Supporting Information

Transcriptome-wide Identification of Single-stranded RNA Binding Proteins

Ruiqi Zhao,‡[a] Xin Fang,‡[a] Zhibiao Mai,‡[b] Xi Chen,[a] Jing Mo,[a] Yingying Lin,[b] Rui Xiao,[c, d] Xichen Bao,*[b] Xiaocheng Weng*[a] and Xiang Zhou*[a, d]

‡ These authors contributed equally to this work.

Table of Contents

Materials

Antibodies used in this work were listed as follows. anti-GRSF1 (A20879, 1:500; Abclonal), anti-PCBP1 (A19276, 1:500; Abclonal), anti-UPF1 (A5071, 1:500; Abclonal), anti-HNRNPR (A19999, 1:500; Abclonal), anti-CSDE1 (A5941, 1:500; Abclonal), anti-SRSF3 (A6067, 1:500; Abclonal), anti-TRUB2 (A15818, 1:500; Abclonal), anti-AQR (A6011, 1:500; Abclonal), anti-GAPDH (AC033, 1:50,000; Abclonal), anti-β-tubulin (A12289, 1:2000; Abclonal), anti-CNBP (67109-1-Ig, 1:3,000; Proteintech), anti-PNPT1 (14487-1-AP, 1:2,000; Proteintech), anti-Rabbit IgG (HRP) (SA00001-2, 1:2,000; Proteintech), anti-Mouse IgG (HRP) (SA00001-1, 1:2,000; Proteintech), anti-PKR (ab32506, 1:5,000; Abcam).

DNA primers used in this work were purchased from Genecreate, the sequences were listed in Table S1. RNAs used in this work were purchased from Takara, the sequences were listed in Table S2.

Methods

Cell culture

HeLa cells were cultured in DMEM high glucose (Genview) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v) 10,000 U/mL penicillin and 10,000 μg/mL streptomycin (Genview), and maintained at 37°C with 5% CO_2 (Thermo, Heracell™ VIOS 160i CO_2).

MTT assay

HeLa cells were seeded in 96-well plates at a density of 10,000 cells per well. After 12 h, the cells were grown to about 80-90% confluence, the medium were discarded and replaced with fresh medium containing different concentrations of N_3-kethoxal for 30 min. In the case of the analysis of N_3-kethoxal incubation time, cells were incubated with 5 mM N_3-kethoxal for different time. Then, the medium was discarded and cells were washed three times with PBS, followed by adding 100 μL fresh medium and 10 μL MTT (5 mg/mL in PBS). The cells were incubated for addition 4 h at 37°C with 5% CO_2. At last, the medium was discarded and 150 μL DMSO was added to each well, and the cells were incubated at 37°C in dark for 10 min. The absorbance at 492 nm was determined using a Multiskan MK3 (Thermo) to evaluate cell viability. The results were shown in Figure S1.

Reaction of RNA and N_3-kethoxal

N_3-kethoxal was synthesized as previous protocol descripted.[1] The oligonucleotides labelled with a HEX or FAM fluorophore at the 5' end were purchased from Takara, the sequences were shown as Table S2. Two complementary RNA strands were mixed in equal concentration in annealing buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl) and heated at 95°C for 3 min, and then gradually cooled to 25°C at a rate of -1°C/min to acquire double-stranded RNA (dsRNA). 5 μM ssRNA (CNTR) and dsRNA (CNTR-ds) were incubated with different concentrations of N_3-kethoxal (0, 1, 2, 5, 10, 50 mM) in reaction buffer (0.1 M sodium cacodylate, 50 mM potassium borate pH 7.0). The reactions were performed at 37°C for 30 min. The RNA was separated by non-denaturing polyacrylamide gel electrophoresis, and visualized by the fluorecence of HEX using Molecular Imager® ChemiDocTM XRS+ Imaging System (Bio-Rad).

