C9orf72 and triplet repeat disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms

XUEYIN WANG,1,2,4 KAREN J. GOODRICH,1,2,6 ERIN G. CONLON,3,5,6 JIANCHAO GAO,3 ANNETTE H. ERBSE,1 JAMES L. MANLEY,3 and THOMAS R. CECH1,2

1Department of Biochemistry, University of Colorado Boulder, Boulder, Colorado 80309, USA
2BioFrontiers Institute and Howard Hughes Medical Institute, University of Colorado Boulder, Boulder, Colorado 80309, USA
3Department of Biological Sciences, Columbia University, New York, New York 10027, USA

ABSTRACT

Some neurological disorders, including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), fragile X syndrome, Huntington’s disease, myotonic dystrophy, and various ataxias, can be caused by expansions of short nucleic acid sequence repeats in specific genes. A possible disease mechanism involves the transcribed repeat RNA binding an RNA-binding protein (RBP), resulting in its sequestration and thus dysfunction. Polycomb repressive complex 2 (PRC2), the histone methyltransferase that deposits the H3K27me3 mark of epigenetically silenced chromatin, binds G-rich RNAs and has especially high affinity for G-quadruplex (G-Q) structures. Here, we find that PRC2 target genes are derepressed and the RNA binding subunit EZH2 largely insoluble in postmortem brain samples from ALS/FTD patients with C9ORF72 (C9) repeat expansions, leading to the hypothesis that the (G4C2)n repeat RNA might be sequestering PRC2. Contrary to this expectation, we found that C9 repeat RNAs (n = 6 or 10) bind weakly to purified PRC2, and studies with the G-Q specific BG4 antibody and circular dichroism studies both indicated that these C9 RNAs have little propensity to form G-Qs in vitro. Several GC-rich triplet-repeat expansion RNAs also have low affinity for PRC2 and do not appreciably form G-Qs in vitro. The results are consistent with these sequences forming hairpin structures that outcompete G-Q folding when the repeat length is sufficiently large. We suggest that binding of PRC2 to these GC-rich RNAs is fundamentally weak but may be modulated in vivo by protein factors that affect secondary structure, such as helicases and other RBPs.

Keywords: ALS; G-quadruplex; epigenetics; neurological disorders; RNA–protein interactions

INTRODUCTION

The expansion of short sequence repeats in chromosomal DNA is known to cause a variety of neurological disorders including spinocerebellar ataxia, fragile X syndrome (FXS), Friedreich’s ataxia, Huntington’s disease, and myotonic dystrophy (Brook et al. 1992; Caskey et al. 1992; Orr and Zoghbi 2007). More recently, expansion of the hexanucleotide repeat (GGGGCC)n in the C9ORF72 (C9) gene has been revealed to be the most common known cause of the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (DeJesus-Hernandez et al. 2011; Renton et al. 2011).

Considering these repeat expansion diseases together, it has variously been proposed that the mechanism of pathology is at the DNA level, the RNA level, or the protein level (e.g., polyglutamine/repeat associated non-AUG translation), and indeed these possibilities are not mutually exclusive (Conlon and Manley 2017). The work presented here relates to the possible structures and functions of the RNA repeats, as they have important implications for many of the proposed pathomechanisms.

Repeated RNA sequences have the potential to sequester RNA-binding proteins (RBPs). For example, in myotonic dystrophy type 1 (DM1), a CUG trinucleotide expansion in the 3′-untranslated region of the mRNA for DMPK (dystrophia myotonica protein kinase) (Brook et al. 1992) forms
hairpin structures (Tian et al. 2000) and sequesters the RBP MBNL1 (muscle-blind-like 1). The reduction in available MBNL1 dysregulates alternative splicing and mRNA translation and stability (Miller et al. 2000; Mankodi et al. 2002; Jiang et al. 2004). Analogously, C9 RNA repeats sequester another RBP, hnRNP H, to disrupt splicing in ALS/FTD patient brains (Conlon et al. 2016, 2018). RNA repeats can also have important interactions with DNA and histone modifying complexes; for example, FXS RNA repeats have been suggested to affect transcription of the FMR1 gene in cis, in a manner that involves R-loop formation (Colak et al. 2014). The ability of CGG RNA hairpins in nascent transcripts to unfold was found to be essential for G-rich RNA to hybridize with the template and direct epigenetic silencing of the locus (Colak et al. 2014). It is also thought that histone methyltransferases can be recruited locus-specifically through binding to nascent RNA (Davidovich and Cech 2015), so the potential for complexes like PRC2 to become sequestered by expanded RNA repeats represents an interesting possibility.

The FXS and C9 RNAs have repeated blocks of G residues, and therefore the potential to form intramolecular G-quadruplex (G-Q) structures. Indeed, we have provided evidence that a portion of C9 repeat RNA can form G-Qs under some circumstances (Conlon et al. 2016). However, both sequences could also fold into intramolecular hairpins (Fig. 1). It is worthwhile to consider the complexity of possible structures, as the folded status of these RNAs may modify protein interactions and predispose non-AUG translation (Wang et al. 2019).

PRC2, the H3K27me3 histone methyltransferase, is important in embryonic development and can contribute to oncogenesis (Margueron and Reinberg 2011; Comet et al. 2016). PRC2 binds G-Q RNAs with particularly high affinity, whereas intramolecular hairpins are weak targets for PRC2 binding (Wang et al. 2017). Several patches of amino acids on the surface of EZH2, the catalytic subunit of PRC2, are major sites of RNA binding (Long et al. 2017). Reduced levels of PRC2 have been implicated in activation of genes that lead to neurodegeneration in mice (von Schimmelmann et al. 2016), and have been suggested to play a role in several neurodegenerative diseases (von Schimmelmann et al. 2016, and references therein). Intriguingly, RNA-sequencing analyses of differentially expressed genes in C9 ALS/FTD have revealed up-regulation of many genes known to be PRC2 targets (Prudencio et al. 2015; Finch et al. 2017). This is consistent with the possibility of reduced PRC2 activity, perhaps due to sequestration by C9 RNA.

