An intrinsically disordered RNA-binding region provides local target selectivity and is essential for LARP6 function.

Authors:

Federica Capraro[1,2], Giancarlo Abis[2], Mehran Karimzadeh[3], Laura Masino[4], Alexander Barley[5], Tam T.T. Bui[2], Hani Goodarzi[3,6], Maria R. Conte[2,✉], Faraz K. Mardakheh[1,7,✉].

Affiliations:

1: Centre for Cancer Cell & Molecular Biology, Barts Cancer Institute, Queen Mary University of London, UK.
2: Randall Centre for Cell and Molecular Biophysics, King's College London, UK.
3: University of California, San Francisco, CA, USA.
4: The Francis Crick Institute, 1 Midland Road, London, UK.
5: Independent researcher
6: Arc Institute, Palo Alto, CA, USA.
7: Department of Biochemistry, University of Oxford, South Parks Road, Oxford, UK.

✉: sasi.conte@kcl.ac.uk

✉: faraz.mardakheh@bioch.ox.ac.uk

Abstract:

In addition to structured RNA-binding domains (RBDs), most RNA-binding proteins (RBPs) contain intrinsically disordered regions (IDRs), which are thought to possess additional RNA-binding activities. Despite their prevalence, the specific RNA-binding functions of IDRs in RBPs have remained largely uncharacterised. Here we examined the ordered and disordered RNA-binding activities of LARP6, an RBP known for its diverse RNA-binding repertoire. Using mass spectrometry-based mapping of RNA-protein interaction sites, we showed that both the La-module RBD and the two IDRs of LARP6 directly interact with RNA in living cells. Mutagenesis coupled with individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) revealed the La-module to be indispensable for LARP6 binding to RNA. Deletion of the IDRs did not diminish the RNA-binding capability of LARP6, but instead broadened its interaction footprints on the target RNAs, and impaired LARP6 function in promoting cancer cell viability and invasion. *In vitro* biophysical analysis corroborated the iCLIP-based findings, demonstrating that rather than independently binding to RNA, IDRs can modulate the RNA-binding patterns of the tethered RBD. Based on these findings, we propose a previously unknown role for IDRs in fine-tuning RNA-RBD interactions through modulating local RNA-binding selectivity, resulting in highly specific RNA-protein interaction patterns that are critical for RBP functions.

Introduction:

RBPs control all aspects of RNA regulation and function (1). Conventional RBPs have been traditionally defined by the presence of one or more structurally conserved RBDs, which mediate high affinity interactions with their target RNA molecules through sequence- or structure-based RNA recognition (2). The human genome encodes hundreds of such RBPs, which can be grouped into diverse super-families, characterised by the presence of specific sets of RBDs (3). Through combinatorial action of multiple RBDs, RBPs are thought to engage with and regulate specific sets of transcripts in context-specific manners, creating highly diverse yet specific post-transcriptional regulatory networks (4,5). In addition to containing well-defined high-affinity RBDs, most conventional RBPs also harbour IDRs that are characterised by an ensemble of interconverting conformations. Despite lacking a clear structure, IDRs can play critical functions in various RBPs (2). Moreover, the majority of germ-line mutations in RBPs that are associated with human Mendelian genetic disorders occur within the IDR coding regions, further indicative of their critical roles (1). Interestingly, unbiased RNA-protein interaction site-mapping studies have revealed that many IDRs can directly engage with RNA in living cells (6,7). However, the exact mechanisms through which these 'non-conventional' RNA recognition motifs combine with RBDs remain largely uncharacterised.

The La-related protein-6 (LARP6) is an evolutionarily conserved RBP found in all multicellular eukaryotes. LARP6 is a member of the LARP superfamily, a group of ancient eukaryotic RBPs involved in various aspects of post-transcriptional gene regulation, including the synthesis, processing, maturation, stability, and translation of diverse RNA targets (8). Similar to all LARPs, LARP6 is characterised by the presence of a conserved bipartite RBD, the La-module, which consists of a La-motif and an RNA Recognition Motif-1 (RRM1), connected by a short linker domain (9). This La-module is N- and C-terminally surrounded by IDRs. In addition, LARP6 possesses a SUZ-C domain in its C-terminal region, a structurally uncharacterised short motif that is known to mediate protein-protein interactions (8). Functionally, LARP6 has been shown to control diverse cellular processes including type-I collagen synthesis (9,10), myogenesis (11-13), oocyte development (14), neural tube development (15), as well as pollen tube guidance in plants (16). Recently, we identified a role for LARP6 in regulating mRNA localisation, cell proliferation, and invasion, in highly invasive breast carcinoma cells that have undergone epithelial–mesenchymal transition (EMT) (17). Transcriptome-wide iCLIP analysis revealed over five thousand mRNAs, spanning diverse functional gene categories, as direct binding targets of LARP6, highlighting a capability to engage with a broad spectrum of RNA targets (17). Consistent with these results, an RNA-based high-throughput systematic evolution of ligands by exponential enrichment (HTR-SELEX) study demonstrated that LARP6 has a highly complex RNA-binding repertoire, capable of interacting with various linear and structured motifs (18). The underlying mechanism by which an RBP with a single RBD (La-module) can selectively recognise such a diverse array of RNA targets remains unclear.

Here we devised an unbiased mass spectrometry approach to map the regions of LARP6 that directly interact with RNA in living cells. Our results reveal that, in addition to the structured La-module, regions within the N- and C-terminal IDRs of LARP6 can bind to RNA. Using a doxycycline-inducible system to express full-length or various deletion mutants of LARP6, coupled with transcriptome-wide iCLIP analysis, we demonstrated that the La-module is essential for the RNA-binding activity of LARP6 in living cells. Removing the IDRs did not abrogate LARP6's ability to interact with RNA. Instead, the deletion of the N-terminal IDR broadened LARP6's interaction footprints on target RNAs, indicating that this IDR modulates

the activity of the adjacent La-module and provides local selectivity in RNA binding. *In vitro* biophysical assessment of the RNA-binding activity of different LARP6 protein regions corroborated our observations in living cells, indicating that the IDR-mediated local RNA selectivity of LARP6 is an intrinsic property of the protein and not dependent on additional cellular factors. Importantly, deletion of the N-terminal IDR compromised the viability and invasiveness of glioblastoma cells, suggesting that the IDR-mediated local RNA-binding selectivity is crucial for LARP6 function. Together, our findings uncover a critical role for IDRs in fine-tuning RNA-RBP interactions, resulting in locally selective RNA-binding patterns that are critical for RBP functions.

Materials and methods:

Cell culture

U-87 MG (ATCC) and HEK293-T cells were grown in DMEM media (Gibco) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. U-87 MG myc-LARP6 inducible cell lines were grown as above, and myc-LARP6 expression was induced with either 60 ng/mL (iCLIP), or 30 ng/mL (all other experiments). MDA-MB231 cells stably expressing GFP-LARP6 cells were generated previously (17), and were cultured as above. All cells were maintained in a humidified incubator at 37°C, 5% CO_2. Cells were passaged consistently before reaching confluence with Trypsin-EDTA (Gibco), and tested regularly for mycoplasma, using the MycoAlert Mycoplasma Detection Kit (Lonza Bioscience).

Cell line generation

The human codon-optimised LARP6 full-length (FL) sequence was cloned in the mammalian doxycycline-inducible lentiviral expression vector pCW57.1 using Gateway cloning (Thermo Fisher Scientific) and tagged at the N-terminus with a myc-tag. Deletion mutations were then introduced in the LARP6 FL sequence using Q5 site-directed mutagenesis (NEB) to generate the following deletion mutants: ΔNTR (N-terminal region, deletion of the 2–83 amino acid region), ΔNTD (N-terminal domain, 2–300 deletion), ΔLa-module (La-module, 84–300 deletion), and ΔCTR (C-terminal region, 403–491 deletion). All constructs were validated by Sanger sequencing. Subsequently, each construct was transfected into HEK293-T cells, alongside lentiviral packaging plasmids, using Lipofectamine 2000 (Thermo Fisher), according to the manufacturer's instructions. The lentiviral particle-containing supernatant was collected after 48 hours and used to transduce U-87 MG (ATCC) parental cells. The resulting transduced cells were subsequently expanded and underwent selection for ten days with 0.4 µg/mL puromycin, to generate stable doxycycline-inducible myc-LARP6 cell lines. Inducible myc-LARP6 expression was validated and optimised with western blotting and immunofluorescence.

Endogenous LARP6 knock-down

U-87 MG cells were seeded in tissue culture plates 24 hours before transfection. An siRNA selectively targeting the 3'UTR of endogenous LARP6, but not the myc-LARP6 mutants, and a non-targeting siRNA control (scramble) were then transfected into the cells with lipofectamine RNAiMAX (Thermo Fisher), following the manufacturer's protocol, using a final siRNA concentration of 40 nM. Cells were split or harvested 72 hours post-transfection for downstream analyses. The knockdown efficiency was validated with western blotting.

Immunoprecipitation coupled with Orthogonal Organic Phase Separation (IPOOPS)

Two replicates of 40 million cells were grown in 15 cm dishes and UV-C crosslinked on ice at 400 mJ/cm^2 in a Hoefer Scientific UV crosslinker. Cells were then harvested by scraping in cold PBS, followed by centrifugation at 500 *g* for 10 minutes at 4°C to obtain cell pellets. Cell pellets were lysed in 500 µL cold RIPA buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate, cOmplete protease inhibitor cocktail (Roche), 1 mM PMSF) and incubated on ice for 30 minutes with extensive pipetting on 10-minute intervals. Cell lysates were then cleared by centrifugation at 15000 *g* for 10 minutes at 4°C. GFP-LARP6 was immunoprecipitated using 30 µL of GFP-Trap beads (ChromoTek) for one hour at 4°C in a rotating wheel, followed by two washes in cold RIPA buffer. Samples were eluted from the beads by direct addition of 1 mL TRIzol (Thermo Fisher) and 5-minute incubation at 65°C, shaking at 1100 rpm. Samples were then placed on a magnet and the eluate was collected. A first OOPS was then performed as described in (19) to isolate the LARP6–RNA adducts. Briefly, phenol-chloroform phase separation was obtained by addition of 200 µL chloroform (SIGMA), extensive mixing, and 15-minute centrifugation at 12000 *g*, at 4°C. After centrifugation, three phases had formed: the upper aqueous phase containing free RNA, the lower organic phase containing free protein, and the interphase containing crosslinked LARP6–RNA adducts. The interphase was isolated and washed three times by addition of 1 mL TRIzol, 200 µL chloroform, mixing, and centrifugation as described above. Finally, 1 mL 100% methanol was added to the interphase to precipitate LARP6–RNA adducts, followed by 5-minute centrifugation at 12000 *g*, at 4°C, and isolation of the LARP6–RNA pellet. Partial protease digestion with Arg-C (Promega) protease was then performed overnight (ON) at 37°C, according to manufacturer's instructions, to narrow down the protein regions involved in RNA binding. A second OOPS was subsequently performed, but with some modifications. The samples were first brought to 250 µL by addition of water. 750 µL of TRIzol LS (Thermo Fisher) was then added to the mixture, with the subsequent phase extraction steps carried out as before. Pellets were then resuspended in 50 µL solubilisation buffer (100 mM Tris-HCl, pH 8, 8 M urea) and sonicated with a probe sonicator (SoniPrep 150, MSE, amplitude 10, 15 sec) to help pellet solubilisation. Samples were reduced by treatment with 10 mM DTT (Thermo Fisher) for 30 minutes at room temperature (RT), followed by alkylation with 55 mM iodoacetamide (G Biosciences) for 30 minutes at RT, while protected from light. The urea concentration was then diluted four folds by addition of 100 mM Tris-HCl pH 8, followed by the second digestion with 2 µg trypsin (SIGMA) ON at RT, with shaking at 1100 rpm. The reaction was then stopped by addition of 2% TFA.

