

RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF

Huilin Huang,^{1,5} Jing Zhang,^{1,2,3,4,5} Samuel E. Harvey,^{1,2,3,4,5} Xiaohui Hu,^{1,2,3,4} and Chonghui Cheng^{1,2,3,4}

¹Division of Hematology and Oncology, Department of Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA; ²Lester and Sue Smith Breast Center, ³Department of Molecular and Human Genetics, ⁴Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

It is generally thought that splicing factors regulate alternative splicing through binding to RNA consensus sequences. In addition to these linear motifs, RNA secondary structure is emerging as an important layer in splicing regulation. Here we demonstrate that RNA elements with G-quadruplex-forming capacity promote exon inclusion. Destroying G-quadruplex-forming capacity while keeping G tracts intact abrogates exon inclusion. Analysis of RNA-binding protein footprints revealed that G quadruplexes are enriched in heterogeneous nuclear ribonucleoprotein F (hnRNPF)-binding sites and near hnRNPF-regulated alternatively spliced exons in the human transcriptome. Moreover, hnRNPF regulates an epithelial–mesenchymal transition (EMT)-associated CD44 isoform switch in a G-quadruplex-dependent manner, which results in inhibition of EMT. Mining breast cancer TCGA (The Cancer Genome Atlas) data sets, we demonstrate that hnRNPF negatively correlates with an EMT gene signature and positively correlates with patient survival. These data suggest a critical role for RNA G quadruplexes in regulating alternative splicing. Modulation of G-quadruplex structural integrity may control cellular processes important for tumor progression.

[*Keywords:* RNA G quadruplex; alternative splicing; hnRNPF; splicing factor; EMT]

Supplemental material is available for this article.

Received August 8, 2017; revised version accepted November 22, 2017.

Nearly all human genes are estimated to undergo alternative splicing, providing the complexity and diverse functions of human proteomes (Graveley 2001; Johnson et al. 2003; Pan et al. 2008; Sultan et al. 2008; Wang et al. 2008; Jangi and Sharp 2014). This conserved post-transcriptional process must be precisely regulated, and accumulating evidence has indicated that aberrant alternative splicing causes human diseases, including cancer (Liu and Cheng 2013; Zhang and Manley 2013; Xu et al. 2014).

The code for regulating alternative splicing resides in the pre-mRNA. The majority of the code identified thus far is comprised of short nucleotide motifs located near exonic splice sites (Fu and Ares 2014; Rosenberg et al. 2015). These consensus sequences recruit RNA-binding proteins to splice sites to promote or inhibit spliceosome assembly. In contrast to these linear consensus motifs, there is a void in our understanding of how RNA secondary structures contribute to the splicing code.

One type of RNA secondary structure is the RNA G quadruplex. RNA G quadruplexes are assembled through the interactions between four guanines that are organized in a cyclic Hoogsteen hydrogen-bonding arrangement (Kumari et al. 2007; Agarwal et al. 2012; Bugaut and Balasubramanian 2012; Millevoi et al. 2012). G quadruplexes functioning as *cis* elements to regulate splicing have been reported in a growing number of genes, such as fragile X mental retardation 1 (*FMR1*) and the tumor suppressor *TP53* (Marcel et al. 2011; Blice-Baum and Mihailescu 2014). Recent profiling of RNA G quadruplexes has revealed widespread and evolutionarily conserved G-quadruplex structures in the human transcriptome (Kwok et al. 2016). Given the prevalence of RNA G-quadruplex structures in the transcriptome, it is essential to define the functions of these G quadruplexes in normal and disease states as well as the factors that recognize them in order to execute splicing regulation. Several RNA-binding

⁵These authors contributed equally to this work.

Corresponding authors: chonghui.cheng@bcm.edu, jing.zhang3@bcm.edu
Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.305862.117>.

© 2017 Huang et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

proteins have been shown to interact with RNA G quadruplexes that impact biological functions (Garneau et al. 2005; Khateb et al. 2007; Lattmann et al. 2010; von Hacht et al. 2014; Zhang et al. 2015; Conlon et al. 2016; Liu et al. 2017). However, the relationship between splicing factor recognition of G quadruplexes and regulation of alternative splicing remains largely unclear.

In exploring the effect of alternative splicing in cellular functions, we and others showed that alternative splicing of the cell adhesion molecule CD44 is dynamically regulated during epithelial–mesenchymal transition (EMT), a developmental program that is abnormally activated during tumor metastasis (Thiery and Sleeman 2006; Yang and Weinberg 2008; Brown et al. 2011; Xu et al. 2014; Hernandez et al. 2015). Switching of splice isoforms from CD44 variants (CD44v) to CD44 standard (CD44s) is necessary for cells to transit from an epithelial cellular state to a mesenchymal phenotype. A *cis*-acting RNA element designated “I-8” is located in the intron immediately downstream from the CD44 variable exon 8 (v8) and was identified to modulate the splicing of CD44 (Xu et al. 2014).

In this study, we found that the I-8 element contains an RNA G-quadruplex structure. The RNA sequence with the capacity of forming the secondary structure of G quadruplex itself, rather than the linear G tract, promotes alternative splicing and the production of the epithelial-specific CD44v isoform. We further demonstrate that heterogeneous nuclear ribonucleoprotein F (hnRNPF) regulates G-quadruplex-associated alternative splicing across the transcriptome. Depletion of hnRNPF stimulates EMT-associated CD44 splice isoform switching and accelerates the EMT phenotype. These results connect RNA G quadruplexes to EMT, a critical process that drives tumor metastasis.

Results

Formation of a G quadruplex within the CD44-splicing *cis* element

The CD44 I-8 *cis* element recruits splicing factors to modulate alternative splicing of CD44, producing splice isoforms that regulate the switch between epithelial and mesenchymal states (Xu et al. 2014). On close examination, we noticed the presence of four G tracts in the I-8 sequence that exhibit high levels of conservation across seven mammalian species (Fig. 1A). To determine whether these G tracts form a G-quadruplex structure (Fig. 1B), we performed circular dichroism (CD) spectroscopy using an I-8-containing 28-nucleotide (nt) RNA fragment. The CD spectrum showed a peak at 264 nm and a trough at 240 nm that was further enhanced in a potassium ion (K⁺) concentration-dependent manner (Fig. 1C). K⁺, compared with other monovalent cations Na⁺ or Li⁺, showed the most enhanced CD spectra (Fig. 1D), a characteristic feature of G quadruplexes (Fahlman and Sen 1998; Davis 2004; Novy et al. 2008). The wavelengths and the enhancement of CD spectrum peaks in response to K⁺ are in agreement with a typical CD spectrum for a parallel RNA

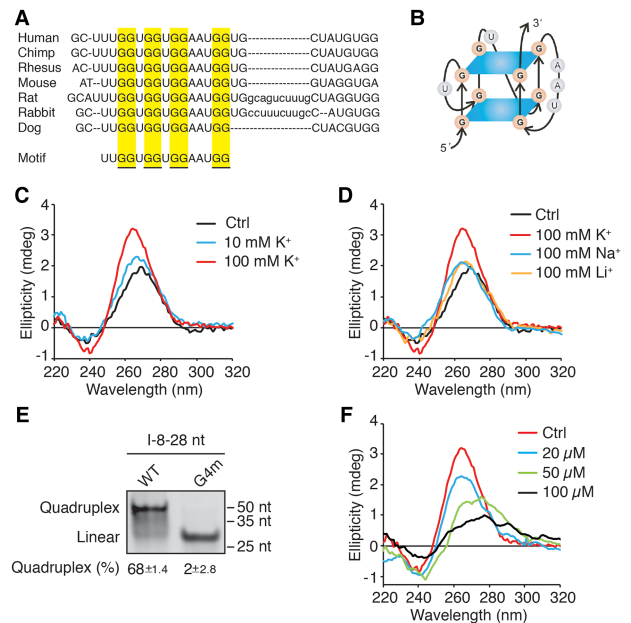


Figure 1. Formation of the I-8 G-quadruplex structure. (A) Conservation of the I-8 G-quadruplex motif across mammalian species. The conserved four G tracts are highlighted in yellow. (B) Proposed parallel G-quadruplex secondary structure of I-8. (C) CD spectrum analysis of I-8 RNA oligonucleotides in the presence of different concentrations of K⁺. (D) CD spectrum analysis of I-8 RNA oligonucleotides in the presence of different monovalent cations. (E) Native gel images showing locations of I-8 wild-type (WT) and G4m mutant RNA oligonucleotides prepared in 100 mM K⁺. (F) CD spectrum showing TMPyP4 disrupting the G-quadruplex structure of I-8 in a dose-dependent manner. RNA oligos were incubated in the presence of 100 mM K⁺ and different amounts of TMPyP4.

G-quadruplex structure (Vorlickova et al. 2012), as shown in CD spectra of two previously published G-quadruplex structures, NRQ and CRQ (Supplemental Fig. S1A,B; Kumari et al. 2007; Weng et al. 2012). Substitution of the four GG to GA of I-8 to destroy the G quadruplex, a mutation termed I-8 G4m, abolished the K⁺-dependent increase in the CD spectrum peak (Supplemental Fig. S1C). Moreover, native gel electrophoresis showed that the 28-nt I-8 RNA migrated at a size of 50 nt as a higher-order RNA structure, more slowly than its predicted linear form, whereas its G4m mutated form migrated at the expected size of 28 nt (Fig. 1E). These results indicate that the I-8 RNA has the capacity to form a G-quadruplex structure.

TMPyP4 is a G-quadruplex ligand that binds to both DNA and RNA. When binding to DNA, TMPyP4 stabilizes G-quadruplex structures (Han et al. 1999; Siddiqui-Jain et al. 2002; Martino et al. 2009). When binding to RNA, however, TMPyP4 disrupts these secondary structures (Ofer et al. 2009; Morris et al. 2012; Zamiri et al. 2014). We found that addition of TMPyP4 perturbed the characteristic CD spectrum of the I-8 RNA in a dose-dependent manner even in the presence of 100 mM K⁺ (Fig. 1F) and shifted the slowly migrating higher-order RNA structure to the linear form (Supplemental Fig. S1D). These results

further support the notion that the sequence-conserved I-8-containing RNA fragment adopts an RNA G-quadruplex structure.

