Italy
[d] Max Planck Institute for Molecular Genetics, Berlin, Germany
[e] University of Copenhagen, Copenhagen, Denmark
[f] University of Rome Sapienza, Rome, Italy

ARTICLE INFO

Keywords:
RNA
RNA binding proteins
G-quadruplex
Protein complexes

ABSTRACT

Here, we show that the histone lysine demethylases KDM5A and KDM5B can bind to RNA through interaction with G-quadruplexes, despite neither being categorized as RNA- nor G-quadruplex binding proteins across numerous experimental large-scale and computational studies. In addition to characterizing the KDM5 G-quadruplex interaction we show that RNA is directly involved in the formation of KDM5-containing protein complexes. Computational predictions and comparisons to other ARID domain containing proteins suggest that the ARID domain is directly interacting with both DNA and RNA across several proteins. Our work highlights that the RNA-binding by KDM5 lysine demethylases is dependent on recognizing G-quadruplex structures and that RNA mediates the formation of alternative KDM5-containing protein complexes.

1. Introduction

RNA-binding proteins are involved in gene expression regulation, by interacting with nascent transcripts and governing their splicing, maturation, localization and translation. Investigations to uncover novel RNA-binding proteins (RBP) have identified numerous proteins that do not harbor canonical RNA-binding domains or proteins that are characterized to carry out cellular functions unrelated to RNA [1–4].

The development of interactome capture demonstrated in two independent studies the widespread impact of RNA-binding proteins [2,5], and this experimental approach has since been widely applied to expand our knowledge of RNA-binding proteins. While the first studies used poly(A) purification from whole cell extracts, more recent work improved the specificity using LNA modified oligo(dT) to reduce unspecific protein contamination [6] and applied phenol-toluol extraction to purify RNA-binding proteins independent of the poly(A) status of the RNA [7]. The method has also been extended by us to study specific cell compartments, describing the first nuclear RNA-protein interactome using tandem purification steps to increase specificity [4]. In addition, RBPmap is a variant of interactome capture aimed at identifying RNA-binding sites on RBPs, where oligo(dT) purification is coupled to protease digestion [8].

In the combined interactome capture studies, an increasing number of nuclear proteins have been shown to interact with both RNA and chromatin, illustrating the diverse functions of RBPs in epigenetic, transcriptional as well as post-transcriptional regulation of gene expression [9, 10]. One of the key findings of our previous work focusing on the nuclear RNA-protein interactome in K562 cells is the identification of a number of dual DNA-RNA binding proteins not previously annotated as RNAbinding. Among those, we found the histone lysine demethylase KDM5A known to bind DNA and promote histone H3K4 demethylation [4]. Currently, only one other report has implicated the KDM5 enzyme in functional RNA-binding, proposing interaction to specific mRNAs through binding to their 3'UTR for the yeast KDM5 and

* Corresponding author at: Italian institute of Technology, Genova, Italy.
** Correspondence to: Aarhus University, Universitetsbyen 81, Aarhus 8000, Denmark.
E-mail addresses: gian.tartaglia@iit.it (G.G. Tartaglia), ulf.orom@mbg.au.dk (U.A.V. Ørom).

https://doi.org/10.1016/j.csbj.2025.06.027
Received 2 April 2025; Received in revised form 13 June 2025; Accepted 13 June 2025
Available online 16 June 2025

the human KDM5B proteins [11]. Identifying the molecular mechanisms involved in determining KDM5 RNAbinding could therefore add to our understanding of how RNA can affect transcription regulation, mRNA processing and epigenetic modifications.

The KDM5 family of lysine demethylases are known for their important roles in transcriptional regulation. The four highly similar paralogous enzymes in humans (KDM5A/B/C/D) are close in their domain architecture [12] (Fig. 1) and function, targeting H3K4me3 and H3K4me2 for demethylation [13]. Despite the similarities, members of the KDM5 family have different roles in transcriptional regulation [14, 15], differentiation [16–18], cell cycle regulation [19], DNA repair [20] and overall chromatin maintenance [14]. The nucleic acid binding ability of KDM5 enzymes could be attributed to the ARID domain, which is located within the catalytic Jumonji domain, splitting it into N- and C-terminal parts. The ARID-containing proteins are known to interact with DNA and facilitate transcription and chromatin regulation [21]. Through DNA pulldown assays, the preferred binding motif for KDM5A was determined as CCGCC [22] and for KDM5B GCACA [23], however, KDM5 ARID domains were initially not reported to display sequence specificity [21]. We propose that the diversity in sequence and nucleic acid specificity among ARID-proteins could form the potential to interact with both DNA and RNA [24].

Proteins capable of recognizing DNA and RNA are often linked to several layers of gene expression regulation, acting as transcription factors as well as post-transcriptional regulators [25–27]. Interaction with DNA and RNA can occur through the same protein domains leading to competitive outcomes [28]. As DNA and RNA share inherent similarities in their composition, protein binding could also occur through structural motifs present in both nucleic acids, such as G-quadruplexes (G4s) [29, 30]. These nucleic acid secondary structures are formed through non-canonical base pairing of guanosines, folding into three-dimensional four-stranded structures. Genomic DNA harbors G4s at important regulatory regions, such as promoters and enhancers [31, 32]. Additionally, G4 formation at the telomeres, both at the level of DNA and RNA, is crucial for chromosome maintenance [33–35]. The G4 forming telomeric repeat RNA (TERRA) facilitates interaction with telomere-binding proteins [36], telomerase [37], and regulates telomere transcription through chromatin modifications [38, 39]. Within the transcriptome, G4 formation is skewed towards 5' and 3'UTR regions of mRNAs [40–42], and they are present in various non-coding RNAs [43, 44]. The function of mRNA 5'UTR G4s has been shown to include translational regulation [45–50], while G4s in the 3'UTR can regulate alternative polyadenylation and microRNA binding [51, 52]. G4s have also been shown to occur in human ribosomal RNA [53], as well as within the 5'UTRs of ribosomal protein genes, where they control ribosomal protein abundance as well as the overall translational output of the cell [54]. Concomitantly, G4s are prevalent at the repetitive ribosomal DNA loci, where they play a role in the maintenance of genomic stability [55, 56]. G4 formation is tightly linked to open chromatin states and is shown to precede transcriptional activity, indicating that G4 folding is a regulator of transcription through establishing accessible chromatin [57]. Additionally, G4 formation is highly specific to cell types and developmental stages [58, 59], supporting the notion that G4s are involved in manifesting regionally open chromatin, rather than being a consequence of transcriptional activity. Although G4s are important in the regulation of transcription and translation, strict maintenance of G4 folding must occur to avoid transcription-replication stress and genome stability [60–63] as well as the accumulation of translationally inactive mRNAs [50, 64].

The structural features of RNA are important for protein interactions [65] and the majority of chromatin-associated proteins are predicted to prefer interaction with a folded G4 structure over unfolded [66]. As an example, the chromatin-modifying enzyme KDM1A has been shown to interact with TERRA, leading to the transcriptional repression of the telomeric region [39]. Another example of G4 recognition by chromatin-modifying enzymes is the PRC2. Multiple reports have proposed that RNA-binding by PRC2 subunits regulate its function [67–69], and to function dependent on the presence of G4s [70, 71]. These examples illustrate how multifaceted regulation can occur through G4 recognition on the level of DNA and RNA at the chromatin regions where G4 formation is permitted.

2. Results

2.1. KDM5 proteins bind RNA and chromatin regions with G4-forming potential

To confirm KDM5A binding to RNA as proposed by our serial interactome capture [4], we used an approach similar to the preparation of an iCLIP library. First, we crosslinked RNA to protein using UV irradiation in K562 cells and isolated KDM5A protein-RNA complexes using a KDM5A-specific antibody. After this, RNA was partially digested with varying dilutions of RNase I and radioactively labeled with ^{32}P using a PNK assay. The immunoprecipitated (IP) protein was then separated on

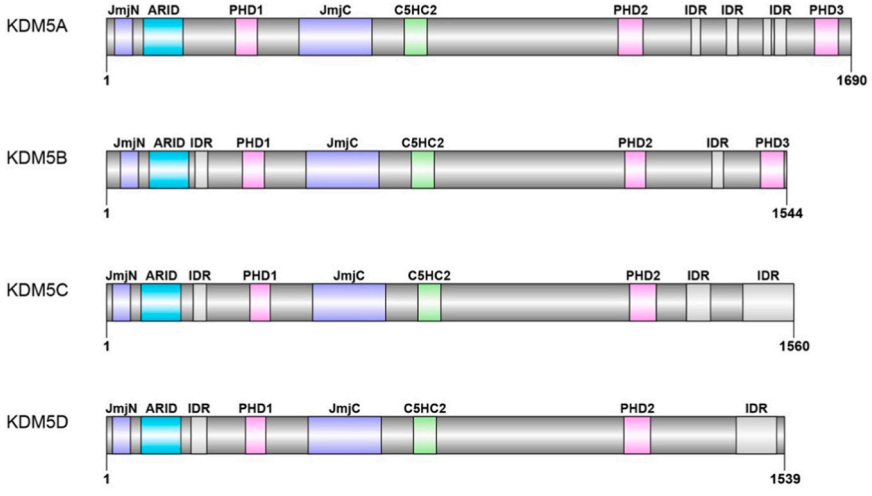


Fig. 1. Domain architecture of KDM5 proteins. Annotated protein domains from Uniprot database from N- to C-terminal direction of KDM5A, KDM5B, KDM5C and KDM5D proteins: N-terminal Jumonji (JmjN), AT-rich interacting domain (ARID) involved in nucleic acid binding, the first plant homeodomain (PHD1), C-terminal part of the catalytic Jumonji domain (JmjC), C5HC2 zinc finger domain, the second PHD2 and the intrinsically disordered regions (IDR) are common for all four proteins. KDM5A and KDM5B contain an additional third PHD3. Figure made with [72].

a gel to distinguish KDM5A from proteins binding in an unspecific way and exposed to a phosphoimager. The size of KDM5A (192 kDa) coincides with the presence of radioactively labeled RNA, demonstrating a direct interaction of cellular RNAs to the KDM5A protein (Fig. 2).

