ARTICLE
PUBLISHED ONLINE: 12 DECEMBER 2016 | DOI: 10.1038/NCHEMBIO.2251

nature chemical biology

Precise small-molecule recognition of a toxic CUG RNA repeat expansion

Suzanne G Rzuczek[1], Lesley A Colgan[2], Yoshio Nakai[1], Michael D Cameron[3], Denis Furling[4], Ryohei Yasuda[2] & Matthew D Disney[1]*

Excluding the ribosome and riboswitches, developing small molecules that selectively target RNA is a longstanding problem in chemical biology. A typical cellular RNA is difficult to target because it has little tertiary, but abundant secondary structure. We designed allele-selective compounds that target such an RNA, the toxic noncoding repeat expansion (r(CUG)^{exp}) that causes myotonic dystrophy type 1 (DM1). We developed several strategies to generate allele-selective small molecules, including non-covalent binding, covalent binding, cleavage and on-site probe synthesis. Covalent binding and cleavage enabled target profiling in cells derived from individuals with DM1, showing precise recognition of r(CUG)^{exp}. In the on-site probe synthesis approach, small molecules bound adjacent sites in r(CUG)^{exp} and reacted to afford picomolar inhibitors via a proximity-based click reaction only in DM1-affected cells. We expanded this approach to image r(CUG)^{exp} in its natural context.

RNA is considered to be an important target for lead therapeutics and chemical probes of function. The most activity in this area has been in the development of small molecules that target the complex three-dimensional folds in the bacterial ribosome[1,2] and in riboswitches[3,4]. Oligonucleotide-based strategies such as antisense or RNA interference are most commonly used to target human RNAs, whereas there are few examples of small molecules that do so. Historically, it has been difficult to develop small molecules that potently and precisely target most cellular RNAs, which usually have extensive secondary structure but limited tertiary structure. RNA repeat expansions, which cause >20 genetically defined diseases including amyotrophic lateral sclerosis, Huntington's disease, fragile X syndrome and various forms of muscular dystrophy, are notable examples of such RNAs.

Myotonic dystrophy type 1 (DM1) is the most common adult-onset form of muscular dystrophy and is caused by an expansion of r(CUG), r(CUG)^{exp}, in the 3′ untranslated region (UTR) of the dystrophia myotonica protein kinase (*DMPK*) mRNA[5]. In the mutant *DMPK* transcript, r(CUG)^{exp} ranges from 100 to several thousand repeats and folds into a hairpin that has gain-of-function activity. In particular, r(CUG)^{exp} binds and sequesters proteins such as muscleblind-like 1 (MBNL1) splicing factor, leading to deregulation of alternative pre-mRNA splicing. Furthermore, mutant *DMPK* mRNAs harboring r(CUG)^{exp} are retained in nuclear foci[6,7] and have markedly reduced nucleocytoplasmic transport in DM1-affected cells[8,9].

Several studies have shown that binding r(CUG)^{exp} with oligonucleotides, peptides or small molecules can improve DM1-associated defects, providing a therapeutic strategy. Of these modalities, small molecules are favored for therapeutic use. The development of selective compounds that target RNA is difficult, owing to RNA's anionic nature and the relatively low abundance of most cellular RNAs[10], among other factors. To identify small molecule leads, we utilized a strategy called Inforna, which identifies highly selective, privileged RNA-motif–small-molecule interactions[11,12]. Inforna has successfully facilitated the design of potent small-molecule modulators of several RNA repeat expansion disorders[13–15] and cancer-related microRNAs[11,16].

We developed approaches to enhance, study and deliver potent and selective designer small molecules that specifically target r(CUG)^{exp}. These strategies include non-covalent binding, covalent binding, cleavage and on-site probe synthesis. Our non-covalent-binding small molecule modulated r(CUG)^{exp} dysfunction in cells derived from patients with DM1. Evolution of our small molecule modulator into a covalent binder, through the attachment of a chemical cross-linker, not only enhanced its potency but also enabled the identification of cellular RNA targets. We found that the mutant *DMPK* allele containing r(CUG)^{exp} was selectively targeted over other mRNAs containing short r(CUG) repeats, including wild-type *DMPK*. Covalent target identification was complemented by small-molecule cleavage, facilitated by the attachment of bleomycin, which selectively reduced mutant, but not wild-type, *DMPK*. On-site probe synthesis established that r(CUG)^{exp} could be used as a catalyst to synthesize potent picomolar inhibitors of DM1-associated cellular defects. The approach was expanded into on-site fluorescent probe synthesis, which enabled the imaging of r(CUG)^{exp} in its natural context in live cells.

RESULTS

Initial compound design

A compound was designed to target r(CUG)^{exp} using an RNA-motif-small-molecule (Inforna) database[11], which identified *bis*-benzimidazole H as a ligand for the 1 × 1 nucleotide UU internal loops found in r(CUG)^{exp} (**Fig. 1a**). Dimeric display of H modules on an *N*-methyl peptide backbone generated compound 2H-K4NMeS (**1**), allowing recognition of two adjacent UU loops and the distance between them (**Fig. 1a**). **1** has improved metabolic stability compared with our previously reported compound, 2H-K4NMe[17] (**Supplementary Results**; **Supplementary Note** and **Supplementary Fig. 1a**,**b**). Notably, treatment of DM1-patient-derived cells with 100 nM 2H-K4NMeS selectively improved the MBNL1-dependent *MBNL1* exon 5 pre-mRNA splicing defect by semiquantitative and quantitative measurements (**Fig. 1b**; **Supplementary Fig. 2** and **Supplementary Table 1**). **1** also improved the alternative splicing of other MBNL1-regulated mRNAs to a similar extent as *MBNL1* exon 5, including

[1]Departments of Chemistry and Neuroscience, The Scripps Research Institute, Jupiter, Florida, USA. [2]Max Planck Florida Institute for Neuroscience, Jupiter, Florida, USA. [3]Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida, USA. [4]Sorbonne Universités UPMC Université Paris 06, INSERM, CNRS, Centre de Recherche en Myologie, Institut de Myologie, Paris, France. *e-mail: disney@scripps.edu

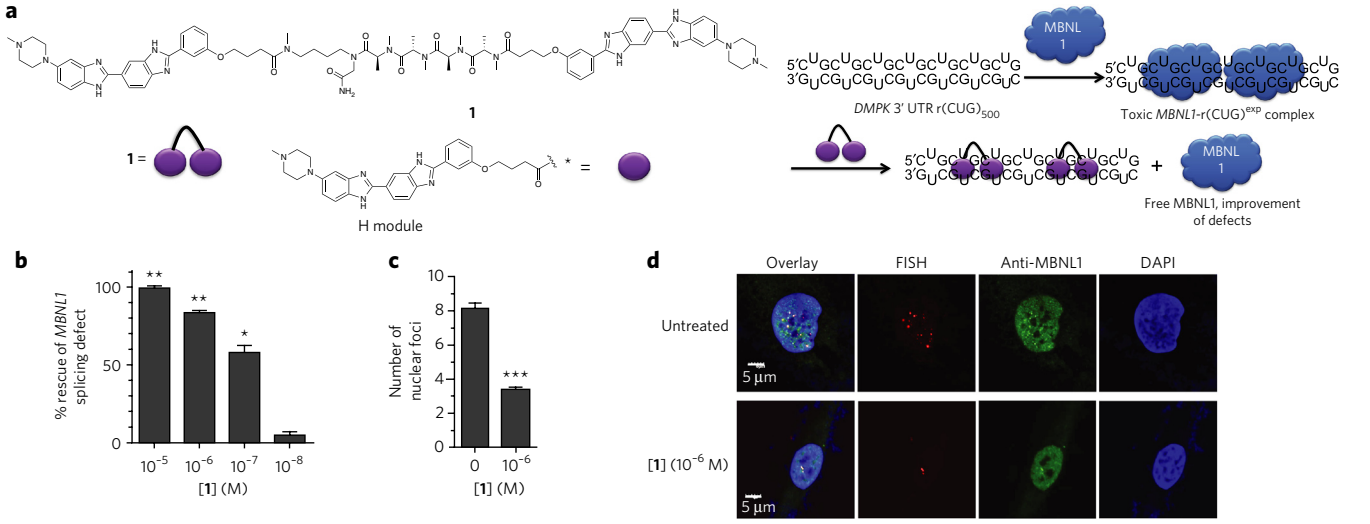

Figure 1 | A rationally designed small molecule improves DM1-associated defects in patient-derived cells. DM1 is caused by a toxic gain of function by r(CUG)^{exp}. We designed small molecules that ameliorate disease-associated phenotypes and that can be used to assess target selectivity in DM1 cells. (**a**) Structure of **1**, a non-covalent binding compound. Multivalent small molecules are represented by purple spheres (RNA-binding modules) connected by a line (*N*-methyl peptide scaffold). Also shown is the secondary structure of r(CUG)^{exp} and binding of MBNL1, which causes disease. Release of MBNL1 by small molecules improved DM1-associated defects. (**b**) **1** rescued *MBNL1* exon 5 splicing defects in DM1-affected cells (n = 6, 6 biological replicates, 2 replicate experiments). (**c**) **1** reduced the number of r(CUG)^{exp} nuclear foci in DM1 cells (n = 100 cells, 5 biological replicates, 2 replicate experiments). (**d**) Representative images from RNA-FISH experiments to assess formation of nuclear foci. Scale bars represent 5 µM. Data represent mean values ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, as determined by a two-tailed Student t test.