Isolation of RNA interactome

The RBPs were captured according to previously described procedure with minor modifications.[2] For a typical proteomics experiment, twenty 15 cm-plates of HeLa cells were used. Cells were grown to about 80-90% confluence, and incubated with 5 mM N_3-kethoxal in fresh medium for 30 min (Block-cCL; for control and cCL, cells were incubated with DMSO in fresh medium). Then the medium was discarded and cells were washed with PBS three times. For Block-cCL and cCL, cells were irradiated with 0.15 J/cm^2 UV at 254 nm on the ice (UVP HL-2000 HybriLinker), while control samples were kept in dark. Then, 2 mL lysis buffer (100 mM Tris–HCl pH 7.5, 500 mM LiCl, 0.5% lithium-dodecylsulfate (LiDS), 1 mM EDTA pH 8.0) supplemented with 1 mM phenylmethanesulfonylfluoride (PMSF) and RiboLock RNase inhibitor (Thermo, cat# EO0381) for one plate were added. Cells were putted on ice for 10 min before harvested by scraping. The lysates were homogenized by a 0.45 mm diameter needle and then centrifuged at 12,000 g for 15 min. The supernatant was collected and diluted with an equal volume of lysis buffer, and 5% of supernatant was taken out as input. Remained supernatant was incubated with oligo(dT) magnetic beads (NEB, cat# S1419S) at 4°C for 1 h under gentle rotation. Then beads were captured by magnetic stand and supernatant was removed to another tubes. The beads were further washed with lysis buffer, washing buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA pH 8.0), washing buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 1 mM EDTA pH 8.0), and washing buffer 3 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA pH 8.0) twice for 10 min at 4°C under rotation. After a brief wash with elution buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0), the beads were resuspended with elution buffer and RNA-protein complexes were eluted from beads by heating at 80 °C for 5 min. The beads were eluted

again and supernatant was combined. At last, RNAs or proteins were released from complexes by proteinase K treatment or RNase A/T1 treatment respectively.

Expression and purification of recombinant proteins

Human cDNAs of TRBP, CNBP, MBNL1, SUPV3L1and SRP19 were reversely transcribed from mRNA of HeLa cells and amplified by PCR using primers designed according to the sequence of corresponding gene in NCBI. The sequnce of primers were showed as Table S1. The cDNAs were cloned into pET-24b (+) vectors. The constructed expression plasmids were transformed into BL21(DE3). Bacteria were grown at 37°C until OD_{600} values reached 0.6, and then induced with 0.5 mM IPTG (for CNBP, cells were incubated with 0.3 mM IPTG and 50 μM $ZnCl_2$) over night at 16°C. Then, cells were lysed in lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl) by sonication. The His6 tagged proteins were purified by Ni^{2+}-NTA Resin (Bio-rad, cat# 7800812) using fast protein liquid chromatography (Bio-rad NGCTM chromatography system). The purified recombinant proteins were stored in 20 mM Tris–HCl pH 7.5 buffer, which contained 300 mM NaCl and 30% glycerol. The purified proteins were determined by SDS-PAGE and stained using coomassie blue (Beyotime, cat# P0017F). The concentrations of proteins were measured by BCA Protein Assay Reagent Kit (Thermo, cat# 23250).

Electrophoretic mobility shift assay (EMSA)

For EMSA, RNA was incubated with different concentrations of proteins in 10 μL reaction buffer (100 mM Tris-HCl pH 7.5, 10 mM KCl, 1 mM $MgCl_2$, 10% glycerol, 0.1μg/μL BSA) at 37°C for 30 min. The RNA was fractionated by a 6% non-denaturing polyacrylamide gel. To examine the affinity of protein and N_3-kethoxal labeled RNA, RNA was reacted with different concentrations of N_3-kethoxal in reaction buffer at 37°C for 30 min firstly, then protein was added and the mixture was incubated for another 30 min. To examine the affinity of RNA and N_3-kethoxal labeled protein, protein was reacted with different concentrations of N_3-kethoxal in reaction buffer at 37°C for 30 min and purified by Micro Bio-Spin P-6 Gel Columns (Biorad, cat# 7326222), then reacted with RNA at 37°C for 30 min.

Western blot assay and silver staining

Proteins were acquired as described in "Isolation of mRNA interactome". For the western blot assay of PNPT1 and PKR, RNA-protein complexes were cross-linked by 1% formaldehyde instead of 254 nm UV irradiation to enhance cross-linking efficiency of double-stranded RNA binding protein as previous work descriped.[3] The isolation of RNA-protein complexes were same as described in "Isolation of mRNA interactome". Proteins were separated by 4-12% Bis-Tris SDS-PAGE Gel (Sangon Biotech, cat# C691103). Silver staining was performed using Fast Silver Stain Kit (Beyotime, cat# P0017S) according to the manufacturer's recommended procedures. For the western blot assay, proteins were transferred to PVDF membrane (Millipore) from SDS-PAGE Gel. The membranes were blocked with 5% BSA in TBST at 37°C for 2 h before incubated with the primary antibodies at 4°C for 12 h. After five times washes with TBST, the membranes were incubated with HRP conjugated secondary antibodies for 2 h at 37°C. After additional five times washes, the signal was generated with enhanced chemiluminescence (SuperSignal™ West Pico Chemiluminescent Substrate, cat# F34077, Thermo) using Molecular Imager® ChemiDocTM XRS+ Imaging System (Bio-Rad).