The G-quadruplex secondary structure modulates alternative splicing

We next explored whether formation of the G-quadruplex structure within I-8 affects alternative splicing. We introduced a minimum fragment of I-8 that contained the 15-nt G-quadruplex sequence in a bichromatic fluorescent splicing construct, RG6 (Orengo et al. 2006), where the fluorescent proteins EGFP and dsRED are produced exclusively from distinct splice isoforms, with EGFP as a readout for cassette exon inclusion and dsRED as a readout for exon skipping (Fig. 2A).

Human embryonic kidney (HEK) 293FT cells transfected with the control minigene showed fluorescence signals of both EGFP and dsRED in cells, with a stronger signal in dsRED (Fig. 2B). RT-PCR analysis showed the percent spliced in (PSI) index as 29%, corroborating the fluorescence signal (Fig. 2C). Insertion of the I-8 15-nt fragment into the RG6 intron led to a shift in color from skipping-associated red to inclusion-associated green in cells, indicating that this 15-nt G-quadruplex sequence shifted isoform production to more exon inclusion (Fig. 2B). Both semiquantitative PCR and quantitative real-time PCR confirmed this result, showing a PSI increase from 29% to 66% (Fig. 2C) and an increase in the ratio of inclusion to skipping by sevenfold (Fig. 2D).

We next disrupted the G-quadruplex structure in the I-8 15-nt fragment by substituting the last guanine nucleotide with an adenosine but keeping the G-tract sequence intact by replacing the nearest adenosine with a guanine to maintain the same number of guanine residues (Fig. 2A). This G-quadruplex mutant, termed I-8 15-nt G4m1,

partially abolished the I-8 15-nt-mediated exon inclusion, resulting in a drop in exon inclusion (Fig. 2B–D). Furthermore, disrupting the G quadruplex by substituting all four GG repeats to GA sequences (I-8 15-nt G4m) (Fig. 2A) caused a more pronounced effect, showing a threefold drop of the inclusion to skipping ratio (Fig. 2B–D).

To further examine the dependence of the G quadruplex in promoting exon inclusion, we treated cells transfected with the I-8 or I-8 G4m mutant splicing minigenes with TMPyP4. Treatment with TMPyP4 decreased the I-8-dependent exon inclusion but had no effect on the G4m splicing minigene (Supplemental Fig. S2A–C). Examination of endogenous CD44 splice isoform expression in human mammary epithelial cell lines HMLE and MCF10A also showed that TMPyP4 treatment resulted in an increased production of the exon-skipping isoform CD44s (Fig. 2E,F). To exclude the possible role of nonsense-mediated decay (NMD) causing the observed changes in endogenous splice isoform ratios, we treated cells with cyclohexamide (CHX) to block NMD and observed the same effect with TMPyP4 treatment (Supplemental Fig. S2D). Thus, these results show that the G quadruplex residing in I-8 modulates alternative RNA splicing to promote exon inclusion.

Bona fide G-quadruplex structures promote exon inclusion

We next sought to determine whether the G-quadruplex structure itself, as opposed to its particular linear sequence, promotes exon inclusion. We examined four additional G-quadruplex sequences, including two G-quadruplex sequences containing GU repeats (G2U1 and G3U1) and two previously identified G-quadruplex sequences (NRQ [Kumari et al. 2007] and CRQ [Weng et al. 2012]) that do not contain GU repeats (Supplemental Fig.

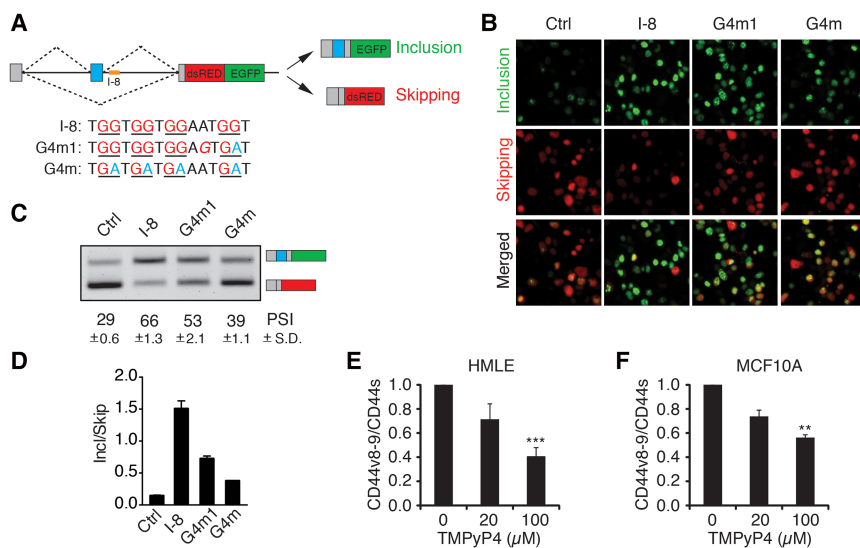


Figure 2. The I-8 G-quadruplex structure promotes alternative splicing. (A) Diagram of a fluorescent splicing reporter construct with I-8 and I-8 mutant sequences inserted downstream from the variable exon. The fluorescent proteins EGFP and dsRED are produced exclusively from the exon inclusion or exon-skipping splice isoforms, respectively. (B) Fluorescence images of cells transfected with the I-8 splicing reporter and its mutants, G4m1 and G4m. GFP and dsRED in live cells were recorded under a 10× fluorescence microscope. (C,D) Semi-quantitative RT-PCR images (C) and quantitative RT-PCR (qRT-PCR) results (D) showing that the I-8 G quadruplex stimulates exon inclusion. (E,F) qRT-PCR results showing that TMPyP4 treatment caused a decrease in exon inclusion of the endogenous CD44 pre-mRNA in HMLE (E) and MCF10 (F)

cells. The exon inclusion isoform CD44v8–9 and exon-skipping isoform CD44s were measured. (**), $P < 0.01$; (***), $P < 0.001$, Student's t -test.

S3A). Notably, insertion of all four G-quadruplex sequences in the minigene construct resulted in enhanced exon inclusion, as shown by increased fluorescence of exon inclusion-associated green (Fig. 3A) and increased PSI values (Fig. 3B). The G2U1 repeats that formed a two-guanine tetrad exerted the weakest effect on stimulating exon inclusion, while G3U1, NRQ, and CRQ, which formed more stable G-quadruplex structures with three guanine tetrads, showed nearly 90% exon inclusion. The degree of G-quadruplex-forming capacity measured by CD (Supplemental Fig. S3B) positively correlated with the levels of exon inclusion.

Introducing mutations into the guanine tetrads in the NRQ, CRQ, or G2U1 constructs abolished the effect on G-quadruplex-mediated exon inclusion (Fig. 3C–E; Supplemental Fig. S3C–F). Of note, we generated a G4m1 mutant of NRQ by shuffling two guanines to destroy the potential to form intramolecular G quadruplexes while maintaining the same compositions of nucleotides. Similar to what was observed for the I-8 G4m1 mutant, this NRQ G4m1 mutant resulted in a decrease in exon inclusion of NRQ. Further mutations of NRQ and CRQ, which mutated two to three guanines but maintained strong G tracts of four G bases, showed significantly diminished activity in exon inclusion (NRQ G4m2 and CRQ G4m2) (Fig. 3C–E; Supplemental Fig. S3C–E). Together, these results indicate that sequences with the capacity to form secondary G quadruplexes as opposed to their primary RNA sequences serve as the principle regulatory elements sufficient to stimulate exon inclusion.

The splicing factor hnRNPF is a potential G-quadruplex-binding protein

Having observed that G quadruplexes promote alternative RNA splicing, we aimed to identify RNA-binding proteins that mediate this regulation. We predicted the ability of RNA-binding proteins to bind to G quadruplexes by analyzing previously published CLIP-seq (cross-linking immunoprecipitation [CLIP] combined with deep sequencing) data (Huelga et al. 2012) for several hnRNPs, an abundant family of proteins that regulate alternative splicing. This study was the largest comparative analysis of CLIP-seq data available at the time, including six hnRNP CLIP-seq data sets. We merged overlapping CLIP sites between each data set and computed pairwise correlations with hierarchical clustering (Supplemental Fig. S4A). The hnRNPH data set was poorly correlated with and distinct from the other hnRNP data sets. This might be caused by a batch effect because the hnRNPH data set was not generated by the study investigators and was analyzed from a previous publication (Katz et al. 2010). We excluded the hnRNPH data set from further analysis.

We used quadruplex-forming G-rich sequence (QGRS) Mapper (Kikin et al. 2006) to score the strength of predicted G-quadruplex (PGQ) structures within the CLIP-seq hnRNP-binding regions. We examined the regions that comprised the defined hnRNP-binding sites and 25 nt of flanking sequence both upstream of and downstream from the binding sites. Bioinformatics analysis showed that PGQs were most significantly enriched in the binding regions of hnRNPF relative to random control sequences,