CAMK2G exon 14 and *NCOR2* exon 45a, and had more modest effects on *NFIX* exon 7 (**Supplementary Fig. 2a–c**). Notably, **1** had no effect on MBNL1-dependent splicing in non-DM1 cells or NOVA-dependent *MAP4K4* splicing in DM1-patient-derived cells (**Supplementary Fig. 2d,e**).

As mentioned above, r(CUG)^{exp} is sequestered in nuclear foci[7,18,19]. Treatment with **1** also improved this hallmark of DM1, reducing the number of foci by ~50% at a dosage of 1 µM (**Fig. 1c,d**). To ensure that the observed reduction was not due to the compound blocking binding of the RNA fluorescent *in situ* hybridization (FISH) probe, we studied whether **1** alleviated another defect caused by the sequestration of r(CUG)^{exp} in the nucleus, impaired nucleocytoplasmic transport of *DMPK* mRNA[20]. If compound binding releases r(CUG)^{exp} from foci or prevents foci formation, nucleocytoplasmic transport could be improved. To investigate this phenomenon, we embedded r(CUG)$_{800}$ in the 3′ UTR of the firefly *luciferase* gene[15]. The presence of r(CUG)$_{800}$ prevents trafficking of *luciferase* mRNA to the cytoplasm and reduces its translation, as measured by enzymatic activity[15]. Consistent with the observed reduction in foci, **1** increased luciferase enzymatic activity, but did not affect *luciferase* mRNA levels (**Supplementary Fig. 2h,i**). Thus, **1** targeted r(CUG)^{exp} and disabled its gain-of-function activities.

A covalent approach to target modulation and validation

Selective recognition is required for generating preclinical probes. Discrimination between r(CUG) repeats of pathogenic and non-pathogenic length was assessed using a covalent approach; **1** was appended with cross-linking (chlorambucil, CA) and purification (biotin) modules, yielding 2H-K4NMeS-CA-biotin (**2**) (**Fig. 2a** and **Supplementary Note**). **2** selectively reacted with r(CUG)$_{109}$ and potently inhibited a model r(CUG)^{exp}-MBNL1 complex *in vitro*, whereas a control compound lacking the RNA-binding modules did not (2NAc-K4NMeS-CA-biotin, compound **2a**; **Supplementary Fig. 3a** and **Supplementary Note**). In addition, the precursor lacking the cross-linking moiety, 2H-K4NMeS-biotin-NH$_2$ (**2b**; **Supplementary Fig. 3b,c** and **Supplementary Note**), had a similar IC$_{50}$ as **1**, indicating that the enhanced potency was due to the compound's inability to dissociate from the target after covalent attachment. In cells, 10 nM of **2** potently and selectively improved DM1 molecular features such as deregulation of MBNL1-dependent splicing and the presence of nuclear foci in two different DM1-patient-derived cell lines (**Fig. 2b,c** and **Supplementary Fig. 4a–h**). As observed for **1**, **2** facilitated nucleocytoplasmic transport of a transcript harboring r(CUG)^{exp} in its 3′ UTR while not affecting the levels of the transcript itself (**Supplementary Fig. 4i,j**).

To validate the targets of **1** and **2** in cells, we used chemical cross-linking and isolation by pulldown (Chem-CLIP) (**Supplementary Fig. 5a**) and its competitive variant (C-Chem-CLIP)[21]. Cells were treated with **2** and then cross-linked targets were isolated. Quantification revealed that, in DM1-patient-derived cells. **2** enriched the pulldown of *DMPK* mRNA by ~13,000-fold (**Fig. 2d**). In contrast, we observed no enrichment in non-DM1-affected cells (**Fig. 2d**). Furthermore, pre-dosing of DM1-affected cells with **1** followed by treatment with **2** (C-Chem-CLIP) showed that **1** effectively competed with **2** for binding to *DMPK* mRNA with a K_i of 2.9 nM (**Supplementary Fig. 5b**). Thus, both compounds selectively and potently bound to r(CUG)^{exp}.

Direct binding to r(CUG)^{exp} in *DMPK* was confirmed by using Chem-CLIP-Map (**Fig. 2e** and **Supplementary Table 2**)[22]. In these studies, total RNA isolated from DM1-patient-derived cells treated with **2** was annealed with antisense oligonucleotides targeting specific portions of *DMPK* mRNA. After cleavage with RNase H, the mRNA was digested into fragments with and without r(CUG)^{exp}. Regions of *DMPK* mRNA bound by the compound were captured and quantified by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) using specific primers for each fragment of the mRNA. Enrichment was only observed for fragments containing r(CUG)^{exp} (primer sets B, binds the r(CUG)^{exp}-containing fragment 5′ of the repeats, and C, binds the r(CUG)^{exp}-containing fragment 3′ of the repeats; **Fig. 2e**), illustrating that **2** recognized the site that it was specifically designed to target in DM1-patient-derived cells.

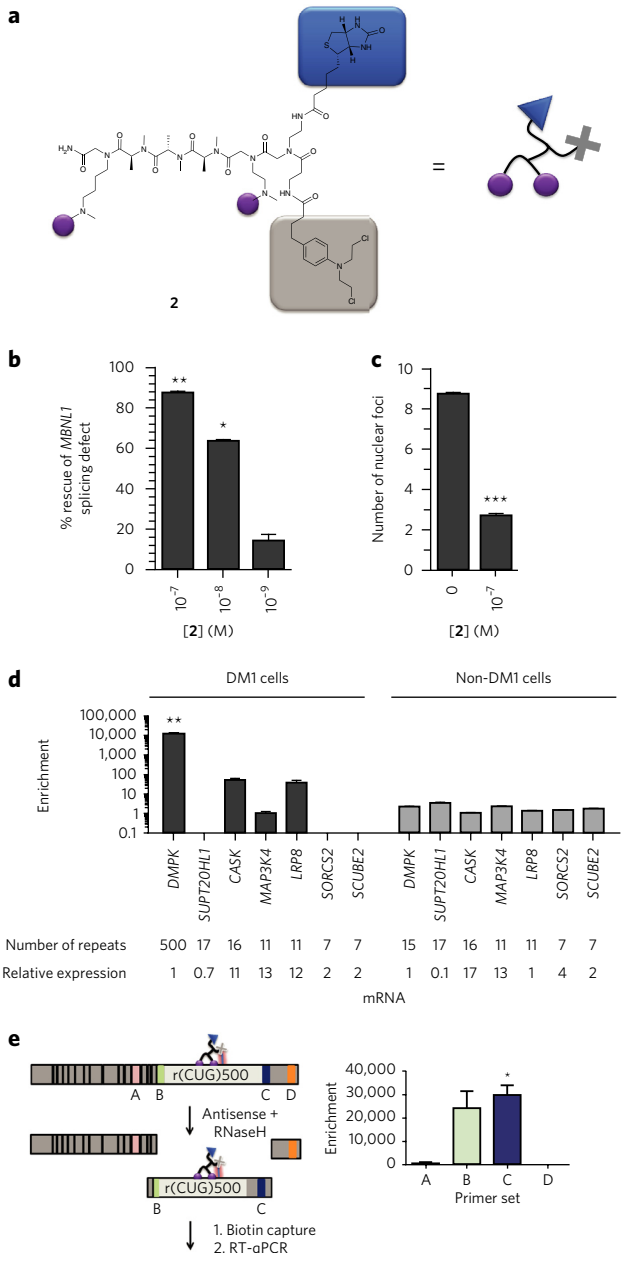

Figure 2 | Covalent (Chem-CLIP) approach to inhibit r(CUG)^{exp} dysfunction and assess cellular selectivity. (**a**) Structure of **2**; purple spheres indicate the RNA-binding modules, the blue triangle is the biotin purification tag and the gray X is the cross-linking module chlorambucil. (**b**) Treatment of **2** in DM1-affected cells rescued *MBNL1* pre-mRNA splicing (n = 6, 6 biological replicates, 2 replicate experiments). (**c**) **2** reduced the number of nuclear foci in DM1 cells (n = 100 cells, 5 biological replicates, 2 replicate experiments). (**d**) **2** reacted selectively with r(CUG)^{exp}-containing *DMPK* mRNA in DM1-affected cells. *SUPT20HL1*, *SORCS2* and *SCUBE2* mRNAs were not detected in the pulled-down fraction. Enrichment in wild-type *DMPK* was not observed in non-DM1 cells. Enrichment was calculated by normalizing to the least-abundant mRNA in each cell type (six biological replicates each with three technical replicates, two replicate experiments). (**e**) Chem-CLIP-Map defined the sites of reaction of **2** in *DMPK* mRNA in DM1-affected cells. RNase H digestion of the *DMPK* mRNA followed by capture and RT-qPCR revealed that the molecule reacted with the 3′ UTR fragment that contained r(CUG)^{exp}. Primer set locations are indicated with A, B, C and D (n = 6, 6 biological replicates each with 3 technical replicates, 2 replicate experiments). Data represent mean values ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$, as determined by a two-tailed Student t test.

We also quantified all other human mRNAs with short r(CUG) repeats pulled down by **2** in DM1-patient-derived and non-DM1-patient-derived cells. In DM1-affected cells, no RNAs were enriched within two orders of magnitude of the enrichment observed with *DMPK* (**Fig. 2d**). In non-DM1 cells, there was minimal enrichment for any RNA containing short r(CUG) repeats (**Fig. 2d**). Furthermore, allele-selective RT-qPCR revealed significant enrichment in the mutant expanded *DMPK* allele (**Supplementary Fig. 6a**). Thus, these designer small molecules discriminated between RNAs on the basis of repeat length in DM1-patient-derived cells. In this particular case, selectivity did not appear to be derived from compound localization, as **1** and **2** were distributed throughout the nucleus and cytoplasm (**Supplementary Fig. 6b**).