RT-qPCR

The samples from elution, input and supernatant (supernatant of input after incubated with oligo(dT) magnetic beads) were treated with DNase I (NEB, cat# M0303S) according to the manufacturer's recommended procedures. Then the samples were incubated with 50 mM GTP at 95 °C for 10 min to remove N_3-kethoxal modifications in RNAs. RNAs were purified by RNA Clean & Concentrator Kit (Zymo research, cat# R1016) and reconstituted in RNase-free H_2O. The RNA was subjected to RT-qPCR using TransScript® Green One-Step qRT-PCR SuperMix (Transgen, cat# AQ211-01) with C1000^{TM} Thermal Cycler equipment. Primers used were listed in Table S3.

Sample preparation for mass spectrometry

Eluted proteins were incubated in reaction solution (1% sodium deoxycholate (Sigma, cat#D6750), 100 mM Tris-HCl pH 8.5, 10 mM tris(2-carboxyethyl) phosphine (Aldrich, cat#C4706), and 40 mM chloroacetamide (Aldrich, cat#C0267)) at 60°C for 1 h, to perform protein denaturation, disulfide bond reduction, and cysteine-SH alkylation in one step. Next, protein concentrations were measured using Bradford method. And equal volume of H_2O was added to dilute the samples. Then, proteins were digested with trypsin at a ratio of 1:50 (enzyme: protein, w/w) overnight at 37°C. After terminating digestion reaction with TFA, the samples were centrifugated at 16,000 g for 15 min at 4°C. Peptides in the supernatant were desalted using desalting columns filled with Poly(styrene-divinylbenzene) copolymer (SDB) materials self-made according to previous protocol.[4] The purified peptide was vacuum dried and stored at -20°C before used.

LC-MS/MS

LC-MS/MS data acquisition was carried out on an Orbitrap Exploris 480 mass spectrometer coupled with an Easy-nLC 1200 system (Thermo). Peptide mixtures were separated in a C18 analytical column (75 μm × 250 mm, 1.9 μm particle size, 100 Å pore size, Thermo). Mobile phase A (0.1% formic acid) and mobile phase B (0.1%

formic acid, 80% acetonitrile) were used to establish a 60 min separation gradient (0 min, 6% B; 42 min, 25% B; 53 min, 45% B; 53.5 min, 80% B; 60 min, 80% B), with a constant flow rate at 300 nL/min. LC-MS/MS raw data was acquired in a data-independent-acquisition (DIA) mode. Each scan cycle was consisted of one full-scan mass spectrum (R = 60 K, AGC = 3e6, Max IT = 30 ms, scan range = 350-1250 m/z) followed by 40 variable MS/MS events (R = 30 K, AGC = 1000%, Max IT = 50 ms). High-field asymmetric waveform ion mobility spectrometry (FAIMS) compensation voltage (CV) was set to -45. High energy collision dissociation (HCD) was operated with a normalized collision energy setting of 30.

LC-MS/MS data analysis

The raw MS data were processed using DIA-NN software (V1.8).[5] Briefly, UniProt database (March 17, 2021) containing 20,381 Swiss-Prot/reviewed human protein sequences were downloaded to generate human proteome database. A spectral library was predicted using deep learning algorithms in DIA-NN. And another spectral library was generated from raw DIA data employing MBR function. The protein quantitation information was acquired using these two spectral libraries. Trypsin/P was set as digestion mode with maximum two missed cleavages and the minimal peptide length was set to 6 aa. Carbamidomethyl cysteine was set as fixed modification. Methionine oxidation and acetyl N-terminal were set as variable modifications. The MS1 match tolerance was set as 20 ppm and 4.5 ppm for the first search and main search respectively. The MS2 tolerance was set as 20 ppm. Proteins with at least two unique peptides were required. The identified proteins with false discovery rate (FDR) less than 0.01 at both protein and peptide levels were used for further analysis.