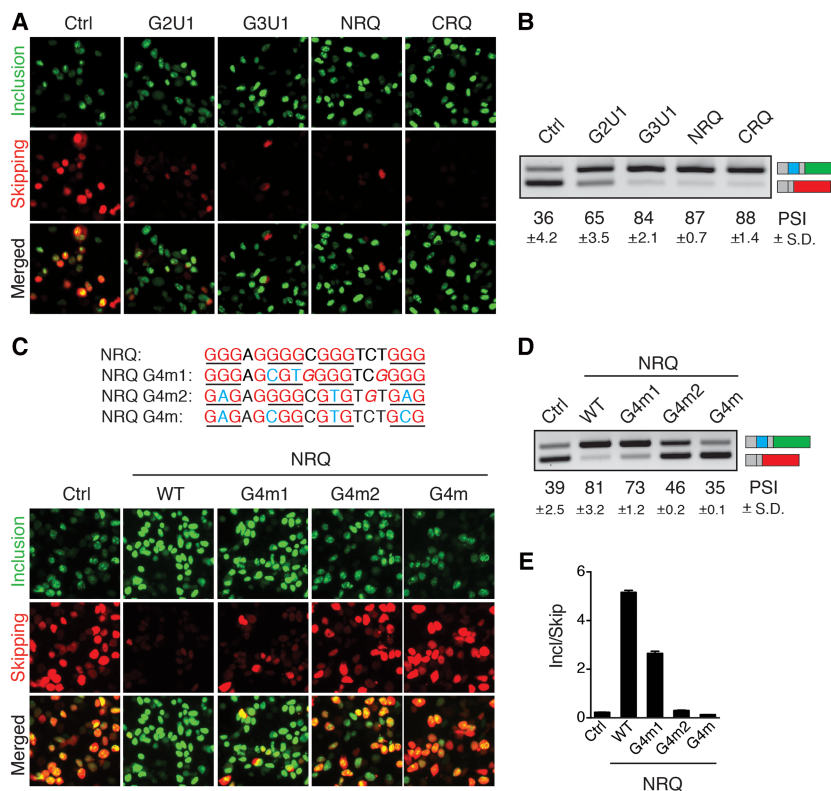


Figure 3. Bona fide G-quadruplex structures promote exon inclusion. (A) Fluorescence images of cells transfected with different G-quadruplex-containing constructs. (B) Semiquantitative RT-PCR results showing that G quadruplexes stimulate exon inclusion. (C, top) The sequences of NRQ and its mutants: G4m, G4m1, and G4m2. (Bottom) Fluorescence images of cells transfected with the indicated splicing reporter constructs showing the effect of NRQ mutations on abrogating exon inclusion. (D,E) Semiquantitative RT-PCR (D) and qRT-PCR (E) analysis of NRQ and its mutants showing that the NRQ G quadruplex stimulates exon inclusion, while NRQ G-quadruplex mutations decrease promotion of exon inclusion.

which exhibited the most significant enrichment out of the five hnRNPs analyzed (Fig. 4A; Supplemental Table S1). There were 3304 PGQ-containing binding sites within 13,616 hnRNPF-binding regions, comprising 24% of total hnRNPF-binding sites ($P = 1.18 \times 10^{-67}$, Fisher's exact test), compared with the second highest hnRNPA1 with 450, representing ~22% of 2043 hnRNPA1-binding sites ($P = 1.20 \times 10^{-13}$, Fisher's exact test). The percentages of PGQ per motif and PGQ per kilobase in the hnRNPF-binding regions were the highest of all hnRNPs analyzed (Supplemental Table S1). As a more stringent analysis, we used the G score provided by QGRS Mapper to obtain higher-confidence G quadruplexes. The G score is a metric to predict how likely a PGQ is to form a stable quadruplex structure and prioritizes shorter loop sequences between each G tract as well as loop sequences that are relatively equal in length (Kikin et al. 2006). To determine a relevant G score threshold unlikely to occur by chance, we quantified the number of G quadruplexes per G score occurring in 10,000 randomly shuffled sequences. G quadruplexes with G scores ≥ 19 had a P -value of <0.05 . Thus, we chose G scores ≥ 19 as an appropriate threshold to identify significant G quadruplexes in the hnRNP-binding sites, labeled PGQ-hi (Fig. 4A). This analysis revealed that hnRNPF had the most significant enrichment of PGQ-hi, compared with shuffled control, among all hnRNPs considered ($P = 1.53 \times 10^{-81}$, Fisher's exact test) (Fig. 4B). These data show that PGQs, especially those with a high probability of forming RNA G-quadruplex structures, are significantly enriched within hnRNPF-binding regions.

hnRNPF and hnRNPA1 had the first and second highest percentage of CLIP sites containing PGQs, respectively. By quantifying the position of the start positions of the 5' end of each PGQ relative to the center of the CLIP site, we observed a stronger central enrichment of PGQs in hnRNPF-bound CLIP sites compared with those of

hnRNPA1 (Fig. 4C). The percentage of PGQs in hnRNPF CLIP sites is highest near the center of hnRNPF CLIP sites and decreases further from the center, showing that G quadruplexes are quite proximal to the hnRNPF-binding site. In contrast, PGQs in hnRNPA1 CLIP sites are more dispersed and shifted upstream of the center of the CLIP sites. The increased central enrichment of PGQs within hnRNPF-binding sites suggests that hnRNPF binding is directly associated with the presence of a PGQ.

QGRS Mapper identifies one G quadruplex per hnRNP-binding site by default. Because PGQs that contain more than four G tracts have the potential to adopt distinct conformations by using different G tracts and thus form multiple overlapping G quadruplexes, we computed all possible G quadruplexes by requiring no more than four G tracts within each PGQ. Among all tested hnRNPs, hnRNPF-binding sites showed the most remarkable enrichment of overlapping PGQs (Fig. 4D; Supplemental Table S1), supporting its greater potential to form G quadruplexes. Taken together, the enrichment of PGQs, especially those with high G scores and overlapping G quadruplexes, within close vicinity of hnRNPF-binding sites suggests that hnRNPF preferentially binds to G-quadruplex-containing RNA regions compared with other hnRNPs. Examples of hnRNPF CLIP sites (Huelga et al. 2012) containing PGQs that are located in introns proximal to hnRNPF-regulated alternatively spliced exons are in Supplemental Figure S4B.

hnRNPF promotes exon inclusion of the I-8 G-quadruplex-containing pre-mRNA

We next sought to experimentally examine whether hnRNPF binds to RNA G quadruplexes to promote exon inclusion. We performed RNA pull-down assays by incubating a biotin-labeled 28-nt I-8 RNA fragment with

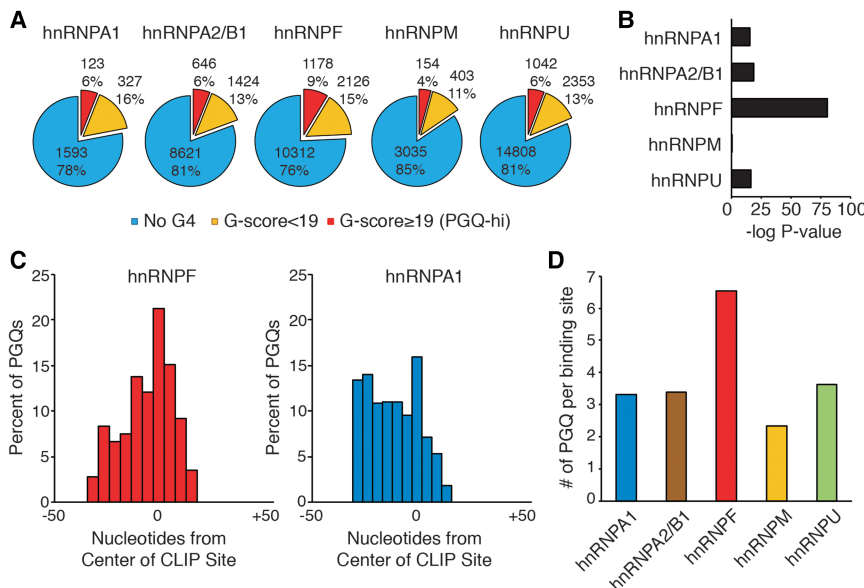


Figure 4. Transcriptome-wide enrichment of PGQs in hnRNPF-binding regions. (A) Percentage of QGRS PGQs in hnRNP CLIP binding sites and 25-nt upstream and downstream flanking sequences. The fraction of CLIP sites containing PGQs with G scores of ≥ 19 (PGQ-hi) are colored in red, while sites containing PGQ with G scores of <19 or sites lacking PGQs are colored in yellow and blue, respectively. (B) $-\log P$ -values for Fisher exact tests comparing enrichment of PGQ-hi in hnRNP CLIP binding sites compared with shuffled control sequences. (C) Histograms displaying the percentage and location of the start of PGQs relative to centers of hnRNPF (left panel) and hnRNPA1 (right panel) CLIP sites. PGQs occurring within ± 25 nt of the CLIP site are shown, but histograms are plotted ± 50 nt of the center of the CLIP site. (D) The average number of overlapping PGQs per binding site within the hnRNP CLIP binding sites, ± 25 nt.

HEK293FT nuclear extracts. The I-8 RNA was found to specifically interact with hnRNPF but not hnRNPA1 (Fig. 5A). When the GG repeats in the I-8 RNA (I-8 mut) were eliminated, hnRNPF lost the ability to bind to this RNA (Fig. 5A). We also found that TMPyP4 abrogated hnRNPF binding to I-8 in a dose-dependent manner (Fig. 5B). Additional examination showed that hnRNPF binds to G-quadruplex-containing RNAs, including the I-8 15-nt and other G quadruplexes (G2U1, NRQ, and CRQ) but not the mutant RNAs (Fig. 5C). In contrast, hnRNPA1 showed binding to only the NRQ probe (Fig. 5C).

To examine whether hnRNPF binds directly to the G-quadruplex-containing I-8, we incubated purified recombinant hnRNPF protein with the biotin-labeled I-8 wild-type and increasing molar ratios of cold unlabeled I-8 and I-8 G4m mutant RNA oligos and found loss of hnRNPF binding only with titration of unlabeled I-8 and

not I-8 G4m (Supplemental Fig. S5A). We also detected recombinant hnRNPF/I-8 complex formation in a native gel shift assay. Similar to what was observed in Figure 1E, the I-8 RNA ran slower than the same length G4m RNA oligo due to formation of the I-8 G quadruplex. Notably, we observed the appearance of supershifted I-8 RNA-protein complexes with hnRNPF at higher concentrations, suggesting the formation of hnRNPF-I-8 G-quadruplex higher-order complexes (Fig. 5D). In aggregate, these results reveal that hnRNPF interacts with the I-8 G-quadruplex structure.