LC-MS/MS mass spectrometry sample preparation, data acquisition, and analysis

The IPOOPS samples were prepared for LC-MS/MS analysis by desalting following the StageTip procedure (20). Samples were reconstituted in 7 µL buffer A* (2% acetonitrile, 0.1% TFA, 0.5% acetic acid, water), and 6 µL were injected into a Q Exactive Plus Orbitrap mass spectrometer coupled with a nanoflow ultimate 3000 RSL nano HPLC platform (Thermo Scientific). Mass spectrometry analysis was carried out as described before (21). The MaxQuant (version 1.6.3.3) (22) was used for all the peptide searches and quantifications. Raw data files were searched against a FASTA file of the human LARP6 protein sequence (Uniprot 2016). Enzyme specificity was set to Trypsin. False discovery rates (FDR) were calculated using a reverse database search approach and were set to 1%. Default MaxQuant parameters were used, except for the enabling of the "Match between runs" and "Re-quantify" options. The downstream data analysis was performed with the software Perseus (version 1.6.2.1) (23). Contaminants were removed by filtering for "Only identified by site", "Reverse", and "Potential contaminants". The remaining identified LARP6 peptides represented the protein regions adjacent to the ones directly involved in RNA binding.

Transcriptome-wide analysis of RNA localisation

Separation of protrusions and cell bodies was carried out as described before (17,24). Briefly, 10 cm polycarbonate membrane transwell filters with 3 µm pores (Costar, catalogue number 3420) were coated on both the top and bottom side with collagen I (PureCol, 3 mg/mL) diluted in water 1:30. Coating was performed ON at 4°C, followed by complete transwell drying. Endogenous LARP6 was depleted as described above, and 10 million cells per condition were seeded on top of each 10 cm transwell. Media was added underneath the transwells and cells were incubated in the 37°C humidified incubator for 90 minutes for protrusion induction. The protrusions that formed on the bottom side of the filters were then separated from the cell bodies at the top of the filters and the two fractions were collected in 4% SDS buffer (100 mM Tris-HCl pH 7.5, 4% SDS, RNase-free). Each sample was split into three replicates and total RNA was extracted using TRIzol LS according to the manufacturer's instructions. The total RNA concentration was measured with the Qubit RNA High-sensitivity assay (Invitrogen) on a Qubit 4 Fluorometer (Invitrogen). The quality of the RNA was validated with the Agilent High Sensitivity RNA ScreenTape assay at an Agilent TapeStation 4000 instrument, with all samples having a quality score RIN ≥ 9.1. Library preparation was performed with the Lexogen QuantSeq 3' mRNA-Seq V2 Library Prep Kit FWD with Unique Dual Indices according to the manufacturer's instructions. Each replicate was processed individually to avoid batch effect. 330 ng total RNA in 5 µL volume was used as input. The amplified libraries were quantified and validated with the Qubit dsDNA High-sensitivity and Agilent D1000 high sensitivity ScreenTape assays, pooled in an equimolar ratio, and deep sequenced on an Illumina NextSeq 2000 instrument, using a single-end 100 nt reverse complement workflow. Sequencing was performed at the Barts and The London Genome Centre. The demultiplexed fastq reads were uploaded on the Lexogen Data Analysis Solution, which allowed the automated pre-processing of the data. Adapters were trimmed and low-quality reads removed using Cutadapt (V1.18) (software developed by Marcel Martin). Trimmed reads were aligned to the human genome (GRCh38, Ensembl release 107) with STAR (V2.6.1a) (25). Gene counting was performed with FeatureCounts (V1.6.4) (26) and only uniquely mapped reads were considered. The analysis of protrusions/cell bodies enrichment was performed using the Perseus software. The gene counts were first $\log_2$ transformed, and ratios were calculated by dividing the $\log_2$ counts for the protrusions over the $\log_2$ counts for the cell bodies ($\log_2$ ratio protrusions/bodies). The median of each column was then subtracted to normalise the data. Transcripts were then annotated with the KEGG database, and 2D annotation enrichment was performed on the averaged $\log_2$ ratios, as described in (27), to identify categories that were significantly changing.

3D collagen I invasion assay

The invasion assay was performed as described (28), with some modifications. µ-Plates 96-well (ibidi) were first coated with 0.2% BSA (diluted in PBS) for at least 30 minutes at 4°C. 200 µL of collagen I mix (18 parts PureCol 3 mg/mL (Advanced BioMatrix), 6.66 parts 5X DMEM (Gibco), 1 part 0.1 M sodium hydroxide (SIGMA)) was prepared on ice for each replicate and 20000 cells were quickly resuspended in the mix and plated. The plates were always kept on ice after cell plating to prevent matrix solidification. At least technical quadruplicates were used in each experiment. After all samples were plated and before the matrix had set, the plates were centrifuged at 1500 rpm for 5 minutes at 4°C to gather the cells at the bottom of the wells. The plates were then incubated in the 37°C humidified cell incubator for three hours to allow the matrix to solidify. 60 µL complete medium was then added to the top of each well, without touching the matrix, to stimulate cell invasion from the bottom of the well towards the top. 1 µM Rho kinase inhibitor 48 (ROCKi) (29) diluted in DMSO was added to the negative control samples. Cells were incubated in the 37°C incubator for 24 hours and allowed to invade into the matrix. After 24 hours, cells were fixed by addition of 18.5% paraformaldehyde (diluted in PBS) to each well. Hoechst 33258 (1:250, Invitrogen) was also added simultaneously to stain

the nuclei. Plates were incubated for four hours at RT, then stored at 4°C until imaging, which was performed within one week. Samples were imaged at a Nikon ECLIPSE Ti Spinning Disk confocal microscope with a plan apochromat VC 20X objective. Hoechst 33258 was excited with a LUN-V 405 nm laser (30% intensity, 300ms exposure). Six z-stacks were acquired 10 µm apart starting from the bottom of the well where cells had gathered until the furthest distance reached by cells, for a total of 50 µm invasion distance. The following settings were used: 2x2 binning, 540 MHz readout rate, 12-bit and gain 4-bit depth, large image 2x2 fields with 15% overlap (blending). At least two separate images were acquired for each replicate. The nuclei present in each stack were then counted using the Fiji software (30). First, each image was converted into binary format with the "threshold" command (default settings), then nuclei that were touching were separated with the "watershed" command. The "analyse particles" function was used to count the nuclei in each stack, with size set to 40-infinity and nuclei excluded from the edges. The bottom two stacks were considered the 0 µm distance, as the matrix was not completely flat in some points. Cells with nuclei in the bottom three stacks (≤ 10 µm distance) were considered non-invading, while cells with nuclei in the top three stacks (≥ 20 µm distance) were considered invading. The invasion capability of the cells was calculated as the invasion index, representative of the percentage of cells that were invading. To calculate the invasion index, the number of nuclei in each stack was first summed, then the percentage of cells that were invading and non-invading was calculated. The percentage of invading cells was then converted into the invasion index by considering the average percentage of invading cells for the parental cells as invasion index = 100 and normalising the other samples accordingly.

Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP)

iCLIP was performed on U-87 MG myc-LARP6 cells to identify the binding sites of LARP6 at nucleotide resolution, transcriptome-wide. The five myc-LARP6 mutants and the parental negative control were used in the experiment, in triplicates. iCLIP was performed as described before (31,32), with specific myc-LARP6 optimisations. Briefly, 20 million cells per condition were grown in 15 cm dishes and myc-LARP6 mutants were induced with 60 ng/mL doxycycline for 24 hours. Cells were crosslinked on ice at 150 mJ/cm^2 at 254 nm in a Hoefer Scientific UV crosslinker to form covalent bonds between RBPs and RNAs that were in direct contact, then scraped in PBS and pelleted. Cell pellets were lysed in iCLIP lysis buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate, cOmplete protease inhibitor cocktail (Roche), 1 mM PMSF), cleared, and diluted to 1 mg/mL in 1 mL volume. Samples were then treated with 0.2 units/mL lysate RNase I (Thermo Scientific, catalogue number EN0602) and 2 µL Turbo DNase (Ambion) for 3 minutes at 37°C to partially digest the RNA. 50 µL Dynabeads Protein G (Invitrogen) per sample were first conjugated with 4 µg (10 µL) mouse anti-myc-tag antibody (Cell Signaling, 9B11) for one hour at RT in a rotating wheel, then added to the samples to perform the immunoprecipitation (IP) for one hour at 4°C in a rotating wheel. After the IP, samples were stringently washed twice with high-salt wash buffer (50 mM Tris-HCl, pH 7.4, 1 M NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate), then once with PNK wash buffer (20 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20). The RNA fragments were dephosphorylated at the 3' end, followed by 3' end ligation of different barcoded adapters, to allow for downstream multiplexing of each replicate into one sample. The samples were then eluted from the beads, run on an SDS-PAGE gel, followed by isolation of the RNA fragments and subsequent library preparation as described previously. The amplified libraries were quantified and validated with the Qubit dsDNA High-sensitivity and Agilent D1000 high sensitivity ScreenTape assays, pooled in an equimolar ratio, and deep sequenced on an Illumina NextSeq 500 instrument, using a single-end 100 nt workflow. Sequencing was performed at the Barts and The London Genome Centre.

iCLIP data analysis

iCLIP data was analysed to identify the RNA-binding sites of LARP6 at nucleotide resolution. Raw fastq reads were first demultiplexed based on the experimental barcodes used, to obtain the data from each separate sample. Ultraplex (version 1.1.4) (33) was used for the demultiplexing with default settings. The demultiplexed reads were then uploaded on the webserver iMAPS (imaps.goodwright.com) for primary iCLIP data analysis. The human genome GRCh38.p13 GENCODE V42 was used with the primary assembly comprehensive annotation. Reads were first trimmed with TrimGalore! (software developed by Felix Krueger), to remove remaining adapters and low-quality reads, then pre-mapped to rRNA and tRNA sequences with Bowtie 2 (34) and discarded. This step ensured that reads mapping to repetitive regions were not misaligned to mRNA. The remaining unmapped reads were then mapped to the genome with STAR (25). PCR duplicates were removed with UMI-tools (35) based on the presence of identical UMIs in reads mapping to the same position. The exact crosslink position was then identified as the first nucleotide after the experimental barcode, and this data was downloaded from the webserver and processed further offline. The crosslink information was used to call the peaks, which represent the binding sites of LARP6, for each individual replicate. The peak caller Clippy (version 1.5.0) (36) was used with the following settings: rolling mean 15; minimum counts per gene 3; minimum counts per peak 3; prominence 1 (default); peak threshold 0.4 (default). The peaks called for the individual replicates were then combined into one file per mutant by merging the peaks and summing their peak scores using bedtools (version 2.30.0) (37), with the peak scores indicative of the binding strength. Background normalisation was performed by subtracting the peaks that were called in the parental negative control from each combined peak file, using bedtools. The normalised peaks were then annotated to the corresponding transcript and genomic region using a custom script. First, the genomic coordinates for the start and end of the peaks were averaged to obtain the centre of the peak, which was the nucleotide considered for the annotation. This choice was made because peaks overlapping different genomic regions would otherwise be misannotated. The following priorities were then set for the annotation in genomic regions: CDS > 5'UTR > 3'UTR > ncRNA > intron > intergenic > transcript > gene. These priorities ensured that the coding region and untranslated regions were prioritised over more general terms. Peaks in transcripts without a defined name were annotated as "Unknown". Category enrichment analysis was performed using the Perseus software. The normalised peak scores (normalisation described below) were first batch corrected using ComBat (38). Data was then annotated with the GO database and underwent a Fisher exact test with Benjamini-Hoechberg FDR cutoff of 0.02. The metagene analysis was performed in R (version 4.3.3) using the metagene2 package (version 1.18.0) (software developed by Eric Fournier, Charles Joly Beauparlant, Cedric Lippens, and Arnaud Droit). First, the nucleotide with the highest crosslink score within each peak, here termed "summit", was identified by comparing each myc-LARP6 FL iCLIP peak (peak score > 3) with its corresponding crosslink signal. The myc-LARP6 FL summits were saved in .bed format and used as the regions of interest in metagene2, with a padding of 50 nt on each side (padding_size = 50). Next, the individual crosslink files for each myc-LARP6 replicate were merged and converted to .bam files with bedtools and used as the alignment files in metagene2. Each dataset was normalised to 100% to obtain the normalised crosslink density for each nucleotide surrounding the summit, also accounting for crosslink score. Metagene2 was run with default parameters, with the exception of: strand_specific = TRUE, bin_count = 100, assay = "RNAseq". Positional motif analysis was performed using PEKA (version 0.1.5) (39) to identify LARP6 binding motifs. First, crosslinks from the three replicate files were merged and sorted by position using bedtools. Next, a genome-wide segmentation file that takes into account the different genomic regions was obtained from the "segment" function in iCount (40). The human genome

GRCh38.p13 GENCODE V42 was used in this analysis. The k-mer length was set to 5 (default). Motifs in the whole gene and in the 3'UTR were investigated.