To investigate the nucleic acid preference of KDM5 enzymes, we analysed ENCODE ChIP-seq data for KDM5A and KDM5B. We assessed the chromatin binding sites of KDM5 proteins for their potential to form G4 structures, scoring each peak sequence using a computational algorithm derived from pqsfinder [73]. An example of KDM5B binding from ChIP-seq in K562 is shown in Supplementary Figure 1a along with predicted G4 forming sequences. Interestingly, we detect G4-forming potential for 50.5 % (8186 out of 16196) of binding sites for KDM5A and 60.7 % (17361 out of 28624) for KDM5B. To assess whether this is a general feature of transcription factors due to the enrichment of G4 forming sequences at promoters, we compared to G4-forming potential of 90 additional transcription factors where ChIP-seq data are available (Supplementary figure 1b, Table S1, Materials and Methods). Our analysis shows the KDM5A and KDM5B proteins at the top of the ranked list compared to other transcription factors, suggesting that KDM5 proteins exert an increased binding preference for G4-forming sequences at promoters.

Based on this analysis, we predicted RNA sequences with a propensity to interact with KDM5 proteins (Fig. 3a), employing the *cat*RAPID approach [74]. *cat*RAPID is a computational algorithm that predicts protein-RNA interactions by analyzing physicochemical properties such as secondary structure, hydrogen bonding, and van der Waals forces. It allows large-scale identification of RNA-binding proteins and their targets with high confidence [65, 75]. We sourced enriched motifs from KDM5A ChIP-Seq peaks via SeaMotE [76] analysis from ENCODE and subjected them to catRAPID omics v2.1 calculations [77] on human mRNA and long ncRNA transcripts (Materials and Methods). Upon applying stringent filters for Z-score and interaction propensity—with thresholds exceeding 3 for the Z-score and 60 for interaction propensity—we collected sequences with motifs no shorter than 6 nucleotides within a 15-nt sequence frame. This process yielded a selection of 162,895 unique 15-nucleotide RNA segments as potential aptamers. Upon evaluating the interaction propensities of various segments against KDM5 proteins, we sorted the segments by their interaction propensity. In our statistical analysis, we utilized the strongest interaction regions for each aptamer (Fig. 3A). Considering that the presence of stable RNA secondary structures promotes stronger protein binding [65], we utilized free energy calculations [78] to refine our selection of aptamers and identified UGGGGGAGGGGGCCG as a high-affinity sequence (interaction propensity with ARID > 16 for KDM5; Fig. 3A; Materials and Methods). We then used this sequence to determine which regions of the protein exhibit the highest propensity for interaction. Our analysis revealed that the N-terminus, particularly the first 200 amino acids, shows the strongest affinity for UGGGGGAGGGGGCCG (Fig. 3B). With the resolution provided by *cat*RAPID, we identified the JmjN domain (amino acids 32–73), the ARID domain (amino acids 97–187), and an intrinsically disordered region (IDR) as key binding region. Additional analysis using the G4-FUNNIES algorithm [66] further supports the domain-specific interactions of KDM5B proteins with G4 structures (Fig. 3b). G4-FUNNIES was developed to differentiate interactions with folded and unfolded G4 configurations by analyzing protein binding preferences under potassium (K^+)-rich and lithium (Li^+)-rich conditions. The results indicate that the ARID domain of KDM5B predominantly binds to folded G4 structures. Additionally, an adjacent intrinsically disordered region (amino acids 201–230) shows a preference for folded G4 binding, along with a polar compositional bias region (amino acids 1374–1390), albeit to a lesser extent. The interaction of disordered domains is particularly novel, as previous studies suggested that these regions primarily interact with single-stranded RNA. However, a subset of these regions has now been found to bind more complex structures such as double-stranded RNA [65, 79]. Conversely, PHD-type zinc finger domains in KDM5B—including PHD-type 1 (amino acids 309–359), PHD-type 2 (amino acids 1176–1224), PHD-type 3 (amino acids 1484–1538), and C5HC2 (amino acids 692–744)—do not preferentially bind folded G4 structures. These domains are primarily known for their roles in chromatin regulation and histone modification recognition [80]. Their limited interaction with G4 structures suggests a broader role in recognizing diverse DNA configurations rather than specifically targeting folded G4 conformations. This behavior could contribute to the regulation of chromatin accessibility and transcription by engaging with alternative DNA structures [81]. Furthermore, to explore the length and structural dependencies of RNA-KDM5 interactions, we analyzed a 21-nucleotide aptamer (GGCUGGGGGAGGGGGCCGGGG). This sequence, characterized by a pronounced pattern of G residues, has the potential to form a G4 [66]. It maintains secondary structure elements and exhibits an interaction propensity greater than 26 even when catRAPID calculations are run without considering the structural G4 state [66]. Using the interaction propensity, we can also show that KDM5 prefers binding to a folded over an unfolded RNA. sequence corresponding to a G4-forming sequence (Fig. 3c).

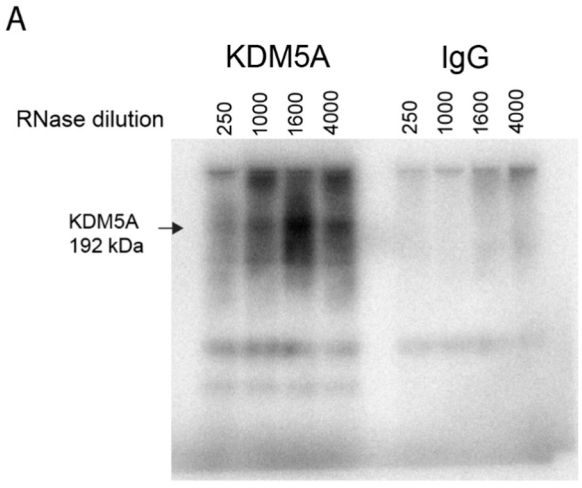


Fig. 2. KDM5A binds RNA in K562 cells. IP of KDM5A-RNA complexes using specific antibodies after UV-crosslinking in K562 cells. The IP is followed by cell lysis and trimming of RNA with RNase I and 5'end labelling of bound RNAs with ATP [γ-^{32}P] by PNK. The IP shows a dominant band corresponding to the size of the KDM5A protein indicated with an arrow.

2.2. Experimental validation of binding to predicted RNA aptamer

Next, we investigated the binding to the predicted RNA sequences *in vitro* using electrophoretic mobility shift assays (EMSA). We used both RNA and DNA aptamers (Fig. 4a) for competitive EMSA to assess the interaction preference. We show using EMSA and oligonucleotide competition that KDM5B binds to the RNA aptamer *in vitro* and that competition with cold RNA aptamer significantly decreases the signal (Fig. 4b). Meanwhile, interaction with a DNA oligonucleotide with the same sequence as the RNA aptamer shows that binding to KDM5B is weaker than with RNA and can be competed off with either RNA or DNA aptamers. The AU-rich control RNA does not result in a band shift (Fig. 4c), demonstrating a specific preference for the RNA aptamer in binding to KDM5B. We note, that the KDM5B protein alone has an apparent molecular weight of 400 kDa, corresponding to a duplex formation as previously suggested for KDM5 proteins by structural studies [82], while the aptamer-bound KDM5B protein appears at a molecular weight above 1000 kDa indicative of aggregation of proteins induced by aptamer binding.

