The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation

Daniel Benhalevy,[1,5] Sanjay K. Gupta,[2,5] Charles H. Danan,[1] Suman Ghosal,[1] Hong-Wei Sun,[3] Hinke G. Kazemier,[4] Katrin Paeschke,[4] Markus Hafner,[1,6,*] and Stefan A. Juranek[2,4,*]

[1]Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD 20892, USA
[2]Department of Biochemistry, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany
[3]Biostatistics and Datamining Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[4]European Research Institute for the Biology of Ageing (ERIBA), University Medical Center Groningen, University of Groningen, 9713 AV Groningen, the Netherlands
[5]Co-first author
[6]Lead Contact
*Correspondence: markus.hafner@nih.gov (M.H.), s.a.juranek@umcg.nl (S.A.J.)
http://dx.doi.org/10.1016/j.celrep.2017.02.080

SUMMARY

The CCHC-type zinc finger nucleic acid-binding protein (CNBP/ZNF9) is conserved in eukaryotes and is essential for embryonic development in mammals. It has been implicated in transcriptional, as well as post-transcriptional, gene regulation; however, its nucleic acid ligands and molecular function remain elusive. Here, we use multiple systems-wide approaches to identify CNBP targets and function. We used photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) to identify 8,420 CNBP binding sites on 4,178 mRNAs. CNBP preferentially bound G-rich elements in the target mRNA coding sequences, most of which were previously found to form G-quadruplex and other stable structures in vitro. Functional analyses, including RNA sequencing, ribosome profiling, and quantitative mass spectrometry, revealed that CNBP binding did not influence target mRNA abundance but rather increased their translational efficiency. Considering that CNBP binding prevented G-quadruplex structure formation in vitro, we hypothesize that CNBP is supporting translation by resolving stable structures on mRNAs.

INTRODUCTION

Gene regulation involves recognition of *cis*-acting sequence elements on both DNA and RNA by transcription factors and RNA-binding proteins (RBPs), respectively (Vaquerizas et al., 2009; Gerstberger et al., 2014). Nucleic acid recognition is mediated by a limited set of protein domains that are highly conserved throughout evolution (Wilson et al., 2008; Gerstberger et al., 2014; Lunde et al., 2007; Finn et al., 2010). One of the canonical nucleic acid-recognizing domains found in both DNA-binding proteins and RBPs is the zinc finger (ZnF) domain. ZnF domains can be broadly subdivided into six distinct folds, of which CCCH and CCHC ZnFs are most common in eukaryotes (Klug and Rhodes 1987). A common feature of nucleic acid-binding domains in general, and ZnF domains in particular, is their frequent occurrence in homo- or heterotypic arrays of multiple nucleic acid-binding domains (Gerstberger et al., 2014). As an example, human ZFP100 contains up to 18 ZnF domains of four different types (Gerstberger et al., 2014). Given that most nucleic acid-binding domains recognize short and degenerate 4- to 6-nt-long segments, it is thought that such a modular design increases affinity and sequence specificity of the protein-nucleic acid interaction (Ascano et al., 2012a; Lunde et al., 2007).

The cellular nucleic acid-binding protein (CNBP/ZNF9) is among the proteins with the highest number of repeats of the same nucleic acid-binding domain. It is evolutionary conserved in eukaryotes and harbors seven CCHC-type ZnF domains in addition to one arginine-glycine (RG/RGG)-rich motif (Figure 1A). *CNBP* is ubiquitously expressed at high levels in adult tissues (Figure S1A) (Gerstberger et al., 2014), and gene knockout in mice is embryonically lethal, presumably due to impaired embryonic forebrain development (Chen et al., 2003; Shimizu et al., 2003). In humans, a CCTG expansion in intron 1 of CNBP causes myotonic dystrophy type 2 (Liquori et al., 2001). Nevertheless, little is known about the function of CNBP, its targets, and its molecular mechanism.

CNBP was first described as a DNA-binding protein and potential regulator for the sterol regulatory element (SRE) (Rajavashisth et al., 1989). It scored high in an in vitro screen for binders of single-stranded DNA (ssDNA) with the defined sequence GTGCGGTG. A role for CNBP in transcriptional regulation was further suggested by its ability to bind the CT element in the promotor of the *c-myc* gene in vitro and activate it after
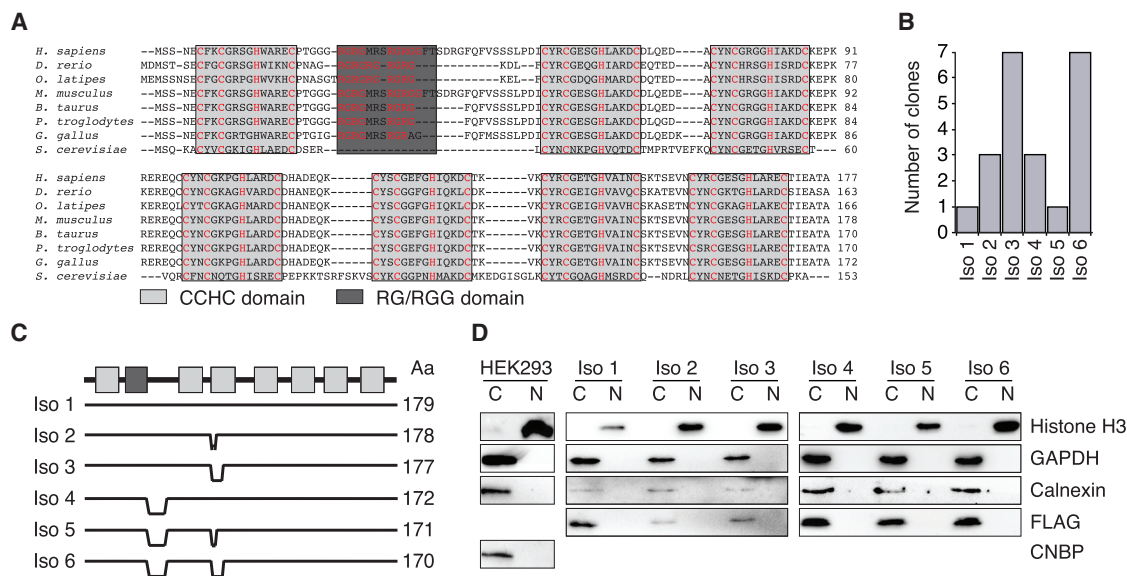
Figure 1. Overview of CNBP Protein

(A) ClustalW multiple sequence alignment of CNBP isoform 1 from various eukaryotes. CCHC domains are indicated by light gray boxes, and the RGG domain is indicated by a dark gray box. Signature C and H amino-acid positions are highlighted (red).

(B) Clone count distribution across the six CNBP isoforms (iso) from 22 sequenced full-length clones from HEK293 cDNA.

(C) Comparison and protein length (amino acids, Aa) of the six different CNBP isoforms.

(D) Analysis of the nucleocytoplasmic distribution of wild-type CNBP and FH-CNBP isoform 1–6 HEK293 cell lines, as indicated. Cytoplasmic (C) and nuclear (N) fractions were probed with anti-CNBP, anti-FLAG (FH-CNBP), anti-Histone H3 (nuclear marker), anti-GAPDH (cytoplasmic marker), and anti-Calnexin (endoplasmic reticulum marker) antibodies.

overexpression in cultured murine teratocarcinoma cells (P19) (Michelotti et al., 1995; Shimizu et al., 2003). In contrast, pull-down of polyadenylated RNA after in vivo UV crosslinking identified CNBP as an RBP in amphibian oocytes and human cell lines (Pellizzoni et al., 1997; Calcaterra et al., 1999; Baltz et al., 2012; Castello et al., 2012). Other studies suggested that human CNBP interacts with G-rich single-stranded RNA (ssRNA) in vitro (Armas et al., 2008) and associates with the terminal 5′ oligopyrimidine (TOP) sequence of specific mRNAs (Iadevaia et al., 2008) to promote their translation (Huichalaf et al., 2009). CNBP was also found to enhance the cap-independent translation of reporter constructs containing an ornithine decarboxylase (ODC) internal ribosome entry site (IRES), suggesting a function as a IRES trans-acting factor (ITAF) (Gerbasi and Link 2007). The yeast homolog of CNBP, Gis2, associates with PABPC1 and the translational machinery and localizes to stress-induced cytoplasmic granules, which further supports a potential function for CNBP in translational regulation (Rojas et al., 2012).

