CRISPR-Assisted Detection of RNA-Protein Interactions in Living Cells

Wenkai Yi

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Jingyu Li

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Xiaoxuan Zhu

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Xi Wang

School of Medicine, Northwest University, Xi'an, China

Ligang Fan

School of Medicine, Northwest University, Xi'an, China

Wenju Sun

School of Medicine, Northwest University, Xi'an, China

Linbu Liao

School of Medicine, Northwest University, Xi'an, China

Jilin Zhang

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

Xiaoyu Li

School of Medicine, Northwest University, Xi'an, China

Jing Ye

Department of Pathology, The Fourth Military Medical University, Xi'an, China

Fulin Chen

School of Medicine, Northwest University, Xi'an, China

Jussi Taipale

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

Kui Ming Chan (✉ ming.chan@cityu.edu.hk)

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Liang Zhang (✉ liangzhang.28@cityu.edu.hk)

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Jian Yan (✉ jian.yan@cityu.edu.hk)

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

Method Article

Abstract

Delineating the protein network associated with long non-coding RNAs (lncRNAs) is fundamental to understanding the functional mechanisms of lncRNAs. Current methods to identify lncRNA-binding proteins either rely on crosslinking-mediated complex co-precipitation or require extensive molecular engineering, leading to drawbacks such as loss of cellular context and low capture efficiency. Here we describe a CRISPR-Assisted RNA-Protein Interaction Detection method (CARPID), which leverages CRISPR/CasRx-based RNA targeting and proximity labeling, to rapidly capture binding proteins of specific lncRNAs in their native cellular context followed by LC-MS/MS identification. Applied to a variety of lncRNAs of different lengths and subcellular localizations, CARPID is proven to be a reliable and robust tool to discover the binding proteins of lncRNAs inside living cells.

Introduction

Although only 2% of the human genome codes for protein sequences[1], more than 70% of the genomic DNA can be transcribed to RNA at various stages during development[2]. For decades, these tremendous numbers of non-coding RNAs (ncRNAs) were recognized as "dark matter" and their function remains underexplored. Recent interests have been raised upon these ncRNAs, especially the long non-coding RNAs (lncRNA, defined as ncRNA of more than 200 nucleotides in length) that become widely accepted as important components participating in epigenetic regulation. Identification of lncRNA associated proteins will reveal the molecular mechanism of its cellular function in complex human disorders.

Despite the growing appreciation of the functional importance of RNA binding proteins (RBPs), technical limitations exist in elucidating lncRNA-protein interactions in living cells. The current methods mostly depend upon chemical or UV-mediated crosslinking between RNA and protein molecules for efficient complex isolation[3-6]. Such procedures may generate non-systematic bias and mask the interacting proteins under physiological circumstances.

To circumvent the limitations and detect RBPs that bind to endogenous lncRNA transcripts in living cells, we developed a method termed CRISPR-Assisted RNA-Protein Interaction Detection (CARPID) (**Figure 1**). A nuclease activity free form of a compact Type VI-D RNA CRISPR single-effector system dCasRx was employed for specific lncRNA targeting[7]. This dCasRx effector protein retained the capability of processing guide RNA (gRNA) arrays to two or more component gRNAs without cleaving targeted RNA transcripts. Taking this advantage, we designed a guide RNA array composed of two gRNA sequences spaced by a 30-nt direct repeat (DR) to target two adjacent loci on the same lncRNA transcript. This theoretically allows augmented targeting specificity and thus reduced background noise.

To identify the RBPs bound to the target lncRNA, we fused the dCasRx with the engineered biotin ligase BASU (Figure 2). To monitor and minimize the variation caused by heterogeneous expression of BASU enzyme among cells, we cloned the BASU-dCasRx in-frame with a self-cleaving T2A peptide and an eGFP cDNA. CARPID was carried out by co-expressing the BASU-dCasRx chimeric protein with gRNAs targeting specific regions of a lncRNA, inducing biotinylation, and subsequent enriching biotinylated proteins for mass spectrometry-mediated protein identification and quantification. Here we provide the protocol that we have optimized for CARPID.