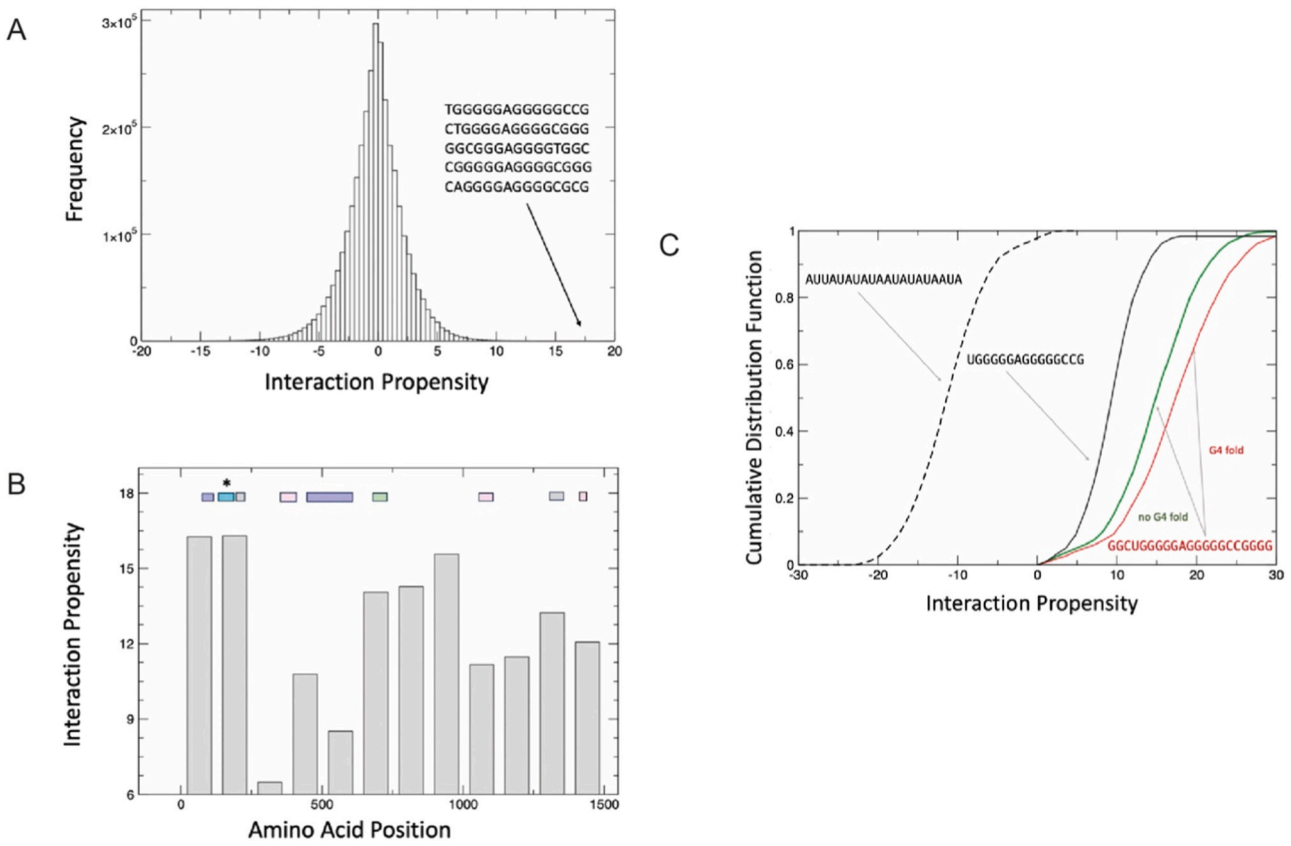


Fig. 3. Predicting aptamers that bind KDM5 proteins. A) Distribution and binding propensities of aptamers targeting the KDM5 proteins predicted by *cat*RAPID. The arrow indicates the scores of the prioritized aptamers (i.e. > 15). We note that for each generated aptamer, the region of KDM5B that exhibited the strongest interaction is used in the statistics. B) Interaction propensity of UGGGGGAGGGGGCCG with regions of the KDM5B protein, indicating that the ARID domain has the strongest binding to the aptamer. The domains defined in Fig. 1 are indicated at the top of the plot. The star indicates the region that is predicted by G4-FUNNIES to have the highest G4 binding propensity. C) The interaction propensity computed by *cat*RAPID for aptamers against KDM5B regions is ranked and displayed for negative (AUUAUAUAUAAUAUAUAAUA), positive (UGGGGGAGGGGGCCG), and extended (GGCUGGGGGAGGGGGCCGGGG) aptamers in both unfolded and folded G4 conformations.

2.3. KDM5 proteins bind in vitro to various DNA and RNA G4s

We further investigated whether the binding preference of KDM5B is due to sequence features or arises from the structure of nucleic acids. To challenge this hypothesis, we conducted EMSAs with both KDM5A and KDM5B proteins using the validated G4-forming sequence G4A4 (AAAA $[(G)_4(A)_4]_4$AAAA) [66], which were designed as both DNA and RNA oligonucleotides. Since G4 formation is strongly influenced by the ionic context, we used either K^+ or Li^+ cations in the binding reaction, where K^+ strongly stabilizes G4 folding, and Li^+ is regarded as destabilizing [83]. For both KDM5A (Fig. 5a) and KDM5B (Fig. 5b), we can detect nucleic acid band shifts for both DNA and RNA G4A4 only in the presence of K^+, suggesting that interaction is specific to G4 folding. While there seems to be a preference for RNA over DNA oligonucleotides with the predicted aptamer sequence, the G4A4 DNA and RNA are equally efficient in interacting with KDM5 proteins (Fig. 5). This is likely due to the distinct properties of the aptamer sequence, that allow G4 formation within the RNA, but not DNA, which would support the notion that KDM5 proteins preferentially interact with the G4 structure.

Next, we investigated whether KDM5 proteins binding occurs to RNA G4 sequences occurring endogenously. For this, we carried out EMSAs with the telomeric repeat containing RNA (TERRA). Indeed, G4-forming TERRA interacts with both KDM5A and KDM5B in EMSA experiments in a K^+-dependent manner (Fig. 6a-b).

2.4. KDM5 proteins form RNA-dependent complexes

To assess an impact of RNA on the KDM5 proteins, we asked if RNA binding could affect binding to other proteins as part of protein complex formation. To do this, we used pull-down of KDM5A with or without RNaseA treatment, to dissociate protein interactions dependent on RNA. We observe many proteins that interact with KDM5A in an RNA-dependent manner (Fig. 7a, Table S2). We used the CRAPome [84] to filter out interacting proteins that are frequently found to bind as unspecific background in co-IP experiments. After this filtering, we see binding to two groups of proteins. First, the canonical histone lysine demethylase complex, including SIN3B and PHF12, whose interaction with KDM5A is independent of RNase A treatment. Second, we also identify proteins not known to participate in chromatin-remodeling complexes, RPL4, DDX21, NCL, TOP2A and RPL3 that lose their interaction with KDM5A upon RNase A treatment (Fig. 7b). Several of the bound proteins are localized to the nucleolus, suggesting a nucleolar localization of KDM5 proteins. Spatiotemporal analysis of the nucleolar proteome shows KDM5 proteins as localized to both nucleoplasm and nucleolus, with KDM5D being particularly enriched in the nucleolus [85]. Interestingly, NCL [30,86–89] and DDX21 [53, 64] are known G4-binding proteins, both related to ribosomal RNA biogenesis [90–93], while RPL3 and RPL4 are ribosomal proteins, that could have a function in recognition of G4 in ribosomal RNA [53]. Therefore, RNA that can form G4s could mediate the interaction between selected proteins in complexes containing KDM5A to facilitate functions in transcription or translation.