To define potential mechanisms endowing discrimination of pathogenic r(CUG) repeats, we performed a series of assays. **1** failed to bind DNA and potential competitor RNAs, but bound r(CUG)$_{12}$ and r(CUG)$_{109}$ with K_d values of 280 and 12 nM, respectively (**Supplementary Fig. 7**). Alteration of the peptide linkages for improved metabolic stability had little effect on the binding of **1**, as it bound r(CUG)$_{12}$ with binding affinity and stoichiometry similar to those of 2H-K4NMe (**Supplementary Fig. 7**)[17]. Analysis of Hill coefficients (HCs) showed that binding to r(CUG)$_{12}$ was negatively cooperative (HC = 0.6), whereas binding to r(CUG)$_{109}$ was not (HC = 1.0). Thus, selective recognition of r(CUG)^{exp} was likely due to a variety of factors that include differences in binding cooperativity, mass action and the fact that shorter r(CUG) repeats in the context of a cellular transcript do not form the repeating 1 × 1 UU internal loops required for binding.

A cleavage approach to target modulation and validation

Small molecules that selectively cleave RNAs could have broad applicability and extend targeting by such probes to a variety of cellular RNAs. Thus, we developed small molecules that cleave r(CUG)^{exp} in cells. The natural product bleomycin cleaves RNA *in vitro*[23], and we conjugated it to **1**, yielding 2H-K4NMeS-bleomycin (**3**) (**Fig. 3a** and **Supplementary Note**). *In vitro*, a model of r(CUG)^{exp} was cleaved by **3** between U and G nucleotides (**Supplementary Fig. 8a–d**). No cleavage was observed when the r(CUG)^{exp} RNA was incubated with a control compound lacking the RNA-binding modules (2NAc-K4NMeS-bleomycin, compound **3a**; **Supplementary Fig. 8c**,**d** and **Supplementary Note**). Notably, treatment of DM1-affected cells with **3** (250 nM) selectively reduced *DMPK* mRNA levels, by ~30%, but not those of other mRNAs with short r(CUG) repeats (**Fig. 3b**,**c**). Furthermore, it markedly improved defects in MBNL1-dependent pre-mRNA splicing while having no effect on non-MBNL1-dependent pre-mRNA splicing (**Fig. 3d** and **Supplementary Fig. 9**). Notably, **3a**, which lacks the RNA-binding modules, had no effect on pre-mRNA splicing (**Supplementary Fig. 9e** and **Supplementary Note**). Allele-selective qPCR confirmed cleavage of the mutant-expanded *DMPK* allele by **3** (**Supplementary Fig. 6c**). Furthermore, the cleavage effects of **3** were specific, as they were effectively competed by co-treatment of DM1-affected cells with **1** (**Fig. 3b**). Finally, cleavage of *DMPK* mRNA and alterations in pre-mRNA splicing were not evident when **3** was added to non-DM1 cells (**Supplementary Figs. 8e** and **9c**). Thus, both target depletion by cleavage, dubbed small–molecule nucleic acid profiling by cleavage applied to RNA (Ribo-SNAP), and covalent capture via Chem-CLIP showed precise, allele-selective recognition of r(CUG)^{exp}.

Potent and selective intracellular probe synthesis

To generate allele-selective compounds and further improve potency, we used *in situ* click chemistry[24] to direct r(CUG)^{exp} to catalyze the synthesis of its own inhibitor (**Fig. 4a**). This approach used a DM1-affected cell as a round-bottomed flask and r(CUG)^{exp} as the catalyst for probe synthesis. **1** was appended with bioorthogonal

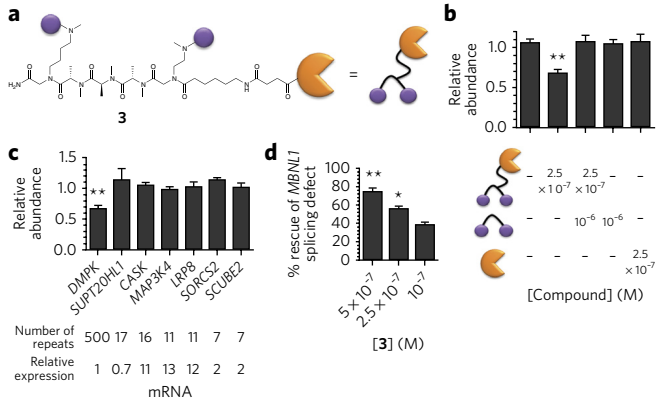

Figure 3 | Cleavage-based (Ribo-SNAP) approach to inhibit r(CUG)^{exp} dysfunction and assess cellular selectivity. (**a**) Structure of **3**; purple spheres indicate the RNA-binding modules and the orange 'Pac-Man' shape represents the bleomycin-cleaving module. (**b**) Evaluation of the cleavage-based approach in DM1-affected cells revealed that **3** cleaved r(CUG)^{exp} *DMPK* mRNA. A competitive experiment with **3** and excess **1** reduced the amount *DMPK* cleavage, indicative of competitive binding. The orange Pac-Man represents bleomycin (n = 6, 6 biological replicates each with 3 technical replicates, 2 replicate experiments). Relative abundance was determined by normalizing to *GAPDH*. (**c**) qPCR analysis of short r(CUG)-repeat-containing mRNAs revealed that **3** selectively cleaved r(CUG)^{exp} found in DM1 *DMPK* (n = 6, 6 biological replicates each with 3 technical replicates, 2 replicate experiments). (**d**) **3** rescued *MBNL1* pre-mRNA splicing defects in DM1 cells (n = 6, 6 biological replicates, 2 replicate experiments). Data represent mean values ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$, as determined by a two-tailed Student t test.

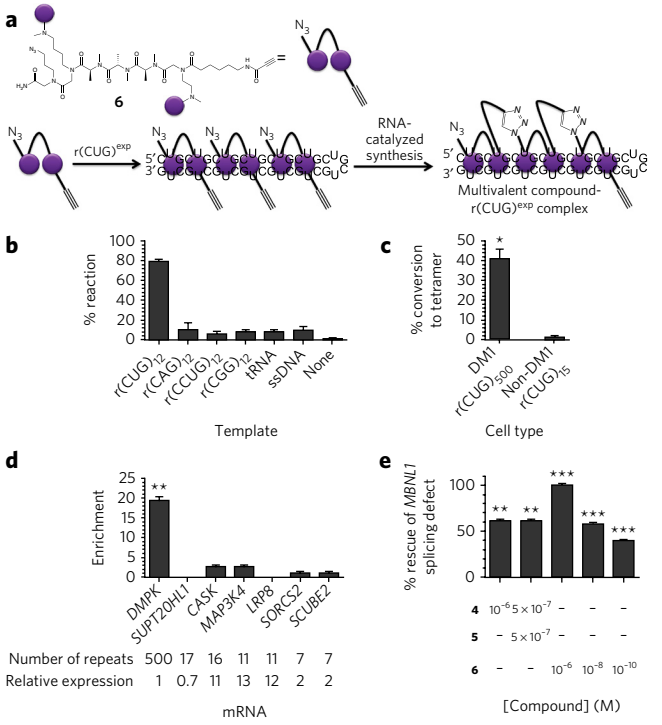

Figure 4 | r(CUG)^{exp}-dependent synthesis of multivalent RNA-binding compounds *in vitro* and in cells. (**a**) Structure of optimal click compound, **6**, and scheme of r(CUG)^{exp}-catalyzed click chemistry between azides and alkynes. (**b**) **4** and **5** only reacted in the presence of r(CUG)$_{12}$ (n = 6, 6 biological replicates, 2 replicate experiments). (**c**) Oligomerization between **4** and **7** was only observed in DM1-affected cells; thus, r(CUG)^{exp} was the catalyst (n = 6, 6 biological replicates, 2 replicate experiments). (**d**) ChemReactBIP identified cellular targets of **8**. The only mRNA significantly enriched was *DMPK* containing r(CUG)^{exp} (n = 6, 6 biological replicates each with 3 technical replicates, 2 replicate experiments). (**e**) **6** rescued the *MBNL1* pre-mRNA splicing defect at picomolar concentrations in DM1 cells (n = 9, 9 biological replicates, 3 replicate experiments). Data represent mean values ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, as determined by a two-tailed Student t test.

azide and alkyne moieties that react to form stable triazole linkages when brought into close proximity by binding adjacent sites in r(CUG)^{exp} (**Fig. 4a**). Using *in vitro* experiments, we identified the optimal distance between azide and alkyne modules incorporated into **1** (**Supplementary Fig. 10a**); 2H-K4NMeS-Aak (**4**) and N_3-2H-K4NMeS (**5**) gave the greatest yield of tetramer when reacted with a model of r(CUG)^{exp} (**Supplementary Fig. 10a** and **Supplementary Note**). This reaction appeared to be selective for r(CUG)^{exp}, as tetramer was not detected in the presence of several other nucleic acids (**Fig. 4b**). On the basis of these results, we synthesized the dual functionalized derivative N_3-2H-K4NMeS-Aak (**6**; **Fig. 4a**).