Reagents

HEK293T cell line (some other cell lines of interest);

Appropriate cell culture media;

Molecular cloning reagents;

Biotin (Sigma #B4639. Dissolve biotin with DMSO to the concentration of 100 mM. Can be stored at -20°C for at least half a year.);

Dynabeads®MyOne™ Streptavidin T1 (ThermoFisher #65601);

Tris-HCl (pH 7.5) (ThermoFisher #15567027);

Tris-HCl (pH 8.0) (ThermoFisher #15568025);

2-Mercaptoethanol (Sigma #M3148);

Protease Inhibitor (Roche #4693132001);

Pierce Protein Quantitation Assay (ThermoFisher #23225);

4-20% Bis-Tris gel (Bio-Rad #4561095);

PVDF membrane (Millipore #ISEQ00010);

Blocker™ BSA (10X) in TBS (ThermoFisher #37520);

DTT (Sigma #43815);

Ammonium bicarbonate (Alfa Aesar #14249);

Sequencing Grade Trypsin (Pierce #90058);

Acetonitrile (LiChrosolv®#75-05-8);

C18 tips (ThermoFisher #87782);

Reverse phase C18 column (ThermoFisher #164568);

Formic acid (Sigma #5330020050);

Iodoacetic acid (Pierce #35603);

PEI (Polysciences, #23966-1. Dissolve PEI with Milli-Q water to the concentration of 0.1%. Can be stored at -20°C for up to one year.);

Lysis buffer for Western Blot (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA supplemented with fresh protease inhibitors);

Elution buffer for Western Blot (2x Laemmli Sample Buffer (Bio-Rad #1610737), 8% 2-Mercaptoethanol, 5 mM Biotin);

Elution buffer I for Mass Spectrometry (50 mM Tris-HCl pH 8.0, 2 M urea, 10 µg/ml Sequencing Grade Trypsin, 1 mM DTT);

Elution buffer II for Mass Spectrometry (50 mM Tris-HCl pH 8.0, 2 M urea, 5 mM iodoacetamide).

Equipment

Standard molecular biology lab equipment;

DynaMag™-2 Magnet (ThermoFisher, #12321D);

Easy-nLC 1200 system (Thermo Scientific);

Q Exactive HF mass spectrometry (Thermo Scientific);

XCalibur 4.0.27 (Thermo Scientific);

Proteome Discoverer 2.2 software (ThermoFisher).

Procedure

1. Design and construct gRNA plasmids for targeting the lncRNA of interest.

1.1 Design the paired gRNA sets with an in-house computational program (gp-designer). Typing in the gene symbol into the program, the full-length sequence of the lncRNA will be extracted for constructing (length=30nt, step=1nt) and filtering (GC%=40%~60%) the sequence substrings. The uniquely aligned (human reference genome, hg19) and paired sequences (spacing, 16~20nt) are exported into a excel sheet with indicated RNAfold structure ratio (DotPercent, high value of "DotPercent" from RNAfold suggests a low chance of forming hairpin structure);

1.2 Choose one or several sets of gRNAs targeting specific regions of lncRNA of interest and clone them into the pre-gRNA plasmid (Addgene #109054) with a Gibson assembly approach (NEB, #E2611S).

2. Cell preparation and transfection.

Before CARPID assay, evaluate the specificity of selected gRNA sets by co-expressing them with wild-type CasRx nuclease (Addgene #109049) and determine the knockdown efficacy via quantitative RT-PCR (**Figure 3**).

2.1 Seed $4×10^6$HEK293T cells in each 10-cm plate one day before transfection. Cells are cultured in DMEM (Life Technologies) supplemented with 10% inactivated FBS (Life Technologies) and 1% penicillin-streptomycin (Life Technologies #15140163) at 37°C with 5% CO_2;

2.2 Co-transfect 4μg of gRNA expressing plasmid and 4μg of BASU-dCasRx plasmid with 24 μg of 0.1% PEI for each 10-cm plate;

2.3 At 24 hours post-transfection, change the medium with fresh medium.

3. Biotin labeling

3.1 At 48 hours post-transfection, change the medium with fresh medium containing 200 μM biotin and incubate for 15 mins at 37°C with 5% CO_2;

3.2 Replace the medium with 10 ml ice-cold PBS to quench the reaction;

3.3 Wash cells with 10 ml ice-cold PBS at least three times to remove the remaining biotin.

4. Cell lysis

4.1 Add 1 ml ice-cold PBS to the plate and scrape the cells off, transfer them to a 15 ml tube and pellet the cells by centrifugation at 1,200 rpm for 5 mins;

4.2 Completely remove the PBS and resuspend the cell pellet in 1 ml cell lysis buffer;

4.3 After end-over-end rotating in cold room for 10 mins, spin the lysate at 15,000 rpm for 10 min at 4°C to remove the cell debris;

4.4 Transfer the supernatant to a new centrifuge tube and measure the protein concentration with Pierce Protein Quantitation Assay;