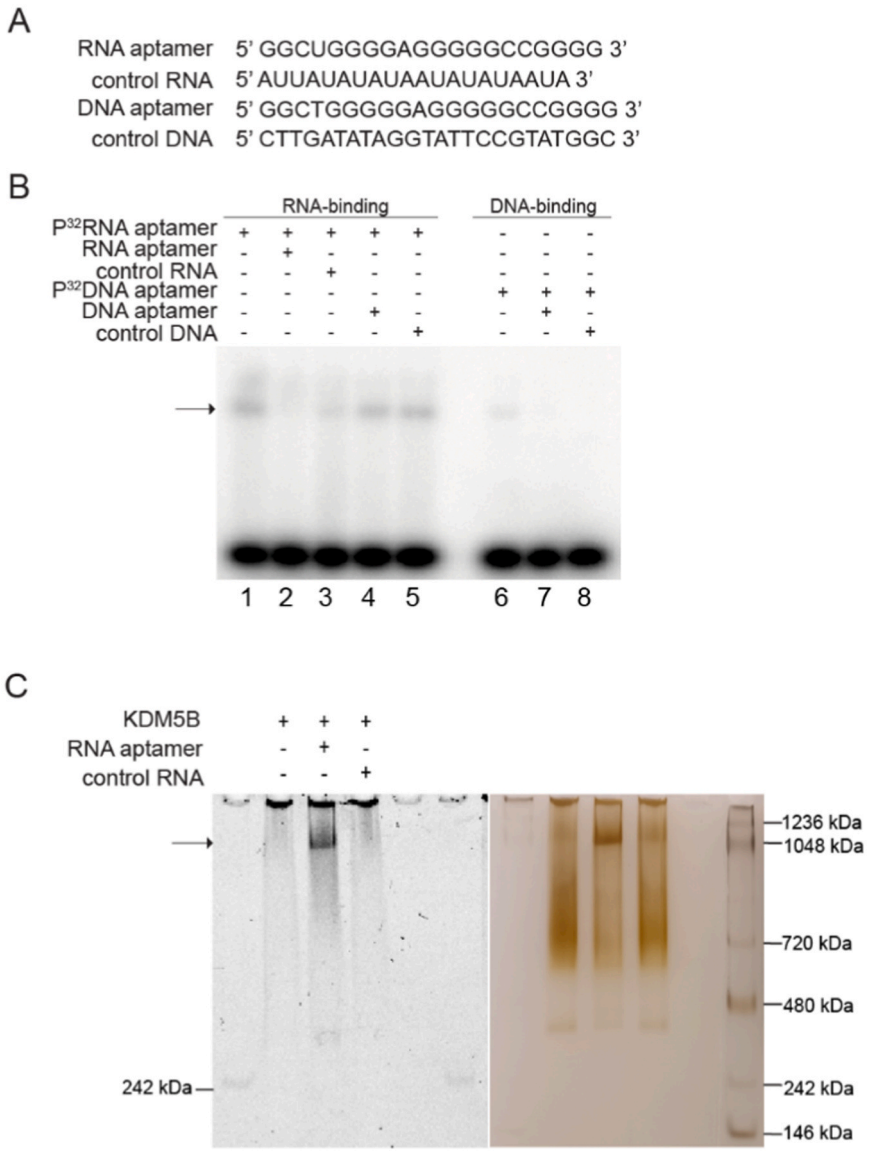


Fig. 4. KDM5B selectively interacts with the predicted RNA aptamer in vitro. A) Oligonucleotide sequences used the following EMSA experiments. B) Isotope-labelled RNA aptamer was incubated with KDM5B in a 30-minute binding reaction (lane 1) indicating a band shift. While addition of 100-fold molar excess of cold RNA aptamer (lane 2) competes off the interaction, DNA aptamer (lane 4) or control DNA (lane 5) are incapable of competition. C) Unlabelled EMSA using SYBR Gold and silver staining to illustrate both nucleic acid and protein migration in native EMSA, signifying the KDM5B-RNA aptamer complexes to cause gel shift.

3. Discussion

In this study, we identify KDM5 proteins as dual DNA-RNA binding proteins through interactions with G4s, revealing unanticipated properties of the histone lysine demethylases as they have not previously been robustly characterized as RNA binding proteins. KDM5A has been found in one nuclear interactome study to bind RNA and has been shown to bind to 3′ UTR of mRNA generally in yeast and validated for a few mRNAs in human [4, 11]. Interestingly, one of the studied mRNAs, CCND1, that was shown to interact with KDM5B [11], has been determined to harbor G4-forming regions in the 3'UTR by rG4-seq and RT-stop profiling experiments [94, 95]. The presence of G4s and their functions in 3'UTRs have been shown for several human transcripts. One study showed that the usage of an alternative polyadenylation sequence in the 3'UTR of the FXR1 transcript was influenced by the presence of a G4, resulting in the shift to a shorter transcript rather than its longer 3'UTR isoform [51]. G4s in the 3′UTR of mRNAs could have an internal regulatory role for determining alternative polyadenylation, and an array of proteins, including KDM5 enzymes, could mediate this function in different cells and for specific transcripts.

In this study we have focused on KDM5A and KDM5B proteins, while both KDM5C and KDM5D would be expected to have similar properties of binding to G4 RNA sequences due to their similar domain structure (Fig. 1). In our recent work, we built a predictor of RNA G4 binding proteins [66] but did not identify KDM5 proteins as G4-binding proteins on the whole-protein level. Our analysis presented here indicates that KDM5 proteins employ domain-specific mechanisms to recognize and interact with G4 structures, where particularly the ARID domain binds to G4 sequences, potentially impacting their function in chromatin remodeling and gene expression regulation (Fig. 3). Our study thus provides insight into an interesting case of a DNA-RNA binding protein that is challenging to assign computationally and identify experimentally due to several domains with opposing function regarding RNA binding. In fact, our data do not exclude the possibility that KDM5 proteins can also interact with other, non-G4, structures. Distinct domains within the protein may exhibit different nucleic acid binding preferences, with the ARID domain standing out as particularly predisposed to G4 interactions. This distinction is crucial, as it underscores the

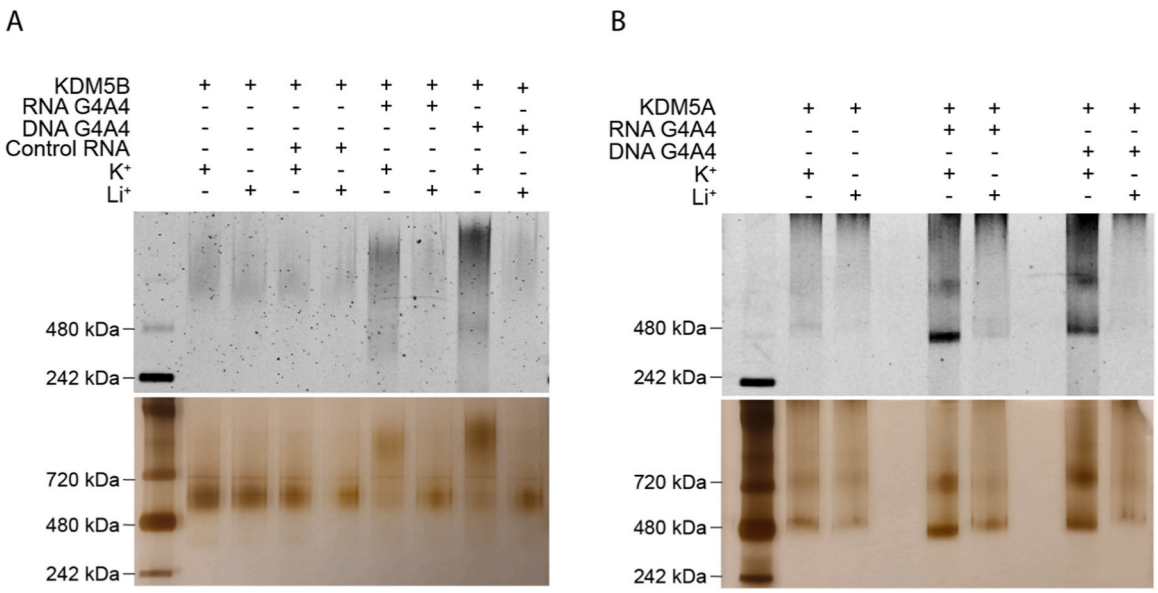


Fig. 5. KDM5A and KDM5B interact with folded DNA and RNA G4s. A) Unlabelled EMSA with KDM5B and G4A4 shows gel shifts for both DNA and RNA oligonucleotides, when incubated with K^+ in the binding buffer, that supports the folding of the G4. B) Unlabelled EMSA with KDM5A and G4A4 shows gel shifts for both DNA and RNA oligonucleotides, when incubated with K^+ in the binding buffer, that supports the folding of the G4. Top panel shows nucleotides stained with Sybr gold, and bottom panel shows silver staining of proteins from the same experiment.