To determine whether r(CUG)^{exp}-catalyzed probe synthesis is manifest in DM1-affected cells, we treated cells with **4** and a biotinylated derivative, N_3-2H-K4NMeS-biotin (**7**) or biotin-N_3-2H-K4NMeS-Aak (**8**), to facilitate isolation of the reacted compound and compound-bound species (**Supplementary Fig. 10** and **Supplementary Note**)[25]. As hypothesized, oligomeric products were detected only in DM1-affected cells and not in non-DM1 cells (**Fig. 4c** and **Supplementary Fig. 10c,d**). Furthermore, enrichment of *DMPK* mRNA, but not other short r(CUG)-repeat-containing transcripts, was observed in the captured fraction from DM1 cells (**Fig. 4d** and **Supplementary Fig. 10e**); thus, r(CUG)^{exp} is the cellular catalyst for the transformation.

We measured and compared the potency of **6** (dual-functionalized N_3-2H-K4NMeS-Aak) for reversing DM1-associated defects to **4** and co-treatment with **4** and **5** (**Fig. 4e** and **Supplementary Figs. 11a,b** and **12a–c**). Treatment with as little as 100 pM of **6** ameliorated the *MBNL1* exon 5 pre-mRNA splicing defect in DM1 cells. Specifically, **6** has an IC_{50} of ~10 nM, corresponding to a 100-fold improvement over **4** and a 50-fold improvement over co-treatment with **4** and **5** (**Fig. 4e**). Notably, **6** had no effect on pre-mRNA splicing in non-DM1 cells (**Supplementary Fig. 11c**). Other MBNL1-dependent splicing defects[26] improved following treatment with **6**, including *NCOR2* exon 45a, *NFIX* exon 7 and *CAMK2G* exon 14 (**Supplementary Fig. 11a,b**). Furthermore, **6** improved DM1-associated pre-mRNA splicing defects in myotubes and myoblasts[7] when dosed at 100 pM, with no effect on non-DM1 myotubes and myoblasts (**Supplementary Fig. 12d**). Treatment with **6** did not affect the splicing of a pre-mRNAs that were not controlled by MBNL1 (**Supplementary Fig. 12e**).

Comparison of small molecules and oligonucleotides

We benchmarked our designer small molecules against RNA-targeting modalities based on Watson–Crick base pairing. Specifically, a morpholino oligonucleotide that improved DM1-associated defects in a mouse model following electroporation[27] was equipped with a guanidinium-rich uptake tag (Vivo-Morpholino, Gene Tools), to engender cellular uptake. This Vivo-Morpholino oligonucleotide improved MBNL1-dependent pre-mRNA splicing defects in DM1-affected cells by about 50% at ~5 μM (**Supplementary Fig. 13a–c**). Thus, **1**, **2**, **3** and **6** were ~65-, ~1,000-, ~100- and ~25,000-fold more potent than the Vivo-Morpholino oligonucleotide, respectively. The improved potencies of the small molecules were likely due to their recognition of r(CUG)^{exp} structure, rather than sequence, and to kinetic issues of oligonucleotide binding to structured RNAs[28].

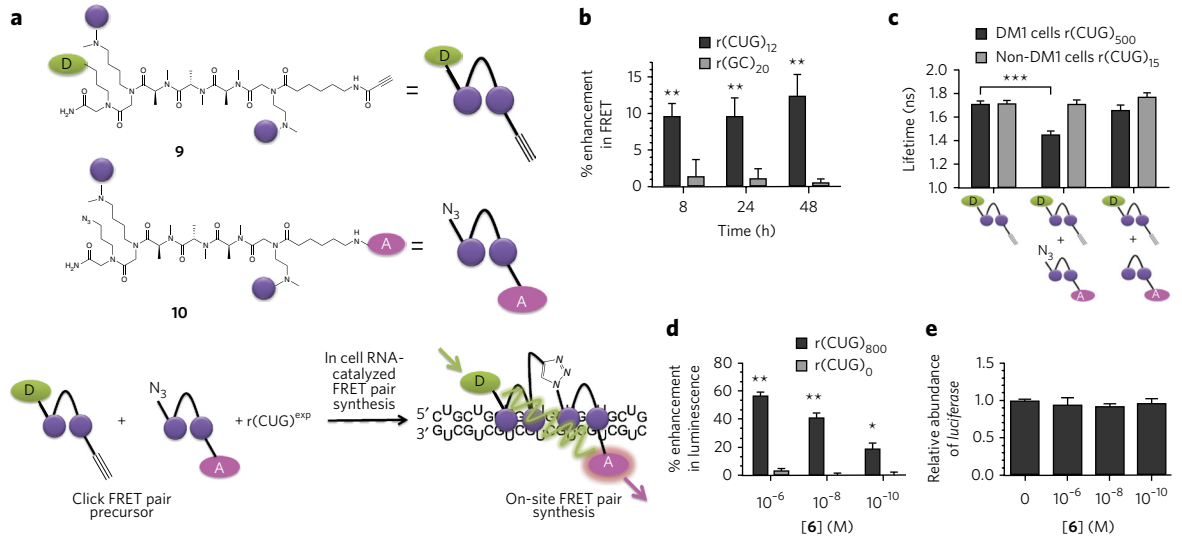

Figure 5 | FRET-based approach to image RNA targets in cells. (**a**) Structures of FRET sensors. r(CUG)^{exp} catalyzed the synthesis of a multivalent compound with FRET donors and acceptors in close proximity. (**b**) Enhanced FRET was observed when compounds were incubated with $r(CUG)_{12}$, but not with the fully base-paired control RNA $r(GC)_{20}$ (n = 10, 10 biological replicates, 2 replicate experiments). (**c**) Fluorescence lifetime measurements in DM1-affected cells revealed that FRET occurred only when both sensors could react via click chemistry (n = 30 cells, 4 biological replicates, 2 replicate experiments). (**d**) **6** rescued nucleocytoplasmic transport of a *luciferase* transcript harboring r(CUG)^{exp} in its 3′ UTR (n = 10, 10 biological replicates, 2 replicate experiments). (**e**) **6** did not change *luciferase* mRNA levels, indicating that the enhancement in luminescence is due to improvement of DM1-associated nucleocytoplasmic transport defects (n = 3, 3 biological replicates each with 3 technical replicates, 1 replicate experiment). Data represent mean values ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, as determined by a two-tailed Student t test.

Imaging of RNA by on-site fluorescent probe synthesis

We expanded our intracellular probe synthesis approach to image r(CUG)^{exp} (**Fig. 5**). We synthesized two molecules that contain fluorescence resonance energy transfer (FRET) pairs: one with an alkyne and fluorescein (FAM) (**9**) and another with an azide and 5-carboxytetramethylrhodamine (TAMRA) (**10**) (**Fig. 5a** and **Supplementary Note**). *In vitro*, these compounds produced a FRET signal following incubation with a model of r(CUG)^{exp} but did not do so in the presence of a fully base-paired control RNA (**Fig. 5b**). In addition, compounds that contained FRET pairs, but could not react, did not produce a FRET signal as a result of their stochastic orientation when bound to r(CUG)^{exp} (**Supplementary Fig. 14a**).

We tested the FRET sensors in DM1-affected cells expressing $r(CUG)_{500}$ and in non-DM1 cells. A decrease in the fluorescence lifetime of the donor, a sensitive measure of FRET[29], was observed only in DM1-affected cells (**Fig. 5c** and **Supplementary Fig. 14b**,**c**). Furthermore, DM1-affected cells co-treated with unclickable derivatives (**9** and **11**, 2H-K4NMeS-TAMRA) failed to produce a FRET signal (**Fig. 5c** and **Supplementary Fig. 14b**,**c**).

Imaging studies revealed that compound binding to r(CUG)^{exp} allowed its cytoplasmic translocation, potentially stimulating translation (**Supplementary Fig. 14b**). Consistent with our imaging findings, the addition of **6** specifically stimulated translation of *luciferase* bearing r(CUG)^{exp} repeats in the 3′ UTR without altering *luciferase* mRNA levels (**Fig. 5d**,**e**). Furthermore, using live cell imaging, we found that small-molecule FRET sensors were mostly localized in the cytoplasm (**Supplementary Fig. 14d**).

DISCUSSION

There has been much speculation that RNA is an 'undruggable' target for small molecules, because most cellular RNAs have extensive secondary structure, but only limited tertiary structure, akin to r(CUG)^{exp}. RNAs with well-defined tertiary structures are well-known targets of antibacterial agents. They bind small molecules in a manner that is similar to how proteins bind small molecules by using complex binding pockets. Collectively, we found that small molecules can selectively target RNAs having extensive secondary, but limited tertiary structure. Given that there are more than 20 diseases that are caused by RNA repeat expansions, our probes may prove to be valuable for studying these varied diseases and may provide key insights into the development of therapeutics.

One major challenge in targeting RNA with small molecules has been developing means for studying target engagement and selectivity. Chem-CLIP and Ribo-SNAP are two synergistic approaches to study target engagement and identify potential off-targets. Using these methods, we found that designer small molecules could selectively recognize larger, disease-associated repeats over shorter, non-pathogenic ones. In complementary studies, we found that small molecules equipped with a natural product that cleaves nucleic acids could selectively reduce RNA levels. If this approach is generalizable, it could increase the number of disease-associated RNAs that are amenable to therapeutic interventions beyond the disruption of protein binding.