Here we find that EZH2 is relatively insoluble in C9 ALS/FTD patient cortical samples, concomitant with derepression of PRC2 target genes, notably a number of Homeobox (HOX) genes, that are epigenetically silenced in normal brains. These data suggested that reduced

**FIGURE 1.** Proposed RNA secondary structures of expansion sequences found in sequence repeat disorders and of a mutated form tested herein. Internal loops are circled. Structures and stabilities predicted by MFOLD. Several of the sequences can fold into alternative G-Q structures (not shown). Nonrepeat nucleotides were added to 5’ ends to facilitate 5’-end labeling with polynucleotide kinase. (A) Two predicted secondary structures of C9 RNA. ΔG° = −22.4 (left) and −16.4 (right) kcal/mol. (B) “Hairpin no-G-Q RNA,” a mutant version of C9 RNA designed to fold into a hairpin structure similar to panel A (left) but with no potential to form alternative G-Q structures. ΔG° = −18.7 kcal/mol. (C) Fragile X RNA. ΔG° = −22.5 kcal/mol. (D) Huntington’s disease RNA. ΔG° = −10.4 kcal/mol. (E) Myotonic dystrophy RNA. ΔG° = −13.0 kcal/mol.
activity of PRC2 could be due to its sequestration by C9 RNA. However, our in vitro binding studies show low affinity of PRC2 for C9 RNA, and little support for C9 RNA forming G-Q structures in isolation. Thus, other factors would be needed to activate C9 RNA for PRC2 binding in vivo. Similarly, we find that several G-rich triplet repeat expansion RNAs have low affinity for PRC2, consistent with their folding into disrupted hairpins rather than G-Q structures.

RESULTS

C9 ALS/FTD cerebellum samples show epigenetic derepression and PRC2 insolubility

In previous studies, we demonstrated that biochemical insolubility of RBPs is directly correlated to changes in gene expression among C9ALS/FTD patient brains (Conlon et al. 2016, 2018). While that work focused on the RBP hnRNP H and alternative splicing, changes in transcript abundance have also been reported (Prudencio et al. 2015; Finch et al. 2017). Intriguingly, gene ontology analysis of differentially expressed transcripts in C9s revealed highly significant enrichment of genes related to development and morphogenesis, including many encoding HOX domain transcription factors (Prudencio et al. 2015). Given that these genes are canonical targets of histone modifying complexes, we hypothesized that the core catalytic and RNA-binding subunits of PRC2—EZH1 and EZH2—might be dysregulated analogously to hnRNP H.

To explore this hypothesis, we performed gene expression and insolubility assays using a cohort of C9+ ALS/FTD-afflicted individuals and age-matched controls. As previously described (Conlon et al. 2016, 2018), we performed $^{32}$P-RT-PCR on RNA that was purified from cerebellum and oligo(dT) selected. To test for elevated expression of PRC2 targets on a patient-by-patient basis, we selected 18 genes, primarily transcription factors, which are established targets of PRC2-mediated repression (Bracken et al. 2006; Margueron and Reinberg 2011). We also tested the EZH1 and EZH2 transcripts themselves, as well as GAPDH for normalization. PCR products were electrophoresed side-by-side by native-PAGE (Fig. 2A). These 21 primer sets gave amplification products of the expected size, and we quantified changes in transcript abundance by normalizing abundance to GAPDH. Out of 20 targets, ten displayed a significant difference between C9 and control ($P < 0.05$), with a range of 2.7–9.3-fold enrichment in C9s versus controls (Fig. 2B). Only one of these ten transcripts, EZH2, displayed lower expression in C9s relative to controls (0.6-fold relative to controls, $P < 0.05$). The transcripts HOXA11, HOXC11, HOXA10, HAND1, and NKX6-1 displayed greater than twofold enrichment in C9s versus controls, but the differences were not statistically significant. We thus confirm that a number of PRC2-target transcripts are indeed highly overexpressed in our cohort of C9 patient brains.

We next tested whether these gene expression changes are reflective of sequestration of EZH1 and/or EZH2 proteins in insoluble aggregates. To assay this, we measured the proportion of sarkosyl-insoluble EZH1 and EZH2 in the motor cortices of control and C9ALS/FTD patients, using 21,000g centrifugation as previously described (Conlon et al. 2018). We performed western blotting of the soluble, sarkosyl-soluble and sarkosyl-insoluble fractions with EZH1 and EZH2 antibodies, as well as histone H3 as a general nuclear control (Fig. 3A). For comparison, we also quantified insolubility of two ALS/FTD-relevant...
RBPs, hnRNP H and TDP-43, that we have previously established (Conlon et al. 2018) to be correlated to each other in their differential insolubility in C9s (Fig. 3A). We found that in nine C9ALS/FTD versus eight controls (including two SOD1ALS), hnRNP H and TDP-43 proteins were 1.5 and 1.9 times, respectively, more insoluble in C9s than in controls (Fig. 3B). While the total proportion of insoluble EZH2 was much less than hnRNP H or TDP-43 (36% versus 72% or 60%, respectively), the differential in insoluble EZH2 in C9s versus controls was the largest of these three proteins, with EZH2 being 2.6 times as insoluble in C9s on average (Fig. 3B). In contrast, no significant portion of EZH1 was detected in the insoluble fraction, nor was histone H3, thus ruling out the possibility of nonspecific nuclear insolubility being a feature of C9 motor cortex samples (Fig. 3A,B). These data reveal that EZH2, but not EZH1, exhibits behavior suggestive of sequestration into insoluble aggregates, providing a possible mechanistic explanation for the increases in expression of the other genes that we describe above.

C9orf72 RNA binds only weakly to PRC2 in both KCl and LiCl

The above results raised the possibility that PRC2 may bind directly to C9 RNA, leading to its sequestration, as we previously demonstrated for hnRNP H (Conlon et al. 2016). Indeed, the C9 RNA repeat has four consecutive guanines, and G-Q formation has been reported (Fratta et al. 2012; Reddy et al. 2013). Furthermore, Haeusler et al. (2014) reported that (GGGGCC)₄ RNA forms a G-Q that binds certain proteins, contributing to ALS pathology, and our previous study also provided support for the existence of G-Qs in C9 ALS patient brains (Conlon et al. 2016). Considering that PRC2 has such a high preference for binding G-Q RNAs, we hypothesized that C9 RNA would have high affinity for PRC2, especially in KCl where G-Qs are stabilized. To test this, we performed gel shift assays (EMSA) with purified PRC2 and a synthetic 40-mer C9 RNA (containing six repeats; Supplemental Fig. 1). Contrary to our expectation, however, binding of PRC2 to this RNA was about 100-fold weaker than binding to G-rich RNAs that can form G-Qs, and it was similarly weak in KCl and LiCl (Table 1; Fig. 4).