Here, we aimed to consolidate these seemingly conflicting hypotheses about CNBP's regulatory function using biochemical and systems-wide approaches. We identified CNBP as a cytoplasmic RBP and profiled its RNA targets on a transcriptome-wide scale at nucleotide-level resolution using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Hafner et al., 2010b). CNBP bound 4,178 mRNAs at 7,545 distinct sites, distributing mainly to G-rich regions in the first 50 nt downstream of the AUG start codon. CNBP binding sites were preferentially found in regions that were recently reported to form G-quadruplex (G4) and other stable structures

(Guo and Bartel, 2016). Interestingly, Guo and Bartel noted that, in eukaryotes, these structures are not detectable in vivo and likely require a dedicated machinery to resolve them and relieve their inhibitory effect on translation. Using a host of genome-wide approaches, including RNA sequencing (RNA-seq), ribosome profiling (Ribo-seq), and quantitative mass spectrometry, we found that CNBP is an example of a sequence-specific RBP promoting the translation of its mRNA targets on a global scale. Considering that CNBP binding prevented G4 structure formation in vitro, we hypothesize that CNBP is supporting translation by resolving stable structures on mRNAs.

RESULTS

CNBP Is a Cytoplasmic Protein Consisting of Multiple Simultaneously Expressed Isoforms

Human CNBP is predicted to encode multiple splice isoforms that change the composition of its nucleic acid-binding domains and potentially modify target specificity (Chen et al., 2013). In HEK293 cells, we detected the expression of six distinct isoforms by sequencing of cDNA (Figure 1B). CNBP isoform 3, which harbors a deletion in the third CCHC domain, and isoform 6, which contains deletions in the third CCHC domain and at the C terminus of the RG-rich domain, were the most abundant, representing more than 60% of the sequenced CNBP clones (Figures 1B and 1C).

RG-rich domains are often targets of arginine methylation (McBride et al., 2009), which could influence CNBP function in two ways: methylated arginine residues can be recognized by
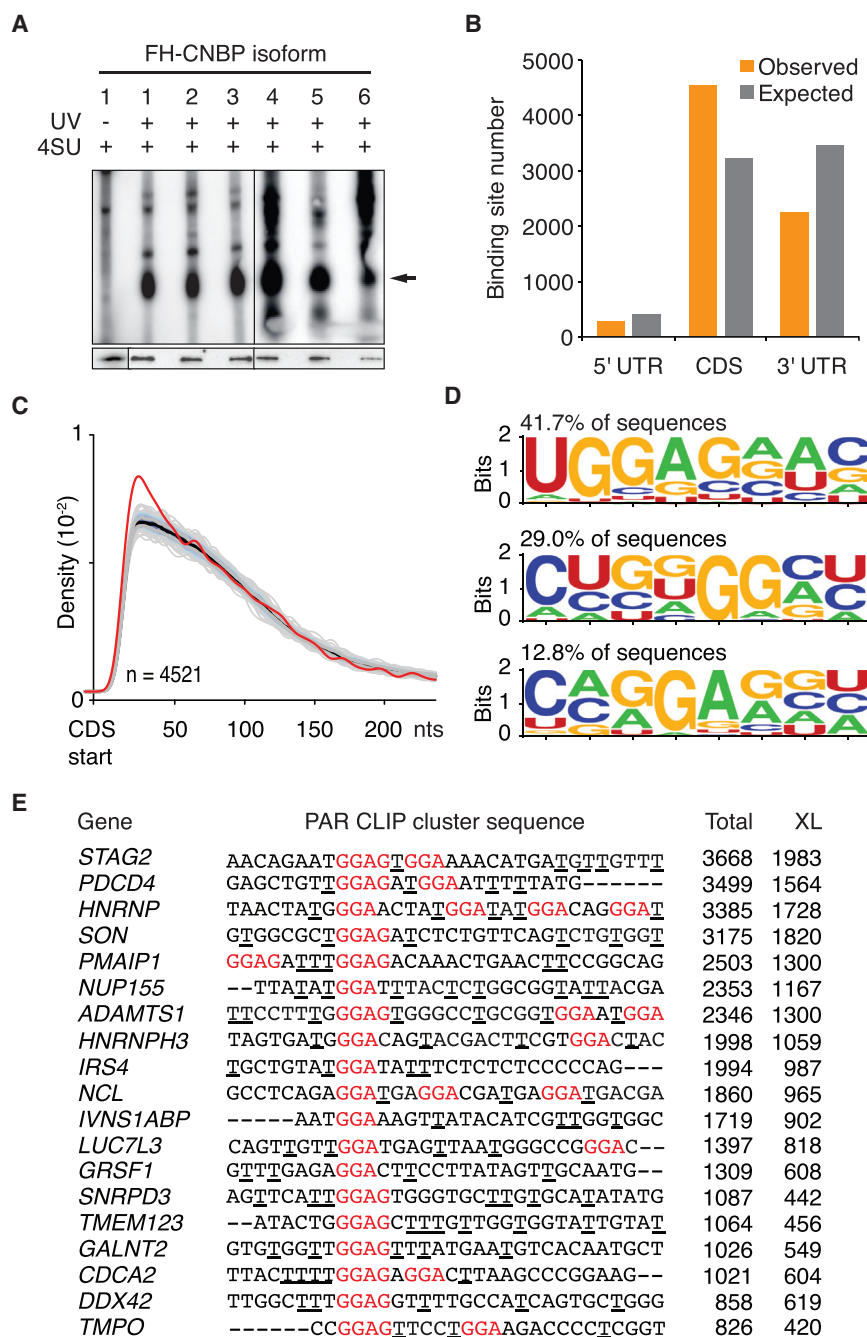
A

B

C

n = 4521

D

41.7% of sequences

29.0% of sequences

12.8% of sequences

Figure 2. CNBP PAR-CLIP

(A) Autoradiograph showing in vivo cross-linked CNBP-RNA RNPs from stably expressing FH-CNBP isoform 1–6 cell lines (arrow indicates crosslinked CNBP-RNPs). Cells not irradiated with UV served as control (lane 1). Western blot analysis of immunoprecipitated CNBP isoforms 1–6 probed with anti-FLAG antibody is shown in the lower panel.

(B) Analysis of target transcript preferences for CNBP. The number of exonic binding sites annotated as derived from the 5′ UTR, CDS, or 3′ UTR of a target transcript is shown (orange bars). Gray bars show the expected location distribution of clusters if CNBP bound without regional preference to the set of target transcripts.

(C) Density of CNBP binding sites (red line) downstream of the mRNA start codon, compared to 1,000 mismatched randomized controls (gray lines).

(D) Weblogo of representative RREs from CNBP PAR-CLIP binding sites generated by HOMER (Heinz et al., 2010). Percentage of CNBP binding sites containing the respective RRE is indicated.

(E) Alignment of top 19 PAR-CLIP clusters with G-rich RREs indicated in red. The number of total (Total) and crosslinked (XL) reads for each cluster is shown. Crosslinking sites determined by the diagnostic T-to-C mutation are underlined.

E

Gene	PAR CLIP cluster sequence	Total	XL
STAG2	AACAGAATGGAGTGGAAAACATGATGTTGTTT	3668	1983
PDCD4	GAGCTGTTGGAGATGGAATTTTTATG------	3499	1564
HNRNP	TAACTATGGGAACTATGGATATGGACAGGGAT	3385	1728
SON	GTGGCGCTGGAGATCTCTGTTCAGTCTGTGGT	3175	1820
PMAIP1	GGAGATTTGGAGACAAACTGAACTTCCGGCAG	2503	1300
NUP155	--TTATATGGATTTACTCTGGCGGTATTACGA	2353	1167
ADAMTS1	TTCCTTTGGGAGTGGGCCTGCGGTGGAATGGA	2346	1300
HNRNPH3	TAGTGATGGGACAGTACGACTTCGTGGACTAC	1998	1059
IRS4	TGCTGTATGGATATTTCTCTCTCCCCCAG---	1994	987
NCL	GCCTCAGAGGATGAGGACGATGAGGATGACGA	1860	965
IVNS1ABP	-----AATGGAAAGTTATACATCGTTGGTGGC	1719	902
LUC7L3	CAGTTGTTGGATGAGTTAATGGGCCGGGAC--	1397	818
GRSF1	GTTTGAGAGGACTTCCTTATAGTTGCAATG--	1309	608
SNRPD3	AGTTCATTGGAGTGGGTGCTTGTGCATATATG	1087	442
TMEM123	--ATACTGGGAGCTTTGTTGGTGGTATTGTAT	1064	456
GALNT2	GTGTGGTTGGAGTTTATGAATGTCACAATGCT	1026	549
CDCA2	TTACTTTTGGAGAGGACTTAAGCCCGGAAG--	1021	604
DDX42	TTGGCTTTGGAGGTTTTGCCATCAGTGCTGGG	858	619
TMPO	------CCGGAGTTCCTGGAAGACCCCTCGGT	826	420