Pythia analysis

Pythia is a deep twin-convolutional neural network (41). One of the arms of the network is a regular trainable convolutional arm capturing sequence context (sequence arm). The other arm is an untrainable fixed dilated convolutional layer capturing base pairing among nucleotides at different distances (structure arm). Pythia integrates the embedding space of the sequence and structure arms to predict RBP binding. We trained Pythia to distinguish peaks that bind LARP6, and its deletion mutant variants, compared to shuffled sequences with a similar dinucleotide distribution. To quantify the structure versus sequence contribution of each peak, we used DeepLIFT (42) to quantify the impact of permuting the input to the sequence and structure arms on the model. We computed the log ratio of the normalized sequencing arm DeepLIFT scores compared to the normalized structure arm DeepLIFT scores. Peaks with higher scores were considered to leverage more from sequence preference of LARP6, while peaks with lower scores leverage more from the structural preference of LARP6.

Total RNA-Seq for iCLIP normalisation

Total RNA-Seq was performed on cells grown under the same conditions as the ones used in the iCLIP experiment to normalise the iCLIP results for transcript abundance. The normalisation was performed to ensure that what was observed in iCLIP was due to true differences in binding and not to differences in RNA abundance. Total RNA was extracted with TRIzol, quantified with the Qubit RNA High-sensitivity assay, and the quality was validated with the Agilent High Sensitivity RNA ScreenTape assay, with all samples having a RIN ≥ 9.3. The whole transcriptome RNAseq library preparation and sequencing was outsourced to Novogene. The obtained demultiplexed reads were then trimmed with TrimGalore! (version 0.6.10) to remove the remaining adapter sequences and low-quality reads. Reads were then mapped to the human genome GRCh38.p13 GENCODE V42 with STAR (version 2.7.10b). Default parameters were used, and gene counts were generated directly from the mapping step by enabling the –quantMode GenerateCounts option. The gene counts from the RNAseq were used for iCLIP data normalisation. The iCLIP peak scores were normalised for transcript abundance according to the following formula: $normalised\ peak\ score = \frac{(peak\ score}{gene\ count)*1000}$.

Generation of LARP6 bacterial expression constructs

The human LARP6 sequence was inserted in the bacterial expression vector pET28 using a cloning-by-PCR strategy, and was engineered to contain an N-terminal 6xHis-SUMO-tag in frame with a Tobacco Etch Virus (TEV) protease cleavage site. The LARP6 constructs generated were the following: FL (2–491), NTR (2–83), and NTD (2–300). The deletions were introduced in the LARP6 FL sequence using Q5 site-directed mutagenesis (NEB). Sequences were validated with Sanger sequencing. The La-Module (70–300) construct was generated as described in (9).

Recombinant protein expression and purification

The LARP6 recombinant proteins were produced in *Escherichia coli* Rosetta 2 cells cultured in Luria broth supplemented with antibiotics. Recombinant protein expression was induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) and bacteria were cultured at 18°C ON in a shaking incubator. Protein purification was performed as described previously (9). Briefly, bacterial pellets were first lysed in bacterial lysis buffer (50 mM Tris-HCl, pH 8, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 5 µM lysozyme, cOmplete protease inhibitor cocktail (Roche),

2 mM PMSF), sonicated with a probe sonicator (Fisherbrand) on ice, and cleared by centrifugation at 38000 *g* at 4°C. Samples were initially purified on a Ni-immobilised affinity chromatography (Ni-IMAC) 5 mL His-Trap FF affinity column (Cytiva) previously equilibrated in 50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole, 5% glycerol. The proteins were eluted using a linear imidazole gradient from 10 to 500 mM. The tags were subsequently cleaved by ON incubation at 4°C with a 6xHis-TEV protease and the untagged proteins were recovered with a reverse nickel affinity chromatography using the Super Ni-NTA affinity resin (Generon). Nucleic acid contaminants were removed by loading the samples on a 5 mL HiTrap Heparin HP affinity column (Cytiva, for LARP6 FL, La-module, and NTD proteins) or a 5 mL HiTrap DEAE FF affinity column (Cytiva, for LARP6 NTR protein), eluting bound species with a linear KCl gradient from 0.1 to 1 M. The final purified products were dialysed in 20 mM Tris-HCl, pH 7.25, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT and concentrated to the desired concentration. The protein concentration was measured at 280 nm absorbance and corrected by the theoretical extinction coefficient derived from the ExPASy ProtParam tool (43). Protein aliquots were snap frozen and stored at -80°C.

Circular dichroism

Recombinant proteins were first diluted in 20 mM Tris-HCl, pH 7.25, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT to a concentration of 0.2 mg/mL. UV and CD spectra were acquired on a Chirascan Plus spectrophotometer (Applied Photophysics) under constant nitrogen flush, as described before (44). Suprasil rectangular cells of 10 mm and 0.5 mm path length (Starna Scientific) were used for the measurements. The experiments were run at 25°C with 1 nm spectral bandwidth, 1 nm data step-size, and 1.5 seconds time-per-data-point. First, UV spectra were acquired between 230–400nm to validate the exact concentration of the recombinant proteins. Next, far-UV CD spectra were acquired between 195–260nm. The UV spectra were corrected for the scattering and smoothened using Savitsky–Golay smoothing with a convolution width of 4 points. The far-UV CD data in millidegrees was then converted to mean residue ellipticity (Θ, $deg*cm^2*dmol^{-1}$), accounting for the exact concentration of each sample calculated from the UV spectra.

Biolayer interferometry (BLI)

BLI experiments were performed on an Octet R8 (Sartorius) instrument, using the Octet BLI Discovery software (12.2.2.20), and Octet BLI Analysis software (12.2.2.4). Recombinant protein concentrations were measured using a V-760 UV-Visible Spectrophotometer (JASCO). The 5' biotinylated RNA oligos and recombinant proteins were diluted in BLI buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.5 mM $MgCl_2$, 1 mM TCEP, 0.05% Tween-20). The assays were carried out at 25°C, in 96-well plates (Greiner), with a sample volume of 200 µL and with 1000 rpm shaking. After hydration of the sensors for at least 20 minutes, the biotinylated RNAs were immobilised on streptavidin-coated biosensors (Octet® SA Biosensors, Sartorius) to a final concentration of 0.10–0.15 µg/mL and incubated with varying concentrations of LARP6 NTD, La-module, and NTR. The protein concentrations used ranged between 0-31.3 nM for the La-module protein, and 0–500 nM for the NTD and NTR proteins. Sensor regeneration in between technical repeats was performed by dipping the sensors in RNase-free 1 M NaCl for 30 seconds, followed by 30 seconds in BLI buffer, three times. Experiments were performed at least in triplicates, using freshly prepared protein solutions. Control experiments to assess non-specific protein interactions with the sensors were performed by repeating the assays with sensors with no RNA loaded. Equilibrium dissociation constants (K_d) for the RNA-protein interactions were estimated by performing steady state analysis and plotting the instrument response at equilibrium as a function of protein concentration. The data was fitted using the Octet BLI Analysis software assuming a 1:1

interaction and using non-linear least squares regression. Association constant values (K_a) were obtained from K_d values using the formula $K_a = \frac{1}{K_d}$.

Computational disorder predictions

The IUPred2A webserver (iupred2a.elte.hu) (45,46) was used to predict the likelihood of intrinsic disorder in LARP6 based on its amino acid sequence using an energy estimation approach. The IUPred2 algorithm was used to identify disordered protein regions. The prediction type default option of "long disorder" was used, to predict global structural disorder of at least 30 consecutive residues. A context-dependent prediction of disorder was performed with the ANCHOR2 algorithm, which identified disordered binding regions.

Western blotting

Cell pellets were lysed in SDS buffer (100 mM Tris-HCl pH 7.5, 4% SDS), sonicated with a probe sonicator (SoniPrep 150, MSE, amplitude 8, 10 sec), and cleared. Total protein concentration was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher). 10–15 µg total protein was prepared for western blotting by addition of 1X NuPAGE LDS Sample Buffer (Thermo Fisher) and 100 mM DTT (Thermo Fisher). Samples were denatured by boiling at 95°C for 5 minutes. Samples were then loaded on 4–12% NuPAGE Bis-Tris Midi protein gels (Thermo Fisher) and run in NuPAGE MOPS running buffer (Thermo Fisher) at 120 V for at least 90 minutes. Transfer was performed using Immobilion-P PVDF membranes (Millipore) in transfer buffer (1X Tris-glycine, 20% SDS) for two hours at 1 A, at 4°C. Membranes were then blocked in 5% milk (diluted in 0.05% PBS-Tween-20) for one hour at RT. Incubation with the primary antibody diluted in 5% milk was performed ON at 4°C, followed by three 10 minutes washes of the membranes with 0.05% PBS-Tween-20. HRP-conjugated secondary antibodies were then diluted in 5% milk and incubated for one hour at RT, followed by three 10 minutes washes with 0.05% PBS-Tween-20. Membranes were probed with SuperSignal West Pico PLUS chemiluminescence substrate (Thermo Fisher) and imaged on an Amersham Imager 600 (GE Healthcare). Band intensities were detected using Fiji.

Immunofluorescence

46000 cells were plated in glass coverslips placed in 12-well tissue culture plates and allowed to grow for 24 hours. Cells were then fixed in 4% paraformaldehyde (diluted in PBS) for 10 minutes and permeabilised in 0.2% Triton X-100 (diluted in PBS) for 15 minutes. Following three washes with PBS, cells were blocked in 4% BSA (diluted in PBS) for 30 minutes at RT. Primary antibodies were diluted in 4% BSA and incubated for one hour at RT, followed by three washes with PBS. Fluorophore-conjugated secondary antibodies and Hoechst 33258 (1:2000, Invitrogen) were diluted in 4% BSA and incubated for one hour at RT, followed by three washes with PBS. Coverslips were then washed in water and mounted on slides using ProLong Diamond Antifade Mountant (Invitrogen). Coverslips were imaged at a Nikon ECLIPSE Ti Spinning Disk confocal microscope with a plan apochromat Lambda 60X/1.40 oil objective. Fluorophores were excited with LUN-V 405 nm, 488 nm, and 640 nm lasers. Three z-stacks were acquired 10 µm apart. The following settings were used: 2x2 binning, 540 MHz readout rate, 12-bit and gain 4-bit depth. At least five separate images were acquired for each replicate. Images were processed using Fiji.

Cell Viability assay

Cell viability was measured with the CTG 2.0 Luminescent Cell Viability assay (Promega). 5000 cells were seeded in 96-well plates at least in quadruplicates in 100 µL volume and allowed to grow for 72 hours. The CTG reagent was diluted 1:2 in PBS and 100 µL of diluted CTG reagent were added to each well and mixed. The plate was then incubated on a shaker

for 2 minutes, followed by 10 more minutes at RT, protected from light. Samples were then transferred from the tissue culture-grade 96-well plate to a white CELLSTAR 96-well plate (Greiner) suitable for luminescence measurements. Luminescence was measured at a FLUOstar Omega Microplate Reader (BMG Labtech). The background luminescence signal from a well containing media only was subtracted from each sample luminescence signal.

Statistical analysis

The statistical test applied to each experiment is indicated in the corresponding figure legend. Analysis of RNA localisation, invasion, colony formation, and CTG assays was performed in GraphPad Prism version 10. The *p*-value cutoff was 0.05, with * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. Normality was not assumed, unless if the data distribution appeared normal-like (e.g. for RP-mRNA protrusion to cell body ratio values).