We next determined the effect of hnRNPF on alternative splicing. Expressing hnRNPF inhibited the production of dsRED from the I-8 15-nt minigene, indicative of a decrease in exon skipping (Fig. 5E). hnRNPF expression had no effect on the I-8 G4m minigene. These results were further confirmed by both semiquantitative RT-PCR and qRT-PCR

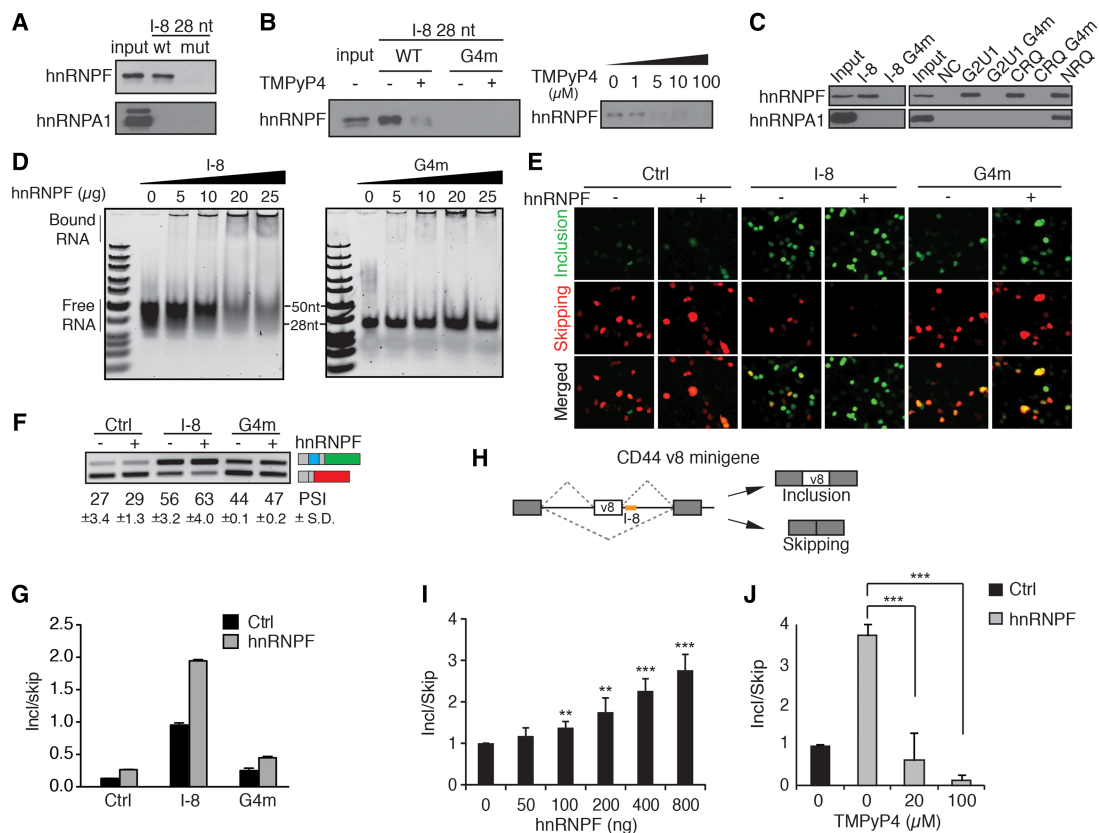


Figure 5. hnRNPF binds to G quadruplexes and promotes exon inclusion in a G-quadruplex-dependent manner. (A) RNA pull-down followed by Western blot analysis showing that the interaction of I-8 and hnRNPF depends on the wild-type G-quadruplex sequence. hnRNPA1 was also examined as a negative control. (B) RNA pull-down followed by Western blot analysis showing that TMPyP4 disrupts I-8 and hnRNPF interaction in a dose-dependent manner. (C) RNA pull-down results showing that hnRNPF binds to RNA G quadruplexes containing different linear sequences. (D) RNA electrophoretic mobility shift assay using 4 μ M I-8 or I-8 G4m and titrations of recombinant hnRNPF protein on native polyacrylamide gels. Only I-8 shows a mobility shift, indicating that I-8 binds to hnRNPF, while I-8 G4m does not. (E) Fluorescent images showing the effect of hnRNPF overexpression on splicing regulation of I-8 and its mutant after transfection into HEK293FT cells. (F, G) Semiquantitative RT-PCR analysis (F) and qRT-PCR analysis (G) of transfections in E showing that hnRNPF promotes exon inclusion in the I-8 G-quadruplex reporter but not the I-8 G4m mutant reporter. (H) Schematic diagram of the CD44v8 minigene illustrating the location of the I-8 G quadruplex downstream from v8 (in yellow). (I) hnRNPF promotes exon inclusion of the CD44v8 minigene in a dose-dependent manner upon cotransfection in HEK293FT cells. (J) Treatment with TMPyP4 antagonizes hnRNPF-mediated exon inclusion of the CD44v8 minigene. (***) $P < 0.001$, Student's *t*-test.

(Fig. 5F,G) and hnRNPF titration assays (Supplemental Fig. S5B,C) showing that hnRNPF promotes exon inclusion. Similar to these observations, cotransfection of hnRNPF with a CD44v8 splicing minigene reporter (Reinke et al. 2012; Xu et al. 2014) that includes the I-8-containing intron and its upstream CD44v8 exon (Fig. 5H) resulted in increased exon inclusion in a dose-dependent manner (Fig. 5I). Conversely, silencing hnRNPF by shRNA resulted in attenuated exon inclusion (Supplemental Fig. S5D). Moreover, addition of TMPyP4 abolished hnRNPF-induced exon inclusion of the CD44v8 minigene (Fig. 5J; Supplemental Fig. S5E). These results show that hnRNPF-mediated exon inclusion depends on the formation of G-quadruplex structures.

In addition to hnRNPF, previous studies have shown that the splicing factor ESRP1 binds to the GU-rich motif in the I-8 sequence and subsequently stimulates exon inclusion (Xu et al. 2014). We next determined whether hnRNPF and ESRP1 depend on each other for their splicing activities. Under the same condition, cotransfecting ESRP1 with the CD44v8 minigene increased the ratio of inclusion to skipping by 1.5-fold, while cotransfecting ESRP1 and hnRNPF increased the ratio to 2.8-fold (Supplemental Fig. S5F), suggesting an additive rather than a synergistic effect. Moreover, silencing ESRP1 or hnRNPF in HCT116 cells where both proteins are expressed did not alter the effect of hnRNPF or ESRP1 on CD44v8 minigene exon inclusion, respectively (Supplemental Fig. S5G,H), suggesting that ESRP1 and hnRNPF do not influence each other in regulating I-8-mediated exon inclusion. Furthermore, we analyzed the effect of hnRNPH, a homolog of hnRNPF, on exon inclusion and found that hnRNPH has a weaker effect on CD44v8 minigene exon inclusion (Supplemental Fig. S5I).

G quadruplexes are enriched near alternative exons regulated by hnRNPF

To examine whether hnRNPF regulates G-quadruplex-mediated alternative splicing globally across the transcriptome, we performed RNA sequencing (RNA-seq) analysis in control and hnRNPF silenced HMLE cells (Fig. 6A,B). hnRNPF knockdown caused significant differential splicing of 190 cassette exons (false discovery rate [FDR] < 0.05, Δ PSI \geq 0.2) (Supplemental Table S2). Among them, hnRNPF promoted inclusion of 136 (71.6%) exons and skipping of 54 (28.4%) exons (Fig. 6C), consistent with a previous exon array analysis showing that hnRNPF predominantly promotes exon inclusion (Huelga et al. 2012).

To test whether hnRNPF mediates alternative splicing of exons containing proximal G quadruplexes, we quantified enrichment of G3N7 PGQ near hnRNPF-dependent cassette exons. We found that within 250 nt upstream of or downstream from the alternative exon, 43 out of 190 (22.6%) hnRNPF-dependent exons contain a PGQ, compared with 303 out of 3471 (8.7%) control cassette exons not differentially spliced upon hnRNPF depletion, representing a statistically significant enrichment ($P = 2.36 \times 10^{-8}$, Fisher's exact test) (Fig. 6D, middle bars). Of these 43 hnRNPF-regulated exons containing PGQs, hnRNPF promoted inclusion of 37 out of 43 (86.0%), consistent with the major role of hnRNPF in promoting exon inclusion. The PGQ enrichment near the cassette exon was more significant than the enrichment near the upstream or downstream exons flanking the cassette exon (Fig. 6D). Furthermore, a more stringent analysis of PGQs within 150 nt upstream of or downstream from the cassette exon also showed significant enrichment of G quadruplexes near hnRNPF-mediated alternative splicing events

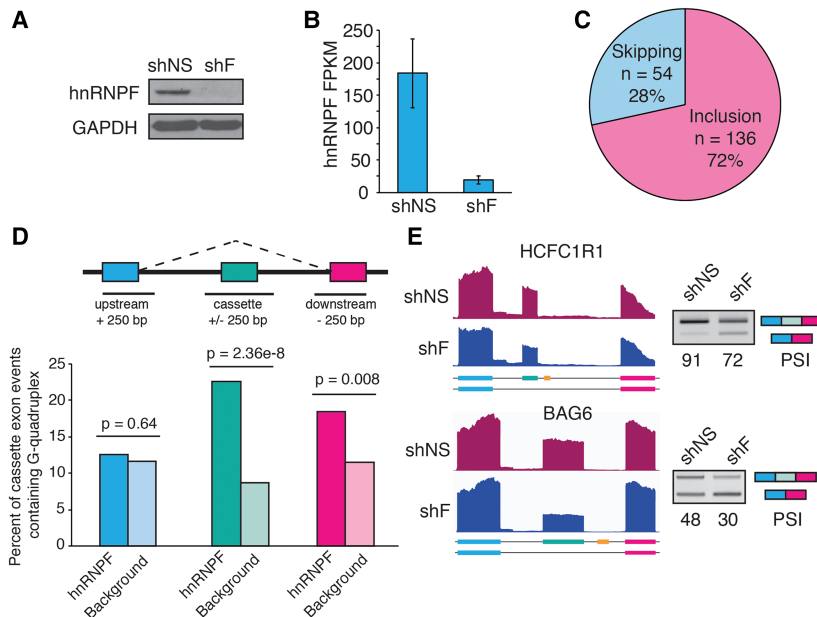


Figure 6. PGQs are enriched near hnRNPF-regulated cassette exons. (A) Western blot images showing hnRNPF expression in HMLE cells expressing control (shNS) and hnRNPF targeting (shF) shRNAs. (B) hnRNPF expression in control and hnRNPF knockdown HMLE cells quantified by RNA-seq. Error bars represent 95% confidence intervals. (C) Proportion of hnRNPF-regulated cassette exon splicing events where hnRNPF promotes exon inclusion (pink) versus exon skipping (blue). (D, top panel) Diagram of cassette exon alternative splicing events and regions scanned for PGQs within exons and 250 nt proximal to splice sites flanking the exon. (Bottom panel) The percentage of hnRNPF-regulated cassette exons (n = 190) containing PGQs upstream of, proximal to, or downstream from the cassette exon compared with the same regions in non-hnRNPF-regulated cassette exons (n = 3471). P-value was calculated by Fisher's exact test. (E, left panels) Genome browser tracks of hnRNPF-regulated exons containing cassette exon-proximal PGQs (shown in yellow bars). (Right panels) RT-PCR images showing increased exon skipping upon knockdown of hnRNPF in HMLE cells.