adaptor proteins shuttling between cytoplasm and nucleus, resulting in changes of subcellular localization (Lee et al., 2012). Alternatively, arginine methylation could influence the nucleic acid-binding properties of the RG-rich domain and/or the neighboring ZnFs (Wei et al., 2014). Because previous reports suggested nuclear as well as cytoplasmic functions for CNBP, we investigated whether differences in the RG-rich domain could influence subcellular localization. We generated stable HEK293 cell lines expressing FLAG/HA (hemagglutinin)-tagged CNBP (FH-CNBP) isoforms 1–6 under control of a tetracycline-induc-

ible promoter (Spitzer et al., 2013). Subcellular fractionation from these six cell lines, as well as from the parental HEK293 cells, revealed that all CNBP isoforms mainly localize to the cytoplasm (Figure 1D). This result favors a predominantly posttranscriptional—rather than transcriptional—gene regulatory function for CNBP.

CNBP Interacts with Mature mRNAs in G-Rich Regions

Next, we investigated the RNA-binding properties of the six CNBP isoforms in living cells. We used 4-thiouridine (4SU) PAR-CLIP to irreversibly crosslink RNA and interacting proteins (Hafner et al., 2010b). Autoradiography of the crosslinked, ribonuclease-treated, immunoprecipitated, and radiolabeled FH-CNBP

isoform 1–6 ribonucleoprotein (RNP) complexes revealed one major band at the expected size of ∼20 kDa molecular mass (Figure 2A). This indicated that all CNBP isoforms maintain the ability to interact with RNA ligands in vivo.

We chose to characterize the targeting profile of one of the two most abundant CNBP isoforms in HEK293, isoform 3 (Figure 1B). We recovered the interacting RNA from the crosslinked FH-CNBP isoform 3 RNP and generated cDNA libraries for next-generation sequencing. Using PARalyzer software (Corcoran et al., 2011), we identified 8,420 high-confidence clusters
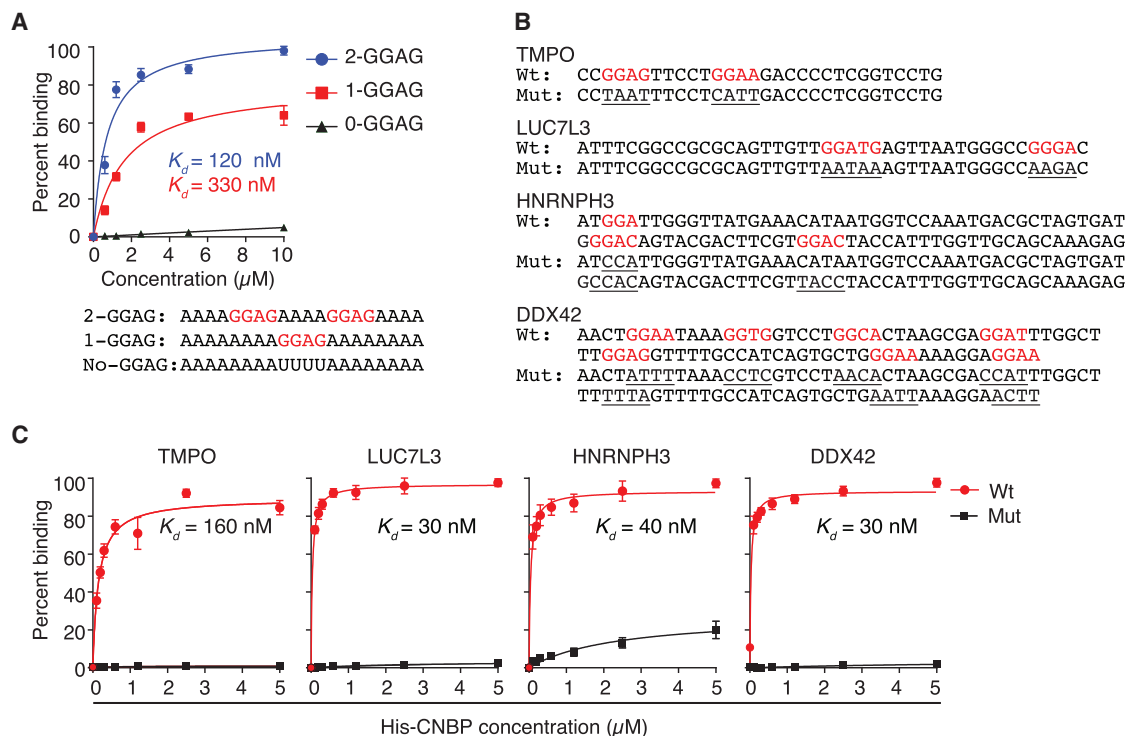
A

2-GGAG: AAAAGGAGAAAAGGAGAAAA
1-GGAG: AAAAAAAAGGAGAAAAAAAA
No-GGAG: AAAAAAAAUUUUAAAAAAAA

B

TMPO
Wt: CCGGAGTTCCTGGAAGACCCCTCGGTCCTG
Mut: CCTAATTTCCTCATTGACCCCTCGGTCCTG

LUC7L3
Wt: ATTTCGGCCGCGCAGTTGTTGGATGAGTTAATGGGCCGGGAC
Mut: ATTTCGGCCGCGCAGTTGTTAATAAAGTTAATGGGCCAAGAC

HNRNPH3
Wt: ATGGATTGGGTTATGAAACATAATGGTCCAAATGACGCTAGTGAT
     GGGACAGTACGACTTCGTGGACTACCATTTGGTTGCAGCAAAGAG
Mut: ATCCATTGGGTTATGAAACATAATGGTCCAAATGACGCTAGTGAT
     GCCACAGTACGACTTCGTTACCTACCATTTGGTTGCAGCAAAGAG

DDX42
Wt: AACTGGAATAAAGGTGGTCCTGGCACTAAGCGAGGATTTGGCT
     TTGGAGGTTTTGCCATCAGTGCTGGGAAAAAGGAGGAA
Mut: AACTATTTTAAACCTCGTCCTAACACTAAGCGACCATTTGGCT
     TTTTTAGTTTTGCCATCAGTGCTGAATTAAAGGAACTT

C

Figure 3. CNBP Interacts with G-Rich Sequence Elements In Vitro
(A) Binding curves and dissociation constants (K_Ds) obtained from filter-binding assay using recombinant CNBP with synthetic RNA targets containing no GGAG (black), one GGAG (red), and two GGAG (blue) sequence elements. Sequences are given below.
(B) Sequences of CNBP binding sites used for filter-binding assays. Nucleotides highlighted in red correspond to the predicted CNBP RREs. Mutations of the putative RRE are underlined.
(C) Binding curves and dissociation constants (K_Ds) obtained from filter-binding assays using recombinant CNBP and sequences from (B).

(binding sites), of which 7,545 distributed over 4,178 mRNAs (Figure S1B; Table S1). Consistent with its cytoplasmic localization, CNBP binding sites were predominantly mapping to mature mRNAs, with 54% of the clusters in coding sequences (CDSs) and 27% in the 3′ UTR (Figure 2B). A metagene analysis across different mRNA features revealed an enrichment of CNBP binding sites within the first 50 nt of the start codon (Figures 2C and S1C–S1G).

We used HOMER (Heinz et al., 2010) to computationally define the preferred RNA recognition element (RRE) on the complete set of PAR-CLIP-defined CNBP binding sites/clusters. This approach identified UGGAGNW as the most common RRE (> 40% of all binding sites). Other G-rich motifs containing a GGA or GGG core also showed strong enrichment in the computational analysis and often occurred multiple times in a single CNBP binding site (Figures 2D and 2E).

