

ARTICLE

DOI: 10.1038/s41467-017-02770-z

OPEN

# Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1

Baixing Wu<sup>1</sup>, Shichen Su<sup>1</sup>, Deepak P. Patil<sup>2</sup>, Hehua Liu<sup>1,3</sup>, Jianhua Gan<sup>3</sup>, Samie R. Jaffrey<sup>2</sup> & Jinbiao Ma<sup>1</sup>

Human hnRNP A2/B1 is an RNA-binding protein that plays important roles in many biological processes, including maturation, transport, and metabolism of mRNA, and gene regulation of long noncoding RNAs. hnRNP A2/B1 was reported to control the microRNAs sorting to exosomes and promote primary microRNA processing as a potential m<sup>6</sup>A “reader.” hnRNP A2/B1 contains two RNA recognition motifs that provide sequence-specific recognition of RNA substrates. Here, we determine crystal structures of tandem RRM domains of hnRNP A2/B1 in complex with various RNA substrates, elucidating specific recognitions of AGG and UAG motifs by RRM1 and RRM2 domains, respectively. Further structural and biochemical results demonstrate multivariant binding modes for sequence-diversified RNA substrates, supporting a RNA matchmaker mechanism in hnRNP A2/B1 function. Moreover, our studies in combination with bioinformatic analysis suggest that hnRNP A2/B1 may mediate effects of m<sup>6</sup>A through a “m<sup>6</sup>A switch” mechanism, instead of acting as a direct “reader” of m<sup>6</sup>A modification.

<sup>1</sup>State Key Laboratory of Genetic Engineering, Collaborative Innovation Centre of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China. <sup>2</sup>Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA. <sup>3</sup>State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Physiology and Biophysics, School of Life Sciences, Fudan University, Shanghai, 200438, China. Baixing Wu and Shichen Su contributed equally to this work. Correspondence and requests for materials should be addressed to J.M. (email: [majb@fudan.edu.cn](mailto:majb@fudan.edu.cn))

Heterogeneous nuclear ribonucleoproteins (hnRNPs) play a variety of roles in regulating transcriptional and post-transcriptional gene expression, including RNA splicing, polyadenylation, capping, modification, export, localization, translation, and turnover<sup>1, 2</sup>. Each hnRNP contains at least one RNA-binding domain (RBD), such as RNA recognition motif (RRM), K-Homology (KH) domain, or an arginine/glycine-rich box<sup>3</sup>. Sequence-specific association between hnRNPs and their RNA targets are typically mediated by one or more RBDs, which usually bind short, single-stranded RNA<sup>4, 5</sup>, but in some instances, also recognize structured RNAs<sup>6</sup>.

As a core component of the hnRNP complex in mammalian cells, hnRNP A2/B1 is an abundant protein and has been implicated in numerous biological processes. The *HNRNPA2B1* gene encodes two protein isoforms, A2 and B1, through alternative splicing. The B1 isoform contains an insertion of 12 amino acids (aa) at its N terminus<sup>7, 8</sup>. Both isoforms have an RNA-binding domain (RBD) composed of tandem RRM domains separated by a 15-aa linker, and a C-terminal Gly-rich low complexity (LC) region that includes a prion-like domain (PrLD), an RGG box, and a PY-motif containing a M9 nuclear localization signal (PY-NLS)<sup>9–11</sup>. These domains are represented schematically with residue numbers based on the hnRNP B1 isoform (Fig. 1a).

hnRNP A2/B1 is linked to several biological processes and diseases, especially neurodegenerative disorders, e.g., mutations in core PrLD in hnRNP A2/B1 cause multisystem proteinopathy and amyotrophic lateral sclerosis (ALS), through promoting excess incorporation of hnRNP A2/B1 into stress granules and driving the formation of cytoplasmic inclusions in animal models<sup>9</sup>. hnRNP A2/B1 also regulates hESC self-renewal and pluripotency<sup>12</sup>.

hnRNP A2/B1 has multiple effects on RNA processing through binding specific sequence. It can bind to HIV-1 RNA, causing nuclear retention of the vRNA, as well as microRNAs, sorting them into the exosomes through binding “EXO-motifs”<sup>13, 14</sup>. A transcriptome-wide analysis of hnRNP A2/B1 targets in the nervous system identified a clear preference for UAG(G/A) motifs confirmed by three independent and complementary *in vitro* and *in vivo* approaches<sup>15, 16</sup>. This is consistent with previous studies indicating that hnRNP A2/B1 binds specifically to UAGGG, GGUAGUAG, or AGGAUAGA sequences<sup>17, 18</sup>. Another recent study demonstrated that hnRNP A2/B1 recognizes a consensus motif containing UAASUUAU (S = G or C) in the 3' UTR of many mRNAs and helps recruiting the CCR4-NOT deadenylase complex<sup>19</sup>. In addition to participating the regulation of mRNAs, hnRNP A2/B1 is also involved in the activities of many other RNA species. For example, hnRNP A2/B1 can promote association of the long noncoding RNA HOTAIR with the nascent transcripts of HOTAIR target genes, thus to mediate HOTAIR-dependent heterochromatin initiation<sup>20</sup>.

Recently, hnRNP A2/B1 was proposed to bind RNA transcripts containing N<sup>6</sup>-methyladenosine, a widespread nucleotide modification in mRNAs and noncoding RNAs<sup>21, 22</sup>. hnRNP A2/B1 was found to mediate m<sup>6</sup>A-dependent nuclear RNA processing events by binding G(m<sup>6</sup>A)C-containing nuclear RNAs *in vivo* and *in vitro*, in which hnRNP A2/B1 associates with a subset of primary microRNA transcripts through binding m<sup>6</sup>A, promoting primary microRNAs processing by recruiting the microprocessor complex Drosha and DGCR8<sup>23</sup>.

Although a series of studies with different approaches pointed out the multitudinous functions of hnRNP A2/B1 mediating by diverse RNA motifs *in vivo*, no detailed mechanism for the binding specificities has been determined. Thus, more biochemistry and structural studies are essential to understand the RNA-binding properties of hnRNP A2/B1 at the molecular level. Here, we report the crystal structures of hnRNP A2/B1 in complex with

variant RNA targets to unveil the RNA-binding specificities and multivariant characteristics. Moreover, our structural data along with RNA-binding and bioinformatic analysis do not support that hnRNP A2/B1 functions as an m<sup>6</sup>A “reader,” since no significant preference for m<sup>6</sup>A modification by either tandem RRMs or full-length protein of hnRNP A2/B1 are observed.

## RESULTS

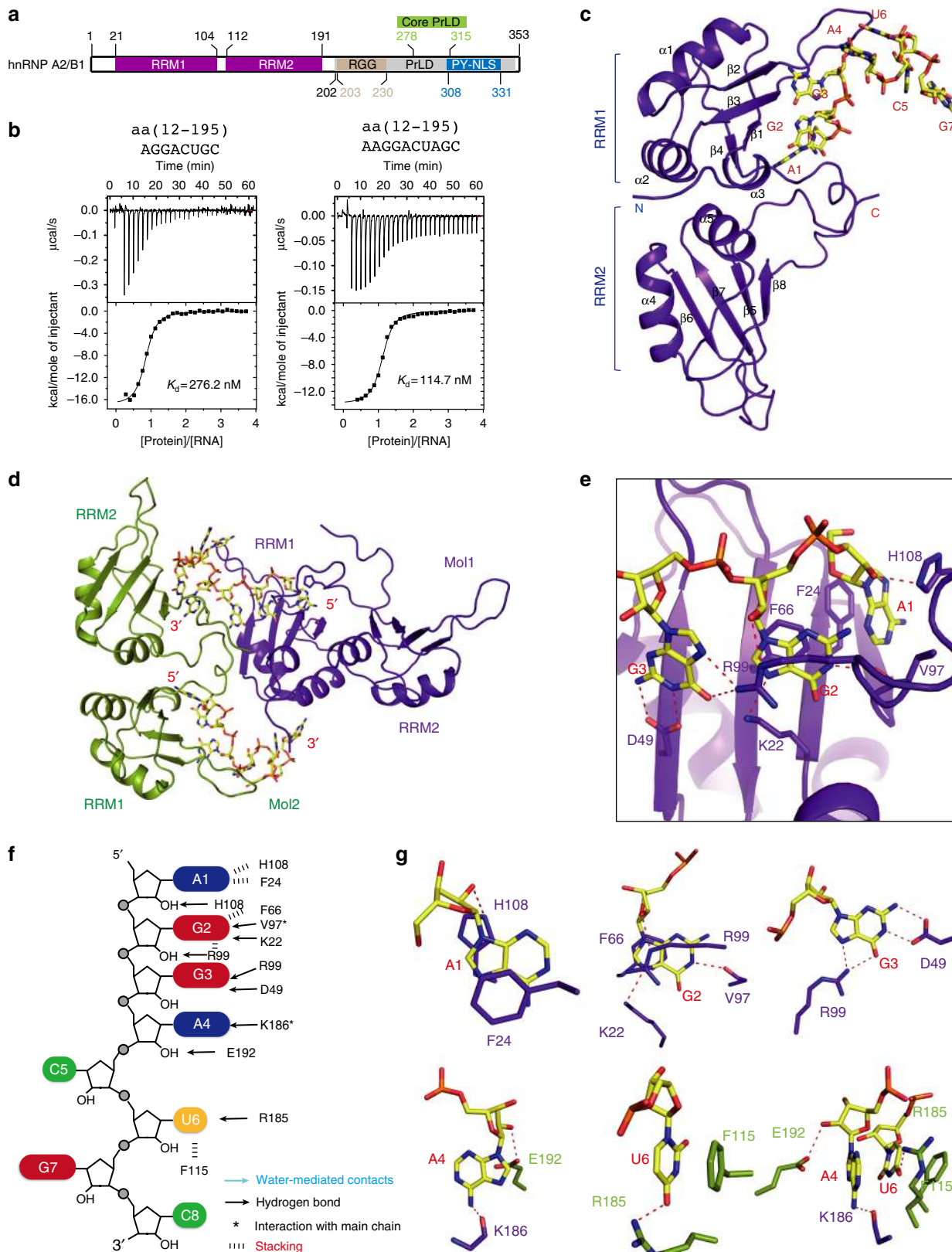
**Crystal structure of hnRNP A2/B1 bound to an 8mer RNA.** To elucidate the RNA-binding properties of tandem RRMs of hnRNP A1/B1, we purified a number of truncations of the hnRNP A2/B1 protein. Using isothermal titration calorimetry (ITC) method, we characterized the RNA-binding activities of each construct with a set of RNA oligonucleotides of sequences according to previously determined binding motifs. ITC results showed that the construct containing the N-terminal fragment (aa 1–11) and two RRM domains, i.e., aa 1–195, can bind target RNAs with high affinity. In addition, deletion of the N-terminal fragment had no obvious impact on the RNA binding. This suggested that the construct containing the tandem RRMs of hnRNP A2/B1 is sufficient for binding target RNAs.