The G-Q specific antibody BG4 recognizes C9 repeat RNAs only weakly

Because the weak binding of C9 RNA to PRC2 seemed inconsistent with G-Q formation, we conducted additional tests of the C9 RNA structure. The BG4 antibody has been previously used to identify G-Q RNA binding (Biffi et al. 2014). We therefore purified BG4 antibody (Supplemental Fig. 2A) and validated its specificity for G-Q RNA binding by EMSA (Supplemental Fig. 2B,C). BG4 clearly bound with higher affinity to (GGAA)₁₀ than to (GA)₂₀; these two RNAs have the same base composition.

### TABLE 1. Repeat-expansion disease RNAs and their affinities for binding PRC2

| RNA                  | Sequence | Kᵅᵣᵖᵖᵖ (nM) |
|----------------------|----------|--------------|
| Fragile X syndrome   | (CGG)₁₂  | 1400 ± 320   |
| Fragile X mutant A   | (AGG)₁₂  | 7.4 ± 3.3    |
| Fragile X mutant U   | (JUG)₁₂  | 2.4 ± 0.7    |
| Myotonic dystrophy 1 | (CUG)₁₂  | 3300 ± 800   |
| Huntington’s          | (CAG)₁₂  | 2800 ± 1100  |
| Friedreich’s ataxia   | (GAA)₁₂  | 1200 ± 160   |
| ALS (C9orf72)         | (G₆C₄)₁₆ | 670 ± 220    |

All binding in 100 mM KCl, mean ± SD, n = 3.
but only the former can form G-Qs (Wang et al. 2017). Furthermore, high-affinity binding to (GGAA)_{10} occurred in KCl but not in LiCl. Another positive control for G-Q structure, TERRA_{40}, bound BG4 in KCl with affinity indistinguishable from that of (GGAA)_{10} (Fig. 5). (TERRA is a telomere repeat RNA [Azzalin et al. 2007] known to form intramolecular G-Qs [Collie et al. 2010; Biffi et al. 2014; Wang et al. 2017]). As a negative control, poly(A) bound to BG4 only at very high concentrations, consistent with nonspecific binding (Fig. 5). Thus, EMSA provides a robust tool for monitoring BG4-G-Q RNA binding. Importantly, when C9 RNA was tested with BG4, we observed $K_d^{app} \geq 3$ µM in KCl, consistent with nonspecific binding (Fig. 5).

Given that C9 RNA could in principle form either a hairpin structure or a G-Q structure, we tested whether different annealing protocols might succeed in trapping a G-Q structure. Twelve annealing protocols were compared. As detailed in Supplemental Figure 3C, none of these protocols gave rise to high-affinity PRC2 binding, so we were unable to find conditions that favored folding of the C9 RNA into G-Qs.

**Circular dichroism spectroscopy confirms folding of C9 RNA into A-form structures**

At this stage, several of our lines of experimentation argued against purified C9 repeat-containing RNA having much G-Q character: C9 RNA bound weakly both to PRC2 and to the BG4 antibody, and its binding was very similar in KCl and LiCl. Nevertheless, we performed an orthogonal test of RNA structure using CD spectroscopy. CD reveals three clear distinguishing features of G-Q RNA: (i) a positive peak at 210 nm, instead of the negative peak observed for A-form RNA; (ii) a major change in the spectrum in KCl versus LiCl, whereas A-form RNA CD spectra are independent of the cation identity; and (iii) stabilization of the G-Q structure from thermal denaturation specifically by KCl, whereas A-form RNA is stabilized similarly by KCl and LiCl. On the other hand, the positions of the positive peak at 260–270 nm and the negative peak at ∼240 nm are not very diagnostic, because RNA G-Qs show only subtle differences from A-form RNA at these wavelengths (Kypr et al. 2009).

CD spectra of the C9 40-mer RNA in three different ionic conditions are shown in Figure 6A. Both the shape of the spectra (note the negative peak at 210 nm) and the small difference in the spectra in KCl versus LiCl indicate little G-Q character. Thermal denaturation analysis showed some stabilization by KCl (Fig. 6B,C). TERRA_{24} RNA (UUAAGG)_{4} was our positive control for a G-Q structure, including a larger dependence of the $T_m$ on the cation (K+ vs. Li+) (Fig. 6D–F). As a further control, we designed a hairpin with internal purine–purine mismatches that recapitulated a hairpin form of C9 RNA but could not form an alternative G-Q structure (Fig. 1B). This “Hairpin no G-Q RNA” had a CD spectrum and absence of cation...
preference very similar to those of C9 RNA (Fig. 6G–I), further cementing the conclusion that C9 40-mer RNA predominantly folds into a hairpin in vitro.

Given the previous work showing G-Q formation by C9 four-repeat RNAs (Fratta et al. 2012; Reddy et al. 2013; Haeusler et al. 2014), we determined the CD spectrum of this shorter 24-mer C9 RNA. In KCl it had a smaller negative peak at 210 nm (Supplemental Fig. 4A) than the C9 40-mer (Supplemental Fig. 4C), consistent with the 24-mer comprising a mixture of structures that is more weighted toward G-Q structures than the 40-mer. This result suggests that the C9 repeats do in fact have some ability to form G-Qs in vitro, perhaps in a manner dependent on RNA length, and we offer an explanation in the Discussion. Notably, unlike the TERRA24 RNA and the Hairpin no-G-Q RNA which gave completely reproducible CD spectra, the C9 24-mer and 40-mer in KCl showed some variability among replicate CD runs done on different days (Supplemental Fig. 4), consistent with different fractions of the molecules forming G-Q structures under the high RNA concentrations required for CD.