CNBP Recognizes G-Rich Sequences In Vitro
We first used electrophoretic mobility shift assays (EMSAs) with recombinant, purified CNBP isoform 3 to evaluate the impact of the putative GGAG containing RRE on the binding affinity. We designed synthetic 20-nt-long ssRNAs with adenosines flanking a centered UUUU, GGAG, or 2xGGAG element (sequences are shown in Figure 3A). The UUUU-containing RNA did not show any appreciable interaction with CNBP. The addition of one

GGAG resulted in a weak shift at high protein concentrations, and the addition of a second GGAG resulted in a complete shift of oligoribonucleotide at CNBP concentrations >0.5 μM (Figure S2A). Accordingly, based on filter-binding assays, CNBP binding constants (dissociation constant, K_D) of UUUU, GGAG, and 2xGGAG oligoribonucleotides were >10 μM, 330 nM, or 120 nM, respectively, confirming that a GGAG motif confers binding to CNBP (Figure 3A). Hill coefficients for these RNA substrates varied around 1, indicating that CNBP binding was noncooperative (Figure S2C).

Next, we validated the binding of CNBP to PAR-CLIP binding sites in vitro. We generated ssRNAs representing four different binding sites (Figure 3B) and quantified their affinity to CNBP by filter binding. The obtained K_D varied between 30 and 160 nM, and mutations in the core G-rich binding motif resulted in loss of CNBP binding (Figure 3C). Finally, we attempted to dissect the contribution of the RG-rich domain and individual CCHC domains using three recombinantly expressed CNBP constructs either lacking the RG-rich domain, the N-terminal CCHC domain, or the two C-terminal CCHC domains. These constructs all had severely reduced binding affinities compared to full-length CNBP, and we concluded that, in addition to the ZnF domains, CNBP also requires the RG-rich domain for high-affinity RNA binding (Figures S2C–S2G). In summary, our in vitro studies validated the PAR-CLIP-derived CNBP binding

sites and emphasized the requirement of G-rich sequences for high-affinity binding.

CNBP Presence Slightly Reduces Target mRNA Abundance

To determine the effect of CNBP on its mRNA targets and to perform loss-of-function studies, we generated CNBP knockout (KO) HEK293 cell lines using Cas9-mediated gene editing. Genomic sequencing of the CNBP locus, as well as RNA-seq of three different CNBP KO clones revealed that all contained frameshift mutations in exon 4, which resulted in loss of detectable protein by western blot (Figures S3A–S3C). CNBP KO did not significantly reduce the growth rate of HEK293 cells. However, when grown in continuous (unpassaged) culture, CNBP KO cells started collapsing after 8–10 days of growth and were completely dead after 15 days, while parental HEK293 readily survived (Figures S3D–S3F). This suggests that CNBP supports cell survival under limiting conditions.

We used these CNBP KO cells to determine whether CNBP is involved in cytoplasmic posttranscriptional gene regulatory processes. We first investigated CNBP target mRNA abundance by RNA-seq (Figure S4A; Table S2). Loss of CNBP led to a marginal, albeit statistically significant increase of target mRNA levels compared to parental HEK293 cells. CNBP PAR-CLIP top targets, binned according to the number of crosslinked reads, increased in abundance by ~2% (Figure 4A). These results indicate that CNBP is most likely not influencing cytoplasmic mRNA turnover.

CNBP Promotes Translation of Its Targets

Given the almost negligible effect on mRNA abundance, we next investigated whether CNBP controls other cytoplasmic post-transcriptional gene regulatory processes beyond mRNA stability. We showed, as described earlier, that CNBP binding was enriched in mRNA CDSs, particularly within 50 nt downstream of the start codon (Figures 2B and 2C). We also demonstrated that CNBP co-sediments quantitatively with the 40S, 80S, and polysomal fractions in polysome profiling experiments (Figure 4B). Taken together, these data pointed to a role for CNBP in translation, and we hypothesized that CNBP might influence translational elongation across G-rich regions.

To directly measure the impact of CNBP on the ribosome occupancy of its mRNA targets, we performed ribosome footprinting (Ribo-seq) and quantified ribosome-protected fragments (RPFs) (Ingolia et al., 2009) in CNBP KO cells and parental HEK293 cells (Table S3). CNBP loss resulted in a robust decrease in RPF numbers on target mRNAs, and this effect depended on the number of crosslinked reads or the number of CNBP binding sites (Figures 4C and S4B), both of which we previously found to correlate well with the occupancy of the RBP on the RNA (Hafner et al., 2010b; Ascano et al., 2012b). We found an ~30% decrease (p < 2.2 × 10^{-16}) in RPFs on the 243 target mRNAs with more than 500 crosslinked reads, compared to mRNAs showing no interaction with CNBP (Figure 4C). Even for the 1,761 targets with only a single CNBP site, RPFs decreased by ~9% (p = 13 × 10^{-10}).

We further investigated the effect of the CNBP binding site location on ribosome density and first focused on mRNAs that

were targeted by CNBP in the 3′UTR. CNBP KO affected neither the RPFs on these mRNAs nor their abundance, suggesting that 3′UTR binding did not contribute to CNBP mediated gene regulation (Figures S4C–S4E). In contrast, CNBP binding anywhere in the CDS was sufficient to increase RPF density, with no preference detectable for sites close to the translational start, despite the enrichment of CNBP binding sites in that region (Figure S4F).

We calculated the density of ribosomes on each mRNA in CNBP KO and control cells by normalizing the number of RPFs with the mRNA abundance. This score, known as the translational efficiency (TE), removes effects of mRNA abundance and approximates the translational output for each mRNA molecule of a given gene (Ingolia et al., 2009; Guo et al., 2010; Bazzini et al., 2012). CNBP KO strongly correlated with decreased TE on CNBP targets (~33% decrease for the 243 top CNBP targets binned by intensity of crosslinking; Figure 4D; p < 2.2 × 10^{-16}).

G-rich sequences are capable of forming G4 and other secondary structures, which act as roadblocks for efficient translation (Rhodes and Lipps, 2015; Guo and Bartel, 2016). We thus reasoned that CNBP possibly facilitated translation by relieving restrictive mRNA structures; elongation rates of ribosomes around CNBP binding sites would, in this case, decrease upon CNBP loss. To test this hypothesis, we plotted RPF density around CNBP binding sites (Figure 4E). Ribosome density around the G-rich CNBP sites was increased in control cells, suggesting reduced elongation rates, even in the presence of CNBP. Upon CNBP loss, we observed a further 60% increase in ribosome density followed by a sharp drop in occupancy >200 nt downstream of the binding sites. Thus, our data suggest that CNBP loss resulted in increased stalling of elongating ribosomes around its G-rich binding sites. The local increase of ribosome density upstream of the CNBP binding sites was compatible with the overall decrease of ribosome footprints on CNBP target mRNAs, considering that CNBP binding sites were predominantly localized 50 nt of the start codon (Figure 2C).

