Identification of G-quadruplex structures in MALAT1 lncRNA that interact with nucleolin and nucleophosmin

Arpita Ghosh[1,2], Satya Prakash Pandey[1,2], Dheeraj Chandra Joshi[1,2], Priya Rana[1,2], Asgar Hussain Ansari[1,2], Jennifer Seematti Sundar[1], Praveen Singh[1,2], Yasmeen Khan[1,2], Mary Krishna Ekka[1,2], Debojyoti Chakraborty [1,2] and Souvik Maiti [1,2,3,*]

[1]CSIR-Institute of Genomics & Integrative Biology, Mathura Road, Delhi 110025, India, [2]Academy of Scientific & Innovative Research (AcSIR), Ghaziabad 201 002, India and [3]CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411 008, India

ABSTRACT

Nuclear-retained long non-coding RNAs (lncRNAs) including MALAT1 have emerged as critical regulators of many molecular processes including transcription, alternative splicing and chromatin organization. Here, we report the presence of three conserved and thermodynamically stable RNA G-quadruplexes (rG4s) located in the 3′ region of MALAT1. Using rG4 domain-specific RNA pull-down followed by mass spectrometry and RNA immunoprecipitation, we demonstrated that the MALAT1 rG4 structures are specifically bound by two nucleolar proteins, Nucleolin (NCL) and Nucleophosmin (NPM). Using imaging, we found that the MALAT1 rG4s facilitate the localization of both NCL and NPM to nuclear speckles, and specific G-to-A mutations that disrupt the rG4 structures compromised the localization of both NCL and NPM in speckles. In vitro biophysical studies established that a truncated version of NCL (ΔNCL) binds tightly to all three rG4s. Overall, our study revealed new rG4s within MALAT1, established that they are specifically recognized by NCL and NPM, and showed that disrupting the rG4s abolished localization of these proteins to nuclear speckles

GRAPHICAL ABSTRACT
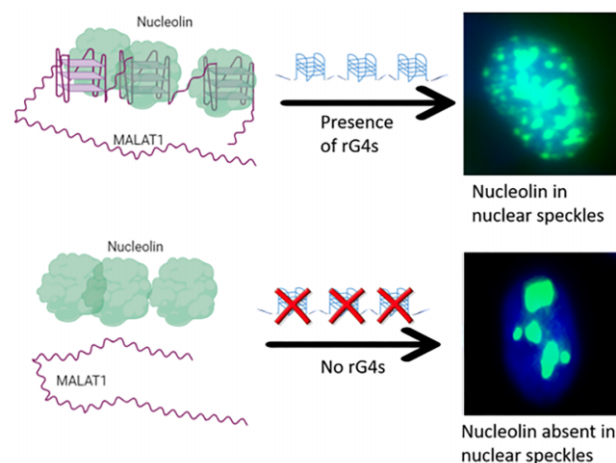
INTRODUCTION

Long non-coding RNAs (lncRNAs) are a class of transcripts defined by their size (>200 nucleotides) and the absence of significant protein-coding potential (1,2). In contrast to small non-coding RNAs, which typically have specific functions within cells based on their class, lncRNAs exhibit remarkable functional versatility (3–7). LncRNAs that predominantly localize in the nucleus can serve as scaffolds for protein complexes, guiding proteins to specific genomic loci, or establish interactions with epigenetic regulators (8). In contrast, cytoplasmic lncRNAs primarily engage in translational and post-translational activities (9). In both cellular compartments, lncRNAs can function as decoys for binding proteins, and also participate in the sequestration of miRNAs (10). The roles of lncRNAs at the cellular and physiological level, include growth, development, cell-cycle

---

*To whom correspondence should be addressed. Tel: +91 11 29879108; Email: souvik@igib.res.in

progression, inflammation, differentiation and even events related to tumor formation, invasion and migration. Dysregulation of lncRNAs is widely observed in various types of cancers, and they can exert either oncogenic or tumor-suppressive actions (11,12). Thus, lncRNAs are key players in the intricate regulatory networks that govern molecular, cellular and physiological activities.