G-Q structures can also be identified by UV-monitored melting, because they have a small but distinctive absorbance at 295–297 nm that decreases upon denaturation (Mergny et al. 1998; Mullen et al. 2010). The TERRA24 RNA showed clear 297 nm absorbance with a K⁺ ion-dependent melting temperature (Supplemental Fig. 5A), but neither the C9 24-mer or 40-mer showed such a signal (Supplemental Fig. 5B,C), further supporting the conclusion that the G-Q subpopulation is low in these C9 RNAs.

When we measured the binding of C9 24-mer to PRC2 using the nanomolar RNA concentrations required for Kd measurements, we saw only weak binding with no evidence of a fraction of the RNA binding tightly (Supplemental Fig. 6); thus, no evidence for a substantial G-Q subpopulation.

### Triplet-repeat disease RNAs bind only weakly to PRC2

We next extended our analysis to examine PRC2 binding to several other disease-associated repeat RNAs with the potential to form G-Qs. The FXS CGG-repeat RNA could potentially form either G-Qs (as the DNA version clearly does [Kettani et al. 1995]) or interrupted hairpin structures as shown in Figure 1C. Zumwalt et al. (2007) performed NMR spectroscopy on RNAs containing six repeats of CGG and found clear evidence of intramolecular hairpins with C-G and G-C base pairs, and no evidence of G-Qs. However, it seems likely that FXS repeat sequences can adopt different structures depending on the cation and on the number of repeats (Malgowska et al. 2014). If the 40-mer FXS RNA formed hairpin structures, it should be a poor substrate for PRC2 binding.

We first measured PRC2 binding to CGG-containing RNAs using gel shift assays. Purified PRC2 5-subunit complex bound (CGG)12 with Kd app = 1400 nM in 100 mM KCl (Table 1; Fig. 7A). This affinity is two orders of magnitude weaker than other 40-mer G-rich RNAs that form G-Q structures, and it is weaker than nonquadruplex G-rich RNAs such as (GA)20 and (GU)20 (Wang et al. 2017). The weak binding of (CGG)12 is consistent with it forming an intramolecular hairpin, as hairpins bind PRC2 weakly (Wang et al. 2017). As a positive control, we mutated all the C’s in (CGG)12 to either A’s or U’s [(AGG)12 and (UGG)12,
These mutants have the same 67% guanine content as \((\text{CGG})_{12}\), but they should not be able to form intramolecular hairpins. PRC2 bound these two mutant RNAs with high affinity (Fig. 7B,C), consistent with our model.

RNA 40-mers comprising triplet repeats associated with myotonic dystrophy 1, Huntington’s disease, and Friedreich’s ataxia (Fig. 1C–E) similarly bound PRC2 with very weak affinity (Table 1; Fig. 7D–F). The first two of these RNAs could form hairpins similar to the FXS RNA, but with U-U or A-A mismatches instead of G-G mismatches (Tian et al. 2000). The Friedreich’s ataxia GAA repeats cannot form any intermolecular Watson–Crick base pairs, so the weak binding is presumably dictated by the fact that PRC2 has exceedingly low affinity for adenine (Wang et al. 2017).

DISCUSSION

This study was stimulated by the possibility that the repeat-sequence expansion RNAs found in a number of genetic diseases might sequester the PRC2 histone methyltransferase complex; the resulting dysregulation of epigenetic gene silencing could then contribute to pathology. Our observations that brains of patients carrying the C9 repeat expansion show derepression of epigenetically silenced genes and insolubility of PRC2 are consistent with formation of C9 RNA-PRC2 RNP assemblies or granules in vivo, although of course there could be other explanations for excess PRC2 insolubility in patient brains. For example, we recently documented RBP insolubility in the brains of sporadic ALS/FTD patients, which lack any known mutations, including expansions. It will be interesting to determine if PRC2 also displays increased insolubility in such samples. The possibility of PRC2 binding directly to the repeat-expansion RNAs was particularly provocative in the case of FXS and C9 RNAs, which have repeated blocks of two or four guanines, respectively, and thus are candidates for G-Q structure formation. G-Qs are exceedingly high-affinity ligands for PRC2 in vitro (Wang et al. 2017). Alternatively, if these RNAs formed hairpin structures they
should be very weak binders (Wang et al. 2017). Thus, we expected the PRC2 binding data to clearly distinguish these RNA structure models, and this expectation was met: All repeat-sequence RNAs tested bound very weakly to PRC2, consistent with the hairpin structure models. In the case of the FXS RNA, its failure to form G-Qs was anticipated (Zumwalt et al. 2007), so weak PRC2 binding was unsurprising. The surprise came with the C9 RNA, which had been shown to have G-Q character (Fratta et al. 2012; Reddy et al. 2013; Haeusler et al. 2014). Importantly, we were able to reproduce the CD spectra and cation specificity reported by these authors for the C9 24-mer RNA. However, we found that the 40-mer and 60-mer RNAs (and presumably longer RNAs) fold differently from the 24-mer, and we provide a simple model to explain this difference, as follows. The energetic stability of a hairpin is the sum of (i) the negative free energy of base-pair formation and base stacking in the stem and (ii) the positive free energy of hairpin loop formation (Tinoco et al. 1973). The first (stabilizing) contribution will scale more-or-less linearly with the number of C9 sequence repeats, while the second (destabilizing) contribution is a constant—one hairpin loop per molecule. For long sequence repeats, the energetic cost of hairpin loop formation is easily compensated by the stability of the stem. But for shorter hairpins, there will be a tipping point at which the folded form becomes destabilized. If there is an alternative G-Q structure, that structure may then predominate in solution. Under our solution conditions, that tipping point appears to be around four repeats or 24 nt (Supplemental Fig. 4). Thus, our 40-mer and 60-mer RNAs exist in the realm where hairpins predominate over G-Qs and PRC2 binding is minimal. The hexanucleotide repeat expansions found in the ALS/FTD disease state are much larger (typically > 250 repeats; Renton et al. 2011), and it is plausible that these RNAs fold heterogeneously; major portions of the RNA could fold into hairpins (PRC2-unbound), while other regions could remain single-stranded or form G-Q structures (PRC2-bound).