CNBP Increases Protein Levels of Target Transcript

CNBP KO cells showed a ~20% decrease in protein content per cell, supporting a role in promoting target mRNA translation (Figure 5A). Nevertheless, such a decrease may also be the result of indirect regulatory effects. To directly test the effect of CNBP binding on the target mRNA translational output, we designed reporter assays using representative PAR-CLIP binding sites from the TMPO, LUC7L3, and DDX42 mRNAs, which showed a reduction in the TE of 40%, 40%, and 50%, respectively. Each of these binding sites was documented to bind CNBP in vitro and was located within less than 50 nt from the start codon in the CDS (Figures 3B and 3C). The sequences were cloned in frame with the firefly luciferase mRNA directly downstream of the AUG codon in the psiCHECK-2 dual luciferase reporter assay system (Figure S5A). Transfection of these reporter plasmids into CNBP KO cells resulted in an approximately 10-fold decrease in both renilla control and firefly luciferase, compared to wild-type cells, while luciferase RNA levels stayed constant (Figure S5B). This reduction in luciferase protein levels was likely due to a concentration of G-rich sequence elements in the CDS of the luciferase genes, making them sensitive to complete loss of CNBP (Figure S5C). We argued that reduction of
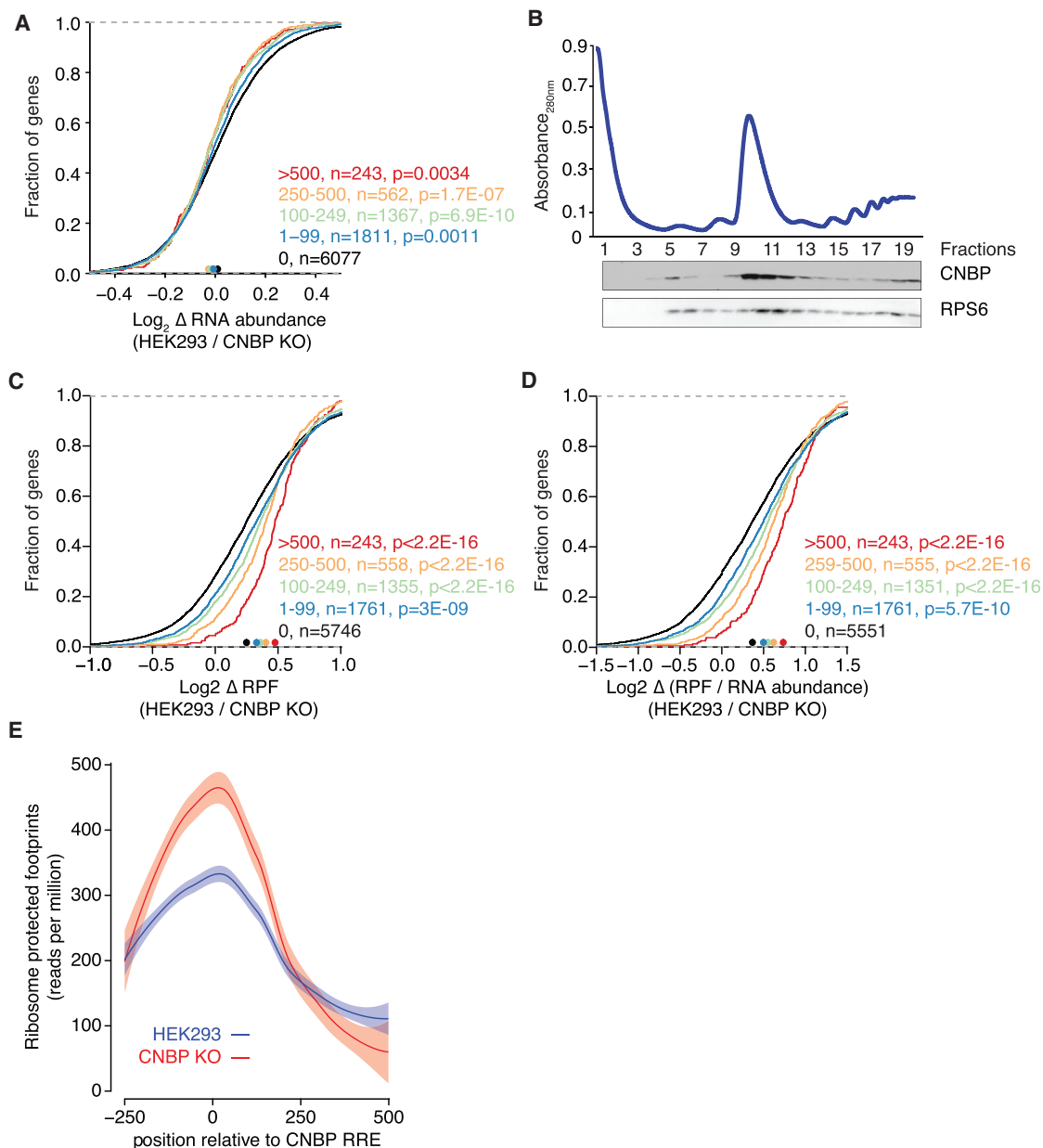
Figure 4. CNBP Increases the Ribosome Density on Its Targets

(A) Cumulative distribution analysis of change in average mRNA expression comparing CNBP KO cell lines (n = 3) with parental HEK293 cells (n = 3). Target mRNAs are binned based on the number of crosslinked reads. Significance was determined using a two-sided Kolmogorov-Smirnov (KS) test. Bin size is indicated.

(B) Sucrose gradient separation profile of HEK293 cell extract. The western blot below shows co-sedimentation of CNBP with free ribosomal subunits (fractions 5–8), monoribosomes (fractions 9–12), and polysomes (fractions 13–19). RPS6 from the 40S ribosomal subunit served as a control protein.

(C) Cumulative distribution analysis of change in ribosome-protected fragments (RPFs) upon CNBP knockout, as determined by ribosome profiling. Target mRNAs are binned based on the number of crosslinked reads. Significance was determined using a two-sided KS test. Bin size is indicated.

(D) Same as in (C), except the cumulative distribution of the translation efficiency (TE, calculated as $\log_2[\mathrm{RPF/RNA}$ abundance]) ratio is plotted.

(E) Distribution of RPF around CNBP binding sites in control (blue) and CNBP KO (red) cells. Curves were fitted using LOESS regression, and the envelope indicates a 95% confidence interval.

CNBP levels, rather than knockout, would preferentially affect the strongest CNBP targets, and we used a siRNA (small interfering RNA)-mediated gene knockdown for our luciferase experiments. Upon knockdown of CNBP, we observed a highly significant

(20%–40%) decrease in firefly luciferase levels for all reporters (Figures 5B and S5D). This decrease was rescued by mutations within the G-rich CNBP binding sites that we had previously shown to abrogate CNBP binding in vitro (Figure 3C).
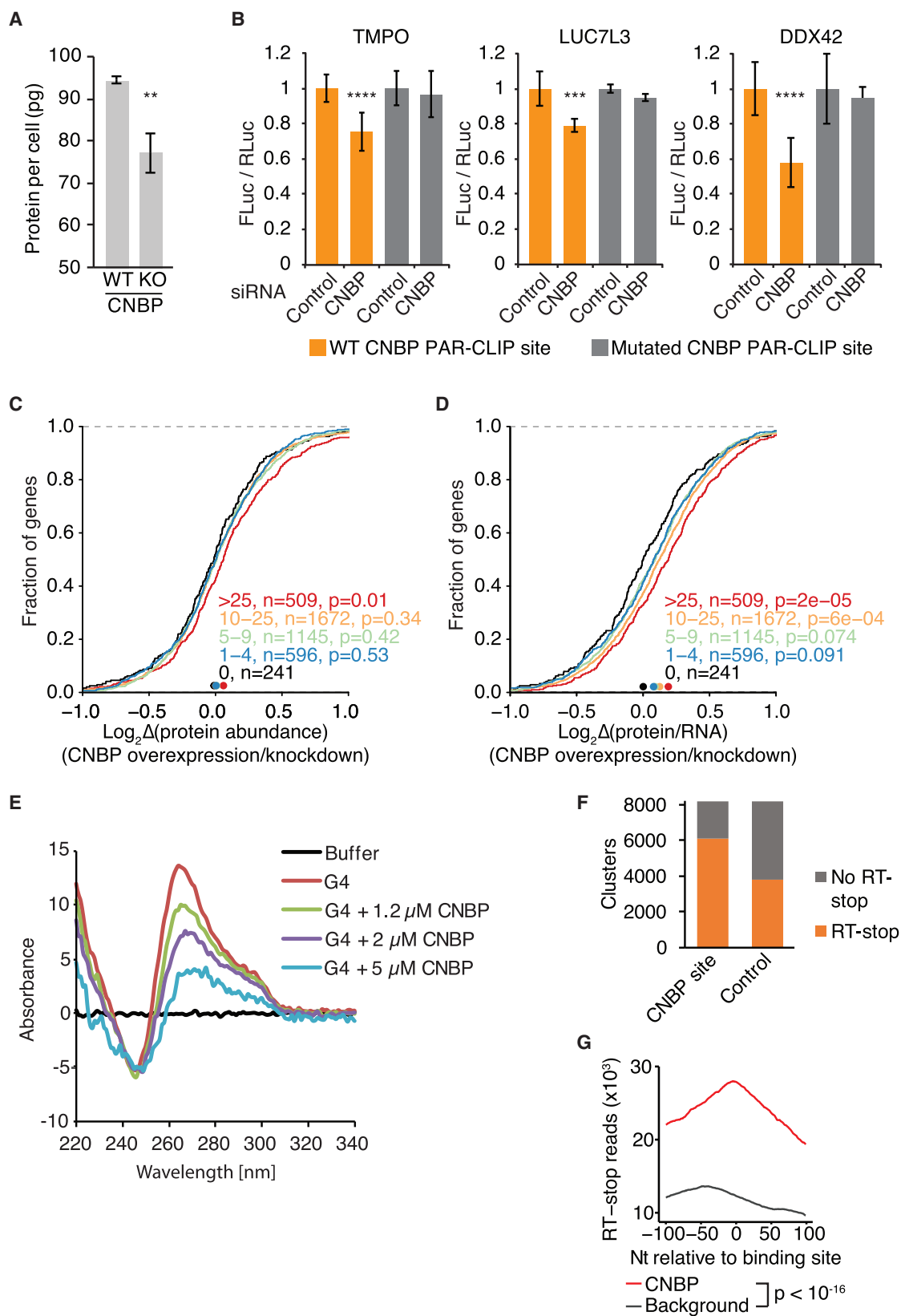
(legend on next page)

To systematically measure the impact of CNBP on target protein levels we performed pulsed stable isotope labeling by amino acids in cell culture (pSILAC) coupled with quantitative proteomics (Ong et al., 2003). We compared the protein expression profile of FH-CNBP expressing HEK293 cells with that of HEK293 cells after CNBP knockdown and could quantify 4,887 proteins (Table S5). Knockdown of CNBP resulted in a small but significant decrease in the abundance of proteins encoded by the top 509 CNBP target mRNAs (Figure 5C). We further normalized the observed changes in protein levels according to changes in mRNA levels (obtained by RNA-seq) for each gene upon CNBP knockdown and overexpression. Consistent with our reporter assays and ribosome-profiling data, we found that CNBP expression directly correlated with increased translational efficiency on CNBP targets (14% increase for the top 509 CNBP targets, p = 2 × 10^{-5}; Figure 5D).