We first determined the crystal structure of RRMs (aa 12–195) of hnRNP A2/B1 bound to the 8-nt RNA oligonucleotide 5'-A<sub>1</sub>G<sub>2</sub>G<sub>3</sub>A<sub>4</sub>C<sub>5</sub>U<sub>6</sub>G<sub>7</sub>C<sub>8</sub>-3' (termed 8mer RNA), which is derived from a recent individual-nucleotide-resolution CLIP study<sup>16, 23</sup>. ITC analysis showed that binding of the 8mer RNA occurs at a 1:1 ratio with a K<sub>d</sub> of 276.2 nM (Fig. 1b). The crystal structure of RRMs (aa 12–195) in complex with the 8mer RNA molecule was determined to 2.60 Å resolutions, details about data collection and structure refinement are summarized in Supplementary Table 1. The tandem RRMs and an RNA molecule are in one asymmetric unit (Fig. 1c); both RRM domains of hnRNP A2/B1 adopt the characteristic RRM fold, which is a typical β1α1β2β3α2β4 topology consisting of an antiparallel four-stranded sheet adjacent to two helices on the opposite side, similar to previously determined RRM structures of other RNA-binding proteins using both crystallographic and NMR methods<sup>24</sup> (Fig. 1c). Each RNA molecule is bound by an RRM1 domain from one hnRNP A2/B1 molecule in an asymmetric unit and an RRM2 domain from another hnRNP A2/B1 molecule in adjacent asymmetric unit (Fig. 1d).

**RRM1 specifically recognizes AGG motif.** The AGG motif of 8mer RNA substrate is specifically recognized mainly by RRM1 (Fig. 1e). For the recognition of the adenine at the first position (A<sub>1</sub>), the 2'-OH group forms a hydrogen bond and π-π interactions with the side chain of His108. Besides hydrogen bonding interactions, base stacking with the Phe24 on the other side also contributes to the definition of the binding environment (Fig. 1f, g). The 2'-OH of G<sub>2</sub> forms a hydrogen bond with the side chain of Arg99, and N<sup>1</sup> groups of G<sub>2</sub> form hydrogen bonds with the carboxyl group of the main chain of Val97 while N<sup>7</sup> interacts with the side chain of Lys22. The base of this guanine G<sub>2</sub> engages in stacking interactions with the base of the benzene ring of Phe66 and guanidyl group of Arg99 (Fig. 1f, g). N<sup>1</sup> and N<sup>2</sup> of G<sub>3</sub>, the last nucleotide in the core recognition AGG motif, are hydrogen bonded to the side chain of Asp49, and O<sup>6</sup> and N<sup>7</sup> are recognized by the side chain of Arg99 (Fig. 1f, g). However, the RNA substrate from A<sub>4</sub> to C<sub>8</sub> are not well specifically recognized (Fig. 1f). The N<sup>6</sup> of A<sub>4</sub> is hydrogen bonded to the main chain of Lys186, and the 2'-OH group is hydrogen bonded to the side chain of Glu192 (Fig. 1f, g). The base of U<sub>6</sub> is sandwiched between Phe115 and U<sub>4</sub> base via π-π stacking, whereas the O<sup>4</sup> of U<sub>6</sub> forms hydrogen bonds with the amino group of the main chain of Arg185 (Fig. 1f, g).

**Both RRM1 and RRM2 are involved in recognition of the 10mer RNA.** The crystal structure of hnRNP A2/B1(aa 12–195) in complex with the 8mer RNA did not provide insight into specific RNA recognition by RRM2. We thus designed another RNA oligonucleotide shown in Supplementary Table 2, based on the speculation that

RRM2 might recognize UAG according to previous sequencing results<sup>17, 18</sup>. This RNA contains both the AGG motif and the UAG motif. ITC results confirmed that the 10-nt RNA oligo 5'-A<sub>0</sub>A<sub>1</sub>G<sub>2</sub>G<sub>3</sub>A<sub>4</sub>C<sub>5</sub>U<sub>6</sub>A<sub>7</sub>G<sub>8</sub>C<sub>9</sub>-3' (termed 10mer RNA) has a higher affinity (Fig. 1b and Supplementary Table 2). We successfully



obtained the crystal of hnRNP A2/B1(12–195) in complex with the 10mer RNA, which was determined to 1.85 Å resolution (Supplementary Table 1). Similar to the previous complex structure with 8mer RNA, there is also only one protein molecule and one RNA molecule in the asymmetric unit. However, this time the RNA molecule is recognized by both RRM1 and RRM2 from two hnRNP A2/B1 molecules, of which the other one is from the asymmetric molecule (Fig. 2a). The 10mer RNA molecule adopts a single-stranded conformation accommodated into a positively charged groove comprised by the canonical RNA-binding surface of the RRM1 and RRM2 from two hnRNP A2/B1 proteins, in a 5′–3′ direction from RRM1 to RRM2 (Fig. 2b). Unlike the complex structure of 8mer RNA in which only RRM1 is involved in specific recognition, the complex structure containing 10mer RNA also shows specific recognition by RRM2 (Fig. 2c). When the structures of RRM1 bound to 8mer and 10mer complex are superimposed, the AGG motifs of the two RNA substrates also superimpose well (Fig. 2d). However, the conformation of the rest of the oligos are dramatically different. Notably, the 10mer RNA is more stretched than the 8mer RNA (Fig. 2d). Moreover, the root-mean-square deviation of protein backbones between the two structures is merely 0.4 Å, suggesting the protein has not changed substantially when binding to the two different RNA targets.

**The RRM2 specifically recognizes UAG motif.** Due to the higher resolution, more detailed interactions are observed in the complex structure of 10mer RNA than in the 8mer RNA complex. Though the recognition of the AGG motif in the two structures is quite similar, more specific recognitions of A<sub>1</sub> and G<sub>2</sub> in the 10mer RNA complex structure are observed. N<sup>1</sup> of A<sub>1</sub> is recognized by the main chain amine group of Val97, whereas N<sup>6</sup> and N<sup>7</sup> form hydrogen bonds with the Lys94 side chain either directly or mediated by a water molecule (Fig. 3a, b). G<sub>2</sub> has the most complicated interacting network in this structure, in which two more G<sub>2</sub> base-specific recognitions are seen in the 10mer complex. These are the side chains of Gln19 and Ser102, which cooperatively bond to O<sup>6</sup> and N<sup>2</sup> of G<sub>2</sub>, respectively (Fig. 3b). In addition to specific recognition of the AGG motif, N<sup>6</sup> of the 5′-end extended adenine (A<sub>0</sub>) is hydrogen bonded to the Glu92, and the N<sup>1</sup> and N<sup>6</sup> of A<sub>4</sub> form hydrogen bonds with the side chain of Lys120 and Asn181 mediated by water, respectively (Fig. 3a, b).

Unlike the 8mer RNA complex structure, the UAG motif in the 10mer RNA is specifically recognized by RRM2. The side chain of Arg185 hydrogen bonds to both the 2′-OH and O<sup>2</sup> group of U<sub>6</sub>, and the N<sup>3</sup> and O<sup>4</sup> groups are hydrogen bonded to the side chain of Glu183 and Arg99, respectively (Fig. 3a, b). The base of A<sub>7</sub> is also clamped by Phe115 and Met193 through hydrophobic interactions; N<sup>7</sup> of A<sub>7</sub> forms hydrogen bond with Arg185; N<sup>1</sup>, N<sup>6</sup>, and the phosphate group are hydrogen bonded to the main chain of Leu188, Lys186, and the side chain of Arg185, respectively (Fig. 3a, b). In addition to the hydrogen bond formed between phosphate group of G<sub>8</sub> and the side chain of Arg153, the O<sup>6</sup> of G<sub>8</sub>