Single-stranded DNA C9 repeats have been reported to form G-Q structures (Grigg et al. 2014; Šket et al. 2015; Zhou et al. 2015). These studies mostly used four-repeat oligonucleotides, so it is possible that a key message of our RNA work—that hairpin structures may outcompete G-Q structures for larger numbers of C9 sequence repeats—might also pertain to DNA. Indeed, Šket et al. (2015) found that longer DNA repeats formed antiparallel G-Q structures as well as “unresolved structures with some Watson–Crick base-pairing.” Of course, with double-stranded DNA the complementary C-rich strand can compete with structure formation in the C9-repeat strand, so the energetic balance must also consider structure formation in the C-rich strand (Zamiri et al. 2015).

Although our biochemical studies show that the various repeat-expansion RNAs are not high-affinity PRC2 binders, a number of factors could overcome this situation in vivo. For example, an RNA hairpin-binding protein with

![FIGURE 7. Triplet repeat RNAs bind weakly to PRC2 and mutating (CGG)n to disrupt its potential to form secondary structure enhances binding. (A–F) Binding of triplet repeat RNAs to PRC2 in protocol #1 binding conditions with 100 mM KCl (see Materials and Methods) was analyzed by EMSA. (G) Binding curves based on three independent experiments (mean ± SD).]
affinity for PRC2 could bridge between the two macromolecules. PRC2 has many accessory proteins, and not all of these have been tested for RNA binding. As another possibility, helicases might unzip RNA hairpins in vivo, allowing transient formation of single-stranded G-tracts or G-Q structures that might be trapped by PRC2 binding. Indeed, the physiological activity of hairpin-specific helicases appears to be consistent with previous studies showing that hairpin melting is a prerequisite for R-loop formation at the expanded \textit{FMR1} locus (Colak et al. 2014). A further obstacle to hairpin formation may also arise from the cotranscriptional association of RBPs, such as hnRNP H. While evidence suggests that hnRNP H, or the highly similar hnRNP F, has a preference for linear G-tract RNA (Dominguez et al. 2010; Samatanga et al. 2013; Conlon et al. 2016), it has also been shown that hnRNP H has greater affinity for G-tract-containing RNAs that are capable of folding into intramolecular G-Qs versus those that are able to form hairpins (Herdy et al. 2018). Thus, depending on the occupancy of RBPs like hnRNP H, there may be unbound regions of RNA that can fold into hairpins or G-Qs, with perhaps only the latter stabilized by adjacent hnRNP H molecules. In any case, if PRC2 is found to be bound to a repeat-expansion RNA in vivo, we suggest that its association requires something beyond the weak binding affinity that we have observed with purified molecules.

\section*{MATERIALS AND METHODS}

\subsection*{Gene expression analysis}

Human patient samples used in this study are identical to those identified previously (Conlon et al. 2018). RNA was extracted, oligo(dT) selected, subject to 32P-RT-PCR, electrophoresed, imaged and quantified as described previously (Conlon et al. 2016, 2018). Plotting and statistical analysis were performed using Graphpad for Prism. PCR primers used in this study are as follows:

| Gene | Sequence 5′–3′ |
|------|---------------|
| GAPDH | CTGTCACCCAGGGCTGCTTTT |
|       | GGGGTCATTGAGGCGAAAT |
|       | GTGTAGGGAGCCATGAGG |
|       | CGTGTCAGGTAGGGACCT |
| HOXA6 | GCAAGGTGAGGAGGGAGC |
| HOXA9 | GGGCTCTCGAGTAAGGTAC |
| HOXA10 | CCAAGATGGGAGGCCAAAT |
| HOXA11 | TCCGGACAGCGATCCTG |
| HOXA13 | GGGAAGGCCAAAGCTCG |
| HOXC13 | AGACGCCGTCCATCAGA |
| HOXD1 | GCCCTACTCGTATAGCCG |
| HOXD2 | GCCGTCTCGAGGAGGAGC |
| HOXD3 | GCTGAAACGCTCAACGGAG |
| HOXD4 | GCAGGTGAGGAGGGAGC |
| HOXD5 | AGAAGGAGGAGCGAAGC |
| HOXD6 | GCAGGTGAGGAGGGAGC |
| HOXD7 | AGAAGGAGGAGCGAAGC |
| HOXD8 | GCAGGTGAGGAGGGAGC |
| HOXD9 | AGAAGGAGGAGCGAAGC |
| HOXD10 | GCAGGTGAGGAGGGAGC |
| HOXD11 | AGAAGGAGGAGCGAAGC |
| HOXD12 | GCAGGTGAGGAGGGAGC |
| HOXD13 | AGAAGGAGGAGCGAAGC |

\subsection*{Biochemical fractionation}

Biochemical fractionation of human patient motor cortex, including simultaneous western blotting, exposure, and quantification, was performed as previously described (Conlon et al. 2016) using a 21,000 g centrifugation to separate sarkosyl soluble from insoluble proteins (Conlon et al. 2018). Antibodies for hnRNP H, TDP-43 and Histone H3 are previously published (Conlon et al. 2018). For EZH2 we used antibody 196-A from Bethyl Laboratories, and for EZH1 we used NB100-1138 from Novus. Plotting and statistical analysis were performed using Graphpad for Prism.

\subsection*{C9 and triplet repeat expansion RNAs}

RNA 40-mers were used for many of our studies because they are small enough for chemical synthesis and large enough to provide multiple protein-binding sites and to give measurable affinity to PRC2, which has a length dependence for RNA binding (Davidovich et al. 2013). All RNAs except (GGGGCC)10 (see next section) were synthesized by GE-Dharmacon, radiolabeled using polyethylene oxide kinase and PAGE purified. C9 40-mer was synthesized in two versions, with either AAAA or UUAA at the 5′ end preceding (GGGGCC)₅; these 5′ nucleotides were added
to facilitate kinase labeling. The counts of purified RNAs were determined by liquid scintillation counting; specific activity was no less than 100,000 cpm/pmol.