There has been much recent activity to develop agents that can image RNAs. Although it remains to be seen, there is potential for the intracellular FRET probe synthesis strategy to complement transformative technologies such as Spinach[30,31], MS2 coat protein–RNA–GFP fusion[32] or oligonucleotide probes that image RNA[33]. The on-site click chemistry approach, as with all FRET sensors, is advantageous, as the FRET signal can be internally controlled by using the emission of the dye itself. This approach also has the potential for multiplexing; that is, multiple transcripts in a single cell could be imaged if they each contained secondary structures recognized by RNA-binding modules that produce a FRET signal when in close proximity.

Received 24 February 2016; accepted 3 October 2016;
published online 12 December 2016

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

References

1. Schlünzen, F. *et al.* Structural basis for the interaction of antibiotics with the peptidyl transferase center in eubacteria. *Nature* **413**, 814–821 (2001).
2. Carter, A.P. *et al.* Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**, 340–348 (2000).
3. Blount, K.F. *et al.* Novel riboswitch-binding flavin analog that protects mice against *Clostridium difficile* infection without inhibiting cecal flora. *Antimicrob. Agents Chemother.* **59**, 5736–5746 (2015).
4. Howe, J.A. *et al.* Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
5. Brook, J.D. *et al.* Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **68**, 799–808 (1992).
6. Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T. & Thornton, C.A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079–3088 (2004).
7. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. & Singer, R.H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* **128**, 995–1002 (1995).
8. Furling, D., Lemieux, D., Taneja, K. & Puymirat, J. Decreased levels of myotonic dystrophy protein kinase (*DMPK*) and delayed differentiation in human myotonic dystrophy myoblasts. *Neuromuscul. Disord.* **11**, 728–735 (2001).
9. Hamshere, M.G., Newman, E.E., Alwazzan, M., Athwal, B.S. & Brook, J.D. Transcriptional abnormality in myotonic dystrophy affects *DMPK* but not neighboring genes. *Proc. Natl. Acad. Sci. USA* **94**, 7394–7399 (1997).
10. Johnson, L.F., Abelson, H.T., Penman, S. & Green, H. The relative amounts of the cytoplasmic RNA species in normal, transformed and senescent cultured cell lines. *J. Cell. Physiol.* **90**, 465–470 (1977).
11. Velagapudi, S.P., Gallo, S.M. & Disney, M.D. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* **10**, 291–297 (2014).
12. Disney, M.D. *et al.* Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).
13. Pushechnikov, A. *et al.* Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.* **131**, 9767–9779 (2009).
14. Lee, M.M., Pushechnikov, A. & Disney, M.D. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem. Biol.* **4**, 345–355 (2009).
15. Childs-Disney, J.L., Hoskins, J., Rzuczek, S.G., Thornton, C.A. & Disney, M.D. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol.* **7**, 856–862 (2012).
16. Velagapudi, S.P. *et al.* Design of a small molecule against an oncogenic noncoding RNA. *Proc. Natl. Acad. Sci. USA* **113**, 5898–5903 (2016).
17. Rzuczek, S.G. *et al.* Features of modularly assembled compounds that impart bioactivity against an RNA target. *ACS Chem. Biol.* **8**, 2312–2321 (2013).
18. Childs-Disney, J.L. *et al.* Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat. Commun.* **4**, 2044 (2013).
19. Wojtkowiak-Szlachcic, A. *et al.* Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy. *Nucleic Acids Res.* **43**, 3318–3331 (2015).
20. Mastroyiannopoulos, N.P., Feldman, M.L., Uney, J.B., Mahadevan, M.S. & Phylactou, L.A. Woodchuck post-transcriptional element induces nuclear export of myotonic dystrophy 3′ untranslated region transcripts. *EMBO Rep.* **6**, 458–463 (2005).
21. Guan, L. & Disney, M.D. Covalent small-molecule-RNA complex formation enables cellular profiling of small-molecule-RNA interactions. *Angew. Chem. Int. Edn Engl.* **52**, 10010–10013 (2013).
22. Yang, W.Y., Wilson, H.D., Velagapudi, S.P. & Disney, M.D. Inhibition of non-ATG translational events in cells via covalent small molecules targeting RNA. *J. Am. Chem. Soc.* **137**, 5336–5345 (2015).
23. Carter, B.J. *et al.* Site-specific cleavage of RNA by Fe(II)-bleomycin. *Proc. Natl. Acad. Sci. USA* **87**, 9373–9377 (1990).
24. Lewis, W.G. *et al.* Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew. Chem. Int. Edn Engl.* **41**, 1053–1057 (2002).
25. Rzuczek, S.G., Park, H. & Disney, M.D. A toxic RNA catalyzes the in cellulo synthesis of its own inhibitor. *Angew. Chem. Int. Edn Engl.* **53**, 10956–10959 (2014).
26. Wang, E.T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710–724 (2012).
27. Wheeler, T.M. *et al.* Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325**, 336–339 (2009).
28. Lima, W.F., Monia, B.P., Ecker, D.J. & Freier, S.M. Implication of RNA structure on antisense oligonucleotide hybridization kinetics. *Biochemistry* **31**, 12055–12061 (1992).
29. Lakowicz, J.R. *Principles of Fluorescence Spectroscopy* (Springer, 2006).
30. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).
31. Strack, R.L., Disney, M.D. & Jaffrey, S.R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).
32. Bertrand, E. *et al.* Localization of *ASH1* mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
33. Baker, M. RNA imaging *in situ*. *Nat. Methods* **9**, 787–790 (2012).

Acknowledgments

We thank T. Kodadek, G. Joyce, W. Ja, J. Childs-Disney, K. Sobczak and J. Cleveland for advice and critical review of the manuscript, and M.D.D. acknowledges J. and H. (nee McDougall) Disney. We also thank the platform for immortalization of human cells from the Institut de Myologie. This work was funded by the US National Institutes of Health (grants DP1NS096898 to M.D.D. and DP1NS096787 to R.Y.) and the Muscular Dystrophy Association (grant 380467 to M.D.D.). S.G.R. was partially supported by a postdoctoral fellowship from the Myotonic Dystrophy Foundation.

Author contributions

M.D.D. directed the study, conceived of the ideas and designed experiments. S.G.R. designed experiments, synthesized all of the compounds and conducted all of the biochemical and cellular studies. Y.N. contributed to the synthesis of compounds. L.A.C. performed the FRET imaging. R.Y. contributed to the FRET studies. M.D.C. contributed to compound stability studies. D.F. provided critical reagents.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to M.D.D.

ONLINE METHODS

Statistical significance. Sample size was determined by first assessing variation in untreated samples. Samples were randomly assigned to treated and untreated groups and were not blinded. Statistical significance was calculated by using a two-tailed Student *t* test. Values reported are the mean ± s.e.m. Variance was similar for groups that were compared. Differences between groups were considered significant when $P < 0.05$ ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). RNA samples isolated from cells were excluded from analysis if they were deemed poor quality by the ratio of Abs_{260}/Abs_{280} (<1.8).

Antisense oligonucleotide. The AG(CAG)$_7$CA Vivo-Morpholino antisense oligonucleotide was purchased from Gene Tools and used according to the manufacturer's recommendation. The sequence of the oligo is 5′-AGCAGCA GCAGCAGCAGCAGCAGCA-3′.

Hepatic microsomal stability. Microsome stability was evaluated by incubating 1 μM compound with 1 mg/ml hepatic microsomes in 100 mM potassium phosphate buffer, pH 7.4. The reaction was initiated by adding NADPH (1 mM final concentration). Aliquots were removed at 0, 5, 10, 20, 40 and 60 min and added to acetonitrile (5×, v/v) to stop the reaction and precipitate the protein. NADPH dependence of the reaction was evaluated using no-NADPH control samples. Precipitated samples were centrifuged through a Millipore Multiscreen Solvinert 0.45-μm low binding PTFE hydrophilic filter plate and analyzed by LC-MS/MS. Data were log transformed and represented as half-life.

Affinity measurements. Affinity measurements of ligands and RNAs were performed by monitoring fluorescence intensity as a function of RNA concentration. Nucleic acids were annealed in 1× Assay Buffer (8 mM Na_2HPO_4, pH 7.0, 185 mM NaCl and 1 mM EDTA) at 60 °C for 5 min and then cooled to room temperature on the benchtop. Then BSA was added to a final concentration of 40 μg/ml. Binding assays with r(CUG)$_{12}$ and r(CUG)$_{109}$ were completed by titrating the folded RNA into 4 μM of 2H-K4NMeS (**1**) in 1× Assay Buffer containing 40 μg/ml BSA. After each addition of RNA, the samples were incubated for 5 min followed by measurement of fluorescence intensity using a BioTek FLX-800 fluorescence plate reader (excitation: 360/40; emission: 460/40; sensitivity = 90). Plots of [RNA]/[ligand] versus change in fluorescence were used to determine stoichiometries. Plots of [nucleic acid] versus change in fluorescence were used to determine binding affinity. Curves were plotted in GraphPad Prism and fit using the equation:

$$y = (B_{max} * x^h)/(K_d^h + x^h)$$

where y is the change in fluorescence; B_{max} is the extrapolated maximum change in fluorescence; x is the concentration of nucleic acid; K_d is the dissociation constant, and h is the Hill slope.

Binding assays with all other nucleic acids were completed by serial dilutions (1:2) of the nucleic acid in a 4 μM solution of **1** in 1× Assay Buffer containing 40 μg/ml BSA. The samples were incubated at room temperature for 1 h and then fluorescence intensity was measured as described above. Plots of [nucleic acid] versus change in fluorescence were used to determine binding affinity.