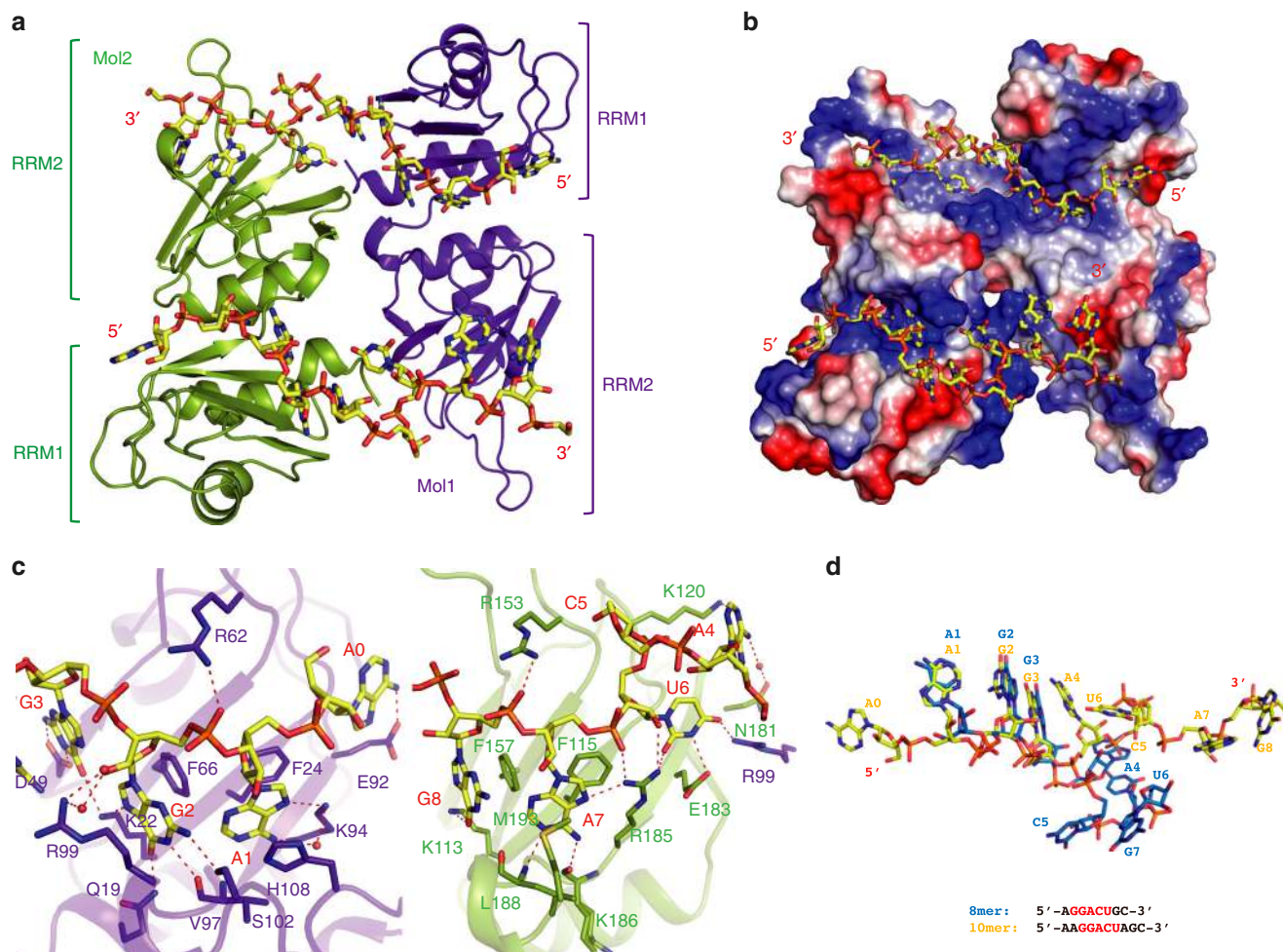
base is recognized by the side chain of Lys113, and the base of G<sub>8</sub> has another  $\pi$ – $\pi$  stacking with Phe157 (Fig. 3a, b). The complete structure of C<sub>5</sub> and C<sub>9</sub> cannot be seen in our structure.

The chains of RRM1 (12–110) and RRM2 (111–195) can be superimposed with a root-mean-square deviation of 0.901 Å with a high sequence identity (Supplementary Fig. 1a). A superimposition of RRM1-AAGG with RRM2-ACUAGC indicated that the recognition of the AG core motif is very similar in RRM1 and RRM2 (Supplementary Fig. 1b, c). We thereafter mutated residues involved in specific recognition of AGG by RRM1 and UAG by RRM2, both of which reduced binding affinities according to ITC (Fig. 3c and Supplementary Fig. 2). Although the results of these amino acid mutations lined with expectations, the nucleotide mutations of 10mer RNA, especially the UAG nucleotides recognized by RRM2, have only moderate effects on the binding affinities (Fig. 3d, Supplementary Fig. 3, and Supplementary Table 2).

**Multivariant RNA recognition modes of RRM1 and RRM2.** In order to understand the molecular basis for hnRNP A2/B1 recognizing different RNA sequences, we grew the crystals of hnRNP A2/B1 (aa 12–195) in complex with different 10mer RNA mutants. Three complex structures containing RNA mutants A1G, U6G, and A7U were determined at high resolution (Fig. 4a, b, c). For the A1G mutant (5′-A<sub>0</sub>G<sub>1</sub>G<sub>2</sub>G<sub>3</sub>A<sub>4</sub>C<sub>5</sub>U<sub>6</sub>A<sub>7</sub>G<sub>8</sub>C<sub>9</sub>-3′), the AGG motif is shifted to 5′-end and recognized by RRM1 in a manner almost identical to the wild-type 10mer RNA structure (Fig. 4d). In addition, the G<sub>3</sub> is recognized by the side chains of Arg99 and Glu18 of RRM1 through hydrogen bond formation (Fig. 4d, e, f). Therefore, RRM1 of hnRNP A2/B1 can specifically recognize an AGGG motif as demonstrated in the A1G mutant structure. Additionally, RRM2 recognizes UAG in the same manner as was seen in the structure of wild-type 10mer RNA (Fig. 4d, e).

In contrast, hnRNP A2/B1 adopts distinct strategies for binding another two RNAs that contained mutations in the UAG motif recognized by RRM2. When U<sub>6</sub> is substituted with G in the U6G complex (5′-A<sub>0</sub>A<sub>1</sub>G<sub>2</sub>G<sub>3</sub>A<sub>4</sub>C<sub>5</sub>G<sub>6</sub>A<sub>7</sub>G<sub>8</sub>C<sub>9</sub>-3′), G<sub>6</sub> lost two hydrogen bonds from Arg99 and Arg185, only keeping hydrogen bonds with Glu182. However, the recognition of the AG core motif by RRM2 is still well maintained (Fig. 4g, h, i). Interestingly, the recognition of AAGG by RRM1 is exactly same as AGGG in the A1G RNA mutant, though their binding modes of AG motif are different, suggesting that RRM1 can accommodate various purine-rich sequences (Fig. 4g, h, i and Supplementary Fig. 4a, b). For the A7U RNA mutant (5′-A<sub>0</sub>A<sub>1</sub>G<sub>2</sub>G<sub>3</sub>A<sub>4</sub>C<sub>5</sub>U<sub>6</sub>U<sub>7</sub>G<sub>8</sub>C<sub>9</sub>-3′), U<sub>7</sub> forms hydrogen bonds with the side chain of Lys113 and Glu142 and forms a  $\pi$ – $\pi$  stacking interaction with Phe157. More interestingly, U<sub>6</sub> adopts a sandwich-like interaction mode with Phe115 and the A<sub>4</sub> base, which is exactly same as the U<sub>6</sub> in the 8mer RNA substrate (Fig. 5j, k, l and Supplementary Fig. 4c, d). This suggests that RRM2 can accommodate the pyrimidine-rich UU sequence.

**Fig. 1** Overview of the structure and ITC of hnRNP A2/B1 in complex with 8mer RNA. **a** Schematic representation of the domain architecture of hnRNP A2/B1. RRM: RNA recognition motif, PrLD: prion-like domain, NLS: nuclear location signal, RGG: arginine-glycine-glycine box. **b** ITC results of hnRNP A2/B1 (12–195) with 8mer and 10mer RNA targets. Solid lines indicate nonlinear least-squares fit the titration curve, with  $\Delta H$  (binding enthalpy kcal mol<sup>-1</sup>),  $K_a$  (association constant), and  $N$  (number of binding sites per monomer) as variable parameters. Calculation values for  $K_d$  (dissociation constant) are indicated. **c** Cartoon representation of RRMs in complex with 8mer RNA. The RNA backbone is colored in yellow shown by stick. The RRMs are colored in purple-blue. **d** Molecules from two adjacent asymmetric units. The molecule from another asymmetric unit is colored in green. **e** Intermolecular contacts between RRM1 and 8mer RNA from A<sub>1</sub> to G<sub>3</sub>. All dashed lines in this study indicate distance <3.2 Å. **f** Schematic showing RRMs interactions with 8mer RNA 5′-AGGACUGC-3′. **g** Close-up view showing the specific recognition of A<sub>1</sub>, G<sub>2</sub>, G<sub>3</sub>, A<sub>4</sub>, and U<sub>6</sub>.



**Fig. 2** Overview of RRMs in complex with 10mer RNA. **a** Cartoon representation of RRMs in complex with 10mer RNA 5'-AAGGACUAGC-3'. The RNA backbone is colored in yellow shown by stick. The molecule from the adjacent asymmetric unit is colored in green. **b** Surface representation of RRMs-10mer complex. **c** Intermolecular contacts between RRM1 and residues of 10mer RNA 5'-AAGG-3', and RRM2 with residues of RNA 5'-ACUAGC-3'. **d** Superposition of 8mer and 10mer RNA substrates. The 8mer RNA 5'-AGGACUGC-3' is colored in marine and the 10mer RNA 5'-AAGGACUAGC-3' is colored in yellow