CNBP Binds Sites Close to G4 and Other Stable RNA Structures

Finally, we asked whether a plausible mechanism for CNBP function would be to promote translation across G-rich elements in the target mRNA CDSs that have the potential to form stable secondary RNA structures. We used SHAPE chemistry to probe for structure in the CNBP PAR-CLIP sites from TMPO and HNRNPH3 and found that the G-rich elements in that oligoribonucleotide were protected from modification in vitro, indicating that they were paired (Figures S5E and S5F). We next examined whether CNBP binding could disrupt formation of RNA secondary structures and performed circular dichroism measurements using a well-characterized G4-forming oligonucleotide (Paeschke et al., 2013) in the presence and absence of recombinant CNBP. The presence of CNBP prevented the formation of the G4 in a concentration-dependent manner (Figure 5E), suggesting that, indeed, it was capable of resolving stable secondary structure. We extended these analyses on a transcriptome-wide scale using data from a recent report, which identified regions in HEK293 cells forming stable structures in vitro (Guo and Bartel, 2016). Guo and Bartel exploited the ability of G4-forming and other structured regions in RNA to stall reverse transcription during cDNA library preparation. Interestingly, 75% of CNBP binding sites overlapped with sequences that stalled reverse transcription (Figure 5F). The localization of CNBP binding sites at and near regions capable of stalling reverse transcription was highly significant in comparison to a set of length-matched sites randomly selected from CDSs

of mRNAs expressed in HEK293 (Figure 5G). In summary, our data are consistent with a role for CNBP in preventing the formation of stable secondary structures in G-rich regions within the mRNA CDS and, thus, promoting translation.

DISCUSSION

Here, we present a comprehensive characterization of the CNBP protein. We identified target RNA-binding sites transcriptome-wide, delineated a consensus binding motif, and defined the effect of CNBP binding on mRNA abundance and translation.

CNBP was described as a multifunctional protein similar to other ZnF-containing proteins (Pace and Weerapana 2014; Hall 2005). It was suggested to interact with both DNA and RNA, leading to conflicting reports claiming a function as a transcriptional as well as post-transcriptional gene regulator (Borgognone et al., 2010; Calcaterra et al., 1999; Sammons et al., 2011; Shimizu et al., 2003; Michelotti et al., 1995). In human cell lines, CNBP encodes at least six protein isoforms, which contain differences in the third CCHC domain as well as in the RG-rich domain conceivably involved in nucleocytoplasmic shuttling. However, our biochemical analyses showed that CNBP could only be detected in the cytoplasm, similar to its homolog in *Xenopus laevis* (Calcaterra et al., 1999). Combined with our finding that PAR-CLIP binding sites mainly localized to exonic mRNA regions and that all six isoforms interacted with RNA in vivo, these data suggest that CNBP is involved in cytoplasmic post-transcriptional gene regulatory processes rather than acting as a transcription factor.

ZnF proteins show diverse binding sequence specificity. Recent structural work for the CCCH-type ZnF of the yeast Nab2 and the *Drosophila melanogaster* Unkempt proteins showed their strong preference for adenine and uridine bases, respectively (Kuhlmann et al., 2014; Murn et al., 2016). In contrast, the two CCHC domains of the LIN28 protein family, which are homologous to CNBP's seven CCHC ZnFs, interact in vitro and in vivo with a GGAG motif (Nam et al., 2011; Wilbert et al., 2012; Piskounova et al., 2008; Graf et al., 2013). This motif is mirrored by the RREs of CNBP found by in vitro selection (Ray et al., 2016), as well as by PAR-CLIP in this study. RNA-protein crosslinking in 4SU-PAR-CLIP, HiTS-CLIP, and other CLIP-seq procedures occurs predominantly at uridines (Kramer et al., 2014; Sharma et al., 2015) and, therefore, requires the presence of uridine bases within a few nucleotides of the binding site. Our unbiased motif enrichment analysis revealed CNBP's G-rich RRE,

---

Figure 5. CNBP Promotes Translation of mRNA Targets
(A) Absolute protein abundance per cell in control (WT) and CNBP KO (KO) cells.
(B) Effect of CNBP knockdown on luciferase reporter gene expression. Firefly luciferase (FLuc) expression is normalized to Renilla luciferase (RLuc) expression and set to 1 for the control knockdown. Results of paired t test are indicated (***p < 0.0006; ****p < 0.0001).
(C) Cumulative distribution analysis of protein level changes upon CNBP knockdown as determined by pSILAC. Significance was determined using a two-sided KS test. Bin size is indicated.
(D) Same as in (C), except the cumulative distribution of a protein (derived by pSILAC) per mRNA (derived by RNA-seq) metric (log$_2$[protein abundance/RNA abundance]) ratio is plotted.
(E) CD measurement of a G4 structure without CNBP and with increasing concentrations of CNBP.
(F) Number of CNBP binding sites overlapping RT-stop reads from Guo and Bartel (2016) compared to a set of background sequence clusters. Background sequences for each CNBP binding site were randomly selected from the same mRNA and the same annotation category (CDS, UTR).
(G) Distribution of RT-stop reads 100 nt up- and downstream of CNBP PAR-CLIP sites (red line, anchored at cluster start coordinate). Distribution of RT-stop reads relative to the background sequence set from (F) (black line).

mitigating concerns that results from UV-crosslinking-based protocols are disproportionately skewed toward U-rich RREs.

While CNBP loss leads to a marginal accumulation of target mRNAs, their translation and protein output—as measured by Ribo-seq and mass spectrometry—is strongly reduced. The TE of the 243 top target mRNAs dropped, on average, by 30%, and that of all ∼4,000 targeted mRNAs dropped by 10%, on average, upon CNBP loss. Interestingly, cells from myotonic dystrophy type 2 (DM2) patients with CCTG expansion in the first intron of *CNBP* have reduced CNBP expression and also show an overall reduced rate of translation, similar to our CNBP KO cells (Huichalaf et al., 2009). Taken together, this represents a massive shift in gene regulation of the large group of target mRNAs, which can explain the embryonic lethality of CNBP knockouts (Chen et al., 2003; Shimizu et al., 2003). Previously, only a handful of RBPs with otherwise well-characterized roles in mRNA processing were implicated as ITAFs, possibly aiding the proper folding of individual IRES elements and thereby promoting translation (Holcik and Sonenberg 2005). In contrast, CNBP represents an RBP enhancing the translation of its target mRNAs on a global scale. We hypothesize that CNBP acts by relieving secondary structures on target mRNAs that exhibit G-rich sequence stretches. Secondary structures are often required for RNA function (Mortimer et al., 2014; Wan et al., 2011), and among them, the G4 structure exhibits particular stability in vitro (Bochman et al., 2012; Millevoi et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases eIF4A and DHX36 in the Aven complex were found to be necessary for the efficient translation of a subset of RNAs with potential G4 structures in the oncogenic gene expression program (Wolfe et al., 2014; Thandapani et al., 2015). A recent survey that mapped G4 and other stable structures on a transcriptome-wide scale found that they are widespread and stable in vitro (Guo and Bartel, 2016). However, they remain unfolded in vivo in eukaryotic cells, suggesting the existence of a specialized machinery supporting resolution of such RNA structures in eukaryotes (Guo and Bartel, 2016). Guo and Bartel excluded the involvement of DHX36 and, using ATP depletion experiments, demonstrated that this machinery likely operates in an ATP-independent manner, precluding helicases altogether. They hypothesized that tight binding by RNA-binding proteins may help to prevent formation of structures. Considering the deep conservation of CNBP in eukaryotes, its absence in prokaryotes, and the inability for G4 structures to form in the presence of CNBP in vitro, we hypothesize that CNBP may represent one of these factors, helping to overcome the deleterious effects of G-rich structures on translation. Our work provides a starting point for the study of CNBP in translational regulation and represents a comprehensive resource for the study of individual target RNAs regulated by CNBP.