Small molecule inhibition of the r(CUG)$_{12}$-MBNL1 complex measured by time-resolved FRET (TR-FRET). The *in vitro* activity of small molecules was assessed by measuring the inhibition of the r(CUG)$_{12}$-MBNL1 complex using a previously reported TR-FRET assay[34]. Biotinylated r(CUG)$_{12}$ was folded at 60 °C in 1× Folding Buffer (20 mM HEPES, pH 7.5, 100 mM KCl, and 10 mM NaCl) and slowly cooled to room temperature. The buffer was then adjusted to 1× TR-FRET Buffer (1× Folding Buffer supplemented with 2 mM $MgCl_2$, 2 mM $CaCl_2$, 5 mM DTT, 0.1% BSA, 0.05% Tween-20) and the small molecule was added. The compounds were incubated with the RNA for 4 h at room temperature followed by addition of MBNL1-His$_6$. The final concentrations of RNA and MBNL1-His$_6$ were 80 nM and 60 nM, respectively. The samples were allowed to equilibrate at room temperature for 15 min, after which 1 μl of antibody solution (1:1 mixture of 8.8 ng/μl Anti-His$_6$-Tb and 800 nM streptavidin XL-665) was added. Controls for maximum TR-FRET (100% complex formation) contained 8 μL of 1× TR-FRET Buffer with RNA and protein and 1 μl water. Controls for minimum TR-FRET (no complex formation) contained 8 μl of 1× Assay Buffer, 1 μl water, and no RNA or protein. Samples were incubated at room temperature for 1 h, and then TR-FRET was measured using a Molecular Devices SpectraMax M5 plate reader using an excitation wavelength of 345-nm and a 420-nm cutoff.

To calculate the percent inhibition of complex formation, the ratio of fluorescence intensity at 545 nm and 665 nm in the presence of compound was compared to the ratio in the absence of small molecule (100% r(CUG)$_{12}$-MBNL1 complex formation) and in the absence of RNA and protein (no complex formation). The resulting curves were fit to the following equation to determine IC_{50} values:

$$y = B + \frac{A - B}{1 + \left(\frac{IC_{50}}{x}\right)^{\text{Hill slope}}}$$

where y is ratio of fluorescence intensities at 545 nm and 665 nm (F545/F665), x is the concentration of small molecule, B is F545/F665 value at max FRET effect (solution has RNA and protein but no small molecule added), A is F545/F665 value at min FRET effect (solution has antibodies but no RNA, protein, or small molecule), and the IC_{50} is the concentration of small molecule where half of the protein is displaced by small molecule.

***In vitro* ChemCLIP.** Growth medium (1X EMEM (Lonza), 10% FBS, 1× Glutagro (Corning), 1× MEM non-essential amino acids (Corning) and 1× antibiotic/antimycotic (Corning)) was inactivated by incubating at 95 °C for 15 min and then slowly cooled to room temperature. Approximately 10,000 counts of 5′-^{32}P-labeled RNA (r(CUG)$_{109}$, r(GC)$_{20}$, or tRNA) were added and folded at 95 °C for 1 min. Then, 1:2 serial dilutions (800 nM initial concentration) of 2H-K4NMeS-CA-Biotin (**2**) were prepared in 50 μl of RNA solution and incubated at 37 °C overnight. A 400-μl slurry of streptavidin-agarose beads (≥15 μg/ml binding capacity, Sigma) was washed with 1× PBS and then re-suspended in 2 ml of 1× PBS. A 30-μl aliquot of this slurry was added to each sample following overnight incubation. After 1 h incubation at room temperature, the samples were centrifuged and the supernatant containing unbound RNA was transferred to a new tube. The beads were washed with 1× PBS containing 0.1% Tween-20 (1× PBST) and centrifuged. The supernatant was added to the tube containing unbound RNA. The total radioactive counts of bound and unbound RNA were measured by scintillation counting.

***In vitro* cleavage of r(CUG)$_{10}$ by 2H-K4NMeS-bleomycin (3).** The r(CUG)$_{10}$ oligonucleotide was purchased from Dharmacon and deprotected according to the manufacturer's standard protocol. The RNA (500 pmoles) was then radiolabeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and purified using a denaturing 20% polyacrylamide gel. The RNA was excised from the gel and tumbled in 300 mM NaCl overnight. Glycogen (0.5 μl; Invitrogen) was added to the resulting solution and the RNA was precipitated with ethanol and resuspended in 50 μl of water. Then 15 μl of the RNA solution was diluted with 150 μl of 5 mM NaH_2PO_4 (pH 7.4) and heated at 95 °C for 30 s. The solution was cooled to room temperature and **3** (2.5, 1, 0.5 and 0.25 μM final concentrations) or **2a** (compound without RNA-binding modules; 2.5, 1, 0.5 and 0.25 μM final concentrations) was added followed by addition of an equimolar amount of freshly prepared $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ in 5 mM NaH_2PO_4, pH 7.4. The solutions were incubated at 37 °C and supplemented with additional equimolar aliquots of $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ in 5 mM NaH_2PO_4, pH 7.4, after 30 min and 1 h. The RNA was treated with compound for a total of 48 h at 37 °C. The reaction was stopped by adding an equal volume of loading buffer (95% formamide, 20 mM EDTA, pH 8.0) and the samples were analyzed using a denaturing 20% polyacrylamide gel run at 60 W for 4 h in 1× TBE buffer.

A T1 ladder was prepared by mixing 4 μl of RNA with 30 μL of T1 buffer (20 mM sodium citrate, 1 mM EDTA, 7 M urea) and heating to 95 °C for 30 s. After cooling to room temperature, RNase T1 (3 units/μl final concentration) was added. The sample was incubated at room temperature for 20 min and stopped by adding an equal volume of loading buffer. A hydrolysis ladder was prepared by mixing 4 μl of RNA with 30 μl of 1× alkaline hydrolysis buffer (50 mM $NaHCO_3$, pH 9.2, and 1 mM EDTA,) and heating at 95 °C for 10 min.

doi:10.1038/nchembio.2251

Gels were exposed overnight and imaged using a Molecular Dynamics Typhoon 9410 variable mode imager.

Evaluation of *in vitro* click by liquid chromatography-mass spectrometry (LC-MS). To determine whether the repeating RNA that causes DM1 serves as a template for oligomerization via Huisgen 1,3-dipolar cycloaddition reaction (HDCR), N_3-2H-K4NMeS (**5**) and each 2H-K4NMeS-activated alkyne derivative (**Supplementary Note**) were incubated with an RNA containing 12 CUG repeats, or r$(CUG)_{12}$. The RNA (100 μM final concentration) was folded in 1× Assay Buffer at 60 °C for 5 min. After cooling to room temperature, azide and alkyne dimers (25 μM final concentration of each) were added and the reaction mixtures were incubated at 37 °C for 24 h. Each sample was analyzed by LC-MS using a Thermo Scientific LTQ-ETD mass spectrometer. A gradient of 0–100% acetonitrile in water plus 0.1% formic acid over 10 min was used for analysis. The alkyne that afforded the most reaction with r$(CUG)_{12}$, 2H-K4NMeS-Aak (**5**), was next evaluated for selectivity by analyzing the amount of dimer formed in the presence of other RNA targets (100 μM) including r$(CCUG)_{12}$, r$(CAG)_{12}$, r$(CGG)_{12}$, and brewer's yeast tRNA (Roche).

***In vitro* evaluation of on-site probe synthesis by FRET.** To further investigate if the repeating RNA that causes DM1 serves as a template for oligomerization via HDCR, derivatives of N_3-2H-K4NMeS-Aak (**6**) containing a FRET donor, FAM-2H-K4NMeS-Aak (**9**), and a FRET acceptor, N_3-2H-K4NMeS-TAMRA (**10**), were incubated with r$(CUG)_{12}$. rCUG_{12} (80 nM final concentration) was folded at 60 °C for 5 min in 1× Folding Buffer (20 mM HEPES, pH 7.5, 100 mM KCl, and 10 mM NaCl) and slowly cooled to room temperature. After cooling, **9** and **10** (50 nM final concentration each) were added and the reaction mixtures were incubated at 37 °C for a total of 48 h. FRET was measured at 8, 24 and 48 h by exciting at 480 nm and monitoring emission at 590 nm using a BioTek FLX-800 fluorescence plate reader. Enhancement in FRET was calculated as the percent change in FRET of the RNA-catalyzed sample compared to non-RNA-catalyzed FRET from **9** and **10** in 1× Folding Buffer. 2H-K4NMeS-TAMRA (**11**), a non-clickable derivative, was also evaluated as described above as well as in control experiments with r$(GC)_{20}$.

Evaluation of compound stability in cell culture. The stabilities of 2H-K4NMe and **1** were evaluated by treating DM1 fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute). Cells were grown as monolayers in six-well plates in fibroblast growth medium (1× EMEM (Lonza), 10% FBS, 1× Glutagro (Corning), 1× MEM non-essential amino acids (Corning) and 1× antibiotic/antimycotic solution (Corning)) to 80% confluence. The cells were then treated with growth medium containing 5 μM final concentration of the compound of interest for 48 h. Following incubation, the growth medium was removed and a 10-μl aliquot was purified using a C_{18} ZipTip (EMD Millipore) and analyzed using an Applied Biosystems MALDI ToF/ToF Analyzer 4800 Pltus using an α-cyano-4-hydroxycinnamic acid matrix.