Supplementary Tables and Figures

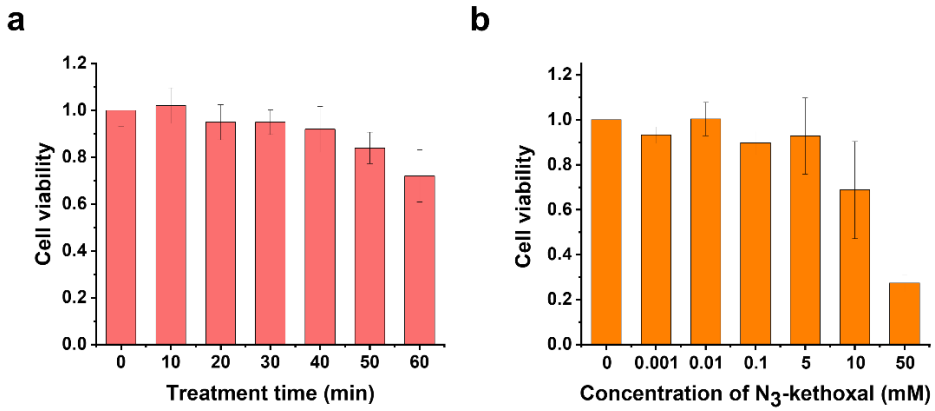

Figure S1. MTT assay of N_3-kethoxal. HeLa cells were seeded in 96-well plates at a density of 10,000 cells per well. (a) After 12 h, cells were treated with 5 mM N_3-kethoxal for different time (0, 10, 20, 30, 40, 50, 60 min) at 37°C. (b) After 12 h, cells were treated with different concentrations of N_3-kethoxal (0, 0.001, 0.01, 1, 5, 10, 50 mM) for 30 min at 37°C. The experiments were progressed in at least three replicates; error bars represented mean ± SD.

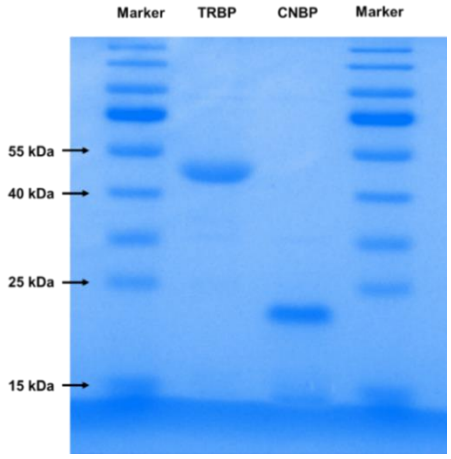

Figure S2. Coomassie blue of purified recombinant TRBP and CNBP. Proteins were separated by SDS-PAGE gel and stained using coomassie blue.