RNA, as a single-stranded molecule, can fold into secondary structures that play a crucial role in carrying out its diverse functions. Regions of RNA enriched in guanine (G) have the propensity to adopt a unique structural motif known as the RNA G-quadruplex (rG4). In an rG4, four guanine bases come together to form a stable arrangement called a G-quartet, where Hoogsteen hydrogen bonds are formed between the guanines. This assembly, facilitated by the presence of cations, results in a square planar structure that deviates from canonical RNA secondary structures composed of Watson-Crick pairs (13,14). While the functions of G-quadruplex structures in DNA have been extensively studied (15), research on their RNA counterparts, rG4s, has been relatively limited (16). Importantly, recent studies have highlighted the critical functional roles played by rG4 domains in both messenger RNAs (mRNAs) and non-coding RNAs (17,18). These rG4 structures have been shown to regulate various processes including transcription and translation, 3′ end processing, alternative splicing, mRNA localization, protein binding, telomere RNA organization, and RNA stability (4,19–27). These findings emphasize the significance of rG4s in governing key aspects of RNA biology and expand our understanding of their functional relevance beyond DNA.

Although long non-coding RNAs (lncRNAs) are generally less abundant and exhibit lower conservation compared to other RNA molecules, they possess highly conserved structural elements. These conserved structural motifs play a crucial role in determining the functional mechanisms of specific lncRNAs across multiple regulatory processes. Among these structural motifs, the presence of rG4 structures has been predicted in a large number of lncRNAs (20,24,25,27,28). One extensively studied lncRNA is Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1), which is highly expressed and predominantly localized within the cell nucleus. MALAT1 is widely associated with cancer and has been implicated in the regulation of alternative splicing and gene expression (6,29–33). Its localization to nuclear speckles (34) enables its interaction with various splicing factors such as SRSF1, SRSF2 (SC35) and SRSF3. Depletion of MALAT1 leads to a reduced association of these splicing factors with nuclear speckles, without affecting the overall formation of these nuclear subdomains (29,35,36). As an 8.7-kilobase-long RNA, MALAT1 can adopt various dynamic conformations within cells, which may facilitate the recruitment of different proteins within its structural scaffold. The secondary and tertiary structures of MALAT1 have been reported to play a role in regulating its stability and function through diverse molecular mechanisms. Notably, recent studies have revealed the existence of a unique bipartite triple-helix structure at the 3′ end of MALAT1, spanning a 74-nucleotide region. This triple-helix structure consists of a U-rich stem-loop that sequesters an A-rich tail, thereby

protecting the RNA from exonucleolytic degradation (37–40). These findings highlight the intricate structural features of MALAT1 and their involvement in modulating its stability and functional properties.

RNA binding proteins (RBPs) exert significant control over various biological processes, most frequently by sequence-specific recognition of short motifs located in single-stranded RNA (41,42). However, RBPs can also recognize RNA structures including rG4s, and these interactions play key roles in multiple cellular processes. Since the discovery of the first rG4-binding protein, FMRP, in 2001, numerous other proteins have been shown to recognize rG4 structures. These proteins can stabilize or destabilize rG4s or interact with other molecules within a complex (43,44). For example, DHX36 facilitates the maturation of human telomerase RNA (hTR) and enhances telomerase function by binding to an rG4 structure located at the 5′ end of hTR (45,46). hnRNP F promotes alternative splicing by binding to an rG4 within the CD44 intron (47). NCL (nucleolin) modulates MYC expression and colorectal cancer cell proliferation by binding to the rG4 structure present in the LUCAT1 RNA (23). These examples highlight the diverse roles of RBPs that associate with rG4 structures.

Here, we report the presence of three stable rG4 structures within the 3′ region of MALAT1 lncRNA and demonstrate that they interact with nucleolin (NCL) and nucleophosmin (NPM) proteins in HeLa cells. These rG4 structures facilitate translocation of NCL and NPM into nuclear speckles, and disrupting them abolishes NCL and NPM localization in speckles. The dynamic rG4-protein interactions identified in our study provide a foundation for further investigation into the multi-faceted roles of MALAT1.

MATERIALS AND METHODS

Bioinformatics

To predict the putative rG4 forming sites in MALAT1 we used a bioinformatics tool, QGRS mapper. It generates information on the composition and distribution of putative Quadruplex forming G-Rich Sequences (QGRS). The program maps QGRS in the entire nucleotide sequence provided in the raw or FASTA format which in this case we provided for MALAT1. The scoring system evaluates QGRS for its chances to form a stable G-quadruplex. A higher G-score for sequences indicates better candidates for G-quadruplex formation. Three putative rG4 sequences were identified in MALAT1, for which the tool showed multiple outputs, depending upon different combinations. Q1, Q2 and Q3 were identified based on the highest G-score for those sequences. To observe conservation in the MALAT1 lncRNA, and especially in the Potential Quadruplex-forming Sequence (PQS), we used UCSC Genome Browser. It uses the PhyloP base-wise conservation scoring system which is derived from the Multiz alignment of 46 vertebrate species.