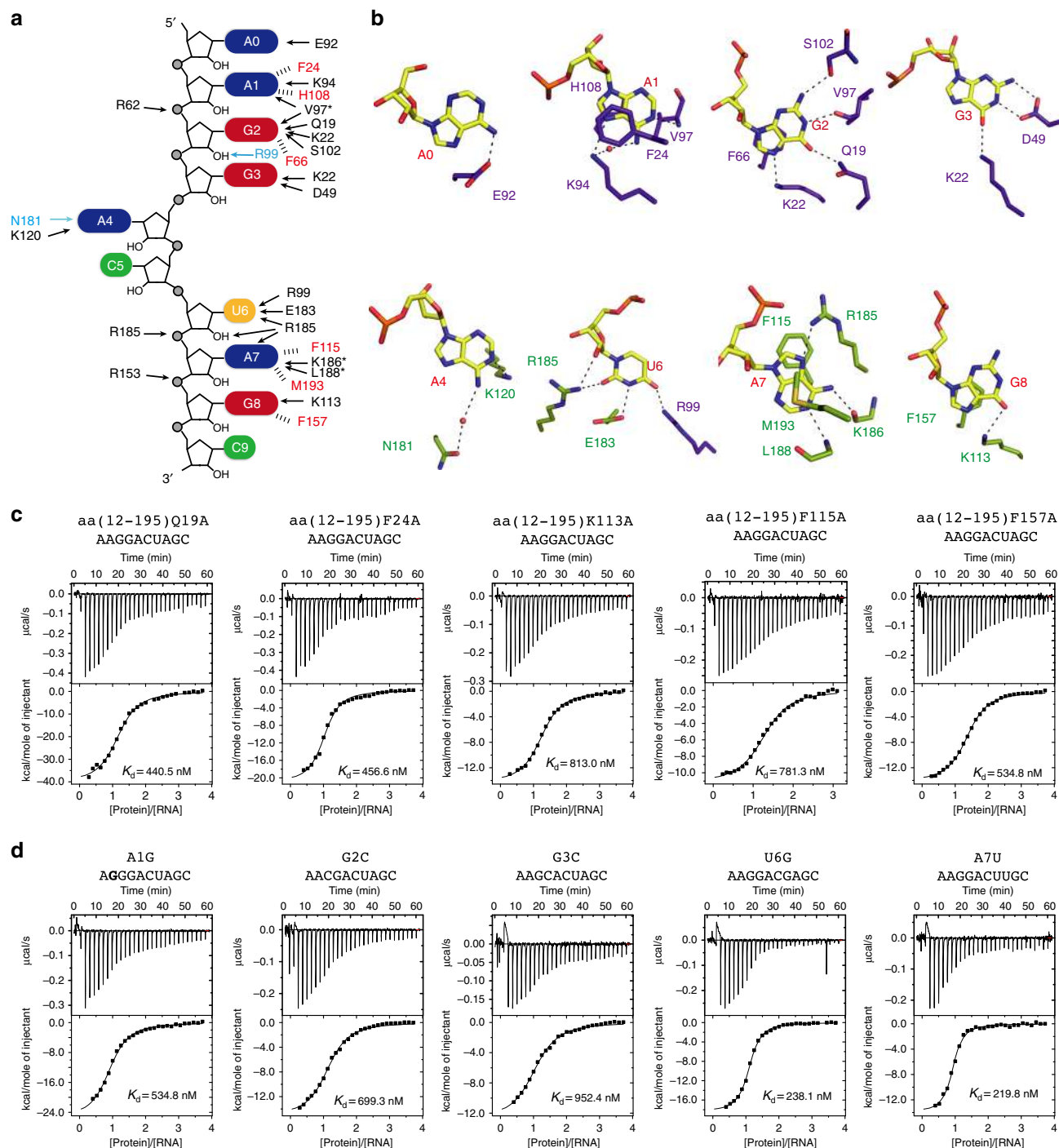
Meanwhile, the AAGG recognition by RRM1 in A7U RNA mutant is very similar to 8mer and 10mer RNAs, but different from A1G and U6G RNA mutant.

Unlike the effect of mutating the UAG motif recognized by RRM2, which just slightly reduced binding affinities, mutation of the AGG motif recognized by RRM1, such as G2C and G3C RNA mutants, have more obvious effects (Fig. 3d). Although we did not obtain crystal structures of these two mutants, our biochemical and structural studies suggested that RRM1 has more stringent recognition for purine-rich AG motif containing RNA sequences, but RRM2 seems to have more broad compatibility to recognize different RNA sequences, including canonical UAG motif, purine-rich GAG, and pyrimidine-rich UU sequences.

**hnRNP A2/B1 binds two antiparallel RNA strands.** A superimposition of all five structures obtained in this study suggested that hnRNP A2/B1 binds two antiparallel RNA strands using RRM1 and RRM2 concurrently (Fig. 5a). The two RRM domains in hnRNP A2/B1, similar to hnRNP A1 both in crystal structure and in solution<sup>25–27</sup>, are held together in a fixed geometry without flexibility (Supplementary Fig. 5a). It is notable that there are extensive interactions between RRM1 and RRM2 from the same hnRNP A2/B1 molecule, including three salt-bridge

interactions of Asp76-Lys168, Arg95-Asp164, Arg82-Asp162, and hydrophobic interactions between Phe20 with Leu171 (Fig. 5b), which are also observed in hnRNP A1 (Supplementary Fig. 5b). In addition, the last  $\beta$ -strand in RRM1 and the first  $\beta$ -strand in RRM2 have the same orientation, which forces the two RRM domains to bind two RNA strands antiparallelly, because the linker between the two domains blocks the binding of RNA targets from the same strand (Fig. 5c, d). In contrast, RNA substrates in most known structures in complex with tandem RRM proteins are bound as single-stranded RNAs and their orientations are from RRM2 to RRM1, such as HuD, HuR, PABP, U2AF65, and TDP-43 (Fig. 5e, f, g, h)<sup>28–32</sup>.

**hnRNP A2/B1 does not specifically recognize m<sup>6</sup>A-modified RNA.** In order to assess the hypothesis that hnRNP A2/B1 might be a direct m<sup>6</sup>A “reader” as proposed in a previous study<sup>23</sup>, the m<sup>6</sup>A motif GGACU is included in the 8mer and 10mer RNAs. As shown in the crystal structures, there is no obvious aromatic cage-like surface that can potentially bind the m<sup>6</sup>A nucleotide (Fig. 6a, b), which was shown to be the key m<sup>6</sup>A-specificity element in previous structural studies of YTHDF1, YTHDC1, and MRB1 complexed with GGm<sup>6</sup>ACU (Fig. 6c and Supplementary Fig. 6)<sup>33</sup>. The crystal structure of hnRNP A2/B1 in complex with GGm<sup>6</sup>ACU could not be obtained. However, we were able to



**Fig. 3** Detailed interactions between RRM domains and 10mer RNA. **a** Schematic showing RRM interactions with 10mer RNA sequence. **b** Close-up view showing the specific recognition from A<sub>1</sub> to G<sub>8</sub>. **c** Mutagenesis study by ITC experiments between protein mutants and 10mer RNA substrate. **d** ITC experiments between hnRNP A2/B1 RRM domains and RNA mutants A1G, G2C, G3C, U6G, and A7U

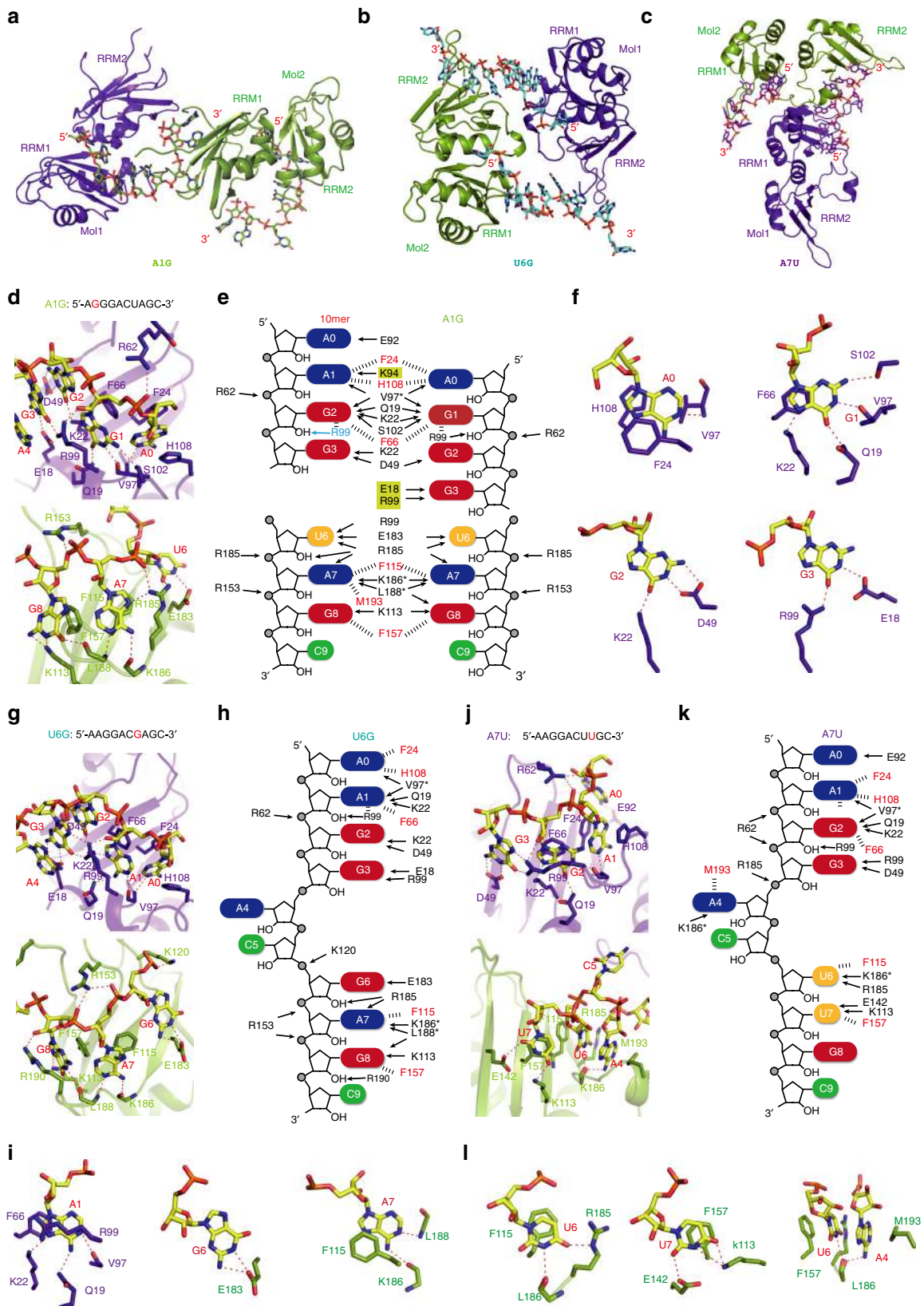
detect binding of hnRNP A2/B1 with the 8mer and 10mer RNA in which an adenosine was replaced with an m<sup>6</sup>A. In both cases, the m<sup>6</sup>A is present within its preferred GGACU sequence context. Notably, the ITC results indicated that the binding affinities of the m<sup>6</sup>A-containing 8mer RNA and 10mer RNA to the tandem RRM (12–195) were reduced onefold and tenfold, respectively, compared to the non-methylated RNA (Fig. 6d).