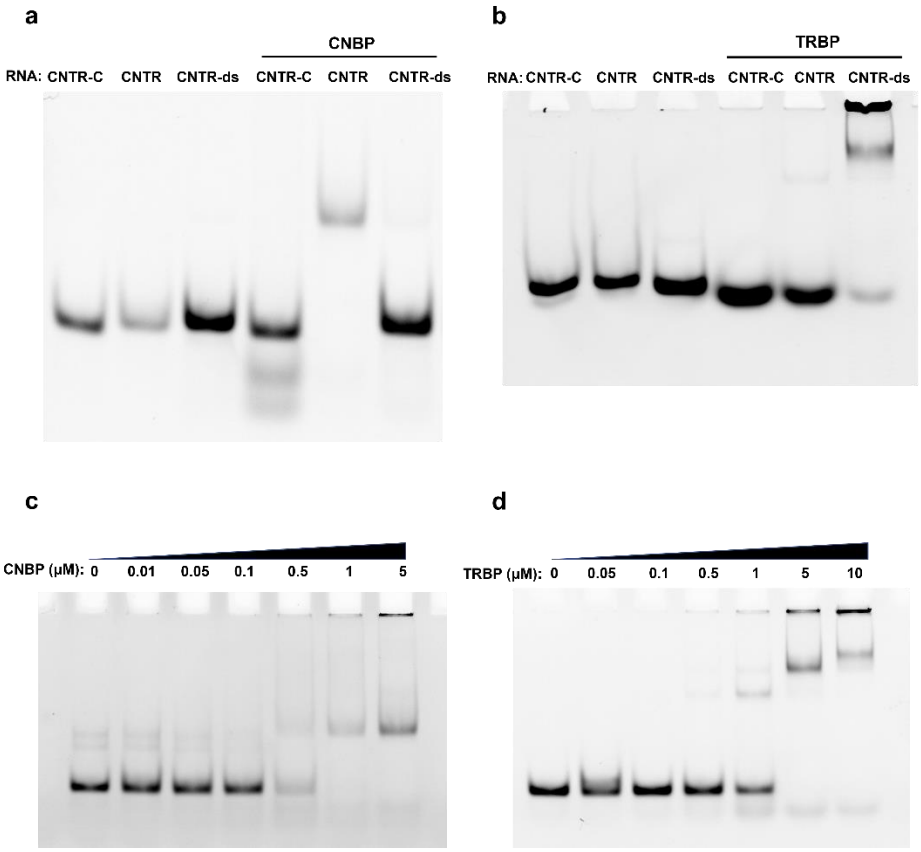

Figure S3. EMSA experiment conformed selective binding affinity of CNBP and TRBP to its target RNA. (a) EMSA of CNBP and different RNAs. 0.5 μM CNTR (ssRNA) and its complementary strand CNTR-C, as well as annealed double-stranded RNA, CNTR-ds, were incubated with 1 μM CNBP for 30 min at 37°C, the RNA was fractionated by non-denaturing polyacrylamide gel. (b) EMSA of TRBP and different RNAs, 5 μM TRBP was used, other procedure was same as (a). (c) EMSA of different concentrations of CNBP and CNTR (0.5 μM). (d) EMSA of different concentrations of TRBP and CNTR-ds (0.5 μM).

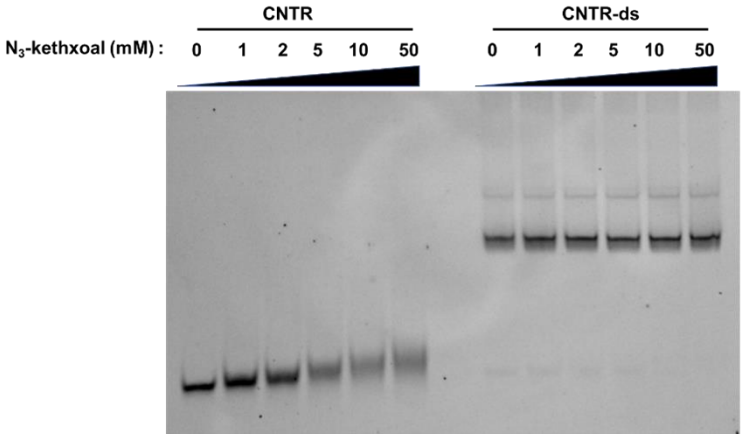

Figure S4. Selective labeling of ssRNA (CNTR) and dsRNA (CNTR-ds) with N_3-kethoxal. 5 μM RNA was treated with different concentrations of N_3-kethoxal (0, 1, 2, 5, 10, 50 mM) for 30 min at 37°C. The RNA was separated by non-denaturing polyacrylamide gel electrophoresis and visualized by the fluorecence of HEX.

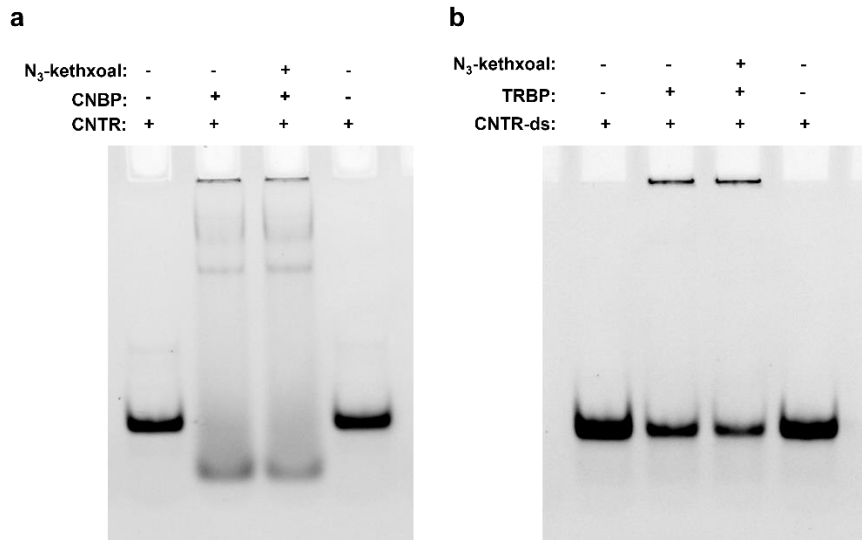

Figure S5. Binding affinity of N_3-kethoxal treated proteins and RNA. 1μM CNBP (a) or 5μM TRBP (b) were treated with 5 mM N_3-kethoxal for 30 min at 37°C and purified, then reacted with 0.5 μM RNA for another 30 min at 37°C. The RNA was fractionated by a 6% non-denaturing polyacrylamide gel.