**In vitro transcription of (GGGGCC)10**

Oligonucleotide 1: `5′-TAA TAC GAC TCA CTA TAG GGG CCG GGG CCG GGG CCG GGG CCG GGG CCG GGG CCG GGG CCG GGG CCG GGG CCG GGG CC-3′`

Oligonucleotide 2: `5′-GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC GGC CCC`  

The above oligonucleotides were annealed and added to a transcription reaction so that the final concentration was 0.1 μg/mL. Final concentrations of other components were 1× transcription buffer (Thermo Fisher), 10 mM DTT, 0.3 mM each ATP, GTP, and UTP, 0.5 pmol/ml alpha-32P CTP (3000Ci/mmol), 0.02 mM CTP, T7 RNA polymerase (Thermo Fisher). After 2 h incubation at 37°C, additional T7 polymerase was added and incubation continued for another 2 h. RO1 DNase buffer was then added to 1× final concentration along with RO1 DNase (Promega) and incubated at 37°C for 20 min. The RNA was ethanol precipitated and resuspended in TE. This was mixed 1:1 with formamide loading dye (93% formamide, 30 mM EDTA, 0.1× TBE, 0.5% bromophenol blue, 0.5% xylene cyanol), heated to 95°C for 5 min and loaded on a 8% acrylamide 7 M urea gel buffered with 1× TBE. After electrophoresis, bands were cut out of the gel and the RNA was extracted with 0.3 M sodium acetate at room temperature for 1 h. Samples were centrifuged briefly, liquid was removed, transferred to a Spin-X column (Corning), centrifuged, and ethanol precipitated. This procedure was then repeated. After ethanol precipitation, the pellet was resuspended in H2O.

**PRC2 purification**

Recombinant human PRC2 5-mer complex (EZH2-SUZ12-EEBD-RBBP4-AEB2 long splice-form) was expressed in insect cells and purified by column chromatography as previously described (Davidovich et al. 2013; Wang et al. 2017). These protein preparations are free of nucleic acids, contain stoichiometric amounts of the five subunits, and are active in histone methyltransferase assays.

**BG4 antibody expression and purification**

BG4-encoding plasmid (pSANG10-3F-BG4) was purchased from Addgene (#55756) and verified by sequencing. The plasmid was transformed into *E. coli* strain BL21-CodonPlus(DE3)-RP (Stratagene), followed by culturing in 2× YT media (1.6% tryptone peptone, 1% yeast extract, and 0.5% NaCl) and 50 μg/mL kanamycin at 250 rpm and 37°C. When the OD600 reached 0.8–1.0, induction was initiated with 0.5 mM isopropyl β-D-thiogalactopyranoside. After 3 h of induction at 37°C, bacterial culture was harvested by spinning at 4000g for 30 min. Samples were taken before and after induction for SDS-PAGE analysis. Pellets were resuspended in 8 mL TES buffer 1 (50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 20% sucrose) per 100 mL volume of expression culture. Twelve milliliters of TES buffer 2 (10 mM Tris-HCl pH 8.0, 0.2 mM EDTA pH 8.0, 4% sucrose) per 100 mL volume of expression culture were added and the mixture was put on ice for 10 min. Cells were centrifuged at 4°C for 10 min at 8000g. Supernatant was collected and filtered via a conventional disposable filter device (0.22 μm). Nickel affinity/anti-His tag resin (Qiagen, 30230) was washed and equilibrated with PBS. Two milliliters of resin per liter of culture was added to filtered supernatant, followed by rotation for 1 h at room temperature. Resin was then washed with 50 CV of wash buffer (PBS, 10 mM imidazole pH 8.0). BG4 protein was eluted with elution buffer (PBS, 250 mM imidazole pH 8.0), buffer-exchanged to PBS, and concentrated using 50 mL Amicon centrifugal devices with 10 kDa cut-off (Millipore, UFC901024). Protein quality was assessed by running SDS-PAGE (4%~12% NuPAGE gel, Invtrogen, NP0322BOX) and gels were visualized with Coomassie staining (Expedeon, ISB11). The concentrated protein was subject to fractionation over Superose 6 increase column (GE, 29091596). Peak protein fractions were confirmed using SDS-PAGE, pooled and concentrated as described above. Final protein concentration was calculated based on absorbance measurements at 280 nm, with an absorbance ratio at 260/280 nm <0.7 indicating no nucleic acid contamination.

**Electrophoretic mobility shift assays (EMSA)**

Two different methods were used for annealing and folding RNAs. In protocol #1 (Wang et al. 2017), a 1.7 mL low-binding Eppendorf tube (OmniSeal) containing 60 μL radiolabeled RNA (1000 cpm/gel lane) in TE (Tris-EDTA), pH 7.5 was heated for 10 min at 95°C and snap cooled on ice for at least 2 min. Binding buffer (4×) was then added to give final concentrations of 50 mM Tris-HCl, pH 7.5 at 25°C, 100 mM KCl, 5 mM MgCl2, 0.5 mM ZnCl2, 0.1 mM CaCl2, 2 mM 2-mercaptoethanol, 0.1 mg/mL bovine serum albumin, 0.1 mg/mL fragmented yeast tRNA (Sigma cat # R5636), 5% v/v glycerol. The RNA was allowed to fold at 37°C for 30 min. RNA was aliquoted into tubes and protein that had been diluted to various concentrations with binding buffer was added and incubated at 30°C for 30 min. Each sample was then loaded onto a 10 × 15 cm, 1× TBE/0.7% SeaKem GTG agarose (Lonzza) gel and electrophoresis preceded for 90 min at 66 V. Next the gel was placed on Hybrid N+ membrane (GE Healthcare) and Whatman 3 mm chromatography paper (GE Healthcare) and dried for 1 h at 80°C. The dried gels were exposed to phosphorimager plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare). For binding of the RNA and protein in LiCl, 100 mM LiCl was substituted for the 100 mM KCl in the binding buffer.

In protocol #2, the RNA was heated in ddH2O for 5 min at 85°C, then vortexed and briefly centrifuged. An equal volume of 300 mM KCl or LiCl was then added (final concentration of 150 mM) and the tube was placed on ice for 10 min. RNA (1 μL) was added to tubes containing 9 μL of protein diluted in binding buffer (50 mM Tris pH 8.0, 50 mM NaCl, 0.2% CHAPS, 2 mM EDTA, 0.5% BSA, 5% glycerol, 0.1 mg/mL fragmented yeast tRNA (Sigma cat # R5636) prepared on ice. After the RNA was added to the tube, it was placed at RT until all samples were complete. Samples were mixed and incubated at 37°C for 30 min. After the incubation the tubes were placed in an aluminum block on
ice. As above, the samples were then loaded onto a 10 × 15 cm, 1× TBE/0.8% GTG agarose gel and run for 90 min at 66 V. The gel was then placed on Hybond N+ membrane and two pieces of Whatman 3 mm chromatography paper and dried for 1 h at 80°C. Dried gels were exposed to phosphorimaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare) using 100-micron pixel size.