(Supplemental Fig. S6A). These observations suggest that primarily G-quadruplex sequences located proximal to the 3' and 5' splice sites of the cassette exon are targets for hnRNPF binding and splicing regulation. Ninety-three percent (40 out of 43) of the cassette exons contained PGQs exclusively in the flanking intronic sequence, with the remaining three events containing PGQs overlapping or entirely within the cassette exon itself, consistent with previously published results identifying that the majority of hnRNPF-binding sites is located in introns (Huelga et al. 2012).

We experimentally validated 12 highly expressed cassette exons containing PGQs within 250 nt of the cassette exon and found that the RNA-seq results successfully predicted changes in alternative splicing upon hnRNPF knockdown in 12 out of 12 exons tested (Fig. 6E; Supplemental Fig. S6B–D). These results indicate that hnRNPF primarily drives inclusion of alternative

splicing events and that G quadruplexes are significantly enriched in close proximity to the hnRNPF-regulated cassette exons.

hnRNPF inhibits EMT and cancer cell invasion

Our next goal was to determine the functional role of hnRNPF. We performed gene set enrichment analysis (GSEA) of genes differentially expressed upon hnRNPF knockdown. We found that hnRNPF depletion resulted in up-regulation of an EMT-associated gene set (Fig. 7A), implying that hnRNPF inhibits EMT. This observation is in line with the results that hnRNPF promotes CD44 variable exon inclusion (Fig. 5I) because our previous study reported that inhibiting CD44 variable exon inclusion, which leads to increased production of CD44s, is essential for EMT (Xu et al. 2014). Further evidence supporting the notion that hnRNPF promotes CD44 variable exon

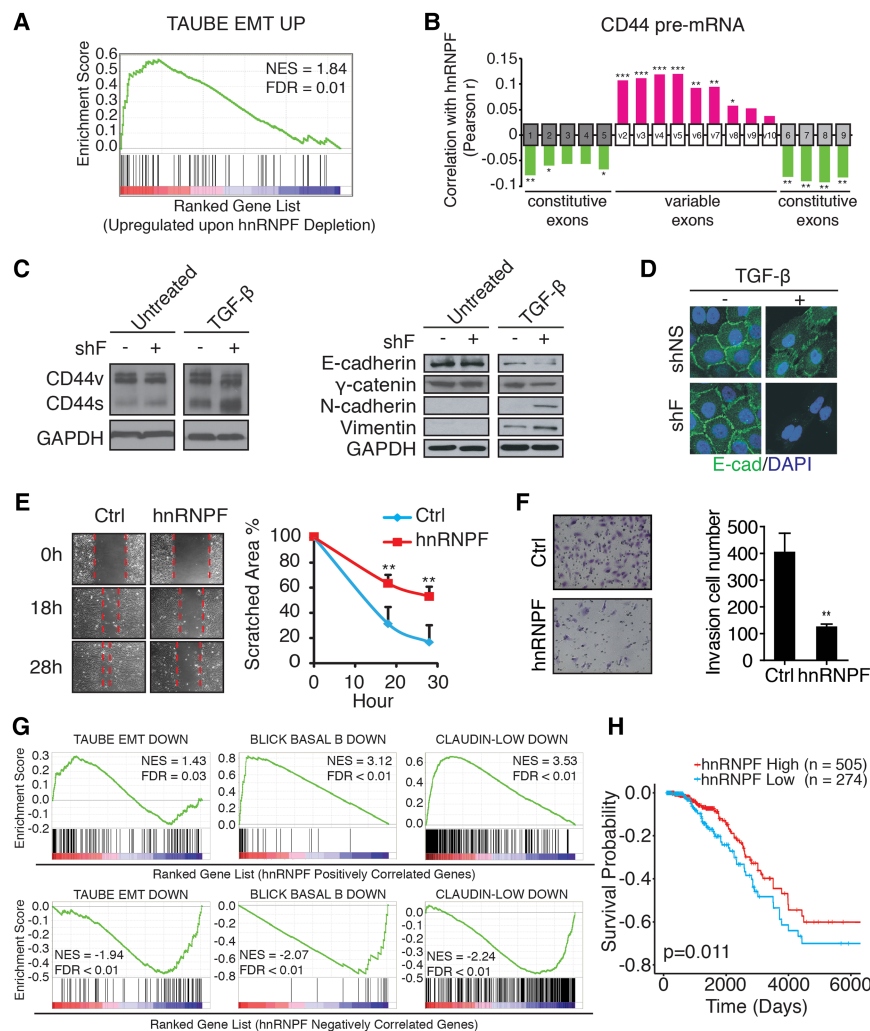


Figure 7. hnRNPF inhibits EMT and positively correlates with breast cancer patient survival. (A) GSEA showing that hnRNPF depletion caused an enrichment in an EMT signature. (NES) Normalized enrichment score. (B) Correlation of hnRNPF gene expression with CD44 exon expression in TCGA BRCA patient RNA-seq data sets showing strong positive correlation with CD44 variable exons (pink) and weaker negative correlation with CD44 constitutive exons (green). (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$, Pearson correlation. (C) Silencing of hnRNPF promoted CD44 isoform switching to CD44s (left panel), decreased the expression of epithelial markers E-cadherin and γ -catenin, and increased the expression of mesenchymal markers N-cadherin and vimentin (right panel) in MCF10A cells during TGF- β -induced EMT. (D) Immunofluorescence images showing that silencing of hnRNPF accelerated the loss of E-cadherin at cell junctions. (E) Wound healing assays showing that ectopic expression of hnRNPF in Mes10A cells inhibits cell migration. Representative images are shown in the left panel, and the percentage of the average scratched area normalized to 0 h is plotted in the right panel. (F) Transwell invasion assays showing that overexpression of hnRNPF in Mes10A cells inhibited cell invasion. Representative images (left panel) and quantification of invading cell numbers (right panel) are shown. Error bars represent SEM. (**) $P < 0.01$, Student's *t*-test. (G) GSEA of genes positively (top panel) and negatively (bottom panel) correlated with hnRNPF in TCGA BRCA patient RNA-seq samples revealed that hnRNPF

positively correlates with genes down-regulated during EMT, in basal subtype breast cancers, and in claudin-low breast cancer, while hnRNPF negatively correlates with genes up-regulated in these same gene sets. (H) Kaplan-Meier survival curves showing a significant overall survival difference between hnRNPF-high-expressing and hnRNPF-low-expressing tumors in TCGA BRCA patients. *P*-value was computed by log rank test.

inclusion came from analysis of The Cancer Genome Atlas (TCGA) breast invasive carcinoma (BRCA) RNA-seq data set. We found that hnRNPF expression significantly positively correlates with the levels of CD44 variable exons while negatively correlating with all constitutive CD44 exons (Fig. 7B).

To directly examine whether hnRNPF inhibits EMT, we used two independent in vitro EMT induction model systems: TGF- β -induced EMT in MCF10A cells and tamoxifen-induced EMT in Twist-ER-expressing HMLE cells. Treatment of MCF10A cells with TGF- β caused cells to undergo a complete EMT in 2 wk. Western blot analysis of cell lysates collected at day 8 after TGF- β treatment showed that hnRNPF silencing promoted CD44s expression. hnRNPF silencing also caused a greater reduction of epithelial markers E-cadherin and γ -catenin and an increase in expression of mesenchymal markers N-cadherin and vimentin (Fig. 7C for hnRNPF shF; Supplemental Fig. S7A for hnRNPF shF2). These hnRNPF silenced cells also displayed a spindle-shaped fibroblast-like appearance (Supplemental Fig. S7B) and a complete loss of E-cadherin at cell-cell junctions (Fig. 7D). Similar results were observed in the HMLE Twist-ER-induced EMT system (Supplemental Fig. S7C–F). Thus, these experimental data in combination with the bioinformatics analysis indicate that hnRNPF inhibits EMT.

A key functional characteristic of EMT is increased migration and invasion. We evaluated the role of hnRNPF in cell migration using a wound healing assay. Overexpression of hnRNPF significantly delayed wound healing in TGF- β -induced mesenchymal MCF10A cells (Mes10A), as shown by decreased closure of scratched areas compared with control cells (Fig. 7E). In contrast, hnRNPF silencing promotes wound closure, suggesting that hnRNPF inhibits cell motility (Supplemental Fig. S7G). Furthermore, overexpression of hnRNPF led to an ~70% reduction in Matrigel invasion of Mes10A cells, as visualized by Crystal Violet staining and quantified by cell counting (Fig. 7F). These hnRNPF-mediated effects were not due to the rate of cell proliferation because growth rate was nearly identical in control and hnRNPF-overexpressed cells (Supplemental Fig. S7H). Thus, these results demonstrate that hnRNPF inhibits cell migration and invasion.

hnRNPF expression correlates with breast cancer patient survival

Abnormal activation of EMT promotes cancer metastasis (Thiery and Sleeman 2006; Yang and Weinberg 2008). To explore the relevance of hnRNPF's function in cancer, we mined the TCGA breast cancer database and performed GSEA on all genes ranked by correlation with hnRNPF gene expression. We found that the hnRNPF gene set positively correlated with gene signatures that were down-regulated during EMT and in basal-like or claudin-low breast cancer subtypes (Fig. 7G). Conversely, hnRNPF gene sets negatively correlated with gene signatures that were up-regulated in these phenotypes. These results, derived from a large data set of human breast cancers,

strongly suggest that hnRNPF antagonizes EMT and invasive breast cancer.

The association of hnRNPF with breast cancer phenotypes associated with patient survival prompted us to examine the relationship between hnRNPF expression in breast cancer and overall patient survival. We performed two-means clustering of the TCGA BRCA samples by hnRNPF expression. Our results showed that patients with hnRNPF highly expressing tumors exhibited a significant survival advantage compared with patients with hnRNPF lowly expressing tumors ($P=0.011$, log rank test) (Fig. 7H). These results support our findings that hnRNPF antagonizes EMT and cancer progression.