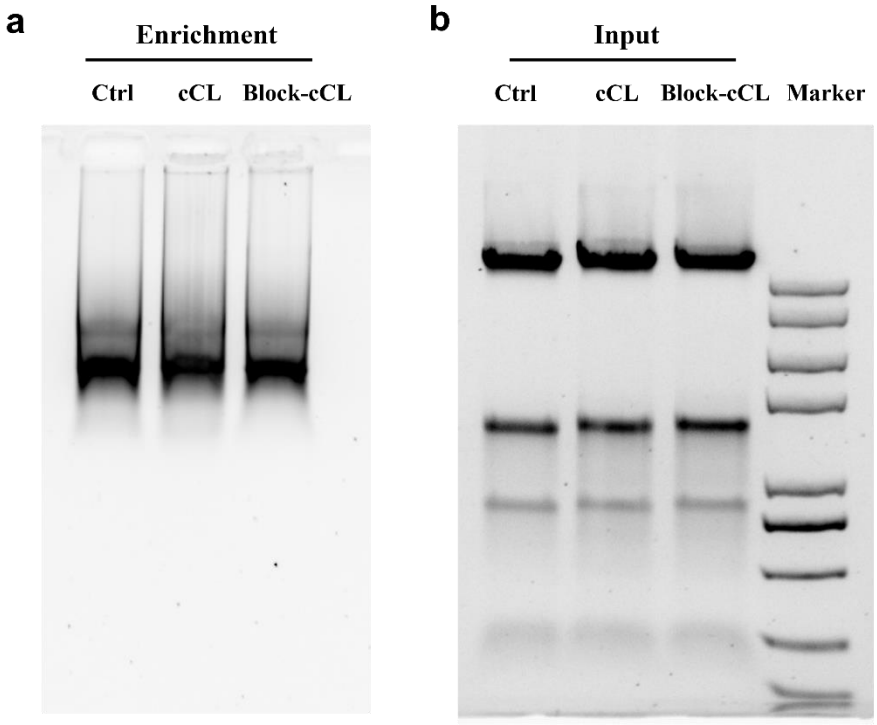

Figure S6. Gel electrophoresis of RNA in enrichment (a) and input (b). RNA contained in cell lysates or isolation by oligo(dT) magnetic beads were separated by 0.8% agarose gel and visualized by GelRed.

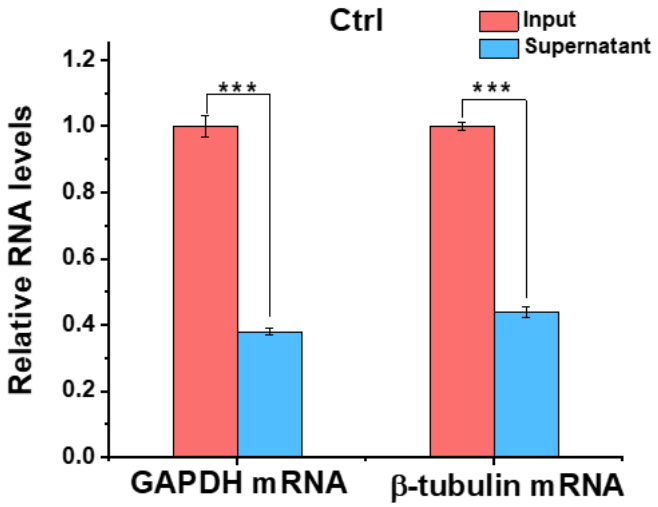

Figure S7. RT-qPCR analysis of GAPDH and β-tubulin mRNA in input and supernatant of Ctrl. Error bars represented mean ± SD from three biological experiments. Significance was assessed by two-tailed Student's t test (***p <0.0001).