Custom oligonucleotides

Oligo ID	Sequence (5'-3')	Company
siRNAs		
Custom LARP6 siRNA	GGGAGAAAGGCCUGGAAUUUU	Horizon Discovery
ON-TARGETplus Non-targeting Control Pool (scramble)	UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUACAUGUUUUCCUA	Horizon Discovery
iCLIP L3-???-App adapters		
L3-ATT-App	/5Phos/WN ATT AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
L3-AGG-App	/5Phos/WN AGG AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
L3-TTA-App	/5Phos/WN TTA AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
L3-TGC-App	/5Phos/WN TGC AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
L3-CTG-App	/5Phos/WN CTG AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
L3-CGT-App	/5Phos/WN CGT AGA TCG GAA GAG CGG TTC AG/3Bio/	IDT
iCLIP reverse transcription primers		
irCLIP_ddRT_12	/5Phos/ WWW GTGGA NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC	IDT
irCLIP_ddRT_13	/5Phos/ WWW TCCGG NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC	IDT
irCLIP_ddRT_14	/5Phos/ WWW TGCCT NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC	IDT
iCLIP PCR primers		
P3 Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGG CATTCCTGCTGAACCGCTCTTCCGATCT	IDT

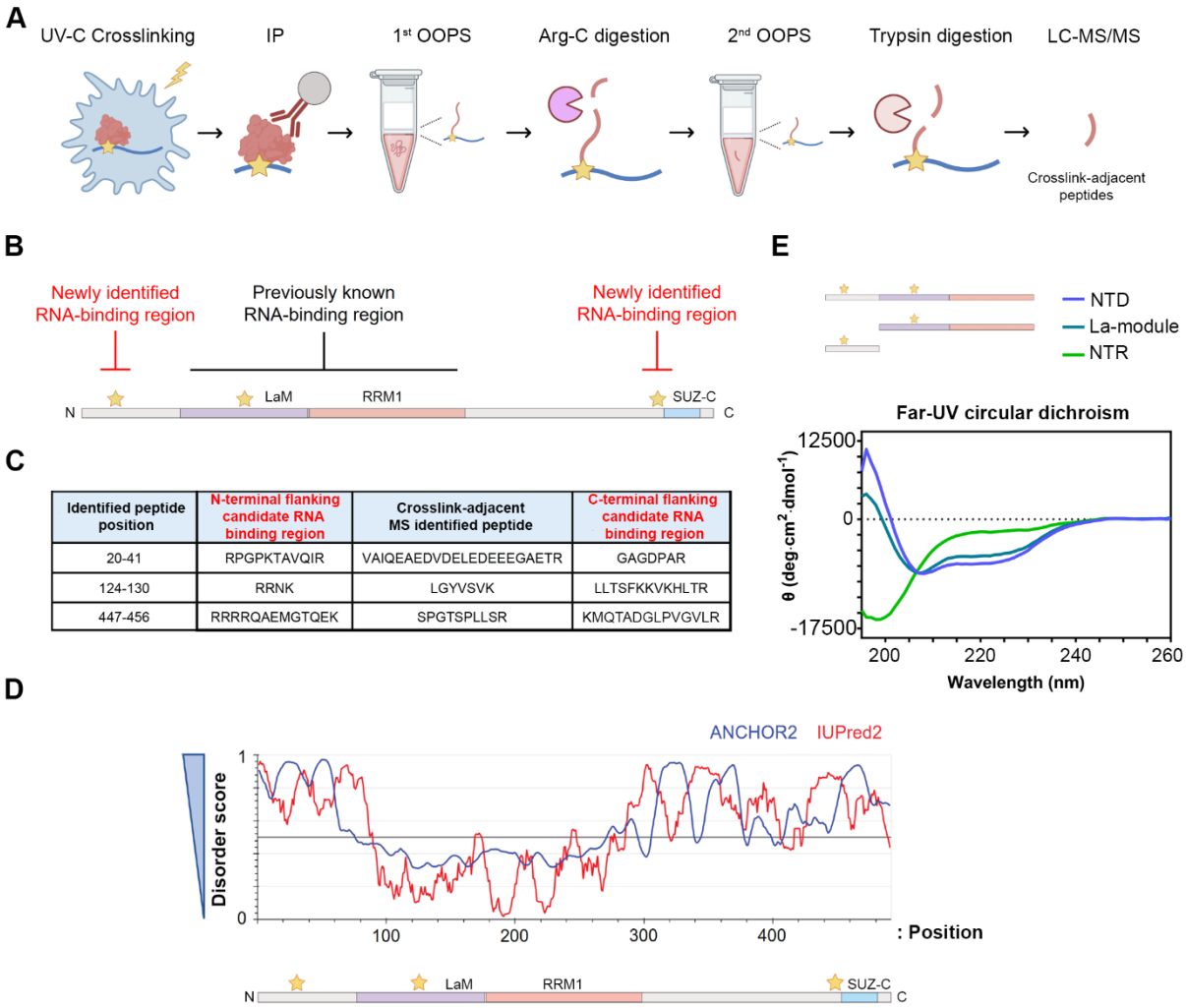
P5 Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	IDT
BLI 5' biotinylated RNA oligos		
Peak 1 (P1)	CUAAAUACAACACUGAUACUAGAUU	Horizon Discovery
Peak 2 (P2)	GGAAAUGGGCAGACUGAACCAGUCC	Horizon Discovery
Upstream 1 (U1)	ACUGUUCGUCACUCAAAUGAAUUUG	Horizon Discovery
Upstream 2 (U2)	CCCUCGGGGCUCCUGAAUAUCAGUC	Horizon Discovery
Downstream 1 (D1)	AGGUGGUGAAUUUUCCAAGAACAUA	Horizon Discovery
Downstream 2 (D2)	GUUUAAGUUGAUUAAAAAUGCUUUU	Horizon Discovery
PolyC	CCCCCCCCCCCCCCCCCCCCCCCCCC	Horizon Discovery

Table 1: List of all custom-made RNA and DNA oligonucleotides used in this study.

Results:

LARP6 interacts with RNA through both ordered and intrinsically disordered regions.
To better understand the complexities of LARP6 RNA-binding activity, we aimed to define its regions of contact with RNA in living cells in an unbiased manner. For this purpose, we developed a mass spectrometry-based approach termed IPOOPS (Immunoprecipitation coupled with Orthogonal Organic Phase Separation) (Figure 1A). This approach combines the isolation of UV-C crosslinked RNA-protein adducts via OOPS (19), with the ability to map protein regions that are attached to RNA by sequential Arg-C and trypsin digestions, as pioneered in the RBDmap method (6). Briefly, GFP-trap beads were used to purify LARP6-RNA complexes from UV-crosslinked MDA-MB231 cells that stably expressed GFP-LARP6 (17) (Supplementary Figure 1A). The RNA-crosslinked LARP6 was then isolated using an initial OOPS, before partial digestion with Arg-C, which generated smaller segments, with or without RNA adducts. The RNA-bound segments were further purified from the unbound segments *via* a second OOPS, followed by trypsin digestion, releasing MS-detectable peptides that were adjacent to a crosslinked region (Figure 1A). The crosslinked peptides themselves are undetectable in this approach, due to the addition of RNA mass. However, through identification of their adjacent peptides, the likely sites of crosslinking can be postulated.

Using IPOOPS, we identified three RNA-bound regions within LARP6 (Figure 1B & 1C). These included an RNA-binding segment within the La-motif, as well as two previously unknown RNA-binding regions in the N- and C-terminal parts of the protein (Figure 1B & 1C). The RNA-binding segment within the La-motif likely contains phenylalanine 135, one of six highly conserved residues in the RNA-binding pocket of LARP6 and shown by mutagenesis to be essential for RNA association (9), providing further confidence to our methodology. The newly discovered C-terminal RNA-binding region is adjacent to the SUZ-C motif, while the newly identified N-terminal RNA-binding region precedes the La-module (Figure 1B & 1C).



Identified peptide position	N-terminal flanking candidate RNA binding region	Crosslink-adjacent MS identified peptide	C-terminal flanking candidate RNA binding region
20-41	RPGPKTAVQIR	VAIQEAEDVDELEDEEEGAETR	GAGDPAR
124-130	RRNK	LGYVSVK	LLTSFKKVKHLTR
447-456	RRRRQAEMGTQEK	SPGTSPLLSR	KMQTADGLPVGVLR

Figure 1: LARP6 harbours two new disordered RNA-binding regions.

A) IPOOPS workflow. The detailed protocol is described in Material and Methods and in the text. Star: crosslinking site; red: protein; blue: RNA.

B) Schematic representation of the RNA-binding regions identified in LARP6 via IPOOPS. The crosslinking-adjacent peptides, directly identified by mass spectrometry, revealed likely areas of RNA crosslinking (marked by stars). LARP6 domains are represented to scale.

C) IPOOPS-identified peptides and their N- and C-terminal flanking candidate RNA-crosslinked regions. The identified crosslink-adjacent peptides are retained in the interface of the 1st OOPS due to their proximity to the RNA crosslink sites but are released in the 2nd OOPS from the crosslink sites following trypsin digestion.

D) Computational disorder prediction for human LARP6 protein sequence. The web server IUPred2A (https://iupred2a.elte.hu/) (45,46) was used to predict the likelihood of disorder for LARP6, based on ANCHOR2 or IUPred2 algorithms. Higher disorder scores denote higher likelihoods of the protein region being intrinsically disordered. The IPOOPS-identified RNA-binding regions are marked by stars.

E) Far-UV CD spectra of LARP6 variants. LARP6 NTD, La-module, and NTR purified proteins were analysed by circular dichroism to investigate their secondary structure content. The negative band

around 200 nm, characteristic of intrinsic disorder, is visible specifically in the NTR CD spectrum. The spectra were normalised for protein concentrations.

Next, we investigated the structural characteristics of the newly identified RNA-binding regions. Computational prediction of disorder using IUPred2A (45) suggested that both the NTR and CTR are likely intrinsically disordered (Figure 1D). We further validated these predictions for the NTR, using circular dichroism (CD), an absorption spectroscopy method that measures the protein secondary structure content (47). In agreement with our solved structure of the La-module of LARP6 (9), purified segments of LARP6 containing the La-module (La-mod), or the La-module + NTR (NTD), exhibited CD spectra indicative of both α-helical and β-sheet secondary structures (Figure 1E). In contrast, the purified NTR alone exhibited a CD profile dominated by an unfolded protein profile (Figure 1E). Together, these results reveal that in addition to the folded La-module, LARP6 contains two previously unknown RNA-binding regions in its N- and C-terminal parts, which are likely to be IDRs.

iCLIP analysis reveals the La-module as essential for LARP6 binding to RNA.
Next, we set out to study the RNA targets of each RNA-binding region of LARP6, in living cells. To reveal the full repertoire of LARP6 binding targets, we chose U-87 MG glioblastoma cells as our cell model, since LARP6 is highly expressed in gliomas (Supplementary Figure 2A), with U-87 MG cells being among the highest LARP6-expressing established cell lines in wide use (Supplementary Figures 2B). LARP6 protein expression levels in U-87 MG cells surpass those of MDA-MB231 cells, which were previously used as a cell model for LARP6 function, making them likely an even more cancer-relevant cell model for investigating LARP6 function (Supplementary Figure 2C). To assess the contribution of each region of LARP6 to its RNA-binding activity, we used a doxycycline-inducible system to express full-length and various deletion mutants of a myc-tagged LARP6, in a controlled manner (Supplementary Figure 2D–G). These cells were then subjected to iCLIP analysis using beads conjugated to a monoclonal anti-myc tag antibody (32) (Figure 2A). The iCLIP experiment was performed in triplicate for each LARP6 variant, with parental U-87 MG cells serving as a negative control.

The resulting iCLIP reads, representing precise RNA-protein crosslink sites, were subsequently mapped to the human transcriptome. We then searched for clusters of reads across the transcriptome, which identified peaks that correspond to the likely LARP6 binding sites (Dataset S1). A total of 14,334 peaks were detected for the full-length (FL) LARP6, most of which were lost upon deletion of the La-module (Figure 2B), demonstrating the La-module's crucial role for the RNA-binding activity of LARP6. In contrast, deletion of the NTR or CTR did not reduce the number of identified peaks. Instead, mapped peak numbers in ΔNTR and Δ CTR mutants increased (Figure 2B, Dataset S2, & Dataset S3). Amongst the identified peaks that were conserved across FL and both mutants, we detected LARP6's well-known binding sites in the 5′UTR of Collagen-I mRNA (10), and the 5′TOP sequences of various ribosomal protein mRNAs (RP-mRNAs) (17) (Figure 2C). Mapping of the peaks to the different genomic regions revealed the vast majority to correspond to mRNAs, with the highest binding density occurring in the 3'UTRs, a binding pattern that was largely unaffected by the NTR or CTR deletions (Figure 2D).