The N<sup>6</sup> atoms of A<sub>4</sub> in both complex structures of 8mer and 10mer RNAs form hydrogen bonds with hnRNP A2/B1 directly or through a water molecule, which may provide a possible

reason why m<sup>6</sup>A modification would reduce the binding affinity (Figs. 1g and 3b). However, these data do not exclude the possibility that full-length hnRNP A2/B1 may form a m<sup>6</sup>A-binding cage through its C-terminal fragment, which contains a RGG box region (Fig. 1a). This domain may also contribute in some way to RNA binding, but was not included in our structural study (Supplementary Fig. 7a). Therefore, we purified the full-length hnRNP A2/B1 with a His<sub>6</sub>-tag or without the tag and a construct containing RGG box (residue 1–249), and used them to measure the binding affinities with various RNA substrates by

EMSA and ITC experiment (Supplementary Fig. 7b, c, d). The EMSA analysis indicated that full-length hnRNP A2/B1 has a slightly weaker binding affinity to the RNA with m<sup>6</sup>A modification than the one without m<sup>6</sup>A (Fig. 6e and Supplementary

Fig. 7e). The ITC results using the full-length hnRNP A2/B1 or RGG box containing construct (residue 1–249) also showed similar trend as the ITC results using tandem RRMs. These ITC data, along with the EMSA results, suggest that full-length



hnRNP A2/B1 does not specifically recognize or show enhanced binding to these m<sup>6</sup>A-modified RNA substrates in vitro (Supplementary Fig. 7f, g).

**m<sup>6</sup>A sites bound by hnRNP A2/B1 in vivo.** To better understand the potential binding interactions of m<sup>6</sup>A and hnRNP A2/B1 in cells, we examined the in vivo binding properties of hnRNP A2/B1. For this analysis, we used a set of 186 nuclear m<sup>6</sup>A sites comprising m<sup>6</sup>A sites in *XIST*, *NEAT1*, and *MALAT1*, which have been mapped at single-nucleotide resolution using miCLIP (m<sup>6</sup>A individual-nucleotide-resolution crosslinking and immunoprecipitation)<sup>34</sup>. In this analysis, we quantified binding at each of these 168 mapped m<sup>6</sup>A residues by assigning each m<sup>6</sup>A residue an “intensity value”, which was the normalized number of miCLIP reads that overlapped each m<sup>6</sup>A residue<sup>34</sup>. The intensity value is influenced by transcript abundance and m<sup>6</sup>A stoichiometry. We next determined the binding of hnRNP A2/B1 at each of these m<sup>6</sup>A sites based on the normalized number of mapped hnRNP A2/B1 HITS-CLIP tags at the m<sup>6</sup>A site. For most m<sup>6</sup>A residues, there was no correlation between m<sup>6</sup>A intensity and hnRNP A2/B1 binding, although 12 m<sup>6</sup>A sites showed proximal hnRNP A2/B1 binding, which might be the result of coincidental proximity between m<sup>6</sup>A and a non-methylated consensus site recognized by hnRNP A2/B1. As a control, we analyzed YTHDC1, a nuclear YTH domain-containing m<sup>6</sup>A reader<sup>35, 36</sup>. A similar analysis of YTHDC1 showed increasing YTHDC1 binding with increased m<sup>6</sup>A levels for essentially all m<sup>6</sup>A sites (Fig. 6f). Thus, unlike hnRNP A2/B1, YTHDC1 appears to function as a general nuclear m<sup>6</sup>A reader.

Our finding that only a small subset of nuclear m<sup>6</sup>A sites are positioned near hnRNP A2/B1, and therefore could potentially be involved in a direct m<sup>6</sup>A-hnRNP A2/B1 interaction, is compatible with the previous analysis by Alarcon et al.<sup>23</sup>. Alarcon et al. reported that only 17% of their total m<sup>6</sup>A-seq clusters overlap with the hnRNP A2/B1 tag clusters. To determine if YTHDC1 shows greater overlap with m<sup>6</sup>A than hnRNP A2/B1 does, we performed a similar cluster overlap analysis. Approximately 43% of the miCLIP clusters from total RNA and 56% clusters from miCLIP of poly(A) RNA (Fig. 6g) showed an overlap with YTHDC1 clusters ( $P < 0.0001$ ). Thus, YTHDC1 has a considerably higher overlap with m<sup>6</sup>A than hnRNP A2/B1. Therefore, this analysis again supports the idea that YTHDC1 is the predominant nuclear m<sup>6</sup>A reader compared to hnRNP A2/B1.

## Discussion

The RNA-binding domain of hnRNP A2/B1 comprises two RNA recognition motifs, RRM1 and RRM2, which is followed by a C-terminal glycine-rich region. hnRNP A2/B1 was previously demonstrated to bind UUAGGG and UAG RNA motifs through various analyses<sup>37, 38</sup>. Recent CLIP-Seq data further showed that hnRNP A2/B1 prefers to bind A/G-rich sequences<sup>17</sup>. However, there was no molecular basis for the recognition of different RNA substrates of hnRNP A2/B1. Here, we determined the crystal

structures of the tandem RRMs of hnRNP A2/B1 in complex with various RNA substrates, revealing the molecular details of specific target RNA recognition and shedding light on the mechanism for hnRNP A2/B1 in various RNA-mediated biological functions.

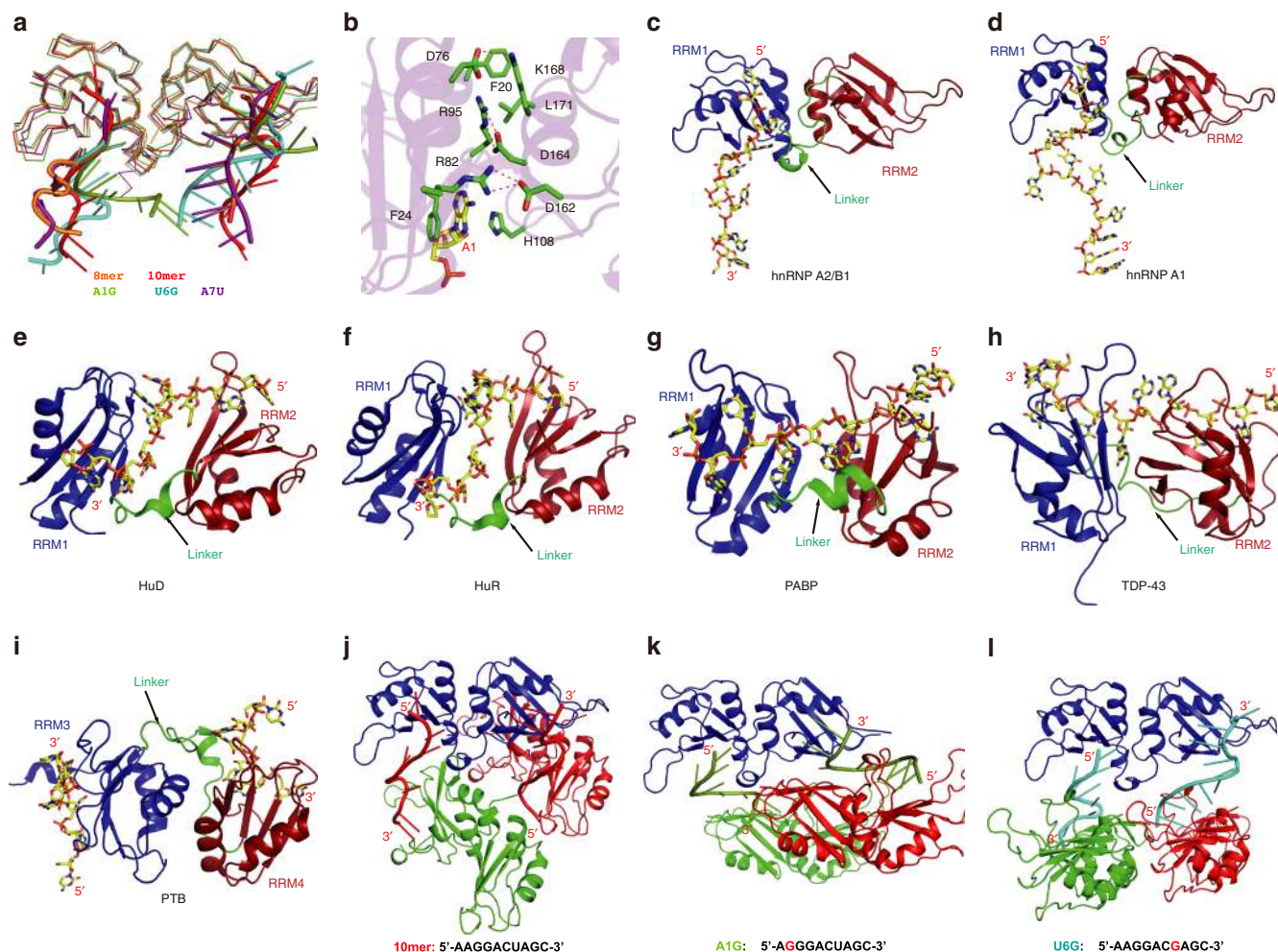
The specific recognition of the AG core motif by both RRM1 and RRM2 is highly consistent with previous studies showing that hnRNP A2/B1 can bind the A2RE sequence<sup>39</sup> and provides an explanation for how sumoylated hnRNP A2/B1 directs the loading of specific EXO-miRNAs into exosomes by binding GAGG, the so-called EXO motif<sup>14</sup>. Furthermore, our results provide the structural basis for hnRNP A2/B1 binding to the UA-rich UAASUUAU motif in the 3' UTR of some mRNAs, which was shown to be necessary for loading the CCR4-NOT complex to mRNAs<sup>19</sup>. Taken together, our structures illustrate the sequence-specific RNA-binding properties of hnRNP A2/B1 and give support to previously reported diverse binding sites<sup>17, 18</sup> (Supplementary Table 3).