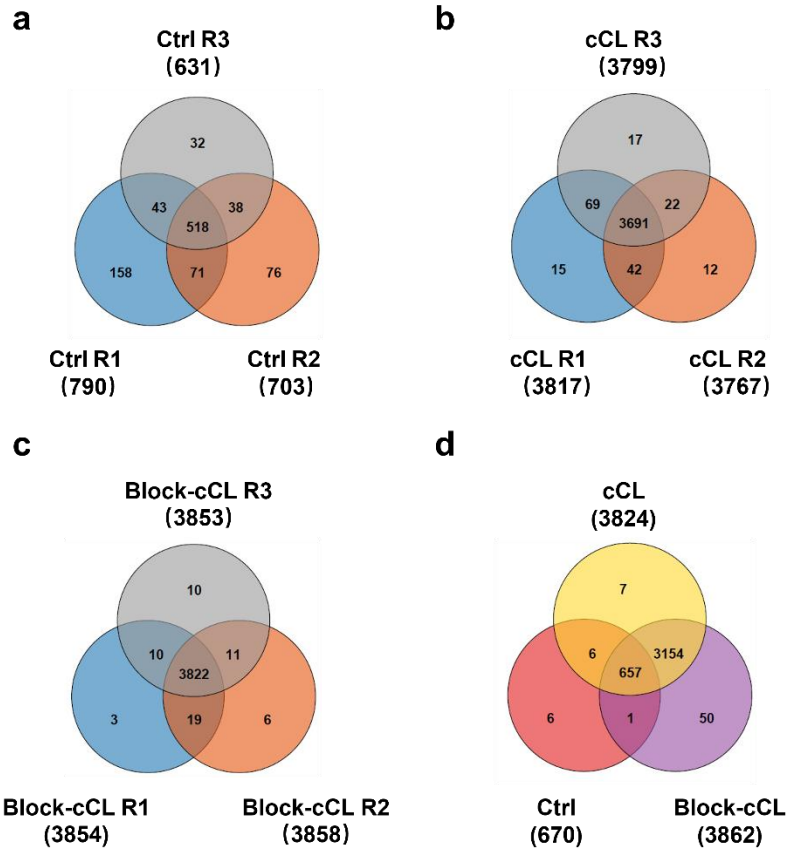

Figure S8. Venn diagram of proteins identified in different samples. The number of proteins identified in the three replications of Ctrl (a), cCL (b) and Block-cCL (c) were represented. (d) Venn diagram comparing proteins identified in Ctrl, cCL and Block-cCL. All shown proteins were quantified at least twice in three replications.

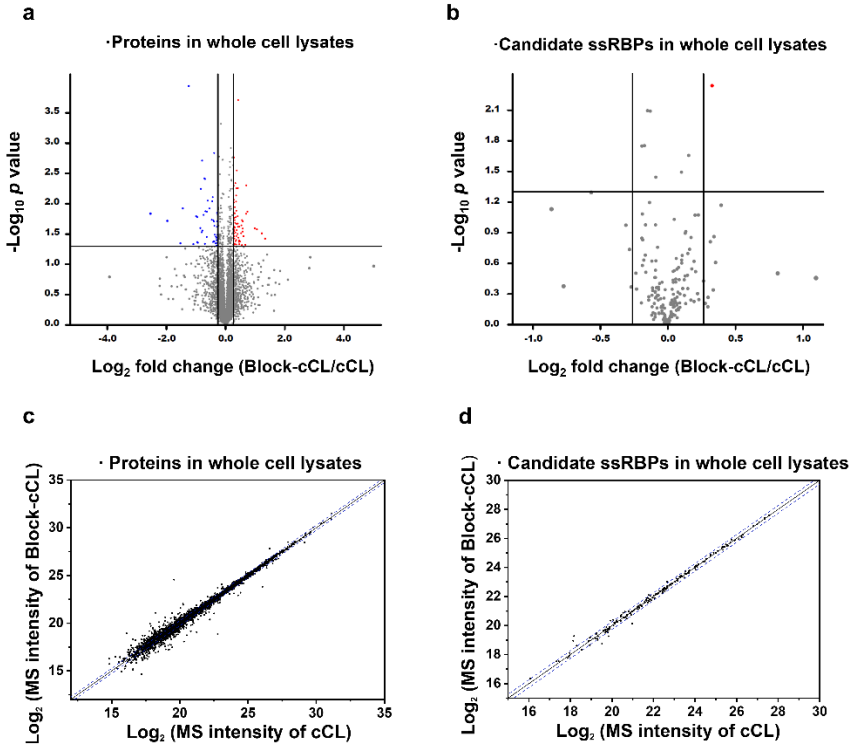

Figure S9. Quantitative proteomic analysis of proteins in cell lysates treated with N_3-kethoxal or not. (a) Volcano plot of all proteins identified in cell lysates. Proteins with p value < 0.05 and a fold change ≥1.2 were considered as significantly changed. (b) Volcano plot of candidate ssRBPs identified in cell lysates. Proteins with p value < 0.05 and a fold change ≥1.2 were considered as significantly changed. (c) Scatter plot comparing abundance of proteins identified in N_3-kethoxal treated cell lysates with untreated cell lysates. (d) Scatter plot comparing abundance of candidate ssRBPs identified in N_3-kethoxal treated cell lysates with untreated cell lysates.