**Calculation of equilibrium binding constants**

Each phosphorimager scan was analyzed in ImageQuant TL (GE Healthcare). Boxes were drawn around each gel band representing Free and Bound RNA, using the rolling ball method for background subtraction. Any signal between the Free and Bound RNA bands was included in the Bound fraction, based on the assumption that any RNA that dissociated from protein during electrophoresis had been bound after incubation. Fraction Bound was calculated as Bound/(Bound + Free). Some RNA samples ran as a discrete band with a lower mobility smear in the zero-protein lane; the fraction of RNA in the smear was then subtracted from the Bound signal in all other lanes and the Bound signal was renormalized. KaleidaGraph was used to plot the binding curves and determine $K_{d,app}$; the binding constants are “apparent” because they depend on the ionic conditions and the amount and type of RNA competitor present in the reaction.

**Additional RNA annealing methods**

The IVT RNA was subjected to different annealing protocols. It was heated in either TE or TE/100 mM KCl at 95°C for 10 min. Then it was either put on ice for ≥2 min, kept at room temperature or cooled slowly to room temperature. These samples then had either protocol #1 buffer added so the final concentration was 50 mM Tris-HCl pH 7.5 at 25°C, 100 mM KCl, 5 mM MgCl$_2$, 0.5 mM ZnCl$_2$, 0.1 mM CaCl$_2$, 2 mM 2-mercaptoethanol, 0.1 mg/mL bovine serum albumin, 0.1 mg/mL fragmented yeast tRNA (Sigma cat # R5636), 5% v/v glycerol or protocol #2 buffer added so the final concentration was 50 mM Tris-HCl pH 7.5 at 25°C, 100 mM KCl, 5 mM MgCl$_2$, 0.5% CHAPS, 2 mM EDTA, 0.5% BSA, 5% glycerol, 0.1 mg/mL fragmented yeast tRNA (Sigma cat # R5636). This was then added to protein, incubated at 37°C for 20 min. After the incubation, the tubes were placed in an aluminum block on ice. As above, the sample was then loaded onto a 10 × 15 cm, 1× TBE/0.8% GTG agarose gel. The gel was run for 90 min at 66 V. When finished running the gel was placed on Hybond N+ membrane and Whatman 3 mm chromatography paper and dried down for 1 h at 80°C. The dried gels were exposed to phosphorimaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare).

**Circular dichroism spectroscopy**

RNA was dissolved in double-distilled water and adjusted to 10 mM Tris (pH 7.5) with or without 100 mM KCl or LiCl as indicated. The 170 µL sample (~0.2 mg/mL, 20 µM RNA) was heated to 95°C for 5 min and allowed to slow cool to room temperature. CD spectra were collected with a Chirascan-Plus spectropolarimeter (Applied Photophysics) equipped with an AP/CD250 Peltier-controlled variable temperature cell holder from Quantum Northwest using a 0.5 mm path length cell. Spectra were recorded from 360 to 200 nm with a 0.5 nm step size and 0.5 sec integration time. All spectra were corrected for background and smoothed using a ten-point Savitzky-Golay filter. Spectra are reported in mean residual molar ellipticity. Thermal unfolding was monitored by ramping the temperature of the cell holder in steps of 10°C from 15°C to 95°C with a 5 min equilibration time after each ramp. Full spectra were recorded at the end of each step. The actual temperature of the sample was recorded for each step using an in-sample temperature probe from Applied Photolysis.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available for this article.

**COMPETING INTEREST STATEMENT**

T.R.C. is on the Board of Directors of Merck, Inc. and a consultant for Storm Therapeutics.

**ACKNOWLEDGMENTS**

We thank members of the Cech and Manley laboratories for help and discussions, Dr. Samie Jaffrey (Weill Cornell Medicine) for stimulating discussions, and Dr. Phil Bevilacqua and Allison Williams (Penn State University) for careful review of the manuscript. We thank Dr. Neil Shneider (Columbia University Medical Center) and the New York Brain Bank for providing us with clinical samples. This work was supported by National Institutes of Health grant R35 GM118136 to J.L.M. T.R.C. is an investigator of the Howard Hughes Medical Institute.

Received March 12, 2019; accepted April 29, 2019.

**REFERENCES**

Azzalin CM, Reichenbach P, Khorouli L, Giulotto E, Lingner J. 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318: 789–801. doi:10.1126/science.1147182

Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. 2007. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem 6: 75–80. doi:10.1038/nchem.1805

Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20: 1123–1136. doi:10.1101/gad.381706

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thrion JP, Hudson T, et al. 1992. 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. doi:10.1016/0092-8674(92)90154-5

Caskey CT, Pizzuti A, Fu YH, Fenwick RG Jr, Nelson DL. 1992. Triplet repeat mutations in human disease. Science 256: 784–789. doi:10.1126/science.1589758
Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR. 2014. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science 343: 1002–1005. doi:10.1126/science.1245831

Collie GW, Haider SM, Neidle S, Parkinson GN. 2010. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. Nucleic Acids Res 38: 5569–5580. doi:10.1093/nar/gkq259

Comet I, Rissing EM, Leblanc B, Helin K. 2016. Maintaining cell identity: PR2C-mediated regulation of transcription and cancer. Nat Rev Cancer 16: 803–810. doi:10.1038/nrc.2016.83

Conlon EG, Manley JL. 2017. RNA-binding proteins in neurodegeneration: mechanisms in aggregates. Genes Dev 31: 1509–1528. doi:10.1101/gad.304055.117

Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA. 2010. Structural basis of alternative splicing in neurons. J Biol Chem 285: 1250–1257. doi:10.1074/jbc.M109.079739

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, DiMauro S, Walter S, Busch MP, Brown RH, Jr, et al. 2010. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9orf72 causes chromosome 9p-linked FTD and ALS. Neuron 72: 245–256. doi:10.1016/j.neuron.2011.09.011

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, et al. 2011. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72: 245–256. doi:10.1016/j.neuron.2011.09.011

Dominguez C, Fisette JF, Chabot B, Allain FH. 2010. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. Nat Struct Mol Biol 17: 853–861. doi:10.1038/nsmb.1814