We then assessed the functional categories of mRNAs that associate with LARP6. Annotation enrichment analysis of LARP6 mRNA targets revealed an overrepresentation of transcripts involved in ribosome biogenesis, cell adhesion and migration, intracellular trafficking, and RNA processing, amongst others (Figure 2E). Importantly, this target repertoire was also unaffected by the NTR or CTR deletions (Figure 2E). Together, these findings suggest that unlike the La-

module, the NTR and CTR are dispensable for the bulk of the RNA-binding activity of LARP6 in living cells.

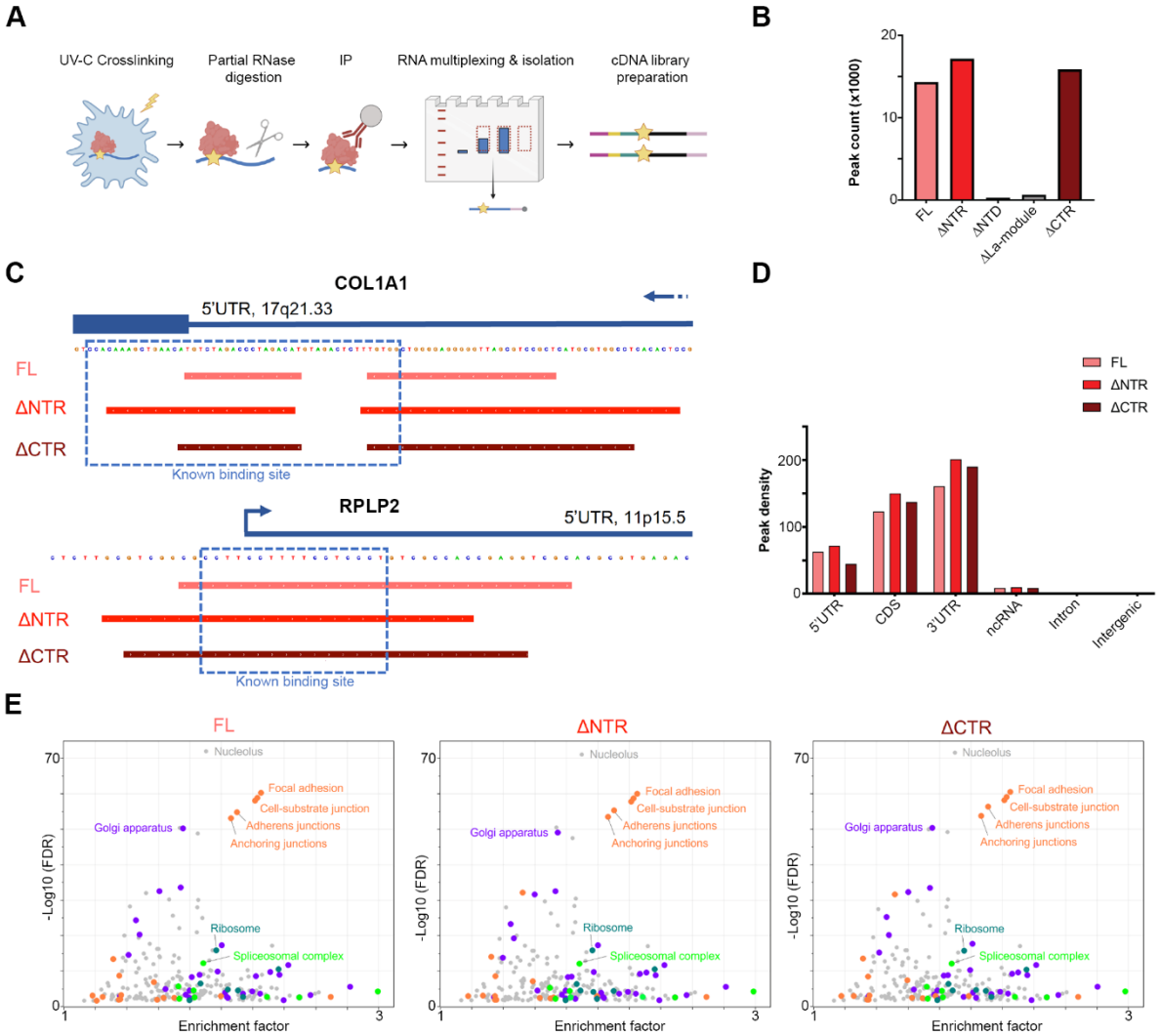


Figure 2: The La-module is essential for the RNA-binding activity of LARP6 in living cells.

A) Schematic diagram of the iCLIP workflow.

B) Total number of iCLIP peaks (representing distinct binding sites), which were identified for each myc-LARP6 variant. The ΔNTD and ΔLa-module mutants did not bind to RNA in cells and were therefore excluded from further downstream analysis.

C) iCLIP binding profile of indicated myc-LARP6 variants in the 5'UTR of COL1A1 mRNA, as well as the 5'UTR of RPLP2 ribosomal protein mRNA. The identified peak regions with respect to the previously known sites of binding are marked.

D) Normalised distribution of iCLIP peaks across different genomic regions. The peak numbers were normalised to the length of each genomic region to calculate peak density. LARP6 mostly binds to coding transcripts, with the majority of binding occurring in the 3'UTR regions.

E) GO cellular component category enrichment analysis of the FL, ΔNTR, and ΔCTR myc-LARP6-bound transcripts. The iCLIP data was normalised to total RNAseq data to account for transcript abundance.

The NTR of LARP6 mediates local RNA-binding selectivity in living cells.
Next, we examined the underlying principles of RNA target recognition by LARP6. For this purpose, we used Pythia, a neural network-based tool designed to model RBP preferences from CLIP data (41). Pythia utilises context-free grammars (CFG) to capture both sequence as well as structural determinants of RNA binding (48-50). We used the identified iCLIP peaks as input to train models for FL, as well as ΔNTR and ΔCTR mutants. The data were divided into training (64%), tuning (16%), and validation (20%) subsets with non-overlapping chromosomes among the 3 data splits. High-quality models were successfully trained for each iCLIP dataset, with the ability to predict each LARP6 variant binding sites with high accuracy and precision (Supplementary Figures 3A & 3B). We then evaluated the performance of each model for predicting the binding sites of the other variants, as an indicator of divergence in their RNA binding. Area under the precision-recall curve (auPRC) was calculated as the performance metric for each model. The ΔCTR model exhibited an auPRC value of 86%, when applied to its own validation dataset, but this performance was slightly reduced when the model was applied to the FL and ΔNTR datasets, with auPRC values dropping to 78% and 80%, respectively (Figure 3A). The ΔNTR model exhibited a similar performance, with an auPRC value of 86% when applied to its own validation dataset, but the performance was slightly more compromised when the model was applied to the FL and ΔCTR datasets, with auPRC values dropping to 76% in both cases (Figure 3B). These results suggest that both CTR and NTR somewhat affect target recognition by LARP6, with the impact of the NTR being slightly more prominent.

Next, we assessed the relative contributions of sequence and structure towards RNA recognition by each LARP6 variant. Consistent with previous SELEX findings (18), FL LARP6 appeared to leverage on both sequence and structural information for identifying its target sites across the transcriptome, with a slight bias towards structure (Figure 3C). A similar profile was detected for the ΔCTR mutant, but this balance was significantly disrupted when the NTR was deleted, with the binding activity becoming slightly more sequence-dependent (Figure 3C). These analyses suggest that the NTR significantly affects how LARP6 recognises its contact sites with RNA, with sequence-based information becoming more prominent in selecting the binding sites.

To gain further insights into the exact consequences of IDR deletions on LARP6-RNA sites of interactions, we also conducted a metagene analysis of LARP6 crosslinking footprints across its mapped binding sites (i.e. peaks) in the transcriptome. The meta-profile of FL LARP6 crosslinks showed a highly narrow interaction footprint, which remained mostly unchanged upon CTR deletion (Figure 3D). In contrast, deletion of the NTR led to a clear broadening of the interaction footprints around the main binding site (Figure 3E), indicating that the NTR is important for enhancing the local RNA-binding selectivity of LARP6. Based on these results, we conclude that the observed increase in sequence-dependent RNA-binding bias by Pythia in the ΔNTR mutant, is likely reflective of this loss in local RNA-binding selectivity, resulting in the emergence of new weaker peripheral RNA-protein interactions, which are likely often devoid of any conserved structural RNA features.

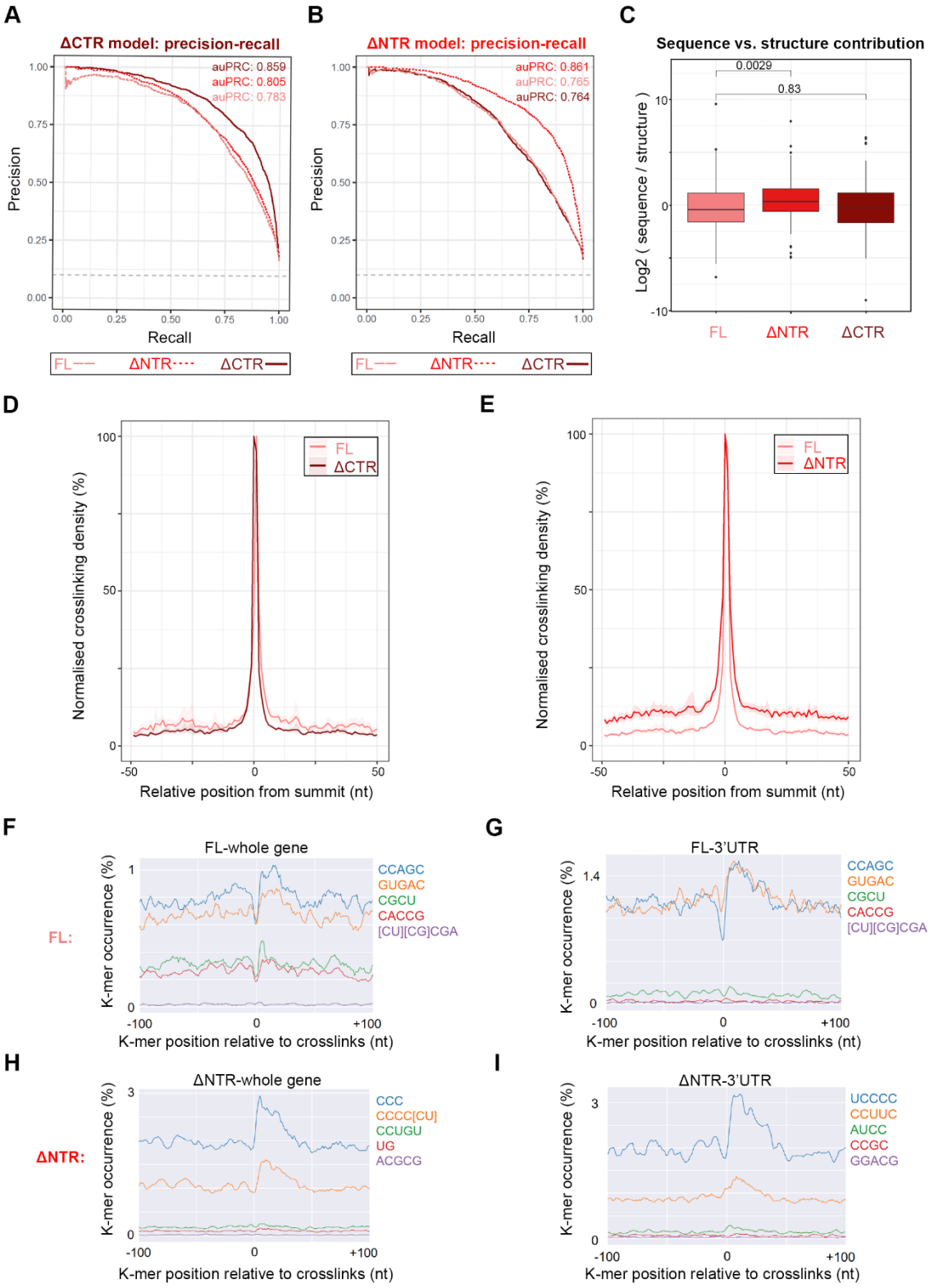


Figure 3: The N-terminal IDR of LARP6 fine-tunes its RNA-binding activity in living cells.

