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

Highlights
- CNBP is a cytoplasmic RNA-binding protein binding mature mRNAs at G-rich elements.
- CNBP binds at sites forming G-quadruplex and other stable structures in vitro.
- CNBP prevents formation of G-quadruplex structures in vitro.
- CNBP increases translational efficiency on its target mRNAs on a global scale.

Authors
Daniel Benhalevy, Sanjay K. Gupta, Charles H. Danan, ..., Katrin Paeschke, Markus Hafner, Stefan A. Juranek

Correspondence
markus.hafner@nih.gov (M.H.), s.a.juranek@umcg.nl (S.A.J.)

In Brief
Benhalevy et al. characterize the RNA-binding protein CNBP/ZNF9 using systems-wide approaches. They find that CNBP preferentially binds at mRNA regions previously found to form G-quadruplex and other structures in vitro. Ribosome profiling revealed that CNBP enhances translation across these sites, which potentially form roadblocks for the ribosome.

Accession Numbers
GSE76193
GSE76627
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. Juranek2,4,*

1Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD 20892, USA
2Department of Biochemistry, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany
3Biostatistics and Data Mining Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
4European Research Institute for the Biology of Ageing (ERIBA), University Medical Center Groningen, University of Groningen, 9713 AV Groningen, the Netherlands
5Co-first author
6Lead 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 evolutionarily 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 GTGCAGTG. A role for CNBP in transcriptional regulation was further suggested by its ability to bind the CT element in the promoter of the c-myc gene in vitro and activate it after
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 CNBP isoforms (iso) from 22 sequenced full-length clones from HEK293 cDNA.
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-inducible 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.
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, KD) 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 non-cooperative (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 KD 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 posttranscriptional 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^{-11}).

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
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).

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$_2$[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.
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, \( p = 2 \times 10^{-5} \); 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 (Borgogone 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 (Kühlmann 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 CCHCs, 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 cross-linking 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). In contrast, the two CCHC domains of the LIN28 protein family, which are homologous to CNBP’s seven CCHCs, 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 cross-linking 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,
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 study 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016). A few proteins, including the ATP-dependent helicases elf4A 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 is 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; Millevoy et al., 2012; Kwok et al., 2016).
Small RNA cDNA libraries from ribosome footprints were constructed as in Hafer 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 (Figures 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., 2008), 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: AAAAAAAAAAGCGC GAGCTGCGCTAGATGCGAATGTGAGCG). The accession numbers for the PAR-CLIP and RNA-seq sequence data reported in this paper are NCBI GEO: GSE76193 and GSE76627, respectively.

**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**


**ACKNOWLEDGMENTS**

We thank the NIAMS Genomics Core Facility and Gustavo Gutierrez-Cruz and Dr. Stefania Dell’Orso (NIAMS/NIH) for sequencing support. We also thank Jyotishman Veepaschit (Biocenter) for help with the purification of CNBP and Dr. Noam Stern-Ginossar (Weizmann Institute of Science) for help with the ribosome-profiling protocol. We also want to acknowledge Wesley Browne (Ferlinga lab) for help with the CD measurement. This work was supported by the Intramural Research Program of the National Institute for Arthritis (grants ZIA AR041200-04 and ZIA AR041205-02) and Musculoskeletal and Skin Disease.

**REFERENCES**