The two RRM domains in hnRNP A2/B1 interact with each other extensively in a fixed antiparallel orientation, and the binding surfaces of RRM1 and RRM2 align in the same plane, which forces hnRNP A2/B1 to bind two antiparallel RNA strands or a single-stranded RNA with a long connecting loop. This binding property of hnRNP A2/B1 offers a molecular basis for the previously described “matchmaking” hnRNP A2/B1-HOTAIR interaction, which requires multiple nucleotide recognition motifs within HOTAIR<sup>20</sup>. hnRNP A2/B1 shares similar antiparallel arrangements of its bound RNA as the polypyrimidine tract-specific splicing regulator PTB, in which the two RRM domains interact each other extensively and bind two antiparallel RNA strands. In the PTB-RNA complex, RRM3 and RRM4 form a heterodimer mediated by a hydrophobic interface, and bring together two remote RNA pyrimidine tracts<sup>40</sup>. Moreover, two hnRNP A2/B1 proteins bound to same RNA strands can adopt various orientations, as seen in the structures of the 10mer RNA and the two 10mer RNA mutants A1G and U6G (Fig. 5j, k, l and Supplementary Fig. 5c, d). These diverse orientations are mainly due to the absence of direct interactions between the RRM domains bound to the same RNA strands. This feature may also be involved in RNA-templated aggregation and the formation of hnRNP A2/B1-containing protein-RNA granules in vivo<sup>16</sup>.

It has recently been demonstrated that hnRNP A2/B1 specifically recognizes m<sup>6</sup>A-modified RNAs<sup>23</sup>. These RNAs share the m<sup>6</sup>A consensus sequence RGM<sup>6</sup>ACH and directly bind to the m<sup>6</sup>A mark with high affinity in vivo and in vitro<sup>23</sup>. Prior to this study, the YTH domain was shown to be a “reader” of m<sup>6</sup>A. However, in addition to directly binding-specific proteins, m<sup>6</sup>A can affect RNA binding through an indirect mechanism. This has been shown with two proteins, HuR and hnRNP C, both of which contain RRM domains and do not directly bind m<sup>6</sup>A. In the case of hnRNP C, m<sup>6</sup>A facilitates hnRNP C binding to a UUUUU-tract in mRNAs and long noncoding RNAs (lncRNAs) by promoting local unfolding of RNA. This unfolding is due to the weaker base pairing of U with m<sup>6</sup>A compared to A<sup>41</sup>. m<sup>6</sup>A induced RNA unfolding and increases the accessibility of hnRNP

**Fig. 4** RNA mutants indicate mutivariant-binding mode of hnRNP A2/B1 RRMs. **a** Structure of RRMs (12–195) in complex with A1G-RNA 5'-AGGGACUAGC-3'. **b** Structure of RRMs (12–195) in complex with U6G RNA 5'-AAGGACGAGC-3'. **c** Structure of RRMs (12–195) in complex with A7U-RNA 5'-AAGGACUUGC-3'. **d** Intermolecular contacts between RNA and RRMs in A1G complex. RRM1 is colored in purple-blue, RRM2 is colored in green. **e** Schematic representation of the comparison of different intermolecular interactions between 10mer RNA 5'-AAGGACUAGC-3' and A1G-RNA 5'-AGGGACUAGC-3'. **f** Close-up view showing the specific recognition from A<sub>1</sub> to G<sub>3</sub>. **g** Intermolecular contacts between RNA and RRMs in U6G complex. **h** Schematic representation of intermolecular interactions in the U6G complex. **i** Close-up view showing the specific recognition of A<sub>1</sub>, G<sub>6</sub>, and A<sub>7</sub> in U6G complex. **j** Intermolecular contacts between RNA and RRMs in A7U complex. **k** Schematic representation of intermolecular interactions in A7U complex. **l** Close-up view showing the specific recognition of U<sub>6</sub>, U<sub>7</sub>, and the stacking interactions involved in A<sub>4</sub>, U<sub>6</sub>, F157, and M193 in A7U complex





**Fig. 5** The RNA binding by RRM adopts an antiparallel mode. **a** Superimposition of different structure complex. 8mer-RNA: 5'-AGGACUGC-3' is colored in orange, 10mer-RNA: 5'-AAGGACUAGC-3' is colored in red, A1G-RNA: 5'-AGGGACUAGC-3' is colored in green, U6G: RNA 5'-AAGGACGAGC-3' and A7U-RNA: 5'-AAGGACUUGC-3' are colored in cyan and purple, respectively. **b** Interactions between RRM1 and RRM2; the amino acids participating in interactions are colored in green and RNA is colored in yellow. **c** The overall structure of hnRNP A2/B1-10mer RNA complex. **d** hnRNP A1 in complex with DNA. **e-h** The overall structure of HuD-RNA, HuR-RNA, PABP-RNA, and TDP-43-RNA. From **c** to **h**, RRM1 is colored in blue and RRM2 is colored in red, the linker is colored in green pointed out with a black arrow, the RNA backbone is colored in yellow shown by stick. **i** The overall structure of PTB-RNA complex with an antiparallel RNA-binding mode. RRM3 is colored in blue and RRM4 is colored in red, other labels are the same as from **c** to **h**. **j-l** Crystal packing interactions in 10mer, A1G and U6G. To illustrate the detailed packing interactions of hnRNP A2/B1 carrying two antiparallel RNA stands with other hnRNP A2/B1 molecules, three hnRNP A2/B1 molecules and two RNA strands of each complex are selected to show

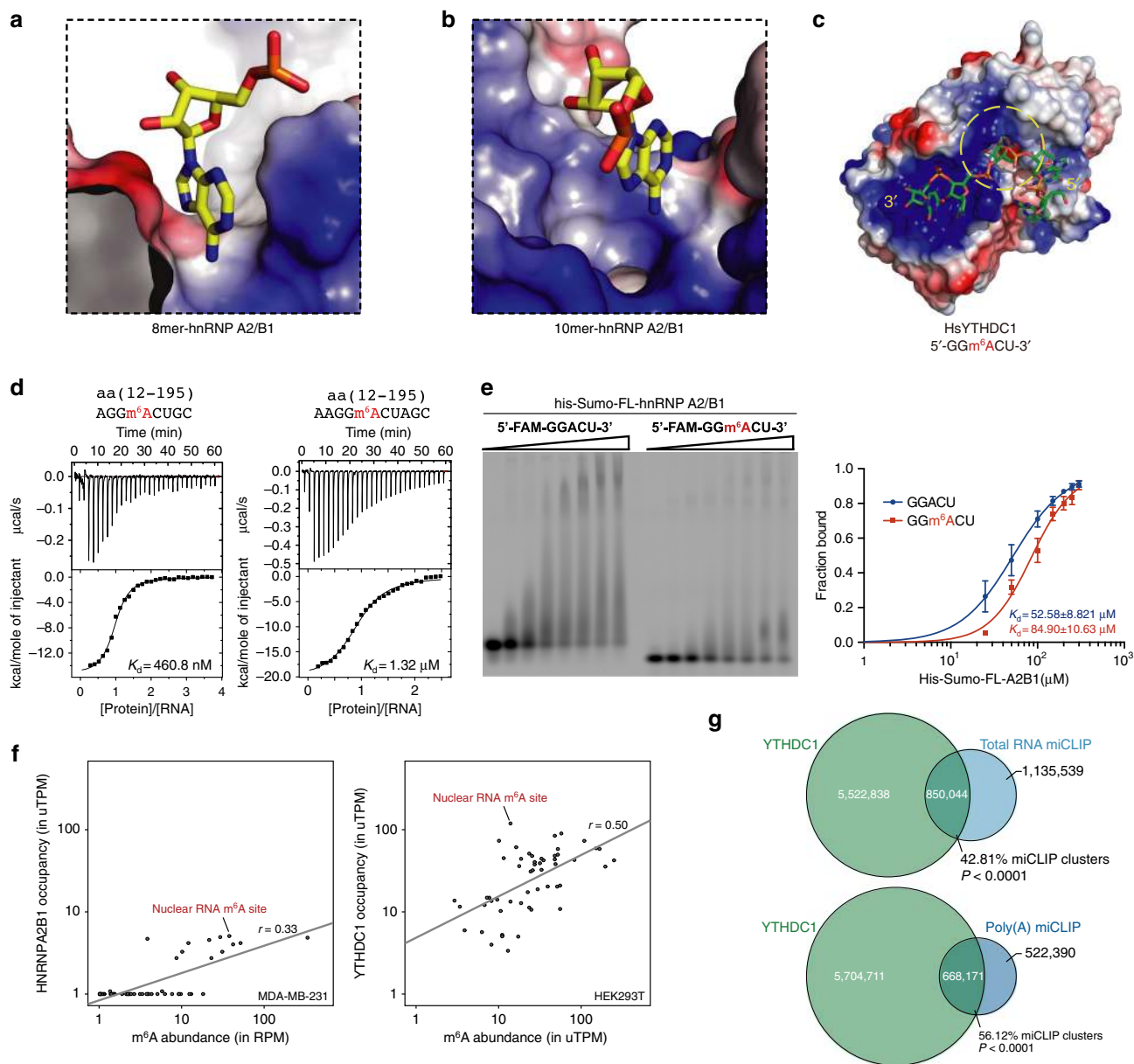
C to single-stranded RNA, and is therefore termed an “m<sup>6</sup>A-switch”<sup>42</sup>. HuR, also known as ELAVL1, has been found to preferentially bind to the 3'-UTR region of mRNAs that lack m<sup>6</sup>A<sup>43</sup>. In this case, m<sup>6</sup>A could impede the formation of a structured RNA motif needed for HuR binding. Our structural study, combined with biochemistry and bioinformatic results, suggest that m<sup>6</sup>A switches may account for the previously seen enhanced hnRNP A2/B1 binding adjacent to m<sup>6</sup>A. Instead of direct binding to m<sup>6</sup>A, m<sup>6</sup>A may promote accessibility of hnRNP A2/B1 to certain binding sites, thereby explaining how m<sup>6</sup>A can facilitate the ability of hnRNP A2/B1 to enhance nuclear events such as pri-miRNA processing. Further in vitro and in vivo investigations will be required to uncover the details of this mechanism.