A) Precision-Recall curves for the ΔCTR Pythia model, applied to the FL, ΔNTR, and ΔCTR myc-LARP6 iCLIP datasets. Area under the Precision-Recall curve (auPRC) values were calculated as measures of the model's performance in predicting the RNA-binding sites for each dataset.

B) Precision-Recall curves for the ΔNTR Pythia model, applied to the FL, ΔCTR, and ΔNTR myc-LARP6 iCLIP datasets. Area under the Precision-Recall curve (auPRC) values were calculated as measures of the model's performance in predicting the RNA-binding sites for each dataset.

C) Comparisons of sequence to structure contributions toward the Pythia model performances for the FL, ΔNTR and ΔCTR myc-LARP6 iCLIP datasets. Log2 of DeepLIFT randomised sequence to structure ratio values for each model are displayed. Deletion of the NTR results in a significant increase in the sequence contributions towards the RNA-binding activity of LARP6. P-values are reported on the plots.

D) Metagene analysis of the iCLIP binding sites from myc-LARP6 FL and ΔCTR. The normalised crosslink densities observed in the 50 nt surrounding each peak summit (the nucleotide with the highest crosslink score within a peak) are displayed. The shaded regions represent the 95% confidence interval.

E) Metagene analysis of the iCLIP binding sites from myc-LARP6 FL and ΔNTR. The normalised crosslink densities observed in the 50 nt surrounding each peak summit (the nucleotide with the highest crosslink score within a peak) are displayed. The shaded regions represent the 95% confidence interval.

F) Positional motif analysis of FL myc-LARP6 FL iCLIP contact sites performed with PEKA (39). The significantly enriched 5-mer motifs in the 100 nt surrounding the crosslink sites in the whole genome are displayed.

G) Positional motif analysis of FL myc-LARP6 FL iCLIP contact sites performed with PEKA (39). The significantly enriched 5-mer motifs in the 100 nt surrounding the crosslink sites in the 3'UTR regions are displayed.

H) Positional motif analysis of ΔNTR myc-LARP6 FL iCLIP contact sites performed with PEKA (39). The significantly enriched 5-mer motifs in the 100 nt surrounding the crosslink sites in the whole genome are displayed.

I) Positional motif analysis of ΔNTR myc-LARP6 FL iCLIP contact sites performed with PEKA (39). The significantly enriched 5-mer motifs in the 100 nt surrounding the crosslink sites in the 3'UTR regions are displayed.

Since Pythia analysis revealed a change in the usage of sequence-based information for target recognition by the ΔNTR mutant, we next looked for significant linear sequence motifs that were present in the vicinity of crosslinking sites for FL LARP6 and the ΔNTR mutant. Positionally enriched k-mer analysis (PEKA) was employed to identify significantly enriched linear motifs, using the default package settings (39). We applied the motif analysis to both the full iCLIP dataset, as well as only the 3'UTR regions, where the majority of LARP6 binding sites reside. The FL exhibited a significant enrichment for several relatively complex sequence motifs, an observation in-line with the previously published SELEX findings (18) (Figure 3F & 3G). Loss of the NTR, however, transformed this pattern, with simpler pyrimidine-rich linear motifs dominating the binding profile (Figure 3H & 3I). A minimal difference was observed between the full dataset and the 3'UTRs, suggesting the observed changes are not restricted to a specific group of binding targets (Figure 3F–I). Collectively, these results suggest that while the La-module is crucial for mediating RNA-LARP6 interactions, the NTR plays a role in fine-tuning the local selectivity of the La-module.

***In vitro* biophysical analysis validates a role for NTR in local RNA-binding selectivity.**

To determine whether the impact of NTR on local RNA-binding selectivity is an intrinsic RNA-binding property of this protein region, rather than a result of additional cellular factors, we next investigated LARP6 interactions with its RNA targets, *in vitro*. A segment within the 3'UTR of the CTNNA1 mRNA, which encodes a cytoskeletal protein that bridges cadherin molecules with actin filaments (51), was selected as the representative target of LARP6 for *in vitro*

analysis. This choice was due to the strong yet highly focused binding of FL LARP6 to this region in the form of two adjacent iCLIP peaks, alongside weaker interactions detected only with the NTR-deleted mutant (Figure 4A), a pattern exemplifying our metagene analysis-based observation. We designed 25 nucleotide-long RNA oligonucleotides corresponding to the two primary peaks (P), as well as four additional surrounding peaks located upstream (U) or downstream (D) of the main binding sites (Figure 4A). As a control, we used a polyC oligonucleotide, which is not expected to bind to LARP6 FL, but could potentially interact with the La-module in the absence of NTR, based on our iCLIP results (Figure 3F–I). These RNA oligos were biotinylated for use in biolayer interferometry (BLI) analysis (52). Biotinylated oligonucleotides were subsequently immobilised on streptavidin-coated biosensors to measure the association-dissociation kinetics with the purified La-module, and the NTD (La-module + NTR), obtained recombinantly from bacteria (Supplementary Figure 4A & 4B). A control experiment without any RNA loaded on the sensors was also conducted to rule out non-specific protein binding to the sensors (Supplementary Figure 4C & 4D). Steady-state analysis was then used to generate equilibrium binding curves for each oligo, by plotting the instrument response against protein concentration. The NTD exhibited varying affinities for the different oligonucleotides representing the CTNNA1 peaks, with low nanomolar affinity towards one of the main peaks (17.3 nM), but weaker high nanomolar binding to the peripheral upstream and downstream peaks (Figure 4B & Table 2). In contrast, the La-module showed comparatively similar binding curves, with affinities in the low nanomolar range (0.9–14 nM), across the different CTNNA1 oligonucleotides (Figure 4C & Table 2). As expected, virtually no binding was detected between the NTD and the polyC oligo, whereas some binding to this oligo, albeit with lower affinity, was measured for the La-module (Figure 4B, 4C, & Table 2). We also assessed the binding of LARP6 NTR region alone to the oligo representing one of the main peaks, which was found to be ~1000 fold weaker than the La-module (Supplementary Figure 4E, 4F, & Table 2), consistent with our iCLIP findings that in the absence of the La-module, the NTR IDR cannot strongly associate with RNA.

LARP6 protein variant	RNA oligonucleotide	*Kd* (nM)
La-module	P1	1.5 ± 0.4
La-module	P2	0.91 ± 0.05
La-module	D1	2.0 ± 0.1
La-module	D2	6.9 ± 0.6
La-module	U1	14.0 ± 0.8
La-module	U2	4.4 ± 0.1
NTD	P1	61.0 ± 5.0
NTD	P2	17.3 ± 1.2
NTD	D1	85.0 ± 4.0
NTD	D2	367 ± 57
NTD	U1	587 ± 29
NTD	U2	183 ± 5.0
NTR	P1	1200 ± 500
La-module	polyC	63.0 ± 3.7
NTD	polyC	>7500

Table 2: List of BLI-estimated steady-state binding *Kd* values for the indicated LARP6 protein variants and RNA oligonucleotides.

Next, we calculated the relative association constants (K_a) of the La-module and NTD to each RNA target. In agreement with our iCLIP results, the NTD exhibited a very selective RNA-binding profile, with the tightest affinity for the oligo corresponding to one of the main peaks, and lower affinities for the surrounding oligo peaks (Figure 4D). Conversely, the La-module

exhibited a less selective affinity profile for all the CTNNA1 peak oligos tested, indicative of its reduced binding discrimination in the absence of the NTR (Figure 4E). Together, these results corroborate our iCLIP based findings from living cells, suggesting that NTR has an intrinsic role in enhancing the local binding selectivity of the La-module.

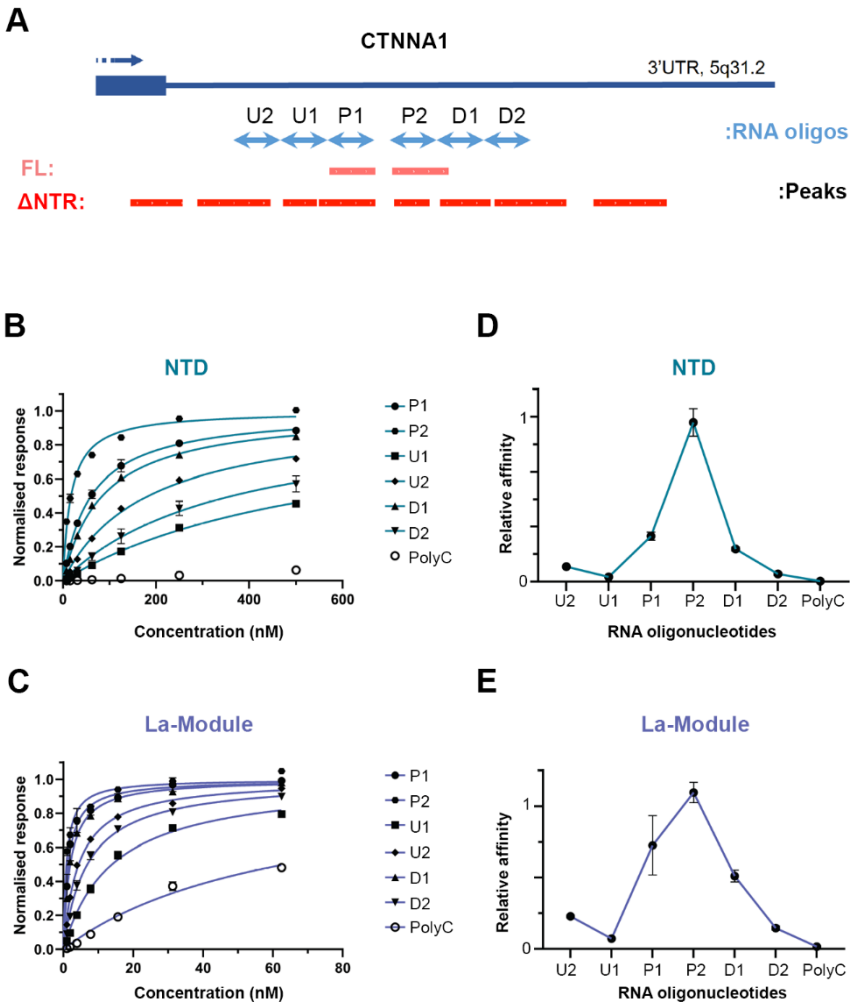


Figure 4: The NTR provides local binding selectivity to the La-module, *in vitro*.

A) The iCLIP binding profile of FL and ΔNTR myc-LARP6 in the 3'UTR of CTNNA1 mRNA. The 25 nt-long RNA oligonucleotides designed to match either the two main FL peaks (P1 and P2) or the 4 peripheral ΔNTR peaks (U1, U2, D1 and D2) are displayed in light blue.

B) Steady-state BLI binding analysis of LARP6 NTD to each indicated 5' biotinylated CTNNA1 3'UTR RNA oligonucleotide. A polyC oligonucleotide was used as control.

C) Steady-state BLI binding analysis of the LARP6 La-module to each indicated 5' biotinylated CTNNA1 3'UTR RNA oligonucleotide. A polyC oligonucleotide was used as control.

D) Relative K_a values, calculated from the BLI steady state analysis of LARP6 NTD (B) against each RNA oligonucleotide.

E) Relative K_a values, calculated from the BLI steady state analysis of LARP6 La-module (C) against each RNA oligonucleotide.

NTR-mediated selective RNA-binding is important for LARP6 function in cells.