Discussion

The critical role for alternative RNA splicing in normal development and pathological processes has been increasingly recognized. New technologies mapping the RNA-binding sites of RNA-binding proteins have greatly improved our understanding of protein–RNA interactions and their biological consequences. In addition to known linear RNA-binding motifs, RNA secondary structures have been increasingly found to serve as RNA-binding protein motifs in their own right in order to mediate RNA processing. G quadruplexes are a particular form of RNA secondary structure and are reported to regulate pre-mRNA processing, RNA turnover, and translation (Kikin et al. 2008; Ji et al. 2011; Millevoi et al. 2012; Conlon et al. 2016). Thus far, denomination of G quadruplexes, their associated RNA-binding proteins, and their biological functions remains largely unexplored.

In this study, we found that RNA sequences that have the capacity to form G quadruplexes serve as bona fide *cis*-regulatory elements to stimulate alternative splicing. We identify one such G-quadruplex structure in the intron of CD44. Using bicolor minigene reporters, the RNA G-quadruplex destabilizer TMPyP4, and biochemical approaches, we show that this G-quadruplex-containing RNA element promotes inclusion of CD44 variable exon v8, a splice isoform that is normally expressed in epithelial cells. We also show that hnRNPF binding potentiates CD44 alternative splicing toward the production of CD44v isoforms, resulting in inhibition of EMT and EMT-associated cell migration and invasion. These results are in line with previous findings that down-regulation of CD44v in epithelial cells and the production of CD44s are necessary for cells to gain invasive and survival properties and undergo EMT (Brown et al. 2011; Xu et al. 2014; Zhao et al. 2016; Liu and Cheng 2017). Importantly, the expression of hnRNPF inversely correlates with breast cancer patient survival.

G tracts that consist of three or more consecutive guanines harboring the potential to form G-quadruplex structures are frequent splicing recognition elements enriched near splice sites (Eddy and Maizels 2008; Huppert et al. 2008). A number of splicing factors were found to interact with RNA motifs with G-quadruplex-forming capacity,

including hnRNPF/H, hnRNPA1, hnRNPA2B1, hnRNPU, and SR proteins (Kiledjian and Dreyfuss 1992; Garneau et al. 2005; Khateb et al. 2007; von Hacht et al. 2014; Conlon et al. 2016; Liu et al. 2017). In this study, we found that potential G-quadruplex sequences are overrepresented in the vicinity of hnRNPF-binding sites compared with other hnRNPs. This observation is congruent with experimental results showing that hnRNPF binds to RNA containing G quadruplexes and stimulates cassette exon inclusion. Another member of the hnRNPF/H family, hnRNPH, is highly homologous to hnRNPF and is found to bind both G-quadruplex and G-tract motifs (Xiao et al. 2009; Conlon et al. 2016). Using the I-8-containing CD44 minigene reporter, we observed a less than twofold effect of hnRNPH on promoting CD44v8 inclusion, implying the specificity of hnRNPF splicing function at least in the context of the I-8 sequences. It is worth noting that the amyotrophic lateral sclerosis (ALS)-associated and frontotemporal dementia (FTD)-associated GGGGCC expansion of the *C9ORF72* gene forms RNA G quadruplexes. The G-quadruplex structure sequesters hnRNPH, resulting in local depletion of hnRNPH and thus disruption of hnRNPH-dependent splicing events that contain G tracts (Conlon et al. 2016). These results together with our findings raise interesting future investigations to determine whether hnRNPF and hnRNPH preferentially bind to G quadruplexes and G tracts, respectively, and whether the interplay between these two homologs may provide a precise mechanism for splicing regulation concerning the dynamics between G-quadruplex formation and the underlying linear G tracts. Another splicing regulator, ESRP1, was shown previously to bind to the I-8 GU-rich motif to regulate CD44 alternative splicing. This GU-rich motif overlaps with the G-quadruplex structure. Interestingly, however, both hnRNPF and ESRP1 stimulate CD44 variable exon inclusion but in an independent manner. Considering the dynamic nature of RNA G-quadruplex secondary structure formation, we speculate that the I-8-containing RNA fragment exists as a mixture of its linear form and G-quadruplex configuration. ESRP1 may occupy the linear GU-rich motif, while hnRNPF may preferentially bind to the G-quadruplex secondary structure, both of which promote inclusion of the CD44 variable exon. This possibility illustrates the complexity of alternative splicing regulation by suggesting a redundant mechanism to maintain CD44v splice isoform expression in epithelial cells and warrants future investigation. hnRNPF has been reported to bind to G quadruplexes as well as linear G tracts (Matunis et al. 1994; Dominguez and Allain 2006; Dominguez et al. 2010; Decorsière et al. 2011; Samatanga et al. 2013). For example, the G quadruplexes at the 3' untranslated region (UTR) of p53 recruit hnRNPF/H to execute its role in regulating p53 3' end processing in response to DNA damage (Decorsière et al. 2011). Interestingly, however, another report suggested that hnRNPH/F may prefer linear G-rich sequences embedded in potential G-quadruplex motifs to modulate epithelial- and myoblast-specific splicing events (Dardenne et al. 2014), but whether hnRNPF preferentially binds to the G-quadruplex secondary structure

or linear G tracts was not tested in this particular study. Moreover, the three-dimensional structure of the quasi-RNA recognition motif (qRRM), one of the three RRM RNA-binding domains of hnRNPF, showed that qRRM binds to single-stranded G tract RNAs to prevent the formation of G quadruplexes (Dominguez et al. 2010; Samatanga et al. 2013). It will be of great interest to determine whether the full-length hnRNPF with three RRMs maintains interaction with G tracts or preferentially interacts with G quadruplexes in a manner different from the isolated qRRM. Our RNA-binding assay showed that hnRNPF exhibited the ability to bind G-quadruplex-forming sequences, supporting the notion that hnRNPF is a G-quadruplex-binding protein. Regulation of the differential ability for hnRNPF to bind to G-quadruplex secondary structure versus linear G tracts would be an interesting topic for future investigation. We speculate that the binding of hnRNPF to G quadruplexes and linear G tracts could be influenced by other RNA-binding proteins and that the preferential binding may determine the outcome of splicing products.

Emerging evidence has indicated the importance of G quadruplexes in regulating fundamental biological activities, including translation regulation (Kumari et al. 2007; Bugaut and Balasubramanian 2012; Wolfe et al. 2014), alternative splicing (Gomez et al. 2004; Marcel et al. 2011; Conlon et al. 2016), and 3' end processing (Decorsière et al. 2011). These findings have stimulated interest in profiling G quadruplexes in mammalian cells. Using K⁺ and the small molecule pyridostatin (PDS), an RNA G-quadruplex-specific stabilizing ligand, it was shown that widespread formation of G-quadruplex structures exists in the human transcriptome (Kwok et al. 2016). Interestingly, when dimethyl sulfate (DMS) was used to modify the N7 position of the guanine residue, which probes for unfolded G quadruplexes, it was observed that most of the PGQ structures were modified by DMS and thus unfolded (Guo and Bartel 2016). These seemingly contradictory results reveal the dynamic nature of RNA G quadruplexes in live mammalian cells. G-quadruplex structures may be toggled between the folded and unfolded states. A transiently unfolded region could be accessed by DMS for N7G modification. The presence of this modification indicates that the G-quadruplex structures can be unfolded but does not exclude the existence of the higher-order G-quadruplex structures. Thus, novel approaches to define G-quadruplex structures in live cells and the regulatory mechanisms underlying the dynamics of G-quadruplex structures are of great interest to elucidate the location and function of these RNA secondary structures.

Taken together, our results identify *cis*-regulatory elements that contain G-quadruplex secondary structures and are critical for alternative splicing regulation by directly recruiting the splicing factor hnRNPF. In addition, hnRNPF, by regulating CD44 alternative splicing, shifts the balance toward the epithelial isoform CD44v and plays a crucial role in maintaining the epithelial phenotype, opening new horizons for cancer therapy by regulating the EMT.

Materials and methods

Cell culture and treatment

The maintenance of immortalized human mammary epithelial cells HMLE, HMLE-Twist-ER, and MCF10A; TGF- β -induced mesenchymal MCF10A (Mes10A) cells; and HEK293 cells was described previously (Brown et al. 2011). Treatment of TGF- β and ectopic expression of the EMT-initiating transcription factor TWIST were performed as described previously to induce EMT in MCF10A and HMLE cells (Brown et al. 2011; Huang et al. 2014). TMPyP4 (Sigma) dissolved in sterile water at a concentration of 40 μ M was stored at -20°C and thawed immediately before use in cell culture experiments.

CD spectroscopy

RNA oligonucleotides were purchased from Integrated DNA Technologies and dissolved in nuclease-free water at a concentration of 100 μ M. The RNA sequences were as follows: I-8 (5'-GC UUUGGUGGUGGAAUGGUGCUAUGUGG-3'), I-8-G4m (5'-GCUUUGAUGAUGAAAUGAUGCUAUGUGG-3'), G2U1 (5'-GGUGGUGGUGG-3'), G3U1 (5'-GGGUGGGUGGGUGGG-3'), NRQ (5'-GGGAGGGGCGGUCUGG-3'), and CRQ (5'-GGG CGGCGGGCGGCGGUGG-3').

For the spectroscopy measurements, RNA oligonucleotides were prepared in buffers containing 10 mM Tris-HCl (pH 7.4) in either the absence or presence of monovalent salt at a concentration of 5 μ M and annealed by heating to 95°C and then cooling slowly to room temperature. CD of RNA oligonucleotides was determined at 20°C by a Jasco J-815 spectropolarimeter equipped with a temperature controller. An average of three CD spectra ranging from 220 to 320 nm was recorded in a 0.1-mm path length cuvette at a scan rate of 50 nm/min with a 2-sec response time, 1.00-nm bandwidth, and continuous scan mode.