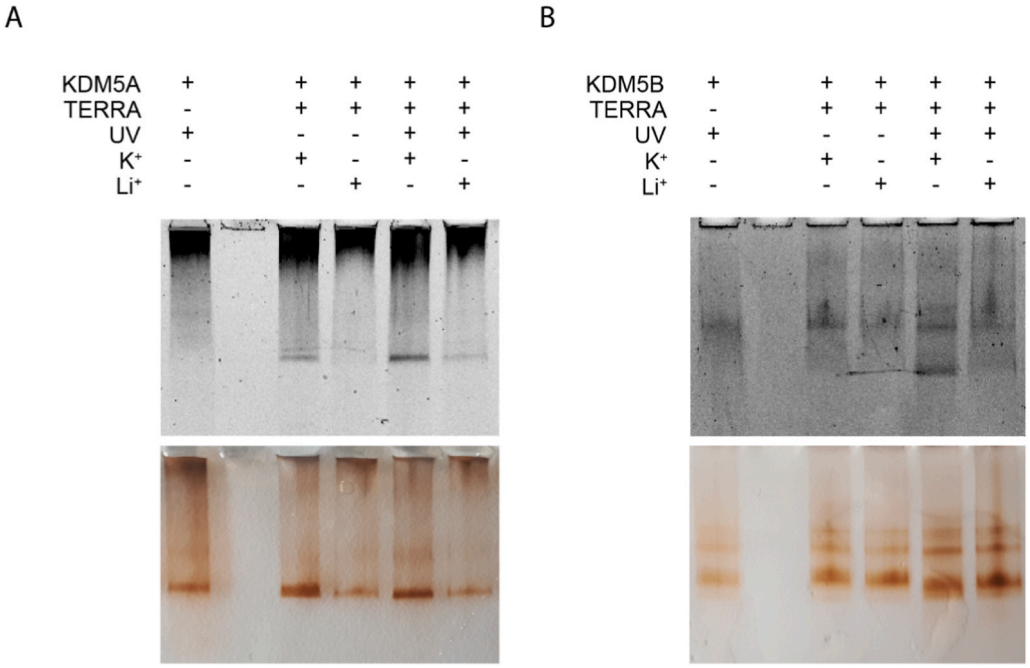


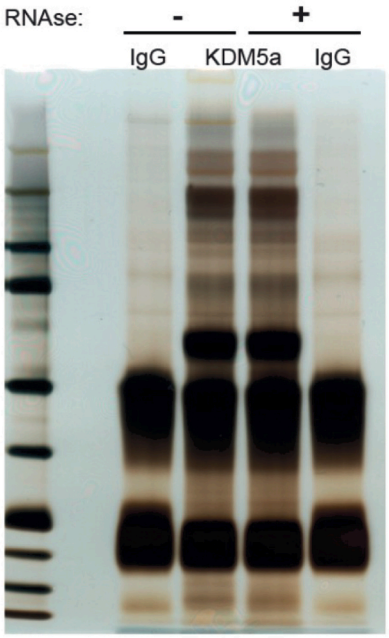
Fig. 6. KDM5A and KDM5B interact with endogenous TERRA in folded state. A) Unlabelled EMSA with KDM5B and TERRA shows gel shifts for both DNA and RNA oligonucleotides, when incubated with K^+ in the binding buffer, that supports the folding of the G4. B) Unlabelled EMSA with KDM5A and TERRA shows gel shifts for both DNA and RNA oligonucleotides, when incubated with K^+ in the binding buffer, that supports the folding of the G4. Top panel shows nucleotides stained with Sybr gold, and bottom panel shows silver staining of proteins from the same experiment.

complexity of KDM5 protein interactions with nucleic acids and suggests that these proteins might be involved in broader regulatory processes than previously appreciated. The agreement between our computational predictions using *cat*RAPID and G4-FUNNIES [66] with experimental data underscores the robustness of our approach and provides a clearer mechanistic view of KDM5 interactions with structured nucleic acids.

We designed G4 forming aptamers and predicted their binding sites within KDM5 proteins, which suggested the interaction to occur through the ARID domain. While ARID domains could mediate binding to both DNA and RNA, we detected preferential interaction with RNA in our study. As ARID5A has recently been shown to bind both AT-rich and GC-rich RNA through the ARID domain [21], we assessed where the top aptamers for KDM5s are predicted to bind and find a preference for binding to the ARID domain of ARID5A as well, albeit with lower predicted interaction propensity.

Finally, we show that several RNA-dependent protein-protein interactions occur for KDM5A, preferentially with other G4-binding proteins, pointing to a role for G4 RNAs in protein complex formation.

a)



b)

Symbol	counts -/IgG	counts +/IgG
KDM5A	104	109
SIN3B	98	110
C11orf30	90	110
PHF12	56	84
RPL4	52	10
DDX21	50	18
C4A	38	34
NCL	38	1
RPL3	32	18
TOP2A	28	1

Fig. 7. KDM5A shows RNA-dependent protein complex formation. A) IP of KDM5A with (+) or without (-) RNase treatment separated by electrophoresis and visualized by silver staining before identification of differentially bound proteins by mass spec. B) Mass spec results showing counts of KDM5A co-IP compared to IgG for both with (+) and without (-) RNase treatment after filtering data using the Crapome database.

Worth noting, we see an apparent oligomerization of both KDM5A and KDM5B proteins *in vitro* when bound to folded G4 aptamers. Both protein-protein interactions and oligomerization of KDM5 proteins suggest a potential role of G4s in aggregating proteins and proposing that RNA G4s bound to proteins could be involved in phase separation, a property of proteins binding to G4 RNA also predicted by our recent work to be a common feature of RNA G4 binding proteins [66]. The observation that G4 aptamers can induce aggregation of KDM5 proteins is supported by similar findings where G4 binding has been proposed to regulate aggregation of the protein Znf706 [96], and where RNA G4s have been shown to form scaffolds and promote a-synuclein aggregation [97].

In conclusion, our findings propose additional functions of KDM5 proteins on top of their role as histone 3 lysine 4 tri- and di- demethylases, suggest that RNA G4 binding can serve as a bridging mechanism to form protein complexes, and provide initial evidence that proteins binding to G4s can form oligomers and aggregates with potential implication for phase separation.

4. Materials and methods

4.1. PNK-assay

K562 cells were grown to a density of 1E6 per ml at 37°C in shaking flasks. Cells were harvested in a centrifuge @ 200 x g for 10 min at 4°C and resuspended in PBS for crosslinking by UV irradiation. Irradiation was done in 15 cm dishes with a total dose of 4,000E. Cells were spun again at 200 x g for 5 min at 4°C and resuspended to 25E6 cells per ml in RIPA buffer containing phosphatase inhibitors (Roche). Samples were sonicated 3 × 10 sec at output 5 and mixed with Turbo DNAse buffer (final concentration 0.5x) as well as 5 mM each of $CaCl_2$ and $MgCl_2$ and Turbo DNAse added before incubation in water bath at 37°C for 5 min. Samples were placed on ice before centrifugation at full speed for 5 min at 4°C. Samples were transferred to a new tube and 50 µl Protein A Dynabeads were added with 2 µg IgG or anti-KDM5A antibody (Abcam) and incubated rotating overnight at 4°C.

IP samples were washed 2 × 1 min with RIPA-HS containing phosphatase inhibitors and 2x with PNK buffer. After the second wash 10 µl RNAse1 (1:8000) dilution was added to 1 ml PNK buffer and incubated 3 min at 37°C shaking at 1100 rpm. Samples were placed on ice for 3 min before washing for 5 min in RIPA-HS with phosphatase inhibitors and 1:1000 RN*ase*In followed by washes 2x in PNK buffer. Samples were labeled by incubating in 20 µl PNK buffer with 1 µl P^{32}-ATP and 0.5 µl RN*ase*IN shaking for 10 min at 37°C and 1100 rpm. Samples were then washed 1x in RIPA-HS with phosphatase inhibitor and 2x in PNK buffer. Then adding 20 µl Turbo DNAse and 0.5 µl RN*ase*In to the samples, shaking for 5 min at 1100 rpm. After incubation samples were washed 1x with RIPA-HS with phosphatase inhibitor and 1x with PNK buffer before being resuspended in 30 µl NUPAGE buffer and incubated shaking 5 min at 75°C and 1100 rpm. Samples were then run on SDS-PAGE gel and exposed to a phospoimager screen to visualize labeled RNA bound to KDM5A protein. KDM5A-RNA complexes were then transferred to PVDF membrane to carry out western blotting against KDM5A protein.

4.1.1. Aptamer design

To design aptamers for KDM5A/B/C (KDM5), we exploited *cat*RAPID omics v2 [77] to assess interactions between KDM5A/B/C and a comprehensive set of nucleic acids, encompassing 16,523 mRNAs (https://tinyurl.com/5n7m3jdt) and 19,170 long non-coding RNAs (https://tinyurl.com/3hy3deuf). The SeaMotE analysis [76] was then applied to sequences featuring nucleic acid motifs from ENCODE ChIP-Seq peaks corresponding to KDM5A (https://tinyurl.com/3fwj7m9s). We set cut-offs for a *cat*RAPID Z-score greater than 3, an *Interaction Propensity* above 60, and the inclusion of motifs at least 6 nucleotides in length. Focusing on RNA fragments of 15 nucleotides, we compiled a collection of 162,895 candidate aptamers. These were then tested against various segments of KDM5A to determine binding sites. The aptamer candidates were evaluated based on their *cat*RAPID predicted interaction protein with the KDM5 protein fragments. For each aptamer, both the highest interaction strength (maximal values) and the average interaction strength (mean values) were considered.

Considering the tendency of RNA sequences with more stable secondary structures to exhibit stronger protein binding [99], we calculated the secondary structure free energies using RNAfold [78]. This approach led us to prioritize those aptamers with the highest *cat*RAPID scores, considering their free energy stability. To further understand the structural basis of RNA interactions with KDM5, we looked at slightly longer aptamers, 21 nucleotides in length, and their capacity to adopt G4 structures. According to *cat*RAPID predictions [66], an enhanced binding propensity is observed with G4 folding.

4.1.2. catRAPID predictions

The *cat*RAPID algorithm estimates the interaction through van der Waals, hydrogen bonding and secondary structure propensities of both protein and RNA sequences [74]. As reported in an analysis of about half a million of experimentally validated interactions [99], catRAPID can separate interacting vs. non-interacting pairs with an area under the curve (AUC) receiver operating characteristic (ROC) curve of 0.78 (with false discovery rate (FDR) significantly below 0.25 when the Z-score values are > 2). The design of aptamers follows our previous work [75], as detailed in the section *Aptamer design.*

4.1.3. Electrophoretic mobility shift assay

For in vitro studies recombinant KDM5A and KDM5B were purchased from Active Motif (31431 KDM5A, 31432 KDM5B). Additionally, KDM5B recombinant protein preparations were obtained from Thomas Boesen's lab as a gift. RNA and DNA oligonucleotides were ordered from Merck. EMSA experiments were performed with two alternative detection approaches.