In vitro transcription (IVT) of RNAs

The IVT of RNAs was performed using the Megascript T7 IVT kit (Ambion®), following the manufacturer's instructions. Briefly, forward and reverse primers were designed for

each domain of MALAT1 containing G-quadruplexes (Q1, Q2 and Q3) along with their corresponding mutant counterparts (Q1M, Q2M and Q3M). These primers were heated and annealed to create a partially complementary double-stranded DNA. Taq polymerase (Geneaid) was then used to extend the partially complementary double-stranded DNA to produce fully complementary DNA containing a T7 promoter. The *in vitro* transcription was carried out overnight at 37°C in a PCR machine. Next, template DNA was digested using TURBO DNase (Ambion®) and the resulting RNAs were purified using NucAway columns (Ambion®). The integrity and purity of the RNAs were assessed using polyacrylamide gel electrophoresis (PAGE) and ethidium bromide staining. Finally, pure RNA species were utilized for further experiments.

rG4 mutant library preparation

The plasmid containing full-length 8.7 kb MALAT1 under a pCMV promoter was obtained as a gift from Dr K.V. Prasanth's lab at the University of Illinois. Overlapping primer sets were designed for the Q1, Q2 and Q3 regions to perform site-directed mutagenesis (SDM) (Supplementary Table S1). G-to-A mutations were introduced in each primer complementary to the rG4 forming region. Long PCR was performed with the MALAT1 cloned pCMV plasmid (FL) as a template using these primers to introduce the mutations into the plasmid as it was amplified. High Fidelity Pfu Polymerase (Agilent) was used for the entire 11.2 kb plasmid amplification at each cycle. To eliminate the parent plasmid without any mutation, Dpn I (New England Biolabs® Inc.) treatment was provided post-PCR for 30 min at 37°C. The reaction was then used for transformation of the new PCR-synthesized plasmid with proper controls. Five colonies were picked from each reaction and amplified to test for positive clones having the mutations introduced. Each of them was verified by Sanger sequencing with primers that could detect Q1, Q2 and Q3 separately. Once Q1m, Q2m and Q3m were verified by sequencing, the second mutation was introduced in their backbone. Q12m, Q23m and Q13m were verified by sequencing, and the third mutation was introduced in the backbone of the double mutant plasmid, which was then Sanger sequenced to obtain Q123m. Primers used to check for mutations in each rG4 by Sanger Sequencing are listed in Supplementary Table S2.

ThT fluorescence titration assay

The *in vitro* transcription of RNAs was performed using forward and reverse primers containing the T7 promoter for each MALAT1 rG4 domain (listed in Supplementary Table S3). Q5 High Fidelity Polymerase (New England Biolabs® Inc.) was used to extend the partially complementary primers spanning the rG4 regions of the plasmid double-stranded DNA to fully complementary DNA containing the T7 promoter. RNAs were synthesized from Q1m, Q2m, Q3m, Q12m, Q13m, Q23m, Q123m and FL plasmids, and then purified using PAGE. The rG4 structures were prepared from these RNAs by slow cooling at a rate of 0.2°C/min after heating to 100°C for 5 min in 100 mM KCl and 10 mM sodium cacodylate buffer (pH 7.4).

Thioflavin T (ThT) compound obtained from Sigma was used in a 96-well microplate from CORNING (Flat Bottom Black Polystyrol). The RNAs, ranging in concentration from 0 to 8 μM, were mixed with ThT at a final concentration of 2 μM, in 10 mM Sodium Cacodylate and 100 mM KCl buffer at pH 7.0. A similar setup was carried out in 10 mM Sodium Cacodylate and 100 mM LiCl buffer as a control. Fluorescence emission was collected at 487 nm with excitation at 440 nm in a microplate reader (Tecan Microplate Reader Life Sciences) at 25°C. Three separate technical replicates were conducted, and S.E.M. was plotted.