Plasmids and shRNAs

The plasmids expressing human hnRNPF shRNAs (shF) or non-specific control shRNA (shNS) were generated by insertion of target sequences (shF, GAGTGATGTTATCTAAGTTTA; shNS, GCCCGAATTAGCTGGACACTCAA) flanked by human miR-30 sequences into the XhoI and EcoRI sites of the TRC lentiviral shRNA vector (Open Biosystems). The shF2 plasmid was purchased from GE Open Biosystems with the target sequence CAATTAAGAGCAGTTATAA. Human hnRNPF cDNA was cloned into BamHI and SalI restriction sites of the pBabe-puro vector for stable overexpression of hnRNPF, while PCMV-SPORT6-hnRNPF was obtained from Harvard Medical School and used for transient expression in reporter assays.

To construct bichromatic fluorescent reporters, the wild-type or mutated G quadruplexes and scrambled sequences were cloned into EcoRI and XhoI sites of the RG6 vector, which was a kind gift from Dr. Thomas A. Cooper, Baylor College of Medicine (Orengo et al. 2006). The luciferase minigene reporter CD44v8 was described previously (Xu et al. 2014). The primers used for plasmid construction are listed in Supplemental Table S3.

Recombinant hnRNPF protein purification

The human recombinant hnRNPF protein was expressed in *Escherichia coli* BL-21. Bacteria were grown in LB medium supplemented with ampicillin at 37°C and induced with 1 mM IPTG at an OD₆₀₀ of 0.6 for 3–4 h at 25°C . Cells were spun down and resuspended in buffer (50 mM NaH_2PO_4 at pH 8.0, 0.5 M NaCl). Cells were lysed with lysozyme at a concentration

of 1 mg/mL on ice for 30 min. After sonication, lysates were centrifuged at 3000g for 15 min to remove cellular debris. The supernatant was incubated with Ni-NTA agarose beads for 3–4 h at 4°C . The beads were washed four times with wash buffer (50 mM NaH_2PO_4 , 0.5 M NaCl, 20 mM imidazole at pH 8.0). Protein was eluted in elution buffer (50 mM NaH_2PO_4 , 0.5 M NaCl, 250 mM imidazole at pH 8.0). Protein was concentrated using Amicon Ultra-4 centrifugal filters 30K. The protein concentration was determined by Bio-Rad protein concentration assay, and purified protein was visualized by SDS-PAGE.

Native PAGE assay

RNA oligonucleotides were incubated in 100 mM KCl and then heated for 3 min at 95°C followed by slow cooling at room temperature. RNA was resolved in a 10% native polyacrylamide gel. For hnRNPF binding to RNA oligonucleotides, 4 μ M RNA was incubated with different amounts of recombinant hnRNPF protein for 2 h on ice. RNA-protein complexes were resolved in an 8% native polyacrylamide gel. RNA species were visualized using SYBR Gold.

CD44 splicing reporter assays

For biochromatic fluorescent reporter assays, 1×10^6 HEK293FT cells were seeded in a 24-well plate 24 h prior to transfection using Lipofectamine 2000 (Life Technologies). GFP and dsRED fluorescence was observed under a $10\times$ microscope 48 h after transfection, and cells were collected for subsequent RNA extraction and RT-PCR analysis.

The luciferase minigene reporter assay was performed as described previously (Brown et al. 2011). Briefly, 2×10^6 HEK293FT cells were seeded in a 24-well plate. Twenty-four hours later, CD44v8 minigene constructs were cotransfected with the indicated amount of pCMV-hnRNPF. Cells were harvested for RNA extraction and RT-PCR analysis 24 h after transfection.

RNA extraction and semiquantitative and real-time RT-PCR

Cellular RNAs were extracted by RNA isolation kit (Omega Bio-Tek). Semiquantitative and real-time RT-PCR were performed as described (Huang et al. 2014; Harvey and Cheng 2016). Briefly, 250 ng of total RNA was used for cDNA synthesis. Hot StarTaq DNA polymerase (Qiagen) was used for semiquantitative PCR, and PCR products were separated on 1.5% agarose gels. The relative signal intensity of each band was quantitated by ImageMaster Totalab software (Nonlinear Dynamics Ltd.). For real-time PCR, Bio-Rad iQ SYBR Green supermix was used.

Western blot and immunofluorescence

Cells were lysed in RIPA buffer and analyzed by Western blotting as described (Huang et al. 2014). Antibodies used for Western blotting were as follows: anti-CD44 (R&D Systems), anti-hnRNPF (Santa Cruz Biotechnology), anti-E-cadherin (Cell Signaling Technology), anti- γ -catenin (Cell Signaling Technology), and anti-vimentin (NeoMarkers). GAPDH (GE) was used as a loading control. For immunofluorescence, cells were fixed in 4% polyformaldehyde and permeabilized with 0.2% Triton X-100 before blocking with 5% BSA. Primary antibody incubations with rabbit anti-E-cadherin (1:200; Cell Signaling Technology) were performed overnight at 4°C followed by incubation with Alexa fluor 488-conjugated anti-rabbit IgG (1:500; Invitrogen) for 2 h at room temperature and DAPI staining. Representative

images were captured under a Nikon C2 laser scanning confocal microscope.

RNA pull-down

The RNA oligonucleotides with 5' biotin modifications were synthesized by Integrated DNA Technologies. The RNA sequences were as follows: Biotin-I-8-wt (5'-biotin-GCUUUGGUGGUGG AAUGGUGCUAUGUGG-3'), Biotin-I-8-mut (5'-biotin-GCUU UGAUGAUGAAAUGAUGCUAUGUGG-3'), Biotin-I-8-15nt (5'-biotin-GACUUGGUGGUGGAAUGGUCA-3'), Biotin-I-8-15nt-G4m (5'-biotin-GACUUGAUGAUGAAAUGAUGCA-3'), Biotin-G2U1 (5'-biotin-GACUGGUGGUGGUGGCA-3'), Biotin-G2U1-G4m (5'-biotin-GACUGAUGCUGAUGCCA-3'), Biotin-CRQ (5'-biotin-GACUGGGCGGGCGGGCGGGCUGGGG CA-3'), Biotin-CRQ-G4m (5'-biotin-GACUGAGCGCCGAGC GCGCUGAGGCA-3'), and Biotin-NRQ (5'-biotin-GACUGGG AGGGGCGGGUCUGGGCA-3').

Biotinylated RNA oligos (400 pmol) were immobilized to streptavidin beads before incubation with 200 μ g of HEK293FT nuclear extracts in 300 μ l of binding buffer (20 mM Tris, 200 mM NaCl, 6 mM EDTA, 5 mM potassium fluoride, 5 mM β -glycerophosphate, 2 mg/mL aprotinin at pH 7.5) overnight at 4°C. Beads were washed three times with binding buffer and boiled in sodium dodecyl sulfate buffer. The samples were analyzed by Western blot.

Cell migration and invasion assays

Mes10A and HMLE cells were seeded at a density of 2×10^5 per well in 24-well plates to allow cells to grow as confluent monolayers in 24 h. After 24 h of serum starvation, the cells were scratched and continuously cultured in complete culture medium, allowing migration for 10–28 h. Images were taken at the indicated time under a Nikon Eclipse TS100 microscope. The scratched areas were calculated using Adobe Photoshop.

Cell proliferation assay

Cell growth/proliferation was detected by CellTiter 96 nonradioactive cell proliferation assay (MTT, Promega) following the manufacturer's instructions. Briefly, 0.5×10^4 Mes10A cells with stable expressed hnRNPF or empty vector (pBabe-puro) were seeded in 96-well plates. Dye solution was added at the indicated time points and incubated for 4 h at 37°C before addition of solubilization/stop solution. The absorbance at 570 nm (with reference at 630 nm) was read at the end of the experiments.

PGQ analysis from CLIP-seq data

High-throughput CLIP-seq data and exon array data of hnRNPs were retrieved from the Gene Expression Omnibus (GEO), and the annotation of RNA-binding motifs were provided by Dr. Gene Yeo. Pairwise correlation of hnRNP RNA-binding motifs was conducted after merging overlapping hnRNP CLIP sites. hnRNP-binding sites ± 25 -nt flanking sequences were analyzed by QGRS Mapper (<http://bioinformatics.ramapo.edu/QGRS/index.php>) with parameters maximum length = 30 nt, minimum G-group = 2, and loop size = 0–36 to identify PGQs. PGQs in randomly shuffled sequences were analyzed as a background control. A Fisher's exact test was used to evaluate the statistical significance of enrichment of PGQs for each set of hnRNP-binding sites compared with the shuffled background control. Central enrichment of PGQs in hnRNP CLIP sites was analyzed by calculating the distance from the 5' end of the PGQ to the center of each corre-

sponding CLIP site. PGQs per CLIP or PGQs per kilobase were calculated by dividing the numbers of PGQs by the total number or length of hnRNP-binding sites, while the enrichment fold change was calculated by normalizing the PGQ per motif of each hnRNP to that of the corresponding shuffled background sequences.