Evaluation of DM1-associated splicing defects in patient-derived cells. The ability of small molecule dimers to improve DM1-associated alternative splicing defects was assessed by using six different cell lines: (i) DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute); (ii) DM1-derived myoblast line (10009) that is heterozygous for *Bpm*1 polymorphism in exon 10 of the *DMPK* gene (a generous gift from K. Sobczak); Adam Mickiewicz University, Poznań, Poland); (iii) non-DM1 fibroblasts (GM07492, Coriell Institute)[18,19]; (iv) non-DM1 fibroblasts (GM07492; Coriell Institute); (v) a wild-type conditional MyoD-fibroblast cell line[35]; and (vi) a DM1 (1300 CUG repeats) conditional MyoD-fibroblast cell line[35]. Conditional MyoD-fibroblast cell lines (generous gifts from D, Furling; Centre de Recherche en Myologie (UPMC/Inserm/CNRS), Institut de Myologie, Paris, France)[35] can be converted differentiated muscle cells by adding doxycycline (2 μg/ml, Sigma).

Cells were grown as monolayers in 12-well plates in fibroblast growth medium. Once cells were ~80% confluent, they were treated with growth medium containing the compound of interest. After 48 h, or 96 h in the case of differentiated muscle cells, the cells were lysed and total RNA was harvested using a Zymo Quick RNA miniprep kit. An on-column DNA digestion was completed per the manufacturer's recommended protocol. Approximately 150 ng of total RNA was reverse transcribed at 50 °C using 100 units of SuperScript III reverse transcriptase (Life Technologies). Next, 20% of the RT reaction was subjected to PCR using GoTaq DNA polymerase (Promega). RT-PCR products were observed after 30 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 5 min. Products were separated on a 2% agarose gel run at 100 V for 1 h in 1× TBE buffer, visualized by staining with SYBR green, and imaged using a Bio-Rad Gel Doc XR+ imaging system. Percent rescue was calculated by dividing the difference between treated and untreated DM1 samples by the difference between untreated DM1 and non-DM1 samples. The resulting curves were fit to the following equation to determine IC_{50} values for splicing correction:

$$y = \min + \frac{\max - \min}{1 + \left(\frac{x}{IC_{50}}\right)^{\text{Hill slope}}}$$

where y is the value of *MBNL1* exon 5 inclusion, x is the concentration of small molecule, min is the value of *MBNL1* exon 5 inclusion in non-DM1 cells and max is the value of *MBNL1* exon 5 inclusion in DM1 patient-derived cells.

MBNL1 exon 5 alternative splicing was quantitatively assessed using RT-qPCR as previously described[36]. Approximately 150 ng of total RNA was used for RT with qScript cDNA synthesis kit (10 μL total reaction volume, Quanta BioSciences). A 2-μl aliquot of the RT reaction was used with each primer set for qPCR with SYBR Green Master Mix (Life Technologies) performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Relative amounts were determined by normalizing to *GAPDH*.

Evaluation of nuclear foci using FISH. FISH was used to determine the effects of small molecules on the formation and disruption of nuclear foci[37]. DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) were grown to ~80% confluence in a Mat-Tek 96-well glass bottom plate in growth medium. Cells were treated with the compound of interest for 48 h in growth medium followed by FISH as previously described[37] using 1 ng/μl DY547-2'OMe-$(CAG)_6$. Immunostaining of MBNL1 was completed as previously described using the MB1a antibody (diluted 1:4) (ref. 38), which was generously supplied by Prof. Glenn E. Morris (Wolfson Centre for Inherited Neuromuscular Disease), and goat anti-mouse IgG-DyLight 488 conjugate (1:2,000 dilution). Untreated controls were stained using a 1μg/μl solution of DAPI in 1× DPBS. Cells were imaged in 1× DPBS using an Olympus FluoView 1000 confocal microscope at 100× magnification.

Target identification and pull down by Chem-CLIP. Target identification of small molecule dimers was assessed using DM1 patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) and healthy fibroblasts (GM07492, Coriell Institute). Cells were grown as monolayers in 100 mm^2 dishes in fibroblast growth medium. Once cells were ~80% confluent, they were treated with growth medium containing **2** (100 nM final concentration). After 48 h, the cells were lysed and total RNA was harvested using Trizol reagent (Life Technologies). Approximately 10 μg of total RNA was incubated with streptavidin-agarose beads (100 μl, ≥15 μg/ml binding capacity, Sigma) for 1 h at room temperature. Then the beads were washed with 1× PBS, and the bound RNA was eluted by adding 100 μl of Elution Buffer (95% formamide, 10 mM EDTA, pH 8.2) for 20 min at 60 °C. Bound RNA was cleaned up using a Zymo Quick RNA miniprep kit. Approximately 100 ng of RNA was reverse transcribed using a qScript cDNA synthesis kit (10-μl total reaction volume); 1 μl of the RT reaction was used with each primer set for qPCR with SYBR Green Master Mix performed on a 7900HT Fast Real-Time PCR System. Relative abundance was determined by normalizing to *GAPDH*. Enrichment was calculated by normalizing to the least abundant mRNA.

Target identification and pull down by C-Chem-CLIP. Competitive Chem-CLIP experiments were conducted as described above. Briefly, 100-mm^3 dishes of DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) were first treated with increasing amounts of **1** (1, 10, 100, 1,000, 10,000 nM final concentrations) for 4 h. Then **2** (100 nM final

concentration) was added to the growth medium. K_i was calculated by fitting the resulting curve to the equation:

$$\log EC_{50} = \log(10 \wedge \log K_i * (1 + [2]/K_d))$$

where K_i is the inhibitory constant, K_d is the dissociation constant, and $\log EC_{50}$ was calculated using the equation:

$$Y = (\text{top} - \text{bottom})/1 + 10 \wedge (X - \log EC_{50}) + \text{bottom}$$

where top and bottom are the plateaus in units of the *y*-axis.

Target identification and pull down by Chem-CLIP-Map. The binding sites of small molecules in mutant *DMPK* was assessed using DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute). Cells were grown as monolayers in 100-mm^2 dishes in fibroblast growth medium. Once cells were ~80% confluent, they were treated with growth medium containing **2** (100 nM final concentration). After 48 h, the cells were lysed and total RNA was harvested using Trizol reagent (Life Technologies). Approximately 6 µg of total RNA was folded individually with each antisense oligonucleotide (8 µM final concentration) in 1× RNase H buffer (Life Technologies) by heating to 95 °C for 1 min and then cooling on ice. Next, 5 units of RNase H (Life Technologies) were added and the reaction was incubated at 37 °C for 1.5 h. The RNase H was then heat inactivated by incubating at 65 °C for 20 min. The cleaved RNA was treated with RQ DNase I (Promega) at 37 °C for 30 min, followed by inactivation of the enzyme by addition of an equal volume of Stop Buffer (Promega) and incubation at 65 °C for 10 min.

The solution containing the cleaved RNA was incubated with streptavidin-agarose beads (100 µl, ≥15 µg/ml binding capacity, Sigma) for 1 h at room temperature. The beads were washed with 1× PBS and the bound RNA was eluted by adding 100 µl of Elution Buffer for 20 min at 60 °C. Bound RNA was cleaned up using a Zymo Quick RNA miniprep kit. Approximately 150 ng of RNA was reverse transcribed using a qScript cDNA synthesis kit (10-µl total reaction volume); 4 µl of the RT reaction was used for each primer pair for qPCR with SYBR Green Master Mix performed on a 7900HT Fast Real-Time PCR System. Relative abundance was determined by normalizing to *GAPDH*. Enrichment was calculated by normalizing to the least abundant segment of *DMPK*. Primer sets represent the positions of the *DMPK* mRNA that were amplified after the mRNA was digested to separate the r(CUG)^{exp}-containing fragments from the rest of the mRNA.

Target identification by cleavage. The cellular targets of small molecules were assessed using DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) and healthy fibroblasts (GM07492, Coriell Institute). Cells were grown as monolayers in 60-mm^2 dishes in fibroblast growth medium. Once cells were ~80% confluent, they were treated with growth medium containing **3** (250 nM final concentration) pre-complexed with equimolar $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, with the addition of **1** (1 µM) for competitive experiments. After 48 h, the cells were lysed and total RNA was harvested using Trizol reagent (Life Technologies). Approximately 800 ng of total RNA was reverse transcribed using a qScript cDNA synthesis kit (20-µl total reaction volume, Quanta BioSciences); 2 µl of the RT reaction was used for each primer pair for qPCR with SYBR Green Master Mix performed on a 7900HT Fast Real-Time PCR System. Relative abundance was determined by normalizing to *GAPDH*.

Analysis of allele selectivity using Chem-CLIP and cleavage approaches in DM1-patient-derived cells. The allele selectivity of **2** and **3** was assessed using DM1 patient-derived fibroblasts containing a *BpmI* polymorphic site in exon 10 of *DMPK* (a generous gift from Krzysztof Sobczak)[19]. Cells were grown as monolayers in 100-mm^2 dishes in *BpmI* growth medium (HAMF10, 20% FBS, 0.39 µg/ml dexamethasone (Sigma), 10 ng/ml epidermal growth factor (Sigma), 25 µg/ml insulin (Sigma), and 1× antibiotic/antimycotic solution (Corning)). Once cells were ~80% confluent, they were treated with *BpmI* growth medium containing 100 nM **2** or 1 µM **3**. After 48 h, the cells were lysed and total RNA was harvested using Trizol reagent. For Chem-CLIP samples, approximately 10 µg of total RNA was incubated with streptavidin-agarose beads (100 µl, Sigma) for 1 h at room temperature. The beads were washed with 1× PBS and the bound RNA was eluted by adding 50 µl of Elution Buffer for 20 min at 60 °C. Bound RNA was cleaned up using a Zymo Quick RNA miniprep kit. Approximately 200 ng of RNA was reverse transcribed using 100 units of SuperScript III reverse transcriptase (Life Technologies) and gene specific primers (10-µl total reaction volume). Then 1.5 µl of RT was used for *DMPK* allele-specific qPCR using a custom TaqMan SNP Genotyping assay according to the manufacturer's recommended protocol using a 7900HT Fast Real-Time PCR System.