Cell culture and transfections

All experiments were conducted on the HeLa cervical cancer cell line, which was authenticated and free of mycoplasma. The MALAT1-null HeLa cell line, along with the normal HeLa control cell line, was obtained from Dr Roderic Guigo and Dr Rory Johnson's lab at CRG Barcelona. The MALAT1-null cells were generated by a dual sgRNA-based CRISPR approach to knockout the promoter region (Supplementary Figure S2A). The promoter deletion was confirmed by genotyping (48,49) (Supplementary Figure S2B), and qRT-PCR was used to check MALAT1 expression in HeLa (wt) and MALAT1 knockout (ko) cells (Supplementary Figure S2C). The cell lines were maintained in DMEM with 10% FBS without antibiotic or anti-mycotic and were incubated in a humidified incubator at 37°C with 5% CO_2. For transfection, the cells were seeded in 6- and 12-well plates at a density of 8 and 4×10^4 cells per well, respectively, and incubated for 24 h. The cells were then transfected with plasmid concentrations 1 μg (for 12 well plates) and 2 μg (for 6-well plates) using Lipofectamine 3000. For siRNA transfections, 25 nM siRNA was transfected in cells seeded in a 12-well plate. The transfected cells were maintained in Opti-MEM™, a reduced serum media, for 4 h, after which the media was changed to complete DMEM with FBS Supplement. The plates were then incubated for 48 h before the cells were harvested for their respective experimental procedures.

qRT-PCR

RNA was isolated from HeLa cells, MALAT1-null cells (see above), or HeLa cells transfected with the specified constructs for 48 h. TRIZOL® reagent (Ambion®) was used to isolate total RNA in the stepwise protocol as per the manufacturer's instructions. cDNA was prepared using a Qiagen cDNA synthesis kit. After the cDNA was prepared (from 1 μg RNA for each sample), real-time qPCR was performed for all the samples and controls in triplicates. The reaction volume for each was 10 μL which included 1 μl of cDNA (after 1:2 dilution). The PCR conditions used were Initial denaturation at 95°C for 3 min followed by 40 cycles of the following conditions:

95°C for 10 s
60°C for 30 s
72°C for 30 s

qRT-PCR primers used in different experiments are listed in Supplementary Table S4. Transcripts were quantified

using a SYBR Green Master Mix: SYBR Premix Ex Taq II (Tli RNase H Plus) (from TaKaRa) in the instrument Light Cycler 480 (Roche). The Ct values obtained for the different transcripts were normalized to that of Beta-Actin. The fold change analysis in the transcript levels for comparative analysis was done using the 2-$\Delta\Delta$Ct method as described before (50). Briefly, fold changes were calculated using the following formula:

$$RelativeFoldchange = \frac{2^{-(avg.Ct[GSP]-avg.Ct[ACT])TEST}}{2^{-(avg.Ct[GSP]-avg.Ct[ACT])MALAT1+/+}}$$

where avg.= average of Ct = Cycle threshold [GSP]= Gene Specific Primer [ACT]= βActin Specific Primer TEST = MALAT+/+; MALAT–/–; or one of the following: MALAT1–/– transfected with plasmids carrying full length MALAT1 gene, or mutations as indicated

RNA FISH

Stellaris® ShipReady RNA FISH probes for MALAT1 were ordered from Biosearch Technologies (LGC) and had Quasar® 570 and 670 dyes. Corning® 22 mm square coverslips were immersed in the 6-well plates before seeding the cells at a density of 8 × 104 cells per well. Transfections were performed as mentioned in the protocol before. After 48 h post-transfection, cells were harvested for RNA FISH. The RNA FISH buffers were purchased from Stellaris® and the Stellaris RNA FISH protocol was performed as prescribed by the manufacturer for adherent cells. All steps were conducted in highly sterile and RNase-free conditions. For immunoFISH experiments, after the entire FISH protocol was conducted, the coverslips were fixed again in 4% formaldehyde (ThermoFisher Scientific), and an immunocytochemistry experiment was carried out in the same coverslips according to the protocol detailed in the next section. The coverslips were mounted on Corning® plain microscope slides, and ProLong® Diamond Antifade from ThermoFisher Scientific was used. The slides were imaged under a 60× objective in a DeltaVision Microscope from GE Healthcare Life Sciences. Image quantification was performed using ImageJ, and co-localization analysis for immune-FISH was carried out with Fiji using the Colloq 2 plugin.