EXPERIMENTAL PROCEDURES

A detailed description of the experimental procedures can be found in the Supplemental Information.

Stable Cell Lines and Their Culture

Stable Flp-In T-REx HEK293 cell lines inducibly expressing FH-CNBP isoforms 1–6 were generated using the Gateway Recombination Cloning technology, as described previously (Spitzer et al., 2013; Walhout et al., 2000), and using the pFRT/TO/FLAG/HA-Dest plasmid (Landthaler et al., 2008).

siRNAs were transfected at a concentration of 50 nM using Lipofectamine RNAiMAX Reagent (Life Technologies);

CNBP siRNA-1: GCAAGGAGCCCAAGAGAGAUU, and
CNBP siRNA-2: CAAGAGAGAGCGAGAGCAAUU.

Cell Fractionation

Subcellular fractionation of FH-CNBP isoform 1–6 HEK293 cells was performed as described by Gagnon et al. (2014) unless otherwise stated.

PAR-CLIP

PAR-CLIP was performed as previously described (Hafner et al., 2010a, 2010b, 2012a), with minor modifications (see Supplemental Experimental Procedures for the complete procedure). PAR-CLIP cDNA libraries were sequenced on an Illumina HiSeq 2500 platform. Clusters of overlapping sequence reads mapped against the human genome version hg19 were generated using the PARalyzer software (Corcoran et al., 2011) incorporated into a pipeline (PARpipe; https://ohlerlab.mdc-berlin.de/software/PARpipe_119/) with default settings. Binding sites were categorized using the Gencode GRCh37.p13 GTF annotation (gencode.v19.chr_patch_hapl_scaff.annotation.gtf; http://www.gencodegenes.org/releases/19.html).

Dimethyl sulfate sequencing (DMS-seq) data for reverse transcription (RT)-stop profiling in HEK293 cells were downloaded from the dataset by Guo and Bartel (2016) (GEO: GSE83617). Reads from a K+ sample were taken for the downstream analysis, as they represent the strongest RT-stops. RT-stop site distribution around CNBP binding clusters was created using SAMtools mpileup. As background clusters for each CNBP cluster, a 40-nt-long sequence was randomly selected within the same transcript and at the same functional region (5′ UTR, CDS, or 3′ UTR).

Recombinant Protein Expression, EMSAs, and Filter-Binding Assays

For the expression of recombinant protein, CNBP isoform 3 was cloned into pET-M11 and was transformed into *E. coli* BL21 DE3 (pLysS). CNBP was purified using standard His-tag affinity purification using nickel-nitrilotriacetic acid (Ni-NTA) agarose beads (QIAGEN), and CNBP was eluted from the beads with 250 mM imidazole.

For EMSAs, 5 nM radioactively labeled oligoribonucleotides (for sequences, see Figure 3B) were incubated with recombinant CNBP in binding buffer (see Supplemental Experimental Procedures) and fractionated on a native 0.4% agarose gel.

For filter-binding assays, radiolabeled oligoribonucleotides were incubated with recombinant CNBP, as described earlier and applied to a dot blot apparatus (BioRad). Unbound RNA was retained by nylon membrane, whereas the RNA-protein complex was retained on a nitrocellulose membrane. The bound and unbound spots were quantified using ImageQuant (GE Healthcare), and their relative K_Ds were calculated using Prism software (GraphPad).

CRISPR/Cas9 Gene Editing

Flp-In T-REx HEK293 cells were generated using CNBP CRISPR (clustered regularly interspaced palindromic repeat)/Cas9 KO plasmids (sc-404090). KO clones were screened by immunoblotting with anti-CNBP antibody (Sigma SAB2100453). Knockout was confirmed by RNA-seq analysis.

RNA-Seq

RNA from three different clones of CNBP KO cells and three biological replicates of the parental Flp-In T-REx HEK293 was isolated using TRIzol reagent according to the manufacturer's instructions. rRNA was removed using the Ribo-Zero rRNA Removal Kit (Illumina), followed by directional cDNA library preparation (New England Biolabs). Sequenced reads were aligned to the human genome version hg19 using TopHat (Trapnell et al., 2012). Cufflinks and Cuffdiff were used to quantify transcripts and determine differential expression (Trapnell et al., 2012).

Ribo-Seq

Ribo-seq was performed as described in Ingolia et al. (2012), from single 10-cm plates of Flp-In T-REx HEK293 (either wild-type [WT] or CNBP KO).

Small RNA cDNA libraries from ribosome footprints were constructed as in Hafner et al. (2012b) and sequenced on an Illumina HiSeq 2500 platform. Sequenced reads were aligned to the human genome version hg19 using TopHat (Trapnell et al., 2012). RNAcounter was used to quantify transcript-mapped footprints. Overlaps of CNBP cluster and different genomic regions were calculated with BEDTools (Quinlan and Hall, 2010). Data extraction, formatting, subgrouping, and preparation for analysis were all facilitated with customized scripts for Bash, Python, Java, and R.

Statistical Methods

Equality of distributions in the empirical cumulative distribution functions for changes in transcript abundance (Figure 4A) and changes in translational efficiency (Figures 4C, 4D, 5C, and 5D) was tested using the two-sample Kolmogorov-Smirnov test (ks.test command in the R statistical software package). Statistical significance of differences in luciferase reporter gene expression upon CNBP knockdowns (Figure 5B) was assessed using a two-tailed t test using the GraphPad Prism software.

Luciferase Reporter Assay

HEK293 cells were cultured at a density of 60%–80% in 12-well plates and transfected with CNBP or control siRNA using Lipofectamine RNAiMax (Life Technologies). 72 hr post-transfection, cells were transfected with a dual-luciferase plasmid (psiCHECK-2) containing WT or mutant binding sequences (Figures S7A and S7B) using Lipofectamine 2000 (Life Technologies). 6 hr later, cell lysates were generated using Promega Passive Lysis Buffer, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) on a TriStar LB941 luminometer (Berthold Technologies).

Selective 2′-Hydroxyl Acetylation Analyzed by Primer Extension

Selective 2′-hydroxyl acetylation analyzed by primer extension (SHAPE) was performed as previously described (Wilkinson et al., 2006), with minor modifications as described in the Supplemental Experimental Procedures.

CD Measurement

Oligonucleotides forming G4 structures were incubated with increasing concentrations of recombinant CNBP isoform 3, and CD measurement was performed using a Jasco J-815 spectropolarimeter with readings that were recorded over a wavelength range of 200–350 nm. As control, a mutant sequence incapable of forming G4 structure was used (G4: AAAAAAAAAAGG GGGAGCTGGGGTAGATGGGAATGTGAGGG; control: AAAAAAAAAAAGCGC GAGCTGCGCTAGATGCGAATGTGAGCG).

ACCESSION NUMBERS

The accession numbers for the PAR-CLIP and RNA-seq sequence data reported in this paper are NCBI GEO: GSE76193 and GSE76627, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.02.080.

AUTHOR CONTRIBUTIONS

Conceptualization, M.H. and S.A.J.; Methodology, S.K.G., D.B., C.H.D., K.P., M.H., and S.A.J.; Software, S.G. and H.-W.S.; Investigation, S.K.G., D.B., C.H.D., and H.G.K.; Formal Analysis, D.B., S.G., H.-W.S., and M.H.; Writing – Original Draft, M.H. and S.A.J.; Writing – Review & Editing, D.B., M.H., S.A.J.; Funding Acquisition, M.H. and S.A.J.; Resources, M.H. and S.A.J.; Supervision, M.H. and S.A.J.

ACKNOWLEDGMENTS

REFERENCES

Armas, P., Nasif, S., and Calcaterra, N.B. (2008). Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. J. Cell. Biochem. 103, 1013–1036.

Ascano, M., Jr., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012a). Identification of RNA-protein interaction networks using PAR-CLIP. Wiley Interdisc. Rev. RNA 3, 159–177.

Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012b). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature 492, 382–386.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell 46, 674–690.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science 336, 233–237.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 13, 770–780.

Borgognone, M., Armas, P., and Calcaterra, N.B. (2010). Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. Biochem. J. 428, 491–498.

Calcaterra, N.B., Palatnik, J.F., Bustos, D.M., Arranz, S.E., and Cabada, M.O. (1999). Identification of mRNA-binding proteins during development: characterization of Bufo arenarum cellular nucleic acid binding protein. Dev. Growth Differ. 41, 183–191.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149, 1393–1406.

Chen, W., Liang, Y., Deng, W., Shimizu, K., Ashique, A.M., Li, E., and Li, Y.-P. (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. Development 130, 1367–1379.

Chen, S., Su, L., Qiu, J., Xiao, N., Lin, J., Tan, J.-H., Ou, T.-M., Gu, L.-Q., Huang, Z.-S., and Li, D. (2013). Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. Biochim. Biophys. Acta 1830, 4769–4777.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. 12, R79.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucleic Acids Res. 38, D211–D222.

Gagnon, K.T., Li, L., Janowski, B.A., and Corey, D.R. (2014). Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. Nat. Protoc. 9, 2045–2060.

Gerbasi, V.R., and Link, A.J. (2007). The myotonic dystrophy type 2 protein ZNF9 is part of an ITAF complex that promotes cap-independent translation. Mol. Cell. Proteomics 6, 1049–1058.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Graf, R., Munschauer, M., Mastrobuoni, G., Mayr, F., Heinemann, U., Kempa, S., Rajewsky, N., and Landthaler, M. (2013). Identification of LIN28B-bound mRNAs reveals features of target recognition and regulation. RNA Biol. *10*, 1146–1159.

Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science *353*, aaf5371.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M., et al. (2010a). PAR-CliP–a method to identify transcriptome-wide the binding sites of RNA binding proteins. J. Vis. Exp. (*41*), 2034.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M., et al. (2010b). Transcriptome-wide identification of RNA-binding protein and micro-RNA target sites by PAR-CLIP. Cell *141*, 129–141.

Hafner, M., Lianoglou, S., Tuschl, T., and Betel, D. (2012a). Genome-wide identification of miRNA targets by PAR-CLIP. Methods *58*, 94–105.

Hafner, M., Renwick, N., Farazi, T.A., Mihailović, A., Pena, J.T.G., and Tuschl, T. (2012b). Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. Methods *58*, 164–170.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. *15*, 367–373.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat. Rev. Mol. Cell Biol. *6*, 318–327.

Huichalaf, C., Schoser, B., Schneider-Gold, C., Jin, B., Sarkar, P., and Timchenko, L. (2009). Reduction of the rate of protein translation in patients with myotonic dystrophy 2. J. Neurosci. *29*, 9042–9049.

Iadevaia, V., Caldarola, S., Tino, E., Amaldi, F., and Loreni, F. (2008). All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5′-terminal oligopyrimidine (TOP) mRNAs. RNA *14*, 1730–1736.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat. Protoc. *7*, 1534–1550.

Klug, A., and Rhodes, D. (1987). Zinc fingers: a novel protein fold for nucleic acid recognition. Cold Spring Harb. Symp. Quant. Biol. *52*, 473–482.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kuhlmann, S.I., Valkov, E., and Stewart, M. (2014). Structural basis for the molecular recognition of polyadenosine RNA by Nab2 Zn fingers. Nucleic Acids Res. *42*, 672–680.

Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Methods *13*, 841–844.

Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA *14*, 2580–2596.

Lee, Y.-J., Hsieh, W.-Y., Chen, L.-Y., and Li, C. (2012). Protein arginine methylation of SERBP1 by protein arginine methyltransferase 1 affects cytoplasmic/nuclear distribution. J. Cell. Biochem. *113*, 2721–2728.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864–867.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

McBride, A.E., Conboy, A.K., Brown, S.P., Ariyachet, C., and Rutledge, K.L. (2009). Specific sequences within arginine-glycine-rich domains affect mRNA-binding protein function. Nucleic Acids Res. *37*, 4322–4330.

Michelotti, E.F., Tomonaga, T., Krutzsch, H., and Levens, D. (1995). Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene. J. Biol. Chem. *270*, 9494–9499.

Millevoi, S., Moine, H., and Vagner, S. (2012). G-quadruplexes in RNA biology. Wiley Interdiscip. Rev. RNA *3*, 495–507.

Mortimer, S.A., Kidwell, M.A., and Doudna, J.A. (2014). Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genet. *15*, 469–479.

Murn, J., Teplova, M., Zarnack, K., Shi, Y., and Patel, D.J. (2016). Recognition of distinct RNA motifs by the clustered CCCH zinc fingers of neuronal protein Unkempt. Nat. Struct. Mol. Biol. *23*, 16–23.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J., and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. Cell *147*, 1080–1091.

Ong, S.-E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. Methods *29*, 124–130.

Pace, N.J., and Weerapana, E. (2014). Zinc-binding cysteines: diverse functions and structural motifs. Biomolecules *4*, 419–434.

Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C., and Zakian, V.A. (2013). Pif1 family helicases suppress genome instability at G-quadruplex motifs. Nature *497*, 458–462.

Pellizzoni, L., Lotti, F., Maras, B., and Pierandrei-Amaldi, P. (1997). Cellular nucleic acid binding protein binds a conserved region of the 5′ UTR of Xenopus laevis ribosomal protein mRNAs. J. Mol. Biol. *267*, 264–275.

Piskounova, E., Viswanathan, S.R., Janas, M., LaPierre, R.J., Daley, G.Q., Sliz, P., and Gregory, R.I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J. Biol. Chem. *283*, 21310–21314.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rajavashisth, T.B., Taylor, A.K., Andalibi, A., Svenson, K.L., and Lusis, A.J. (1989). Identification of a zinc finger protein that binds to the sterol regulatory element. Science *245*, 640–643.

Ray, D., Ha, K.C., Nie, K., Zheng, H., Hughes, T.R., and Morris, Q.D. (2016). RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. Methods *16*, 30472–30478.

Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. *43*, 8627–8637.

Rojas, M., Farr, G.W., Fernandez, C.F., Lauden, L., McCormack, J.C., and Wolin, S.L. (2012). Yeast Gis2 and its human ortholog CNBP are novel components of stress-induced RNP granules. PLoS ONE *7*, e52824.

Sammons, M.A., Samir, P., and Link, A.J. (2011). Saccharomyces cerevisiae Gis2 interacts with the translation machinery and is orthogonal to myotonic dystrophy type 2 protein ZNF9. Biochem. Biophys. Res. Commun. *406*, 13–19.

Sharma, K., Hrle, A., Kramer, K., Sachsenberg, T., Staals, R.H.J., Randau, L., Marchfelder, A., van der Oost, J., Kohlbacher, O., Conti, E., et al. (2015). Analysis of protein-RNA interactions by UV-induced cross-linking and mass spectrometry. Methods *89*, 138–148.

Shimizu, K., Chen, W., Ashique, A.M., Moroi, R., and Li, Y.-P. (2003). Molecular cloning, developmental expression, promoter analysis and functional characterization of the mouse CNBP gene. Gene *307*, 51–62.

Spitzer, J., Landthaler, M., and Tuschl, T. (2013). Rapid creation of stable mammalian cell lines for regulated expression of proteins using the Gateway®

recombination cloning technology and Flp-In T-REx® lines. Methods Enzymol. *529*, 99–124.

Thandapani, P., Song, J., Gandin, V., Cai, Y., Rouleau, S.G., Garant, J.-M., Boisvert, F.-M., Yu, Z., Perreault, J.-P., Topisirovic, I., and Richard, S. (2015). Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. eLife *4*, e06234.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol. *328*, 575–592.

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., and Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. Nat. Rev. Genet. *12*, 641–655.

Wei, H.-M., Hu, H.-H., Chang, G.-Y., Lee, Y.-J., Li, Y.-C., Chang, H.-H., and Li, C. (2014). Arginine methylation of the cellular nucleic acid binding protein does not affect its subcellular localization but impedes RNA binding. FEBS Lett. *588*, 1542–1548.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol. Cell *48*, 195–206.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1*, 1610–1616.

Wilson, D., Charoensawan, V., Kummerfeld, S.K., and Teichmann, S.A. (2008). DBD–taxonomically broad transcription factor predictions: new content and functionality. Nucleic Acids Res. *36*, D88–D92.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature *513*, 65–70.