RNA-seq, alternative splicing, G-quadruplex enrichment analysis, and GSEA

RNA was extracted from HMLE cells stably expressing shNS or shF from two biological replicates using Trizol, and poly-A-selected RNA-seq libraries were generated using TruSeq stranded mRNA library preparation kits (Illumina) and subjected to 100-base-pair PE stranded RNA-seq on an Illumina HiSeq 4000. RNA-seq reads were aligned to the human genome (GRCh37, primary assembly) and transcriptome (Gencode version 24 backmap 37 comprehensive gene annotation) using STAR version 2.5.3a (Dobin et al. 2013) using the following parameters: STAR --runThreadN 16 --alignEndsType EndToEnd --quantMode GeneCounts --outSAMtype BAM SortedByCoordinate. Differential gene expression was quantified using Cuffdiff version 2.2.1 (Trapnell et al. 2010) using the following nondefault parameters: --max-bundle-frags 100000000 --library-type fr-firststrand. Differential alternative splicing was quantified using rMATS version 3.2.5 (Shen et al. 2014) using the following nondefault parameters: --t paired --len 100 --analysis U --libType fr-firststrand. Significant differentially spliced cassette exons were quantified using only unique junction reads and the following cutoffs: FDR < 0.05, Δ PSI ≥ 0.2 , and average junction reads per cassette event per replicate ≥ 20 . Control cassette exons were identified by the following filters: FDR > 0.5, minimum PSI for shNS or shF < 0.85, maximum PSI > 0.15, and average junction reads per cassette event per replicate ≥ 20 . These filters were selected to identify cassette exons with evidence of alternative splicing but were not differentially spliced upon hnRNPF depletion. G-quadruplex enrichment analysis was conducted by curating a strand-specific list of all sequences in the GRCh37 version of the human genome matching the regular expression GxNy, where N is any nucleotides other than G, x is ≥ 3 , and $1 \leq y \leq 7$ to identify all sequences with G tracts of three or more separated by one to seven of any other base. Bedtools version 2.26.0 and custom Bash and Python scripts were used to intersect hnRNPF-dependent alternative splicing events and the 250 or 150 nt flanking each splice site with these G quadruplexes. Twelve highly expressed cassette exons near PGQs were selected for semiquantitative PCR validation per the following criteria: all exon junctions present in Gencode gene annotation, FPKM (fragments per kilobase per million mapped fragments) for the associated gene ≥ 10 , junction reads ≥ 70 per cassette exon event per sample, and junction reads ≥ 10 per inclusion or skipping isoform. Processed TCGA BRCA level 3 data for gene and exon expression were downloaded from the Genomic Data Commons Legacy Archive. GSEA was performed using the Broad Institute javaGSEA desktop application. GSEA using hnRNPF depletion RNA-seq was conducted with all genes with a minimum expression of FPKM ≥ 1 in control or hnRNPF knockdown data sets. GSEA using TCGA BRCA gene and exon expression data included all genes in all BRCA samples for which exon expression data were available ranked by Pearson correlation coefficient with hnRNPF expression. RNA-seq data are available under Gene Expression Omnibus (GEO) accession number GSE107542.

Acknowledgments

We thank Dr. Thomas Cooper for providing the RG6 plasmid. We thank Dr. Gene Yeo for sharing the CLIP data. This research was

supported in part by grants from the National Institutes of Health (F30CA196118 to S.E.H., and R01GM110146 and R01CA182467 to C.C.). C.C. is a Cancer Prevention Research Institute of Texas Scholar in Cancer Research. J.Z., H.H., S.E.H., X.H., and C.C. designed experiments. J.Z., H.H., S.E.H., and X.H. performed research. J.Z., H.H., S.E.H., X.H., and C.C. analyzed the data. J.Z., H.H., S.E.H., and C.C. wrote and edited the manuscript.

References

- Agarwal T, Jayaraj G, Pandey SP, Agarwala P, Maiti S. 2012. RNA G-quadruplexes: G-quadruplexes with 'U' turns. *Curr Pharm Des* **18**: 2102–2111.
- Blice-Baum AC, Mihailescu MR. 2014. Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. *RNA* **20**: 103–114.
- Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, Cheng C. 2011. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest* **121**: 1064–1074.
- Bugaut A, Balasubramanian S. 2012. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**: 4727–4741.
- Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Manley JL. 2016. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife* **5**: e17820.
- Dardenne E, Polay Espinoza M, Fattet L, Germann S, Lambert MP, Neil H, Zonta E, Mortada H, Gratadou L, Deygas M, et al. 2014. RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep* **7**: 1900–1913.
- Davis JT. 2004. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* **43**: 668–698.
- Decorsière A, Cayrel A, Vagner S, Millevoi S. 2011. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev* **25**: 220–225.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Dominguez C, Allain FH. 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res* **34**: 3634–3645.
- Dominguez C, Fisetite 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.
- Eddy J, Maizels N. 2008. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res* **36**: 1321–1333.
- Fahlman RP, Sen D. 1998. Cation-regulated self-association of 'synapsable' DNA duplexes. *J Mol Biol* **280**: 237–244.
- Fu XD, Ares M Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**: 689–701.
- Garneau D, Revil T, Fisetite JF, Chabot B. 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* **280**: 22641–22650.
- Gomez D, Lemarteleur T, Lacroix L, Mailliet P, Mergny JL, Riou JF. 2004. Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res* **32**: 371–379.
- Graveley BR. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**: 100–107.
- Guo JU, Bartel DP. 2016. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**: aaf5371.
- Han FX, Wheelhouse RT, Hurley LH. 1999. Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *J Am Chem Soc* **121**: 3561–3570.
- Harvey SE, Cheng C. 2016. Methods for characterization of alternative RNA splicing. *Methods Mol Biol* **1402**: 229–241.
- Hernandez JR, Kim JJ, Verdone JE, Liu X, Torga G, Pienta KJ, Mooney SM. 2015. Alternative CD44 splicing identifies epithelial prostate cancer cells from the mesenchymal counterparts. *Med Oncol* **32**: 159.
- Huang H, Xu Y, Cheng C. 2014. Detection of alternative splicing during epithelial-mesenchymal transition. *J Vis Exp* **9**: e51845.
- Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiu L, Hoon S, Brenner S, et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**: 167–178.
- Huppert JL, Bugaut A, Kumari S, Balasubramanian S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* **36**: 6260–6268.
- Jangi M, Sharp PA. 2014. Building robust transcriptomes with master splicing factors. *Cell* **159**: 487–498.
- Ji X, Sun H, Zhou H, Xiang J, Tang Y, Zhao C. 2011. Research progress of RNA quadruplex. *Nucleic Acid Ther* **21**: 185–200.
- Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141–2144.
- Katz Y, Wang ET, Airoidi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–1015.
- Khateb S, Weisman-Shomer P, Hershco-Shani I, Ludwig AL, Fry M. 2007. The tetraplex (CGG)_n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X pre-mutation mRNA. *Nucleic Acids Res* **35**: 5775–5788.
- Kikin O, D'Antonio L, Bagga PS. 2006. QGRS Mapper: a Web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* **34**: W676–W682.
- Kikin O, Zappala Z, D'Antonio L, Bagga PS. 2008. GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res* **36**: D141–D148.
- Kiledjian M, Dreyfuss G. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J* **11**: 2655–2664.
- Kumari S, Bugaut A, Huppert JL, Balasubramanian S. 2007. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* **3**: 218–221.
- Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. 2016. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* **13**: 841–844.
- Lattmann S, Giri B, Vaughn JP, Akman SA, Nagamine Y. 2010. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res* **38**: 6219–6233.

- Liu S, Cheng C. 2013. Alternative RNA splicing and cancer. *Wiley Interdiscip Rev RNA* **4**: 547–566.
- Liu S, Cheng C. 2017. Akt signaling is sustained by a CD44 splice isoform-mediated positive feedback loop. *Cancer Res* **77**: 3791–3801.
- Liu X, Ishizuka T, Bao H-L, Wada K, Takeda Y, Iida K, Nagasawa K, Yang D, Xu Y. 2017. Structure-dependent binding of hnRNPA1 to telomere RNA. *J Am Chem Soc* **139**: 7533–7539.
- Marcel V, Tran PL, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou MP, Hall J, Mergny JL, Hainaut P, Van Dyck E. 2011. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32**: 271–278.
- Martino L, Pagano B, Fotticchia I, Neidle S, Giancola C. 2009. Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *J Phys Chem B* **113**: 14779–14786.
- Matunis MJ, Xing J, Dreyfuss G. 1994. The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic Acids Res* **22**: 1059–1067.
- Millevoi S, Moine H, Vagner S. 2012. G-quadruplexes in RNA biology. *Wiley Interdiscip Rev RNA* **3**: 495–507.
- Morris MJ, Wingate KL, Silwal J, Leeper TC, Basu S. 2012. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res* **40**: 4137–4145.
- Novy J, Bohm S, Kralova J, Kral V, Urbanova M. 2008. Formation and temperature stability of G-quadruplex structures studied by electronic and vibrational circular dichroism spectroscopy combined with ab initio calculations. *Biopolymers* **89**: 144–152.
- Ofer N, Weisman-Shomer P, Shklover J, Fry M. 2009. The quadruplex r(CG_n)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res* **37**: 2712–2722.
- Orengo JP, Bundman D, Cooper TA. 2006. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Research* **34**: e148.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413–1415.
- Reinke LM, Xu Y, Cheng C. 2012. Snail represses the splicing regulator epithelial splicing regulatory protein 1 to promote epithelial-mesenchymal transition. *J Biol Chem* **287**: 36435–36442.
- Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. 2015. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**: 698–711.
- Samatanga B, Dominguez C, Jelesarov I, Allain FHT. 2013. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Research* **41**: 2505–2516.
- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci* **111**: E5593–E5601.
- Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci* **99**: 11593–11598.
- Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, et al. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**: 956–960.
- Thiery JP, Sleeman JP. 2006. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol* **7**: 131–142.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- von Hacht A, Seifert O, Menger M, Schutze T, Arora A, Konthar Z, Neubauer P, Wagner A, Weise C, Kurreck J. 2014. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* **42**: 6630–6644.
- Vorlickova M, Kejnovska I, Sagi J, Rencuk D, Bednarova K, Motlova J, Kypr J. 2012. Circular dichroism and guanine quadruplexes. *Methods (San Diego, Calif)* **57**: 64–75.
- Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Weng H-Y, Huang H-L, Zhao P-P, Zhou H, Qu L-H. 2012. Translational repression of cyclin D3 by a stable G-quadruplex in its 5' UTR: implications for cell cycle regulation. *RNA Biol* **9**: 1099–1109.
- Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, et al. 2014. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**: 65–70.
- Xiao X, Wang Z, Jang M, Nutiu R, Wang ET, Burge CB. 2009. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* **16**: 1094–1100.
- Xu Y, Gao XD, Lee JH, Huang H, Tan H, Ahn J, Reinke LM, Peter ME, Feng Y, Gius D, et al. 2014. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev* **28**: 1191–1203.
- Yang J, Weinberg RA. 2008. Epithelial–mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**: 818–829.
- Zamiri B, Reddy K, Macgregor RB Jr, Pearson CE. 2014. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)_n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J Biol Chem* **289**: 4653–4659.
- Zhang J, Manley JL. 2013. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* **3**: 1228–1237.
- Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, Daley EL, Miller SJ, Cunningham KM, Vidensky S, et al. 2015. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**: 56–61.
- Zhao P, Xu Y, Wei Y, Qiu Q, Chew TL, Kang Y, Cheng C. 2016. The CD44s splice isoform is a central mediator for invadopodia activity. *J Cell Sci* **129**: 1355–1365.