4.1.4. G4 prediction in ChiP-seq data

We obtained IDR thresholded ChiP-seq peaks for HepG2 cell line from ENCODE project [100], both for KDM5A and KDM5B proteins. We used pqsfinder (version 2.8.0) to detect G-Queadruplex motifs with default parameters considering hits with score > 52 to estimate the fraction of the experimental peaks with a predicted G4 motif.

4.1.5. Isotope-labelled RNA EMSA

Oligonucleotides were end-labelled with ATP [γ-32P], using T4 Polynucleotide Kinase (PNK) (EK0032, Thermo Fisher Scientific). First 25 pmol of oligonucleotide was denatured at 95°C for 30 s and placed on ice. Next 2 μl PNK buffer A, 1 μl of ATP [γ-32P], and 10 U PNK was added to a total volume of 20 μl. Reaction was incubated at 37°C for 30 min. After labelling 30 μl of nuclease- free water was added, and reaction was purified through Microspin G-50 gel filtration columns (GE Healthcare). Labelled oligonucleotides were further diluted to 25 nM concentrations. Nucleic acid and protein binding reactions were performed in binding buffer, containing 20 mM Tris- HCl (pH 8.0), 150 mM NaCl, 10 % glycerol and 1 mM DTT. Binding was carried out with 1 μg KDM5B and labelled oligonucleotides in the final concentration of 10 nM. For competition experiments 100-fold molar excess of unlabelled oligonucleotide (1 μM final concentration) was added to the binding reaction with an additional 30 min incubation on ice.

Native polyacrylamide gel electrophoresis was performed with Novex 4–12 % Tris-Glycine gels (Thermo Fisher Scientific) in Tris-Glycine running buffer (25 mM Tris, 192 mM Glycine) at 110 V for 1.5–2 h. Radioactive gels were exposed on phosphoimager screens and visualized with the Typhoon FLA 9500 imager (GE healthcare).

4.1.6. Unlabelled RNA EMSA

As an alternative approach, 5 μg KDM5B was mixed with 1 μM unlabelled RNA. Binding reaction and gel electrophoresis was carried out identically to the aforementioned conditions. After electrophoresis, gels were first stained with 1X SYBR gold stain (S11494, Thermo Fisher Scientific) in 1X TBE buffer (15581044, Thermo Fisher Scientific) for 10 min, followed by silver staining (24600, Thermo Fisher Scientific).

4.1.7. Sample preparation for co-IP and proteome profiling

For co-IP 3.5 M K-562 cells were lysed in WCE buffer (50 mM Tris, pD 7.4; 150 mM NaCl; 0.5 % NP-40; 1 mM EDTA; phosphatase inhibitor (Roche)) by sonication. anti-KDM5a antibody was coupled to epoxy dynabeads at 2 μg/100 μl WCE buffer (Life). K-562 lysates were incubated with coupled beads and 5 μl turbo DNase/ml and in the absence or presence of 2 μl RNase A/T1 for 3 min at 1100 rpm in a shaking incubator. Samples were washed 3 times in WCE buffer before separation on PAGE and silver staining.

Triplicates of silver-stained gel samples with either a KDM5A IP with RNAse treatment or without treatment were prepared for mass spectrometric analysis by tryptic in-gel digestion according to the previously described protocol [4]. The extracted peptides were dissolved in 5 % acetonitrile and 2 % formic acid before injection for LC-MS/MS analysis.

4.1.8. LC-MS settings for proteomics

LC-MS/MS was performed by nanoflow reversed-phase liquid chromatography (Dionex Ultimate 3000, Thermo Scientific, USA) coupled online to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Scientific), as previously described by us [101]. Briefly, LC separation was performed using a PicoFrit analytical column (75 μm ID × 40 cm long, 15 μm tip ID (New Objectives, Woburn, MA, USA)) packed in-house with 3 μm C18 resin (Reprosil-AQ Pur, Dr. Maisch, Germany). Peptides were eluted using a gradient from 3.8 % to 98 % solvent B over 45 min at a flow rate of 266 nL/min (solvent A: 0.1 % formic acid in water; solvent B: 80 % acetonitrile and 0.08 % formic acid). 3.5 kilovolts were applied for nanoelectrospray generation. A cycle of one full FT scan mass spectrum (300–1750 m/z, resolution of 35,000 at m/z 200) was followed by 12 data-dependent MS/MS scans at a normalized collision energy of 25 eV. Target ions already selected for MS/MS were dynamically excluded for 30 s.

4.1.9. Proteomics data analysis and statistics

Raw MS data were processed with MaxQuant (v1.5.0.0) and searched against the Homo sapiens database GRCh37.70.pep.all, published in 2013. A false discovery rate (FDR) of 0.01 for proteins and peptides, a minimum peptide length of 7 amino acids, and a mass tolerance of 4.5 ppm for precursor ions and 20 ppm for fragment ions were required. A maximum of two missed cleavages was allowed for the tryptic digest. Cysteine carbamidomethylation was set as a fixed modification, while N-terminal protein acetylation and methionine oxidation were set as variable modifications.

Author contributions

JL: Conceived experiments, performed experiments of Figs. 4a-c, 5, 6, analyzed data, drafted and approved the manuscript.

TC: Conceived original project plan and experiments, performed experiments of Figs. 2, 7, analyzed data, and approved the manuscript.

AA: Designed aptamers and performed computational analysis in Figs. 3, 4d, analyzed data, and approved the manuscript.

AL: Performed experiments leading to the working hypothesis not included in the manuscript, approved the manuscript.

AV: Performed analysis of G4s guiding the project development not included in the manuscript, approved the manuscript.

DM: Performed mass spectrometry and analyzed data in Fig. 7. Approved the manuscript.

MG: Provided KDM5B recombinant protein, discussed results and interpreted data. Approved the manuscript.

GGT: Conceived computational analysis strategies, designed aptamers, performed computational analysis in Figs. 3, 4d, analyzed data. Drafted initial manuscript and approved the manuscript.

UAVØ: Conceived original project plan and experiments, analyzed data, supervised research project, secured funding. Drafted initial manuscript and approved the manuscript.

CRediT authorship contribution statement

Johanna Luige: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Thomas Conrad:** Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Alexandros Armaos:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis. **Annita Louloupi:** Methodology, Investigation. **Anna Vincent:** Methodology, Investigation. **David Meierhofer:** Methodology, Formal analysis. **Michael Gajhede:** Resources. **Gian Gaetano Tartaglia:** Writing – review & editing, Visualization, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Ulf Ørom:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Thomas Boesen, Taner Drace, Annalisa Marsico and Sabina Krakau for preliminary work related to this project. Work in the author's laboratories is funded by The Alexander von Humboldt Foundation, Novo Nordisk Foundation, Lundbeck Foundation, Danish Cancer Society, Independent Research Fund Denmark and Aage and Johanne Louis-Hansen's Foundation (to UAVØ). The research leading to this work was supported by the ERC ASTRA_855923 (G.G.T.), EIC Pathfinder IVBM4-PAP_101098989 (G.G.T.) and PNRR grant from National Centre for Gene Therapy and Drugs based on RNA Technology (CN00000041 EPNRRCN3 (G.G.T.).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.06.027.

Data availability

The mass spectrometry data have been deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [98] with the data set identifier PXD050767.