## Methods

**Preparation of protein samples.** Plasmids encoding different fragments of hnRNP A2/B1 were PCR amplified from the human cDNA. PCR products were double digested with restriction endonuclease *Bam*HI and *Xho*I, then ligated into a modified pET-28a plasmid carrying the Ulp1 cleavage site. Mutations were generated based on the overlap PCR. Recombinant plasmids were confirmed by DNA

sequencing and transformed into *Escherichia coli* BL21 (DE3) to produce target proteins with N-terminal hexahistidine-sumo fusions. *E. coli* cells were cultured in LB medium at 37 °C with 50 mg/l kanamycin until the OD<sub>600</sub> reached 0.6–0.8, then the bacteria were induced with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) at 18 °C for 16 h. Bacteria were collected by centrifugation, resuspended in buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, and lysed by high pressure. Cell extracts were centrifuged at 38,758 × g for 1 h at 4 °C. Supernatants were purified with Ni-NTA (GE), the target protein was washed with lysis buffer and then eluted with a buffer containing 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 500 mM imidazole. Ulp1 protease was added to remove the N-terminal tag and fusion protein of the recombinant protein and dialyzed with lysis buffer 3 h. The mixture was applied to another Ni-NTA resin to remove the protease and uncleaved proteins. Eluted proteins were concentrated by centrifugal ultrafiltration, loaded onto a pre-equilibrated HiLoad 16/60 Superdex 75-pg column in an Äkta-purifier (GE Healthcare), eluted at a flow rate of 1 ml/min with the same buffer containing 10 mM Tris-HCl pH 8.0, 100 mM NaCl. Peak fractions were analyzed by SDS-PAGE (15%, w/v) and stained with Coomassie brilliant blue R-250. Purified fractions were pooled together and concentrated by centrifugal ultrafiltration. The concentration was determined by A<sub>280</sub>. The protein was concentrated to 10 mg/ml for crystallization trials.

**RNA oligonucleotides.** The RNA oligonucleotides (8mer-m<sup>6</sup>A: 5'-AGGm<sup>6</sup>A-CUGC-3', 10mer-m<sup>6</sup>A: 5'-AAGGm<sup>6</sup>ACUAGC-3') with m<sup>6</sup>A modification were ordered from Dharmacon (Thermo Scientific.), and the other unmodified or



**Fig. 6** hnRNP A2/B1 does not specifically recognize m<sup>6</sup>A-modified RNA. **a** Surface representation of the environment around A<sub>4</sub> in 8mer RNA complex. **b** Surface representation of the environment around A<sub>4</sub> in 10mer RNA complex. **c** Surface representation of the canonical N<sup>6</sup>-methylated adenosine binding mode in HsYTHDC1. The aromatic cage is circled with yellow dashline. **d** ITC data of hnRNP A2/B1(12–195) with 8mer and 10mer RNA targets carried N<sup>6</sup>-methylated adenosine. **e** EMSA experiment shows the binding affinity of full-length hnRNP A2/B1 with 5'-FAM-labeled RNA substrates with or without m<sup>6</sup>A modification. Uncropped gel image is shown in Supplementary Fig. 7e. The data represent the mean of three independent experiments, with standard deviation (SD) values indicated by error bars. **f** YTHDC1 shows preferential binding to m<sup>6</sup>A sites in nuclear RNA compared to hnRNP A2/B1. For hnRNP A2/B1, the m<sup>6</sup>A-Seq reads that overlapped with each m<sup>6</sup>A site was plotted on the x-axis, and the HITS-CLIP reads that overlap with each site were plotted on the y-axis. A similar analysis was used to examine YTHDC1 binding at these m<sup>6</sup>A sites. miCLIP reads that overlapped with the m<sup>6</sup>A sites were plotted on the x-axis, and the YTHDC1 iCLIP reads that overlapped with the m<sup>6</sup>A sites were plotted on the y-axis. **g** YTHDC1-m<sup>6</sup>A tag cluster overlap. A Venn diagram indicating the cluster overlap is shown. Roughly, 43 and 56% of miCLIP tag clusters from total cellular RNA and poly(A) RNA showed a significant overlap with the YTHDC1 iCLIP clusters, respectively

unlabeled RNA oligonucleotides were synthesized by the IDT-394 synthesizer in our own lab. The 5'-FAM-labeled RNA chains (5'-FAM-GGACU-3' and 5'-FAM-GGm<sup>6</sup>ACU-3') were ordered from Bioneer Corporation. All the RNA oligonucleotides used for crystallization and biochemical experiments in this study are summarized in Supplementary Table 2.

**Crystallization and data collection.** hnRNP A2/B1 RRM (12–195) in complex with 8-nt-RNA 5'-AGGACUGC-3' was crystallized using the hanging drop vapor diffusion method by mixing 1  $\mu$ l of protein–RNA mixture (molar ratio 1:1.2) and 1  $\mu$ l of reservoir solution at 20 °C. The crystals suitable for X-ray diffraction were grown in reservoir solution consisting of 0.1 M Tris pH 8.5 and 25% (w/v)

polyethylene glycol 3,350 (Hampton Research). hnRNP A2/B1 RRM (12–195) in complex with 10-nt-RNA 5'-AAGGACUAGC-3' was screened as above. The crystal suitable for X-ray diffraction was grown in reservoir solution consisting of 0.2 M Tri-sodium citrate and 20% (w/v) polyethylene glycol 3,350 (Hampton Research). A1G, U6G, and A7U were crystallized as the methods mentioned above in solution containing 20% PEG 3,000, 0.1 M sodium citrate pH 5.5; 20% PEG 3,350, 0.2 M lithium sulfate, 0.1 M Bis-Tris pH 6.5; 25% PEG 1,500, 0.1 M MMT pH 9.0, respectively. Data collection was performed at 100 K with cryo-protectant solution (reservoir solution supplemented with an additional 20% (v/v) glycerol). Diffraction data were collected using a wavelength of 0.97776 Å at beamline BL18U1 of the Shanghai Synchrotron Radiation Facility (SSRF).

**Structure determination and refinement.** For hnRNP A2/B1 RRM1 (12–195)-8-nt complex, the diffraction data set was processed and scaled using HKL3000. The phase was determined by molecular replacement using the program Phaser with the structure of UPI (PDB code: 1U1Q) as the search model<sup>44</sup>. Cycles of refinement and model building were carried out using REFMAC5 and COOT until the crystallography  $R_{\text{factor}}$  and  $R_{\text{free}}$  converged to 19.16% and 23.62%, respectively<sup>45, 46</sup>. Ramachandran analysis showed that 96.0 of the residues were in the most favored region, with 4.0% in the additionally allowed region. For hnRNP A2/B1-10-nt complex, the diffraction data set was processed and scaled using the HKL3000 package. The phase was determined by molecular replacement using the program Phaser with the hnRNP A2/B1(12–195) model collected before as the search model. Cycles of refinement and model building were carried out using REFMAC5 and COOT until the crystallography  $R_{\text{factor}}$  and free  $R_{\text{free}}$  converged to 18.39% and 22.27%, respectively. Ramachandran analysis showed that similarly to hnRNP A2/B1-8-nt, 99% of the residues were in the most favored region, with 1% in the additionally allowed region. For another three complex structures, the same methods were used to solve the structures as mentioned above. The details of data collection and processing are presented in Supplementary Table 1. All structure figures were prepared with PyMOL (DeLano Scientific).

**ITC measurements.** ITC assays were carried out on a MicroCal ITC200 calorimeter (Malvern) at 25 °C. The buffer used for proteins and RNA oligomers was 10 mM HEPES pH 8.0, 50 mM KCl, 1 mM EDTA, and 1 mM BME. The concentrations of proteins were determined spectrophotometrically. The RNA oligonucleotides were diluted in the buffer to 5–15  $\mu\text{M}$ . The ITC experiments involved 20–30 injections of protein into RNA. The sample cell was loaded with 250  $\mu\text{l}$  of RNA at 5  $\mu\text{M}$  and the syringe with 80  $\mu\text{l}$  of protein at 100  $\mu\text{M}$ ; for weak complexes, the measurement was repeated with increased concentrations. Reference measurements were carried out to compensate for the heat of dilution of the proteins. Curve fitting to a single binding site model was performed by the ITC data analysis module of Origin 7.0 (MicroCal) provided by the manufacturer.  $\Delta G^\circ$  of protein–RNA binding was computed as  $RT \ln(1/K_D)$ , where  $R$ ,  $T$ , and  $K_D$  are the gas constant, temperature and dissociation constant, respectively.