4.5 Adjust the volume the lysates to equalize the protein concentration;

4.6 Set aside 50 µl of lysate as Input, and use the remaining lysates for the following immunoprecipitation.

5. Streptavidin immunoprecipitation

5.1 Add 30 µl of MyOne T1 streptavidin beads to the lysate and incubate in cold room for 2 hours with end-over-end rotation;

5.2 Precipitate the streptavidin beads with DynaMag™-2 Magnet stand and remove the supernatant completely;

5.3 Resuspend the beads with 1 ml ice-cold lysis buffer, and perform end-over-end rotation for 5 mins;

5.4 Repeat Step 5.2 and 5.3 twice and place the tube on the DynaMag™-2 Magnet stand.

6. CARPID-Western Blot (CARPID-WB)

6.1 Remove the supernatant completely and add 30 µl elution buffer to heat elute the biotinylated proteins from the streptavidin beads at 95°C for 10 min and label it as IP;

6.2 Load the Input and IP samples to 4-20% Bis-Tris gel and transfer proteins to a PVDF membrane;

6.3 Block the membrane with BSA blocker for 1 hour at room temperature (RT) and subsequently incubate the blocked membrane with indicated primary antibodies in cold room overnight;

6.4 After primary antibody incubation, wash the membrane with 1× TBST for three times at RT;

6.5 Incubate the membrane with the corresponding secondary antibodies for 1 h at RT;

6.6 Wash the membrane with 1× TBST for three times and apply appropriate amount of ECL substrate to the washed membrane at RT;

6.7 Detect the signal with the Bio-Rad ChemiDoc Imaging System;

6.8 For signal quantification, use ImageJ software (version 1.8.0_172) to quantify and normalize the WB signals (**Figure 4**).

7. CARPID-Mass Spectrometry (CARPID-MS)

7.1 For CARPID-MS assay (follow step 5.4), wash the streptavidin beads further with 1 ml of 50 mM ammonium bicarbonate (pH 8.0) at 4°C for three times;

7.2 Resuspend the beads in 50 µl Elution buffer I and incubate at 30°C, mixing at 400 rpm for 60 mins and collect the supernatant to a fresh tube;

7.3 Resuspend the beads with 25 µl Elution buffer II twice and combine all three batches of elutes. Protect samples from light during this procedure;

7.4 Add additional 0.25 µg trypsin into the combined elutes and incubate at 37°C overnight;

7.5 Add 10% formic acid at a ratio of 1:25 (v/v) to stop the digestion;

7.6 Desalt the digested samples with C18 tips following manufacturer's instruction and reconstitute in 20 µl of 0.1% formic acid;

7.7 Perform the LC-MS/MS analysis using an Easy-nLC 1200 system coupled to a Q Exactive HF mass spectrometry. Inject 6 µl of samples to a reverse phase C18 column (75 µm i.d. × 15 cm, 3 µm particle size, Thermo Scientific #164568) at a flow rate of 250 nl/min. Mobile phase A (0.1% formic acid in ultrapure water) with an eluting buffer as mobile phase B (0.1% formic acid in 80% acetonitrile) are used together to establish a linear 50-min gradient of 7−25% mobile phase B;

7.8 Peptides are ionized by electrospray at 2.3 kV. The mass spectrometer is operated in positive ion mode acquiring at a resolution of 120,000, with a full MS spectrum (m/z = 350-1800) using an automatic gain control (AGC) target of $3×10^6$. The top 12 most intense ions are selected for higher-energy collisional dissociation (HCD) fragmentation (normalized collision energy 27) and MS/MS spectra are generated with an AGC target of $1×10^5$ at a resolution of 30,000. The dynamic exclusion time is set to 30 s.

8. LC-MS/MS Data Processing and Analysis

8.1 Analyze all the raw files created by XCalibur (version 4.0.27) software together with the Proteome Discoverer version 2.2 software, against the UniProt human protein database in Sequest HT node.

8.2 Set the precursor and fragment mass tolerances to 10 ppm and 0.02 Da, respectively. The maximum of two missed cleavage sites of trypsin is allowed;

8.3 Set carbamidomethylation (C) as static modification and set oxidation (M) and acetyl (protein N-terminal) as variable modifications. False discovery rate (FDR) of peptide spectrum matches (PSMs) and peptide identification are determined using the Percolator algorithm at 1% based on q-value;

8.4 For label-free quantification (LFQ), use the Minora Feature Detector node in the processing workflow, and use the Precursor Ions Quantifier node and the Feature Mapper node in the consensus workflow;

8.5 Apply enrichment analysis for proteins with more than one peptide detected. Human keratins are included. Normalize the LFQ abundances across the pulldowns and logarithmized. Impute the missing values with values representing the detection limit of the mass spectrometer. Apply the rank product test to determine proteins statistically enriched in the gRNA transfected samples compared to the pre-gRNA controls. Proteins with an adjusted p-value of ≤ 0.05 and ≥ 2-fold change of abundance are considered significant RBPs of the lncRNA of interest (**Figure 5**).