Finch NA, Wang X, Baker MC, Heckman MG, Gendron TF, Bieniek KF, Wu J, DeJesus-Hernandez M, Brown PH, Chew J, et al. 2017. Abnormal expression of homeobox genes and transthyretin in C9ORF72 expansion carriers. Neurol Genet 3: e161. doi:10.1212/NXG.0000000000000161

Fratto P, Miziołek S, Niccoli AJ, Zloh M, Fisher EM, Parkinson G, Isaacs AM. 2012. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep 2: 1016. doi:10.1038/srep01016

Grigg JC, Shumayrijk N, Sen D. 2014. G-quadruplex structures formed by expanded hexanucleotide repeat RNA and DNA from the neurodegenerative disease-linked C9orf72 gene efficiently sequester and activate heme. PLoS One 9: e106449. doi:10.1371/journal.pone.0106449

Haeuvels AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, et al. 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 507: 195–200. doi:10.1038/nature13124

Herdy B, Mayer C, Varshney D, Marsico G, Murat P, Taylor C, D’Santos C, Tannahill D, Balasubramanian S. 2018. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. Nucleic Acids Res 46: 11592–11604. doi:10.1093/nar/gky861

Jiang H, Mankodi A, Swanson MS, Mixley RT, Thornton CA. 2004. 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. doi:10.1093/hmg/ddh327

Kettani A, Kumar RA, Patel DJ. 1995. Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. J Mol Biol 254: 638–656. doi:10.1016/j.jmb.1995.0644

Kyrp J, Kojasevskova I, Rencuk D, Vorlickova M. 2009. Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res 37: 1713–1725. doi:10.1093/nar/gkp026

Long Y, Bolanos B, Gong L, Liu W, Goodrich KJ, Yang X, Chen S, Gooding AR, Maegley KA, Gajwala KS, et al. 2017. Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. eLife 6: e31558. doi:10.7554/eLife.31558

Małgowska M, Gudanis D, Kierzek R, Wyszko E, Gabelica V, Gadanick Z. 2014. Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. Nucleic Acids Res 42: 10196–10207. doi:10.1093/nar/gku710

Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Mixley RT, Cannon SC, Thornton CA. 2002. Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell 10: 35–44. doi:10.1016/S1097-2765(02)00563-4

Margueron R, Reinberg D. 2011. The polycomb complex PRC2 and its mark in life. Nature 469: 343–349. doi:10.1038/nature09784

Mergy NL, Phan AT, Lacroix L. 1998. Following G-quartet formation by UV-spectroscopy. FEBS Lett 435: 74–78. doi:10.1016/S0014-5793(98)01043-6

Miller JW, Urbanić CR, Teng-Umuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. 2000. Recruitment of human muscleblind proteins to (CUG), expansions associated with myotonic dystrophy. EMBO J 19: 4439–4448. doi:10.1093/embj/19.17.4439

Mullen MA, Olson KJ, Dallaire P, Major F, Assmann SM, Bevilacqua PC. 2010. RNA G-quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. Nucleic Acids Res 38: 8149–8163. doi:10.1093/nar/gkq804

Ort HT, Zoghbi HY. 2007. Trinucleotide repeat disorders. Annu Rev Neurosci 30: 575–621. doi:10.1146/annurev.neuro.29.051605.113042

Prudencio M, Belizl VV, Batra R, Ross CA, Gendron TF, Pregent LJ, Murray ME, Overstreet KK, Piazza-Johnston AE, Desaro P, et al. 2015. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat Neurosci 18: 1175–1182. doi:10.1038/nn.4065

Reddy K, Zamiri B, Stanley SY, Macgregor RB Jr, Pearson CE. 2013. The disease-associated r(GGGGCC), repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem 288: 9860–9866. doi:10.1074/jbc.C113.452532

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollison S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllärykas L, et al. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72: 257–268. doi:10.1016/j.neuron.2011.09.010

Samata B, Dominguez C, Jelesová I, Allain FH. 2013. The high kinetic stability of a G-quadruplex limits the hnRNP F qRMM3 binding to G-tract RNA. Nucleic Acids Res 41: 2505–2516. doi:10.1093/nar/gks1289

Šhek P, Pohleven J, Kovanda M, Župunski V, Zalar M, Plavec J, Rogelj B. 2015. Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Neurobiol Aging 36: 1091–1096. doi:10.1016/j.neurobiolaging.2014.09.012

Tian B, White RJ, Xia T, Welle S, Turner DH, Mathews MB, Thornton CA. 2000. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. RNA 6: 79–87. doi:10.1015/s1097-0185(00)91544
Tinoco I Jr, Borer PN, Dengler B, Levin MD, Uhlenbeck OC, Crothers DM, Gralla J. 1973. Improved estimation of secondary structure in ribonucleic acids. Nat New Biol 246: 40–41. doi:10.1038/newbio246040a0
von Schimmelmann M, Feinberg PA, Sullivan JM, Ku SM, Badimon A, Duff MK, Wang Z, Lachmann A, Dewell S, Ma’ayan A, et al. 2016. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. Nat Neurosci 19: 1321–1330. doi:10.1038/nn.4360
Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C. 2017. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. Mol Cell 65: 1056–1067 e1055. doi:10.1016/j.molcel.2017.02.003
Wang ZF, Ursu A, Childs-Disney JL, Guertler R, Yang WY, Bernat V, Rzuczek SG, Fuerst R, Zhang YJ, Gendron TF, et al. 2019. The hairpin form of r(G4C2)exp in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules. Cell Chem Biol 26: 179–190. doi:10.1016/j.chembiol.2018.10.018
Zamin B, Mirceta M, Bomsztyk K, Macgregor RB Jr, Pearson CE. 2015. Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)8*(GGCCCC)8 repeat: effect of CpG methylation. Nucleic Acids Res 43: 10055–10064. doi:10.1093/nar/gkv1008
Zhou B, Liu C, Geng Y, Zhu G. 2015. Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD. Sci Rep 5: 16673. doi:10.1038/srep16673
Zumwalt M, Ludwig A, Hagerman PJ, Dieckmann T. 2007. Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. RNA Biol 4: 93–100. doi:10.4161/ma.4.2.5039