Sequencing analysis of the samples was also performed. Briefly, 4 µl of the RT reaction was used for PCR amplification with Phusion DNA polymerase (50-µl total reaction volume, Thermo Fischer Scientific). PCR products were analyzed after 19–22 cycles of 95 °C for 30 s, 72 °C for 1.5 min and a final extension at 72 °C for 1 min. The PCR products from samples before and after pull-downs were purified on a denaturing 15% acrylamide gel run at 200 V for 4 h in 1× TBE buffer. Products were visualized by staining with SYBR green and imaged using a Bio-Rad Gel Doc XR+ imaging system. PCR products were excised from the gel and tumbled in 300 mM NaCl overnight. After removal of residual gel, the DNA was precipitated with ethanol and 2 µl of glycogen. Then, approximately 150 ng DNA was polyA-tailed using GoTaq (Promega) according to the manufacture's protocol; 3 µl of the polyA reaction was ligated overnight with pGEM-T vector (Promega). Approximately 1 µl of the ligation reaction was used to transform DH5-α *Escherichia coli* and colonies were grown on LB ampicillin agar plates containing IPTG and X-gal to facilitate blue/white screening. Positive colonies from each sample were picked and grown overnight in Terrific Broth. Plasmids were isolated using a Zymo Zyppy plasmid miniprep kit. Samples were sequenced by Eton Bioscience using M13F(-21) primer. The sequence of the *BpmI* polymorphic site is C(G)TGGAG.

Identification of *in cellulis* clicked products by mass spectrometry, ChemReactBIP. To determine whether the repeating RNA that causes DM1 serves as a template for oligomerization in cells, DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) and healthy fibroblasts (GM07492, Coriell Institute) were grown to ~80% confluence in fibroblast growth medium in T-25 dishes and treated with equimolar amounts of **4** and **7** (500 nM each) for 48 h. The cells were washed with 1× DPBS, trypsinized, and pelleted. Cells were washed twice with 1× DPBS and then lysed by adding 0.25 ml of Lysis Buffer (2 ml of Lysis Buffer contains: 2% Triton X-100, 2% NP40, 80 µl RNAsecure (1/25) and 50 µl DNase) for 5 min at room temperature and then 5 µl DNase Stop Solution (Promega) was added.

A 200 µL aliquot of lysate was incubated with streptavidin-agarose beads (100 µl, ≥15 µg/ml binding capacity, Sigma) for 1 h at room temperature. The beads were then washed with 1× PBST and the bound RNA was eluted by adding 20 µl of Elution Buffer for 20 min at 60 °C. Approximately 4 µl of each sample was diluted in 20 µl of water plus 0.1% formic acid and analyzed by LC-MS using a Thermo Scientific LTQ-ETD mass spectrometer. A gradient of 0–100% acetonitrile in water plus 0.1% formic acid over 10 min was used for analysis. Background subtraction of untreated samples was performed before analysis.

Target identification of *in cellulis* clicked products by qPCR, ChemReactBIP. To determine if the cellular target of **6** is indeed r(CUG)^{exp}, DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) and healthy fibroblasts (GM07492, Coriell Institute) were grown to ~80% confluence in fibroblast growth medium in 100-mm^2 dishes and treated with 1 µM **8** for 48 h. The cells were washed with 1× DPBS, trypsinized, and pelleted. Cells were washed twice with 1× DPBS and then lysed by adding 500 µl of Lysis Buffer for 5 min at room temperature and then 5 µl DNase Stop Solution (Promega) was added.

A 200-µl aliquot of lysate was incubated with streptavidin-agarose beads (100 µl, ≥15 µg/ml binding capacity, Sigma) for 1 h at room temperature. The beads were then washed with 1× PBST and the bound RNA was eluted by adding 50 µl of Elution Buffer for 20 min at 60 °C. Bound RNA was cleaned up using a Zymo Quick RNA miniprep kit. Approximately 200 ng of RNA was used for RT qScript cDNA synthesis kit (10-µl total reaction volume); 2 µl of the RT reaction was used with each primer set for qPCR with SYBR Green

doi:10.1038/nchembio.2251

Master Mix performed on a 7900HT Fast Real-Time PCR System. Relative abundance was determined by normalizing to *GAPDH*. Enrichment was calculated by normalizing to the least abundant mRNA.

Evaluation of translational defects using a luciferase reporter assay. C2C12 cells that stably express 800 or 0 CUG repeats in the 3′ UTR of luciferase[15] were grown as monolayers in 96-well plates in C2C12 growth medium (1× DMEM, 10% FBS, 1× Glutagro, (Corning) and 1× antibiotic/antimycotic (Corning)). Once the cells were ~70% confluent, the compound was added in 100 µl of C2C12 growth medium. Cells were treated with compound of interest for 48 h. Cell count was normalized using WST-1 reagent (Roche) as previously described[15]. After removal of the WST-1 reagent and washing with 1× DPBS, the cells were lysed by treating with 50 µl of PPBT Lysis Buffer (100 mM K_2HPO_4, pH 7.8 and 0.2% Tween)[39] at room temperature for 10 min. Then, 50 µl of luciferase substrate was added and luminescence was measured using a Bio-Tek FLX-800 plate reader.

Evaluation of small molecule treatment on *luciferase* mRNA levels. C2C12 cells that stably express 800 CUG repeats in the 3′ UTR of *luciferase*[15] were grown as monolayers in 48-well plates in C2C12 growth medium. Once the cells were ~70% confluent, the compound was added in 100 µl of C2C12 growth medium. Cells were treated with compound of interest for 48 h after which total RNA was harvested using a Zymo Quick RNA miniprep kit. An on-column DNA digestion was completed per the manufacturer's recommended protocol. Approximately 300 ng of RNA was used for RT qScript cDNA synthesis kit (10-µl total reaction volume); 2 µl of the RT reaction was used with each primer set for qPCR with SYBR Green Master Mix performed on a 7900HT Fast Real-Time PCR System. Relative abundance was determined by normalizing to *GAPDH*.

Imaging RNA in cells by using on-site synthesis with two-photon fluorescence microscopy and two-photon fluorescence lifetime imaging (FLIM). Small molecules were used to image RNA in DM1 patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) while healthy fibroblasts with 15 CUG repeats (GM07492, Coriell Institute) were used as a control. Cells were grown as monolayers in 40 mm^2 dishes in fibroblast growth medium. Once cells were ~80% confluent, they were treated with either: 1) 1 µM of **9**; 2) 1 µM of **10**; 3) 1 µM of **9** + 1 µM of **10**; 4) 1 µM of **11**; or 5) 1 µM **9** + 1 µM **11** in growth medium for 3 d. The cells were then imaged at room temperature in a HEPES-buffered imaging solution using a custom-built two-photon microscope equipped with a fluorescence lifetime imaging system. Fluorophores were excited using a two-photon excitation at 920 nm by a Coherent chameleon pulsed laser. Laser intensity was controlled with electro-optical modulators (350–80 LA; Conoptics) before being scanned by galvano-scanning mirrors focused into the sample by an objective (60×, 0.9 numerical aperture; Olympus). Emitted fluorescence was divided with a dichroic mirror (565 nm; Chroma) and detected with cooled photomultiplier tubes (H7422–40 for green and R3896 for red; Hamamatsu) after wavelength filters (HQ510/70–2p for green and HQ620/90–2p for red; Chroma Technology). Fluorescence intensity images were acquired by ScanImage using a data acquisition board (PCI-6110; National Instruments) and fluorescence lifetime images of **9** were acquired using time-correlated single photon counting (SPC-150; Becker and Hickl). To determine cellular localization, cells were stained with DAPI after FLIM imaging.

Imaging of compound localization in live DM1 cells using confocal microscopy. DM1-patient-derived fibroblasts containing 500 CUG repeats (GM03987, Coriell Institute) were grown to ~80% confluence in a Mat-Tek 96-well glass bottom plate in fibroblast growth medium. Cells were treated with the compound of interest for 48 h in fibroblast growth medium. They were then washed and imaged in 1× DPBS using an Olympus FluoView 1000 confocal microscope at 40× magnification.

34. Chen, C.Z. *et al.* Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **402**, 1889–1898 (2012).
35. François, V. *et al.* Selective silencing of mutated mRNAs in DM1 by using modified hU7-snRNAs. *Nat. Struct. Mol. Biol.* **18**, 85–87 (2011).
36. Shapiro, I.M. *et al.* An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.* **7**, e1002218 (2011).
37. Warf, M.B., Nakamori, M., Matthys, C.M., Thornton, C.A. & Berglund, J.A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. USA* **106**, 18551–18556 (2009).
38. Holt, I. *et al.* Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **12**, 1035–1048 (2007).
39. Hampf, M. & Gossen, M. A protocol for combined Photinus and Renilla luciferase quantification compatible with protein assays. *Anal. Biochem.* **356**, 94–99 (2006).