Having revealed the N-terminal IDR of LARP6 as a key mediator of local RNA-binding selectivity, we next sought to determine whether this activity was important for LARP6 function in cells. To investigate this, we developed a knock-down-rescue system in which endogenous LARP6 could be selectively depleted using an siRNA oligonucleotide targeting its 3'UTR, followed by rescue with our previously described doxycycline-inducible myc-tagged LARP6 expression system that is siRNA-resistant due to lacking the endogenous LARP6 3'UTR (Supplementary Figure 5A). Our previous work had demonstrated a role for LARP6 in localising RP-mRNAs to the protrusive fronts of invasive breast cancer cells, as well as

supporting cancer cell viability, and promoting invasion (17). We therefore assessed whether LARP6 played a similar role in U-87 MG cells. Using a microporous filter-based method for isolating protrusions and cell bodies coupled with RNA sequencing (17), we first profiled RNA localisation to protrusions of U-87 MG cells in a global unbiased manner (Figure 5A & Dataset S4). As expected, RP-mRNAs were enriched in protrusions, but this localisation was abrogated following LARP6 depletion (Figure 5B & Dataset S4). Notably, rescue of LARP6 expression with either FL or NTR-deleted mutant myc-LARP6 restored the localisation of RP-mRNAs to protrusions (Figure 5B), demonstrating that this localisation does not significantly depend on the NTR function.

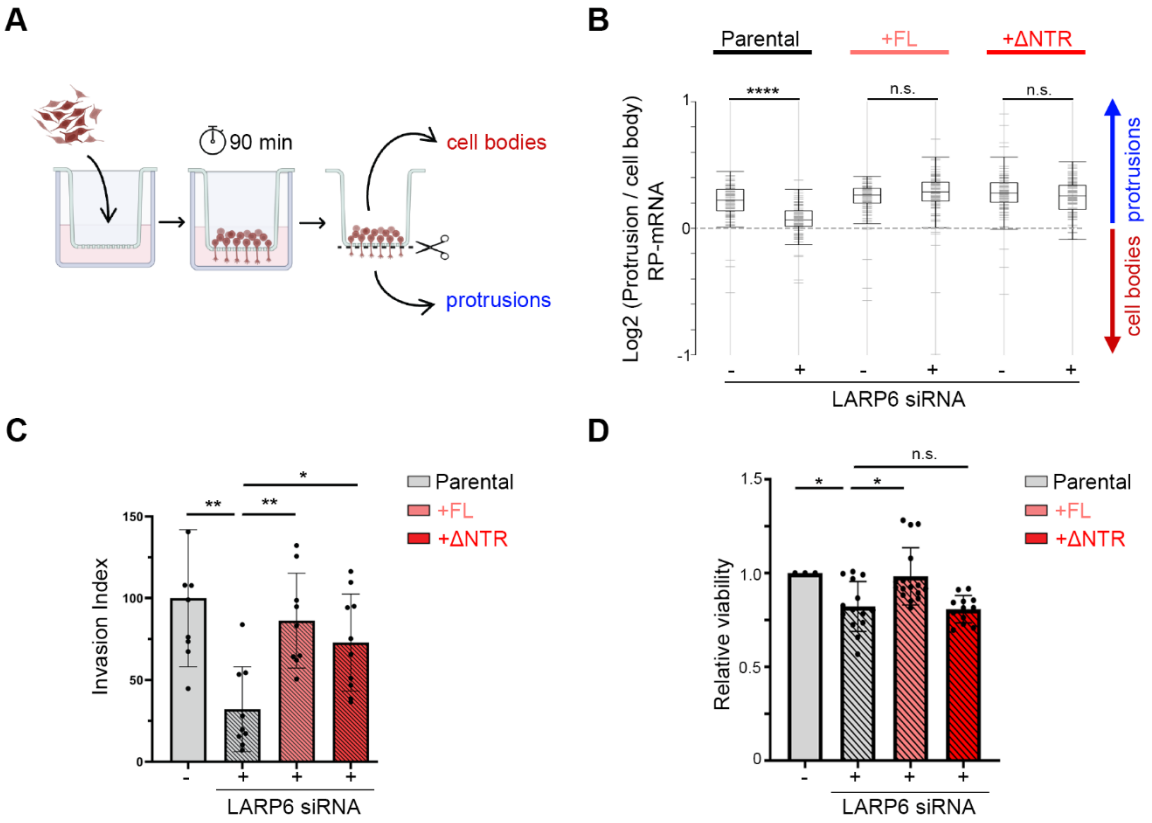


Figure 5: LARP6 NTR is critical for regulation of cancer cell viability and invasion.

A) Experimental workflow for the separation of protrusions and cell bodies using 3µm transwell filters. RNA was subsequently isolated from each fraction and subjected to 3' end RNAseq, to quantify mRNA enrichment in protrusions of the U-87 MG cells.

B) Profile plot of log2 protrusions/body ratio values for all ribosomal protein encoding mRNAs, in the indicated U-87 MG cells. Positive values indicate enrichment in protrusions (blue arrow), while negative values indicate enrichment in cell bodies (red arrow). An unpaired two-sample t-test was used for the calculation of significance (**** : $p \leq 0.0001$, n.s.: $p > 0.05$).

C) Normalised invasion index for the indicated U-87 MG cell populations upon endogenous LARP6 knockdown and rescue with doxycycline-inducible myc-LARP6 FL or ΔNTR. The values were normalised to the non-targeting parental control condition. A Kolmogorov–Smirnov test was applied for the significance calculation (* $p \leq 0.05$; ** $p \leq 0.01$).

D) Normalised cell viability for the indicated U-87 MG cell populations, upon endogenous LARP6 knockdown and rescue with doxycycline-inducible FL or ΔNTR myc-LARP6. Cell viability was assessed five days after siRNA transfection, using CTG assay. The values were normalised to the non-targeting

parental control condition. A Kolmogorov–Smirnov test was applied for the significance calculation (* $p \leq 0.05$, n.s.: not significant).

Next, we assessed the role of NTR in the regulation of invasion, using a 3-dimensional (3D) collagen-I invasion assay (28) (Supplementary Figure 5B). Depletion of LARP6 resulted in a strong reduction in the ability of parental U-87 MG cells to invade into the 3D collagen-I matrix (Figure 5C, and Supplementary Figure 5C). The effect of LARP6 depletion was comparable to inhibition of Rho Kinase, a potent regulator of cell invasion (29) (Supplementary Figure 5D & 5E), highlighting a key role for LARP6 in mediating U-87 MG cell invasion. Crucially, while rescuing LARP6 expression with FL myc-LARP6 could almost fully recover the invasiveness of U-87 MG cells, this rescue was only partial with the NTR-deleted mutant (Figure 5C & Supplementary Figure 5C), suggesting that the NTR is required for the full functionality of LARP6 in the context of invasion. We also assessed the effect of LARP6 depletion on the viability of U-87 MG cells. Knock-down of LARP6 resulted in a small but significant reduction in cell viability, an effect that was rescued by the re-introduction of myc-LARP6 FL (Figure 5D). The ΔNTR mutant, however, was not capable of rescuing this effect (Figure 5D), suggesting a key role for the NTR in the regulation of LARP6-mediated cancer cell viability. Collectively, our depletion-rescue experiments reveal that while the NTR is dispensable for some functions of LARP6 in cells, it is required for the full functionality of the protein in mediating cancer cell viability and invasion.

Discussion

IDRs have emerged as novel RNA-binding protein modalities in recent years, but their exact contributions towards the RNA-binding activity of most RBPs have remained enigmatic. In this study, we devised an unbiased mass spectrometry-based mapping approach to map the RNA interacting regions of LARP6, an RBP with diverse RNA-binding repertoire that is evolutionary conserved in all multicellular organisms (8,17,18). We reveal that, in addition to its well-characterised RBD, LARP6 makes direct contact with RNA in living cells via two IDRs in its N- and C-terminal regions. However, when these IDRs were deleted, the RNA-binding activity of LARP6 remained largely uncompromised. Through neural network modelling and metagene analysis of the binding patterns of wild-type and various deletion mutants of LARP6 across the transcriptome, we identified a role for the N-terminal IDR in fine-tuning the local RNA-binding selectivity of LARP6. Specifically, loss of this region resulted in the acquisition of weaker, peripheral RNA associations around the primary binding sites of LARP6, in living cells. Biophysical *in vitro* analysis validated a role for the N-terminal IDR in modulating the local RNA-binding selectivity of the tethered structured La-module. This reveals an intriguing interplay between tethered ordered and disordered regions in fine-tuning the interactions of LARP6 with RNA. Importantly, while this local decrease in binding selectivity did not prevent LARP6 from recognising its full repertoire of RNA targets, or mediating RNA localisation to cellular protrusions, it was important for promoting cancer cell viability and invasion. These results suggest that even subtle alterations in the local RNA-binding patterns of an RBP can exert significant effects on post-transcriptional regulation. This notion may provide some explanation for the prominence of germ-line mutations that are associated with human genetic disorders within the IDRs of RBPs (1). Such mutations are unlikely to completely abrogate the RNA-binding activity of an RBP, which may be developmentally lethal, but instead can impact distinct downstream post-transcriptional activities, resulting in emergence of specific disease phenotypes. As demonstrated here, the integration of in-cell approaches (e.g. iCLIP) with

machine learning techniques (e.g. Pythia) can uncover nuanced changes in the RNA-protein interaction patterns, and will likely be essential for fully understanding the influence of such disease-relevant IDRs, in future mechanistic studies.

Our previous research had identified LARP6 as a key regulator of cancer cell viability and invasion in highly invasive breast cancer cells that have undergone epithelial–mesenchymal transition (EMT) (17). Additionally, LARP6 plays a key role in fibrosis through its binding to and promotion of type I collagen mRNA translation (53). Given its critical involvement in these pathological processes, and its restricted expression pattern in human tissues, LARP6 has emerged as a promising therapeutic target (54,55). Our findings here suggest that pharmacological inhibition of the La-module of LARP6, which harbours its primary RNA-binding interface, would likely provide the most potent therapeutic effect. This conclusion stems from the observation that the deletion of La-module abrogates the RNA-binding activity of LARP6 in living cells, with the IIDR-mediated RNA-protein interactions unlikely to compensate for the loss of RNA-binding on their own. Therefore, rational drug design strategies targeting LARP6 should prioritise the inhibition of the La-module to achieve maximal therapeutic efficacy.

In addition to its RNA-binding capabilities, LARP6 interacts with several protein partners, the most prominent of which is STRAP (56). STRAP is an adaptor protein known for its role in the negative regulation of TGF-beta signalling, and for its interaction with the Survival of Motor Neuron (SMN) complex, where it influences the subcellular localisation the complex (57,58). The interaction between STRAP and LARP6 is mediated by the SUZ-C domain, located immediately downstream of the C-terminal RNA-binding IDR identified in this study. Within the context of LARP6, STRAP binding has been shown to inhibit the translation of type I collagen mRNA, with this activity being important for ensuring a balanced expression of different Collagen-I subunits, although the molecular mechanism remains unclear (56). Our results here demonstrate that the deletion of the C-terminal region of LARP6, encompassing both the SUZ-C domain and the newly identified IDR, did not significantly impact the RNA-binding activity of LARP6. These results are consistent with STRAP's role as a negative regulator of LARP6 function. However, the deletion mutant used in this study did not allow us to fully dissect the independent contributions of STRAP binding and the IDR, in the context of LARP6 RNA-binding and its cellular activities. It is possible that STRAP binding directly interferes with the adjacent IDR's ability to engage with RNA. To clarify the functional interplay between these closely positioned RNA and protein interacting regions, further studies employing more targeted individual deletion mutants—lacking either the SUZ-C or the IDR—will be necessary.

It is important to note that the role uncovered here for the N-terminal IDR of LARP6 in RNA-binding markedly differs from the N-terminal IDR of its paralogue LARP4A (44). In both cases the IDRs do not conform to emerging notions of RNA-binding, and their RNA-related activity entails crosstalk with the adjacent folded La-module. However, while the NTR of LARP4A contains the most determinants of binding affinity to RNA, the NTR of LARP6 operates by enhancing the RNA-binding selectivity of the entire NTD RNA-binding platform, through modulation of the La-module's local RNA selectivity. The exact mechanism by which these disordered regions execute distinct RNA-binding functions in LARPs may reside in their specific sequence-related conformational and dynamics ensembles, as well as their physical coupling with the tethered structured domain (59). To clarify this, structural dynamics investigations of topological arrangements and motions in both apo and RNA-bound states will be required in future studies (59).