References

[1] Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol 2018;19:327–41.
[2] Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 2012. https://doi.org/10.1016/j.cell.2012.04.031.
[3] Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume A-M, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun 2015;6:10127.
[4] Conrad T, Albrecht AS, Costa VRDM, Sauer S, Meierhofer D, Ørom UA. Serial interactome capture of the human cell nucleus. Nat Commun 2016. https://doi.org/10.1038/ncomms11212.
[5] Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 2012. https://doi.org/10.1016/j.molcel.2012.05.021.
[6] Perez-Perri JI, Rogell B, Schwarzl T, Stein F, Zhou Y, Rettel M, Brosig A, Hentze MW. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nat Commun 2018;9:4408.
[7] Urdaneta EC, Vieira-Vieira CH, Hick T, Wessels H-H, Figini D, Moschall R, Medenbach J, Ohler U, Granneman S, Selbach M, Beckmann BM. Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. Nat Commun 2019;10:990.
[8] Castello A, Frese CK, Fischer B, Järvelin AI, Horos R, Alleaume A-M, Foehr S, Curk T, Krijgsveld J, Hentze MW. Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. Nat Protoc 2017;12:2447–64.
[9] Hendrickson DG, Kelley DR, Tenen D, Bernstein B, Rinn JL. Widespread RNA binding by chromatin-associated proteins. Genome Biol 2016;17:28.
[10] Xiao R, Chen J-Y, Liang Z, Luo D, Chen G, Lu ZJ, Chen Y, Zhou B, Li H, Du X, Yang Y, San M, Wei X, Liu W, Lécuyer E, Graveley BR, Yeo GW, Burge CB, Zhang MQ, Zhou Y, Fu X-D. Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. Cell 2019;178:107–121.e18.
[11] Blair LP, Liu Z, Labitigan RLD, Wu L, Zheng D, Xia Z, Pearson EL, Nazeer FI, Cao J, Lang SM, Rines RJ, Mackintosh SG, Moore CL, Li W, Tian B, Tackett AJ, Yan Q. KDM5 lysine demethylases are involved in maintenance of 3′UTR length. Sci Adv 2016;2:1–13.
[12] Dorosz J, Kristensen LH, Aduri NG, Mirza O, Lousen R, Bucciarelli S, Mehta V, Sellés-Baiget S, Solbak SMØ, Bach A, Mesa P, Hernandez PA, Montoya G, Nguyen TTTN, Rand KD, Boesen T, Gajhede M. Molecular architecture of the Jumonji C family histone demethylase KDM5B. Sci Rep 2019;9:1–13.
[13] Klose RJ, Yan Q, Tothova Z, Yamane K, Erdjument-Bromage H, Tempst P, Gilliland DG, Zhang Y, Kaelin WG. The retinoblastoma binding protein RBP2 Is an H3K4 demethylase. Cell 2007;128:889–900.
[14] Outchkourov NS, Muiño JM, Kaufmann K, van IJcken WFJ, Koerkamp MJG, van Leenen D, de Graaf P, Holstege FCP, Grosveld FG, Timmers HTM. Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. Cell Rep 2013;3:1071–9.
[15] Kurup JT, Campeanu IJ, Kidder BL. Contribution of H3K4 demethylase KDM5B to nucleosome organization in embryonic stem cells revealed by micrococcal nuclease sequencing. Epigenetics Chromatin 2019;12:1–18.
[16] Beshiri ML, Holmes KB, Richter WF, Hess S, Islam ABMMK, Yan Q, Plante L, Litovchick L, Gévry N, Lopez-Bigas N, Kaelin WG, Benevolenskaya EV. Coordinated repression of cell cycle genes by KDM5A and E2F4 during differentiation. Proc Natl Acad Sci 2012;109:18499–504.
[17] Vallianatos CN, Iwase S. Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. Epigenomics 2015;7:503–18.
[18] Guo L, Guo YY, Li BY, Peng WQ, Tang QQ. Histone demethylase KDM5A is transactivated by the transcription factor C/EBPβ and promotes preadipocyte differentiation by inhibiting Wnt/β-catenin signaling. J Biol Chem 2019;294:9642–54.
[19] Wong P-P, Miranda F, Chan KV, Berlato C, Hurst HC, Scibetta AG. Histone Demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor p21 cip (CDKN1A). Mol Cell Biol 2012;32:1633–44.
[20] Gong F, Clouaire T, Aguirrebengoa M, Legube G, Miller KM. Histone demethylase KDM5A regulates the ZMY ND8-NuRD chromatin remodeler to promote DNA repair. J Cell Biol 2017;216:1959–74.
[21] Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic Acids Res 2005;33:66–80.
[22] Tu S, Teng Y-C, Yuan C, Wu Y-T, Chan M-Y, Cheng A-N, Lin P-H, Juan L-J, Tsai M-D. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. Nat Struct Mol Biol 2008;15:419–21.
[23] Scibetta AG, Santangelo S, Coleman J, Hall D, Chaplin T, Copier J, Catchpole S, Burchell J, Taylor-Papadimitriou J. Functional analysis of the transcription repressor PLU-1/JARID1B. Mol Cell Biol 2007;27:7220–35.
[24] Von Ehr J, Oberstrass L, Yazgan E, Schnaubelt LI, Blümel N, McNicoll F, Weigand JE, Zarnack K, Müller-McNicoll M, Korn SM, Schlundt A. Arid5a uses disordered extensions of its core ARID domain for distinct DNA- and RNA-recognition and gene regulation. Biochemistry 2024. https://doi.org/10.1101/2024.02.29.582703.
[25] Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. Nat Rev Mol Cell Biol 2014;15:749–60.
[26] Hamilton DJ, Hein AE, Wuttke DS, Batey RT. The DNA binding high mobility group box protein family functionally binds RNA. Wiley Interdisciplinary Reviews RNA 2023;14:e1778.
[27] Wang C, Zong X, Wu F, Leung RWT, Hu Y, Qin J. DNA- and RNA-binding proteins linked transcriptional control and alternative splicing together in a two-layer regulatory network system of chronic myeloid leukemia. Front Mol Biosci 2022;9:920492.
[28] Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA Gas5 is a growth arrest– and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 2010;3. ra8.
[29] Lyonnais S, Tarrés-Solé A, Rubio-Cosials A, Cuppari A, Brito R, Jaumot J, Gargallo R, Vilaseca M, Silva C, Granzhan A, Teulade-Fichou M-P, Eritja R, Solà M. The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. Sci Rep 2017;7:43992.
[30] Lago S, Tosoni E, Nadai M, Palumbo M, Richter SN. The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. Biochim Et Biophys Acta (BBA) Gen Subj 2017;1861:1371–81.
[31] Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Antonio MD, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S. G-