**EMSA.** Aliquot of 0.5  $\mu\text{M}$  of FAM-labeled RNA was mixed with increasing concentrations of full-length hnRNP A2/B1 proteins in a buffer containing 10 mM HEPES pH 8.0, 50 mM KCl, 1 mM EDTA, and 5 mM beta-mercaptoethanol in a total volume of 10  $\mu\text{l}$  and incubated at room temperature for 30 min. The electrophoresis was performed with 6% native-PAGE at 4 °C in running buffer containing 0.5 $\times$  Tris-borate-EDTA (TBE) buffer. The gel was visualized by using a Typhoon FLA-9000 (GE Healthcare) using a method for FAM (Laser 488 nm). Bound and free RNA were quantified using ImageJ. Binding curves were fit individually using GraphPad Prism 6.0 software fitting with “One site – Specific binding with Hill slope” (GraphPad Software). Curves were normalized as percentage of bound oligonucleotides and reported is the mean  $\pm$  SD of the interpolated  $K_d$  from three independent experiments.

**Analytical gel filtration.** Proteins of RRM1 and RRM2 domains of hnRNP A2/B1 were purified using the same procedure as hnRNP A2/B1 (12–195). Analytical gel filtration chromatography was performed at 4 °C using a Superdex 75 10/300 global column (GE Healthcare) pre-equilibrated in 20 mM Tris, pH 8.0, 100 mM NaCl. Aliquots of 100  $\mu\text{l}$  of samples including various hnRNP A2/B1 constructs were injected at a flow rate of 0.3 ml/min. To study complex formation, RRM1 and RRM2 proteins were mixed and incubated on ice for at least 1 h prior to loading.

**Next-generation sequencing data analysis.** Nuclear hnRNP A2/B1-HITS-CLIP sequence data were obtained from a previously published study<sup>23</sup> (GEO accession number: GSE70061, SRA accession numbers: SRR2071655 and SRR2071656, last update date: 21 Jun 2015). In addition to the raw data, the author uploaded sequence alignment files, GSM1716539\_A2B1\_HITS\_CLIP\_1.bedgraph.gz and GSM1716539\_A2B1\_HITS\_CLIP\_2.bedgraph.gz were also obtained from the GEO database for comparison purposes. Robust crosslinking-induced mutation sites (CIMS) (FDR  $\leq$  0.001) in hnRNP A2/B1 HITS-CLIP data were called using a method published elsewhere<sup>47</sup>. UV-induced deletion sites<sup>48</sup> were used as hnRNP A2/B1-binding sites.

Nuclear m<sup>6</sup>A-seq data from MDA-MB-231 cells were obtained from a previously published study<sup>49</sup> (GEO accession number: GSE60213, SRA accession numbers: SRR1539129 and SRR1539130, last update date: 15 Nov 2016). Adapter-free, high-quality sequence reads were aligned to the hg19 genome build using bowtie2 according to the source publication<sup>49</sup>. RPM (reads per million mapped reads) was calculated using bedtools.

For YTHDC1 binding at m<sup>6</sup>A sites in HEK293T cells, miCLIP sequencing data<sup>34</sup> (GEO accession number: GSE63753) and YTHDC1 iCLIP data<sup>35</sup> (GEO accession number: GSE78030) were obtained from the GEO database. Sequence alignments were carried out according to the respective publications. Images of genome alignments were prepared using IGV genome browser and Adobe Illustrator.

**Comparison of hnRNP A2/B1 and YTHDC1 binding at m<sup>6</sup>A sites.** hnRNP A2/B1 or YTHDC1 binding and m<sup>6</sup>A stoichiometry at 10 bp flanking miCLIP sites<sup>34</sup> on nuclear RNAs such as *MALAT1*, *NEAT1*, and *XIST* was compared using an XY-scatter plot in R. Only m<sup>6</sup>A sites conforming a non-BCANN consensus were considered for this analysis. These represent unique sites obtained from merging (mergeBed -s -d 2) of CIMS- and CITS-based m<sup>6</sup>A site calls from ref. <sup>34</sup>. All the rRNA, tRNA, and mitochondrial genomic miCLIP sites were removed. Tag counting was performed using the bedtools suite. Tag counts (uTPM + 1) were compared using scatter plots and Pearson correlation coefficients ( $r$ ) were determined in R. Cluster overlap analysis was carried out using bedtools intersect tool (intersectBed -s -u).

**Data availability.** The coordinates that support the findings of this study have been deposited in the Protein Data Bank with accession codes 5EN1 for RRM1-8mer-RNA complex and 5HO4 for RRM1-10mer RNA structures. A1G, U6G, and A7U are 5WWE, 5WWF, and 5WWG, respectively. Other data in this study are available from the corresponding author on reasonable request.

Received: 30 June 2017 Accepted: 22 December 2017

Published online: 29 January 2018

## References

- Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* **8**, 533–543 (2007).
- Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **582**, 1977–1986 (2008).
- He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell. Mol. Life Sci.* **66**, 1239–1256 (2009).
- Gabut, M., Chaudhry, S. & Blencowe, B. J. SnapShot: the splicing regulatory machinery. *Cell* **133**, 192 e1 (2008).
- Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–D308 (2011).
- Auweter, S. D., Oberstrass, F. C. & Allain, F. H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–4959 (2006).
- Kozu, T., Henrich, B. & Schafer, K. P. Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1. *Genomics* **25**, 365–371 (1995).
- Burd, C. G., Swanson, M. S., Gorlach, M. & Dreyfuss, G. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc. Natl Acad. Sci. USA* **86**, 9788–9792 (1989).
- Kim, H. J. et al. Mutations in prion-like domains in hnRNP A2B1 and hnRNP A1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
- Lee, B. J. et al. Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**, 543–558 (2006).
- Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474**, 1417–1438 (2017).
- Choi, H. S., Lee, H. M., Jang, Y. J., Kim, C. H. & Ryu, C. J. Heterogeneous nuclear ribonucleoprotein A2/B1 regulates the self-renewal and pluripotency of human embryonic stem cells via the control of the G1/S transition. *Stem Cells* **31**, 2647–2658 (2013).
- Beriault, V. et al. A late role for the association of hnRNP A2 with the HIV-1 hnRNP A2 response elements in genomic RNA, Gag, and Vpr localization. *J. Biol. Chem.* **279**, 44141–44153 (2004).
- Villarroya-Beltri, C. et al. Sumoylated hnRNP A2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* **4**, 2980 (2013).
- Hutten, S. & Dormann, D. hnRNP A2/B1 function in neurodegeneration: it's a gain, not a loss. *Neuron* **92**, 672–674 (2016).
- Martinez, F. J. et al. Protein-RNA networks regulated by normal and ALS-associated mutant HNRNP A2B1 in the nervous system. *Neuron* **92**, 780–795 (2016).
- Huelga, S. C. et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).
- Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
- Geissler, R. et al. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev.* **30**, 1070–1085 (2016).
- Meredith, E. K., Balas, M. M., Sindy, K., Haislop, K. & Johnson, A. M. An RNA matchmaker protein regulates the activity of the long noncoding RNA HOTAIR. *RNA* **22**, 995–1010 (2016).
- Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326 (2014).
- Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).

23. Alarcon, C. R. et al. HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. *Cell* **162**, 1299–1308 (2015).
24. Daubner, G. M., Clery, A. & Allain, F. H. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.* **23**, 100–108 (2013).
25. Ding, J. et al. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115 (1999).
26. Barraud, P. & Allain, F. H. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. *J. Biomol. NMR* **55**, 119–138 (2013).
27. Morgan, C. E. et al. The first crystal structure of the UP1 domain of hnRNP A1 bound to RNA reveals a new look for an old RNA binding protein. *J. Mol. Biol.* **427**, 3241–3257 (2015).
28. Wang, X. & Tanaka Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141–145 (2001).
29. Wang, H. et al. The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 373–380 (2013).
30. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845 (1999).
31. Mackereth, C. D. et al. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408–411 (2011).
32. Lukavsky, P. J. et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449 (2013).
33. Wu, B., Li, L., Huang, Y., Ma, J. & Min, J. Readers, writers and erasers of N6-methyladenosine modification. *Curr. Opin. Struct. Biol.* **47**, 67–76 (2017).
34. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
35. Patil, D. P. et al. m6A RNA methylation promotes XIIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
36. Xiao, W. et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).
37. McKay, S. J. & Cooke, H. hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. *Nucleic Acids Res.* **20**, 6461–6464 (1992).
38. Hutchison, S., LeBel, C., Blanchette, M. & Chabot, B. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J. Biol. Chem.* **277**, 29745–29752 (2002).
39. Munro, T. P. et al. Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking. *J. Biol. Chem.* **274**, 34389–34395 (1999).
40. Oberstrass, F. C. et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–2057 (2005).
41. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).
42. Liu, N. et al. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
43. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16**, 191–198 (2014).
44. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
46. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
47. Moore, M. J. et al. Mapping argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* **9**, 263–293 (2014).
48. Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29**, 607–614 (2011).
49. Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482–485 (2015).

## Acknowledgements

We thank the staff from BL18U1 beamline of the National Facility for Protein Science in Shanghai (NFPS) at the Shanghai Synchrotron Radiation Facility for assistance during data collection. We thank Dr Jinzhong Lin (Fudan University) for helpful discussions and critical manuscript reading. This work was supported by grants from the National Natural Science Foundation of China (31230041) and the National Basic Research Program of China (2011CB966304 and 2012CB910502) to J.M., and the National Institutes of Health (R01 CA186702) to S.R.J.

## Author contributions

B.W. and S.S. expressed, purified, and grew crystals of the hnRNP A2B1-RNA complex and performed the biochemical assays. B.W., S.S., and H.L. synthesized the RNA oligonucleotides. B.W., J.M., and J.G. collected X-ray diffraction data and solved the hnRNP A2B1-RNA complex structure. D.P. carried out the bioinformatic analysis. B.W., J.M., D.P., and S.J. wrote and revised the manuscript. J.M. supervised the structural and biochemical study. S.J. supervised the bioinformatic study.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02770-z>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018