Troubleshooting

Time Taken

Anticipated Results

References

1   Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74, doi:10.1038/nature11247 (2012).

2   Djebali, S. et al. Landscape of transcription in human cells. Nature 489, 101-108, doi:10.1038/nature11233 (2012).

3   Chu, C. et al. Systematic discovery of Xist RNA binding proteins. Cell 161, 404-416, doi:10.1016/j.cell.2015.03.025 (2015).

4   Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. Nat Protoc 8, 491-500, doi:10.1038/nprot.2013.020 (2013).

5   Selth, L. A., Gilbert, C. & Svejstrup, J. Q. RNA immunoprecipitation to determine RNA-protein associations in vivo. Cold Spring Harb Protoc 2009, pdb prot5234, doi:10.1101/pdb.prot5234 (2009).

6   Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464-469, doi:10.1038/nature07488 (2008).

7   Konermann, S. et al. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell 173, 665-676 e614, doi:10.1016/j.cell.2018.02.033 (2018).

Acknowledgements

Figures

Figure 1

Figure 1: Scheme of the CARPID workflow. A lncRNA of interest is targeted by a set of gRNAs. The catalytically dead RNA nuclease CRISPR/CasRx (dCasRx) fused with the engineered biotin ligase (BASU) is recruited to the specific locus. Following biotin treatment, the adjacent RNA binding proteins (RBPs) would be biotinylated by BASU. The labelling radius of the BASU-dCasRx is indicated by the red shading. Biotinylated proteins are enriched by streptavidin beads (MyOne T1) for subsequent mass spectrometric (MS) identification and western blot (WB) analysis.
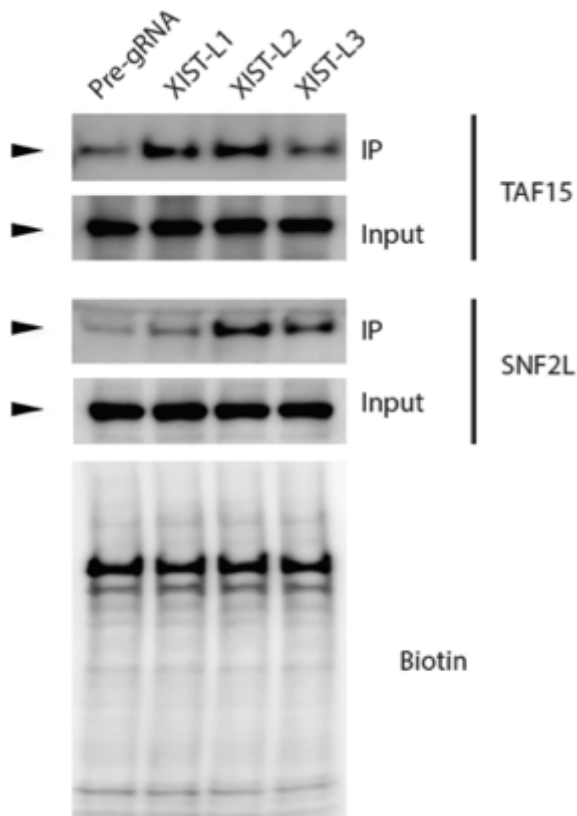