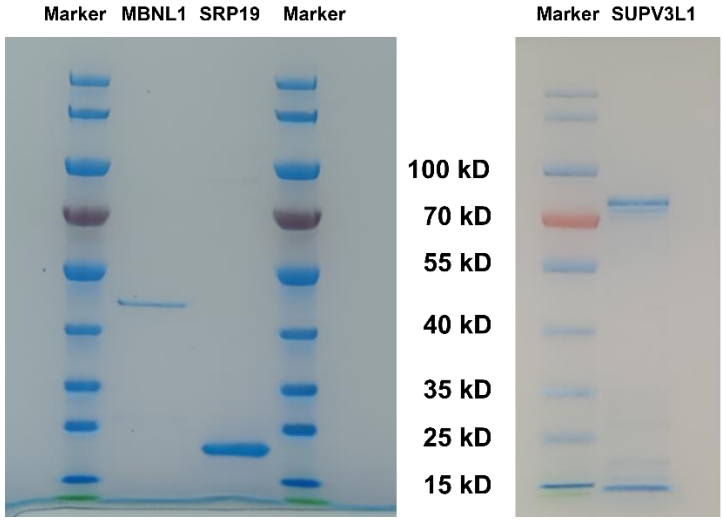

Figure S10. Coomassie blue of purified recombinant MBNL1, SRP19, and SUPV3L1. Proteins were separated by SDS-PAGE gel and stained using coomassie blue.

Table S1. Primers used in plasmid construction. The corresponding restriction enzymes were indicated in the primer names.

Primer name	Sequence (5'-3')
CNBP-FP-BamHI	AAATTTGGATCCAATGAGCAGCAATGAGTGCTTCAAGTGT
CNBP-RP-HindIII	AAATTTAAGCTTGGCTGTAGCCTCAATTGTGCATTCC
TRBP-FP-HindIII	AAATTTAAGCTTATGAGTGAAGAGGAGCAAGGCTCCG
TRBP-RP-NotI	AAATTTGCGGCCGCCTTGCTGCCTGCCATGATCTTGAG
MBNL1- FP-BamHI	AAATTTGGATCCGATGGCTGTTAGTGTCACACCAATTCG
MBNL1-RP-NotI	AAATTTGCGGCCGCCATCTGGGTAACATACTTGTGGCTAGT
SRP19-FP-EcoRI	AAATTTGAATTCGATGGCTTGCGCTGCCG
SRP19-RP-NotI	AAATTTGCGGCCGCCTTCTTTTTCTTTCCTTTCCCTTTTTTACTTCCC
SUPV3L1-FP-EcoRI	AAATTTGAATTCGATGCAGTCTCATTCCCTGGATGTG
SUPV3L1-RP-NotI	AAATTTGCGGCCGCGTCCGAATCAGGTTCCTTCTTCTTTCT

Table S2. Sequence of RNA used in EMSA experiment. RNA was labelled with HEX or FAM fluorophore at 5' end.

RNA name	Sequence (5'-3')
CNTR	HEX-GAGAGAAAGGGUAAAGGGAGAGA
CNTR-C	HEX-UCUCUCCCUUUACCCUUUCUCUC
MBP-SS	FAM-GGAGGAGGAGGGAGGGGAG
MBP-C	CUCCCCUCCCUCCUCCUCC

Table S3. Sequence of primers used in RT-qPCR experiment.

Primer name	Sequence (5'-3')
GAPDH-FP	AGCCACATCGCTCAGACAC
GAPDH-RP	GCCCAATACGACCAAATCC
β-tubulin-FP	CGGATCAGCGTCTACTACAAC
β-tubulin-RP	CCAAAGATGAAATTGTCAGGCC

1. T. Wu, R. T. Lyu, Q. C. You and C. He, *Nature Methods*, 2020, **17**, 749-749.
2. A. Castello, R. Horos, C. Strein, B. Fischer, K. Eichelbaum, L. M. Steinmetz, J. Krijgsveld and M. W. Hentze, *Nature Protocols*, 2013, **8**, 491-500.
3. Y. Kim, J. Park, S. Kim, M. Kim, M. G. Kang, C. Kwak, M. Kang, B. Kim, H. W. Rhee and V. N. Kim, *Molecular Cell*, 2018, **71**, 1051-1063.e1056.
4. J. Rappsilber, M. Mann and Y. Ishihama, *Nature Protocols*, 2007, **2**, 1896-1906.
5. V. Demichev, C. B. Messner, S. I. Vernardis, K. S. Lilley and M. Ralser, *Nature Methods*, 2020, **17**, 41-44.