quadruplex structures mark human regulatory chromatin. Nat Genet 2016;48:1267–72.
[32] Lago S, Nadai M, Cernilogar FM, Kazerani M, Moreno HD, Schotta G, Richter SN. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. Nat Commun 2021;12:3885.
[33] Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res 2015;43:8627–37.
[34] Arora R, Lee Y, Wischnewski H, Brun CM, Schwarz T, Azzalin CM. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nat Commun 2014;5:5220.
[35] Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat–containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 2007;318:798–801.
[36] Biffi G, Tannahill D, Balasubramanian S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. J Am Chem Soc 2012;134:11974–6.
[37] Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, Lovrecz GO, Beck JL, Bryan TM. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nat Commun 2015;6:7643.
[38] Oyoshi T, Masuzawa T. Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins. Biochem Biophys Res Commun 2020;531:39–44.
[39] Hirschi A, Martin WJ, Luka Z, Loukachevitch LV, Reiter NJ. G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. RNA 2016;22:1250–60.
[40] Kwok CK, Marsico G, Balasubramanian S. Detecting RNA G-quadruplexes (rG4s) in the transcriptome. Cold Spring Harb Perspect Biol 2018;10:a032284.
[41] Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res 2008;36:6260–8.
[42] Lee DSM, Ghanem LR, Barash Y. Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. Nat Commun 2020;11:527.
[43] Rouleau S.G., Garant J., Bisaillon M., Perreault J. 2018. G-Quadruplexes in fl uence pri-microRNA processing. 15, 198–206.
[44] Simko EAJ, Liu H, Zhang T, Velasquez A, Teli S, Haeusler AR, Wang J. G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. Nucleic Acids Res 2020;48. gkaa475-.
[45] Morris MJ, Negishi Y, Pazsint C, Schonhoft JD, Basu S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. J Am Chem Soc 2010;132:17831–9.
[46] Stoneley M, Paulin FE, Quesne JPL, Chappell SA, Willis AE. C-Myc 5′ untranslated region contains an internal ribosome entry segment. Oncogene 1998;16:423–8.
[47] Cammas A, Dubrac A, Morel B, Lamaa A, Touriol C, Teulade-Fichou MP, Prats H, Millevoi S. Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. RNA Biol 2015;12:320–9.
[48] Miglietta G, Cogoi S, Marinello J, Capranico G, Tikhomirov AS, Shchekotikhin A, Xodo LE. RNA G-Quadruplexes in Kirsten Ras (KRAS) oncogene as targets for small molecules inhibiting translation. J Med Chem 2017;60:9448–61.
[49] Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol 2007;3:218–21.
[50] Sauer M, Juranek SA, Marks J, De Magis A, Kazemier HG, Hilbig D, Benhalevy D, Wang X, Hafner M, Paeschke K. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. Nat Commun 2019;10:2421.
[51] Beaudoin J-D, Perreault J-P. Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic Acids Res 2013;41:5898–911.
[52] Rouleau S, Glouzon J-PS, Brumwell A, Bisaillon M, Perreault J-P. 3′ UTR G-quadruplexes regulate miRNA binding. RNA 2017;23:1172–9.
[53] Mestre-Fos S, Penev PI, Suttapitugsakul S, Hu M, Ito C, Petrov AS, Wartell RM, Wu R, Williams LD. G-Quadruplexes in Human Ribosomal RNA. J Mol Biol 2019;431:1940–55.
[54] Varshney D, Cuesta SM, Herdy B, Abdullah UB, Tannahill D, Balasubramanian S. RNA G-quadruplex structures control ribosomal protein production. Sci Rep 2021;11:22735.
[55] Datta A, Pollock KJ, Kormuth KA, Jr RMB. G-quadruplex assembly by ribosomal DNA: emerging roles in disease pathogenesis and cancer biology. Cytogenet Genome Res 2021;161:285–96.
[56] Wallgren M, Mohammad JB, Yan K-P, Pourbozorgi-Langroudi P, Ebrahimi M, Sabouri N. G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. Nucleic Acids Res 2016;44:6213–31.
[57] Shen J, Varshney D, Simeone A, Zhang X, Adhikari S, Tannahill D, Balasubramanian S. Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. Genome Biol 2021;22:143.
[58] Zyner KG, Simeone A, Flynn SM, Doyle C, Marsico G, Adhikari S, Portella G, Tannahill D, Balasubramanian S. G-quadruplex DNA structures in human stem cells and differentiation. Nat Commun 2022;13:142.
[59] Lyu J, Shao R, Yung PYK, Elsässer SJ. Genome-wide mapping of G-quadruplex structures with CUT&Tag. Nucleic Acids Res 2021. https://doi.org/10.1093/nar/gkab1073.
[60] Puget N, Miller KM, Legube G. Non-canonical DNA/RNA structures during transcription-coupled double-strand break repair: roadblocks or Bona fide repair intermediates? DNA Repair 2019;81:102661.
[61] Hamperl S, Cimprich KA. Conflict resolution in the genome: how transcription and replication make it work. Cell 2016;167:1455–67.
[62] Bossaert M, Pipier A, Riou J-F, Noirot C, Nguyên L-T, Serre R-F, Bouchez O, Defrancq E, Calsou P, Britton S, Gomez D. Transcription-associated topoisomerase 2α (TOP2A) activity is a major effector of cytotoxicity induced by G-quadruplex ligands. eLife 2021;10:e65184.
[63] Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S, Jackson SP. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat Chem Biol 2012;8:301–10.
[64] McRae EKS, Booy EP, Moya-Torres A, Ezzati P, Stetefeld J, McKenna SA. Human DDX21 binds and unwinds RNA guanine quadruplexes. Nucleic Acids Res 2017;45:6656–68.
[65] Groot NS de, Armaos A, Graña-Montes R, Alriquet M, Calloni G, Vabulas RM, Tartaglia GG. RNA structure drives interaction with proteins. Nat Commun 2019;10:3246.
[66] Luige J, Armaos A, Tartaglia GG, Ørom UAV. Predicting nuclear G-quadruplex RNA-binding proteins with roles in transcription and phase separation. Nat Commun 2024;15:2585.
[67] Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. Mol Cell 2017;65:1056–1067.e5.
[68] Beltran M, Yates CM, Skalska L, Dawson M, Reis FP, Viiri K, Fisher CL, Sibley CR, Foster BM, Bartke T, Ule J, Jenner RG. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res 2016;26:896–907.
[69] Long Y, Bolanos B, Gong L, Liu W, Goodrich KJ, Yang X, Chen S, Gooding AR, Maegley KA, Gajiwala KS, Brooun A, Cech TR, Liu X. Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. eLife 2017;6:e31558.
[70] Beltran M, Tavares M, Justin N, Khandelwal G, Ambrose J, Foster BM, Worlock KB, Tvardovskiy A, Kunzelmann S, Herrero J, Bartke T, Gamblin SJ, Wilson JR, Jenner RG. G-tract RNA removes Polycomb repressive complex 2 from genes. Nat Struct Mol Biol 2019;26:899–909.
[71] Song J, Gooding AR, Hemphill WO, Love BD, Robertson A, Yao L, Zon LI, North TE, Kasinath V, Cech TR. Structural basis for inactivation of PRC2 by G-quadruplex RNA. Science 2023;381:1331–7.
[72] Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: illustrator of protein domain structures. Cell Res 2009;19:271–3.
[73] Hon J, Martínek T, Zendulka J, Lexa M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. Bioinformatics 2017;33:3373–9.
[74] Bellucci M, Agostini F, Masin M, Tartaglia GG. Predicting protein associations with long noncoding RNAs. Nat Methods 2011;8:444–5.
[75] Zacco E, Kantelberg O, Milanetti E, Armaos A, Panei FP, Gregory J, Jeacock K, Clarke DJ, Chandran S, Ruocco G, Gustincich S, Horrocks MH, Pastore A, Tartaglia GG. Probing TDP-43 condensation using an in silico designed aptamer. Nat Commun 2022;13:3306.
[76] Agostini F, Cirillo D, Ponti RD, Tartaglia GG. SeAMotE: a method for high-throughput motif discovery in nucleic acid sequences. BMC Genom 2014;15:925.
[77] Armaos A, Colantoni A, Proietti G, Rupert J, Tartaglia GG. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. Nucleic Acids Res 2021;49:W72–9.
[78] Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. Algorithms Mol Biol 2011;6:26.
[79] Vandelli A, Cid Samper F, Torrent Burgas M, Sanchez de Groot N, Tartaglia GG. The interplay between disordered regions in RNAs and proteins modulates interactions within stress granules and processing bodies. J Mol Biol 2022;434:167159.
[80] Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature 2006;442:100–3.
[81] Sanchez R, Zhou M-M. The PHD finger: a versatile epigenome reader. Trends Biochem Sci 2011;36:364–72.
[82] Johansson C, Velupillai S, Tumber A, Szykowska A, Hookway ES, Nowak RP, Strain-Damerell C, Gileadi C, Philpott M, Burgess-Brown N, Wu N, Kopec J, Nuzzi A, Steuber H, Egner U, Badock V, Munro S, LaThangue NB, Westaway S, Brown J, Athanasou N, Prinjha R, Brennan PE, Oppermann U. Structural analysis of human KDM5B guides histone demethylase inhibitor development. Nat Chem Biol 2016;12:539–45.
[83] Miserachs HG, Donghi D, Börner R, Johannsen S, Sigel RKO. Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. J Biol Inorg Chem 2016;21:975–86.
[84] Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY, Halim VA, Bagshaw RD, Hubner NC, Al-Hakim A, Bouchard A, Faubert D, Fermin D, Dunham WH, Goudreault M, Lin Z-Y, Badillo BG, Pawson T, Durocher D, Coulombe B, Aebersold R, Superti-Furga G, Colinge J, Heck AJR, Choi H, Gstaiger M, Mohammed S, Cristea IM, Bennett KL, Washburn MP, Raught B, Ewing RM, Gingras A-C, Nesvizhskii AI. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods 2013;10:730–6.
[85] Stenström L, Mahdessian D, Gnann C, Cesnik AJ, Ouyang W, Leonetti MD, Uhlén M, Cuylen-Haering S, Thul PJ, Lundberg E. Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. Mol Syst Biol 2020;16:e9469.

[86] González V, Guo K, Hurley L, Sun D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding Protein. J Biol Chem 2009;284:23622–35.
[87] Lista MJ, Martins RP, Billant O, Contesse M-A, Findakly S, Pochard P, Daskalogianni C, Beauvineau C, Guetta C, Jamin C, Teulade-Fichou M-P, Fåhraeus R, Voisset C, Blondel M. Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. Nat Commun 2017;8:16043.
[88] Santos T, Miranda A, Campello MPC, Paulo A, Salgado G, Cabrita EJ, Cruz C. Recognition of nucleolin through interaction with RNA G-quadruplex. Biochem Pharmacol 2021;189:114208.
[89] Masuzawa T, Oyoshi T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. ACS Omega 2020;5:5202–8.
[90] Allain FH -T, Bouvet P, Dieckmann T, Feigon J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. EMBO J 2000;19:6870–81.
[91] Ghisolfi-Nieto L, Joseph G, Puvion-Dutilleul F, Amalric F, Bouvet P. Nucleolin is a sequence-specific RNA-binding Protein: characterization of targets on Pre-ribosomal RNA. J Mol Biol 1996;260:34–53.
[92] Durut N, Sáez-Vásquez J. Nucleolin: Dual roles in rDNA chromatin transcription. Gene 2015;556:7–12.
[93] Calo E, Flynn RA, Martin L, Spitale RC, Chang HY, Wysocka J. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 2015;518:249–53.
[94] Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 2016;13:841–4.
[95] Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science 2016;353:aaf5371.
[96] Sahoo BR, Kocman V, Clark N, Myers N, Deng X, Wong EL, Yang HJ, Kotar A, Guzman BB, Dominguez D, Plavec J, Bardwell JCA. Protein G-quadruplex interactions and their effects on phase transitions and protein aggregation. Nucleic Acids Res 2024;52:4702–22.
[97] Matsuo K, Asamitsu S, Maeda K, Suzuki H, Kawakubo K, Komiya G, Kudo K, Sakai Y, Hori K, Ikenoshita S, Usuki S, Funahashi S, Oizumi H, Takeda A, Kawata Y, Mizobata T, Shioda N, Yabuki Y. RNA G-quadruplexes form scaffolds that promote neuropathological α-synuclein aggregation. Cell 2024;187:6835–6848.e20.
[98] Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 2022;50:D543–52.
[99] Sanchez de Groot N, Armaos A, Graña-Montes R, Alriquet M, Calloni G, Vabulas RM, Tartaglia GG. RNA structure drives interaction with proteins. Nat Commun 2019;10:3246.
[100] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
[101] Aretz I, Hardt C, Wittig I, Meierhofer D. An impaired respiratory electron chain triggers down-regulation of the energy metabolism and De-ubiquitination of solute carrier amino acid transporters. Mol Cell Proteom 2016;15:1526–38.