Figure 2

Figure 2: Scheme of BASU-dCasRx construct. BASU was subcloned from the BASU RaPID plasmid (Addgene #107250) and cloned in-frame into the EF1a-dCasRx-2A-EGFP plasmid (Addgene #109050). LTR, long terminal repeat; T2A, self-cleaving peptide; eGFP, enhanced GFP.
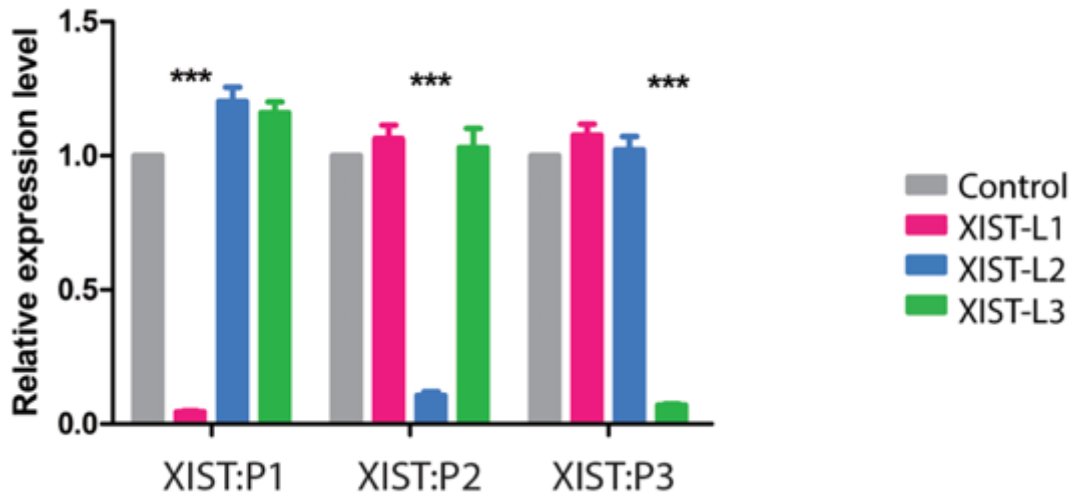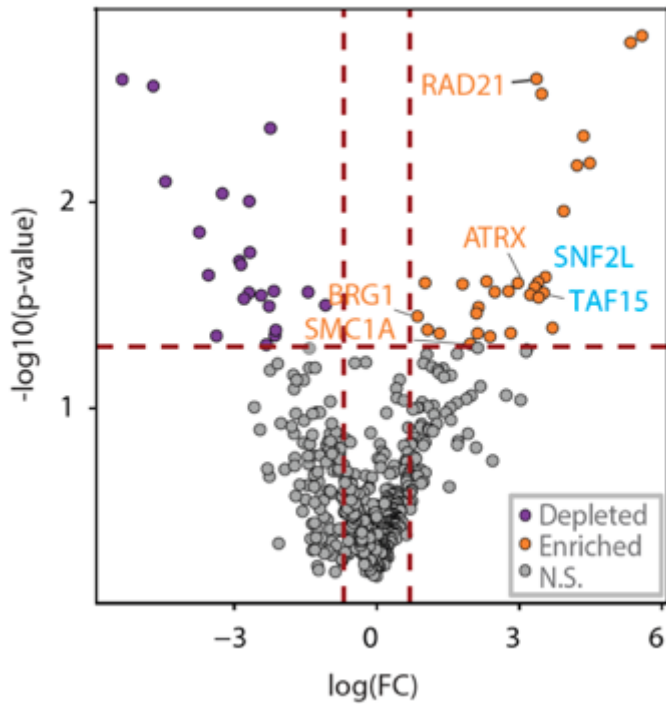
Figure 3

Figure 3: Reverse transcription qPCR shows the specificity of the CRISPR/CasRx system in XIST. HEK293T cells were co-transfected with CasRx and individual gRNA (empty vector control, XIST-L1, XIST-L2 and XIST-L3), respectively. Total RNA was extracted from treated cells followed by reverse transcription and qPCR assay to quantify the level of specific loci of XIST. GADPH was used to normalize the level of XIST. Note that active CasRx together with a set of gRNAs specifically decreased the level of RNA transcript at its targeting locus. Data are represented as mean ± SD, n = 3, ***p < 0.001, unpaired Student's t-test.

Figure 4

Figure 4: CARPID followed by Western blot using antibody against human TAF15 and SNF2L. Precipitations of biotinylated proteins were conducted by applying streptavidin beads to the lysates of HEK293T cells transfected with pre-gRNA (Control), gRNA set specific for locus 1 (XIST-L1), locus 2 (XIST-L2), and locus 3 (XIST-L3) together with BASU-dCasRx, respectively. Three biological replicative experiments were carried out, and a representative result is shown. ImageJ software (version 1.8.0_172) was employed to quantify the WB signals and normalized targets IP signals with biotin-BASU-dCasRx (biotin).

Figure 5

Figure 5: Identification of XIST-associated proteome with CARPID. The volcano plot shows the enrichment of XIST-associated proteins in HEK293T cells. The x-axis indicates the log2 fold change of protein levels in CARPID combining all three sets of XIST gRNAs relative to control. The y-axis shows the negative logarithm transformed p value (the non-parametric rank product test). Significantly enriched proteins are labelled as orange dots. Proteins previously validated to interact with XIST and identified by CARPID are labelled in orange typeface. SNF2L and TAF15, two novel XIST-associated proteins, are indicated by blue typeface.