Acknowledgements:

We would like to acknowledge the BCI mass spectrometry core facility for their support with the MS experiments, and the Barts Genome Centre for their support with next-generation sequencing work. We would also like to thank the past and present members of the Mardakheh and Conte labs for their useful comments and suggestions on various aspects of the project. This work was supported by a Medical Research Council (MRC) project grant (ref: MR/W001500/1) and a Biotechnology and Biological Sciences Research Council (BBSRC) project grant (ref: BB/X007820/1) to F.K.M, and a Leverhulme Trust grant (ref: RPG-2020-264) to M.R.C. F.C. was supported by a BBSRC London Interdisciplinary Biosciences Consortium (LIDo DTP) PhD studentship.

Author Contributions:

F.K.M. and M.R.C conceived the study, acquired the funding, and supervised the work. F.C. performed the experiments and analysed the data. F.K.M, M.R.C, F.C, and G.A wrote the manuscript. G.A. and L.M. assisted with the design, execution, and interpretation of the biophysical experiments. M.K. and H.G. developed Pythia and performed the LARP6 Pythia analysis. T.T.T.B assisted with the CD analysis. A.B. assisted with the iCLIP data analysis.

References:

1. Gebauer, F., Schwarzl, T., Valcarcel, J. and Hentze, M.W. (2021) RNA-binding proteins in human genetic disease. *Nature reviews. Genetics*, 22, 185-198.
2. Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nature reviews. Molecular cell biology*, 19, 327-341.
3. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nature reviews. Genetics*, 15, 829-845.
4. Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nature reviews. Genetics*, 8, 533-543.
5. Khoroshkin, M., Buyan, A., Dodel, M., Navickas, A., Yu, J., Trejo, F., Doty, A., Baratam, R., Zhou, S., Lee, S.B. *et al.* (2024) Systematic identification of post-transcriptional regulatory modules. *Nat Commun*, 15, 7872.
6. Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J. and Hentze, M.W. (2016) Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell*, 63, 696-710.
7. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A. and Bonasio, R. (2016) High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular cell*, 64, 416-430.
8. Maraia, R.J., Mattijssen, S., Cruz-Gallardo, I. and Conte, M.R. (2017) The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. *Wiley interdisciplinary reviews. RNA*, 8.
9. Martino, L., Pennell, S., Kelly, G., Busi, B., Brown, P., Atkinson, R.A., Salisbury, N.J., Ooi, Z.H., See, K.W., Smerdon, S.J. *et al.* (2015) Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module. *Nucleic acids research*, 43, 645-660.

10. Cai, L., Fritz, D., Stefanovic, L. and Stefanovic, B. (2010) Binding of LARP6 to the conserved 5' stem-loop regulates translation of mRNAs encoding type I collagen. *J Mol Biol*, 395, 309-326.
11. Valavanis, C., Wang, Z., Sun, D., Vaine, M. and Schwartz, L.M. (2007) Acheron, a novel member of the Lupus Antigen family, is induced during the programmed cell death of skeletal muscles in the moth Manduca sexta. *Gene*, 393, 101-109.
12. Wang, Z., Glenn, H., Brown, C., Valavanis, C., Liu, J.X., Seth, A., Thomas, J.E., Karlstrom, R.O. and Schwartz, L.M. (2009) Regulation of muscle differentiation and survival by Acheron. *Mech Dev*, 126, 700-709.
13. Glenn, H.L., Wang, Z. and Schwartz, L.M. (2010) Acheron, a Lupus antigen family member, regulates integrin expression, adhesion, and motility in differentiating myoblasts. *Am J Physiol Cell Physiol*, 298, C46-55.
14. Hau, H.T.A., Ogundele, O., Hibbert, A.H., Monfries, C.A.L., Exelby, K., Wood, N.J., Nevarez-Mejia, J., Carbajal, M.A., Fleck, R.A., Dermit, M. *et al.* (2020) Maternal Larp6 controls oocyte development, chorion formation and elevation. *Development*, 147.
15. Manojlovic, Z., Earwood, R., Kato, A., Perez, D., Cabrera, O.A., Didier, R., Megraw, T.L., Stefanovic, B. and Kato, Y. (2017) La-related protein 6 controls ciliated cell differentiation. *Cilia*, 6, 4.
16. Billey, E., Hafidh, S., Cruz-Gallardo, I., Litholdo, C.G., Jean, V., Carpentier, M.C., Picart, C., Kumar, V., Kulichova, K., Marechal, E. *et al.* (2021) LARP6C orchestrates posttranscriptional reprogramming of gene expression during hydration to promote pollen tube guidance. *Plant Cell*, 33, 2637-2661.
17. Dermit, M., Dodel, M., Lee, F.C.Y., Azman, M.S., Schwenzer, H., Jones, J.L., Blagden, S.P., Ule, J. and Mardakheh, F.K. (2020) Subcellular mRNA Localization Regulates Ribosome Biogenesis in Migrating Cells. *Developmental cell*, 55, 298-313 e210.
18. Jolma, A., Zhang, J., Mondragon, E., Morgunova, E., Kivioja, T., Laverty, K.U., Yin, Y., Zhu, F., Bourenkov, G., Morris, Q. *et al.* (2020) Binding specificities of human RNA-binding proteins toward structured and linear RNA sequences. *Genome Res*, 30, 962-973.
19. Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.M., Ramakrishna, M., Harvey, R.F., Dezi, V. *et al.* (2019) Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol*, 37, 169-178.
20. Rappsilber, J., Ishihama, Y. and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry*, 75, 663-670.
21. Dodel, M., Guiducci, G., Dermit, M., Krishnamurthy, S., Alard, E.L., Capraro, F., Rekad, Z., Stojic, L. and Mardakheh, F.K. (2024) TREX reveals proteins that bind to specific RNA regions in living cells. *Nat Methods*, 21, 423-434.
22. Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*, 11, 2301-2319.
23. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M. and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*, 13, 731-740.
24. Dermit, M. and Mardakheh, F.K. (2021) Purification and quantitative proteomic analysis of cell bodies and protrusions. *STAR Protoc*, 2, 100462.
25. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
26. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923-930.
27. Cox, J. and Mann, M. (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics*, 13 Suppl 16, S12.

28. Mardakheh, F.K., Paul, A., Kumper, S., Sadok, A., Paterson, H., McCarthy, A., Yuan, Y. and Marshall, C.J. (2015) Global Analysis of mRNA, Translation, and Protein Localization: Local Translation Is a Key Regulator of Cell Protrusions. *Developmental cell*, 35, 344-357.
29. Sadok, A., McCarthy, A., Caldwell, J., Collins, I., Garrett, M.D., Yeo, M., Hooper, S., Sahai, E., Kuemper, S., Mardakheh, F.K. *et al.* (2015) Rho kinase inhibitors block melanoma cell migration and inhibit metastasis. *Cancer research*, 75, 2272-2284.
30. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9, 676-682.
31. Lee, F.C.Y., Chakrabarti, A.M., Hänel, H., Monzón-Casanova, E., Hallegger, M., Militti, C., Capraro, F., Sadée, C., Toolan-Kerr, P., Wilkins, O. *et al.* (2021) An improved iCLIP protocol. *bioRxiv*, 2021.2008.2027.457890.
32. Azman, M.S., Alard, E.L., Dodel, M., Capraro, F., Faraway, R., Dermit, M., Fan, W., Chakraborty, A., Ule, J. and Mardakheh, F.K. (2023) An ERK1/2-driven RNA-binding switch in nucleolin drives ribosome biogenesis and pancreatic tumorigenesis downstream of RAS oncogene. *The EMBO journal*, 42, e110902.
33. Wilkins, O.G., Capitanchik, C., Luscombe, N.M. and Ule, J. (2021) Ultraplex: A rapid, flexible, all-in-one fastq demultiplexer. *Wellcome Open Res*, 6, 141.
34. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9, 357-359.
35. Smith, T., Heger, A. and Sudbery, I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*, 27, 491-499.
36. Capitanchik, C., Jones, M. (2022). 1.5.0 ed. https://github.com/ulelab/clippy.
37. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-842.
38. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8, 118-127.
39. Kuret, K., Amalietti, A.G., Jones, D.M., Capitanchik, C. and Ule, J. (2022) Positional motif analysis reveals the extent of specificity of protein-RNA interactions observed by CLIP. *Genome Biol*, 23, 191.
40. Curk, T. (2019). https://github.com/tomazc/iCount.
41. Karimzadeh, M., Goodarzi, H. (2022). https://github.com/goodarzilab/pythia.
42. Shrikumar, A., Greenside, P. and Kundaje, A. (2017), *International conference on machine learning*. PMlR, pp. 3145-3153.
43. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31, 3784-3788.
44. Cruz-Gallardo, I., Martino, L., Kelly, G., Atkinson, R.A., Trotta, R., De Tito, S., Coleman, P., Ahdash, Z., Gu, Y., Bui, T.T.T. *et al.* (2019) LARP4A recognizes polyA RNA via a novel binding mechanism mediated by disordered regions and involving the PAM2w motif, revealing interplay between PABP, LARP4A and mRNA. *Nucleic acids research*, 47, 4272-4291.
45. Meszaros, B., Erdos, G. and Dosztanyi, Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*, 46, W329-W337.
46. Erdos, G. and Dosztanyi, Z. (2020) Analyzing Protein Disorder with IUPred2A. *Curr Protoc Bioinformatics*, 70, e99.
47. Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*, 1, 2876-2890.
48. Rabani, M., Kertesz, M. and Segal, E. (2008) Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 14885-14890.

49. Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M. and Tavazoie, S. (2012) Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature*, 485, 264-268.
50. Fish, L., Khoroshkin, M., Navickas, A., Garcia, K., Culbertson, B., Hanisch, B., Zhang, S., Nguyen, H.C.B., Soto, L.M., Dermit, M. *et al.* (2021) A prometastatic splicing program regulated by SNRPA1 interactions with structured RNA elements. *Science*, 372.
51. Escobar, D.J., Desai, R., Ishiyama, N., Folmsbee, S.S., Novak, M.N., Flozak, A.S., Daugherty, R.L., Mo, R., Nanavati, D., Sarpal, R. *et al.* (2015) alpha-Catenin phosphorylation promotes intercellular adhesion through a dual-kinase mechanism. *Journal of cell science*, 128, 1150-1165.
52. Martin, S.R., Ramos, A. and Masino, L. (2021) Biolayer Interferometry: Protein-RNA Interactions. *Methods in molecular biology*, 2263, 351-368.
53. Zuo, S., Wang, B., Liu, J., Kong, D., Cui, H., Jia, Y., Wang, C., Xu, X., Chen, G., Wang, Y. *et al.* (2021) ER-anchored CRTH2 antagonizes collagen biosynthesis and organ fibrosis via binding LARP6. *The EMBO journal*, 40, e107403.
54. Stefanovic, B., Manojlovic, Z., Vied, C., Badger, C.D. and Stefanovic, L. (2019) Discovery and evaluation of inhibitor of LARP6 as specific antifibrotic compound. *Sci Rep*, 9, 326.
55. Stefanovic, B., Michaels, H.A. and Nefzi, A. (2021) Discovery of a Lead Compound for Specific Inhibition of Type I Collagen Production in Fibrosis. *ACS Med Chem Lett*, 12, 477-484.
56. Vukmirovic, M., Manojlovic, Z. and Stefanovic, B. (2013) Serine-threonine kinase receptor-associated protein (STRAP) regulates translation of type I collagen mRNAs. *Molecular and cellular biology*, 33, 3893-3906.
57. Seong, H.A., Jung, H., Choi, H.S., Kim, K.T. and Ha, H. (2005) Regulation of transforming growth factor-beta signaling and PDK1 kinase activity by physical interaction between PDK1 and serine-threonine kinase receptor-associated protein. *J Biol Chem*, 280, 42897-42908.
58. Chari, A., Golas, M.M., Klingenhager, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. and Fischer, U. (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell*, 135, 497-509.
59. Holehouse, A.S. and Kragelund, B.B. (2024) The molecular basis for cellular function of intrinsically disordered protein regions. *Nature reviews. Molecular cell biology*, 25, 187-211.