Immunocytochemistry

HeLa cervical cancer cells for slide preparation were seeded at a density of 8 × 10^4 cells in a 6-well plate that contained Corning 22 mm sq. square coverslips in each well to which cells were left for 24 h incubation at 37°C with 5% CO_2 in a humidified environment. Transfections for desired experiments were performed according to the protocol mentioned before. 48 h post-transfection, the cells were washed with 1x PBS (Gibco) after removing DMEM. Cells on coverslips from culture wells were then fixed using a buffer containing 4% formaldehyde, 5 μM EGTA pH 8.0 (Sigma), 1 μM $MgCl_2$ (Sigma), and incubated for 7 min. Post-fixation, cells were washed twice with washing buffer containing 30 μM glycine (Sigma) in PBS, 5 μM EGTA, and 10 μM $MgCl_2$, and then permeabilized using buffer having 0.2% Triton X-100 (Sigma) in PBS, 5 μM EGTA,

10 μM $MgCl_2$. Cells were incubated with the permeabilization buffer for 7 min. After permeabilization, the same washing step was performed as before. The wells containing the cover slip adhered cells were treated with a blocking buffer (0.5% BSA (HiMedia) in PBS, 5 μM EGTA, 10 μM $MgCl_2$) for 30 min. After blocking was performed the cells were incubated overnight with primary antibody (listed in Supplementary Table S5) at 1:500 dilution in the blocking buffer. For co-immunocytochemistry studies, two antibodies raised in different animals were mixed in 1:500 dilution (in blocking buffer) and incubated with the fixed, permeabilized cells. Post-incubation with primary antibody, cells were washed thrice for 5 min each using a blocking buffer followed by incubation with secondary antibody Alexa Fluor 488 (ThermoFisher Scientific) at 1:1000 dilution for 2 h. For co-immunocytochemistry experiments, Alexa Fluor 488 and Alexa fluor 647 fluorescent secondary antibodies from ThermoFisher Scientific were used, compatible with the animals in which the primary antibody was raised. After incubation with a secondary antibody, cells were washed thrice for 5 min each using a blocking buffer. The coverslips were then mounted on glass slides (Corning®) with a drop of Prolong® Gold Antifade Mountant with DAPI (ThermoFisher) and then viewed and analyzed with EVOS Cell Imaging System from ThermoFisher Scientific and DeltaVision Microscope from GE Healthcare Life Sciences under 60x objective. Quantification of images was performed using ImageJ. Colocalization analysis for immunoFISH was carried out with Fiji using the Colloq 2 plugin.

RNA pull-down

In vitro transcribed RNAs Q1, Q2, Q3 and their mutated counterparts were biotin-labeled using the Pierce RNA 3′ end biotinylation kit. The labeling efficiency was determined using the manufacturer's instructions with the Chemiluminescent detection kit module. To quantify the total protein content of the cells, 10 × 10^6 HeLa cells were lysed using commercially available RIPA buffer (Thermo Fisher Scientific). The cells were washed with filtered PBS and then lysed in RIPA buffer on ice for 30 min. After centrifugation at 12 000 rpm for 15 min, the supernatant was collected and stored at –80°C. The lysate was subjected to BCA protein assay. For the RNA pull-down assay, 100 pmol of biotinylated RNAs (Q1, Q2 and Q3 and their mutant counterparts Q1m, Q2m and Q3m) were incubated with 3 mg of cell extract (treated with RNase inhibitor) in a rotator at 4°C overnight. UV crosslinking was performed to strengthen the protein-RNA interactions at 254 nm for 15 min. The biotinylated RNA-protein complex was pulled down using magnetic Streptavidin beads (DynaBeads™ MyOne™ Streptavidin from Invitrogen) as per the manufacturer's instructions. The beads were incubated with the binding reaction for one hour at room temperature, washed thrice, and boiled in SDS to break the biotin-streptavidin interaction. Isolated proteins were identified using Mass spectrometry (QTOF 6600 (SCIEX)) or used to perform a western blot to detect specific proteins using their respective antibodies (listed in Supplementary Table S5).

Mass spectrometry

The proteins interacting with rG4s and their mutated counterparts obtained from the pulldown experiment were loaded in 10% SDS-PAGE gel and stained with Coomassie blue G250 (Biorad). For each sample, bands higher than 10 kDa were excised into small pieces and placed into a 1.5 ml tube. Sample preparation for mass spectrometry was performed according to the standard protocol (51). Briefly, gel pieces were destained and shrunk, protein reduction was performed by treatment with 25 mM dithiothreitol (DTT) at 60°C for 30 min followed by alkylation with 55 mM iodoacetamide (IAA) by incubation in dark for 30 min at room temperature. Protein digestion was performed using trypsin (V511A, Promega) at 37°C overnight and extraction of tryptic peptides was carried out using 60% acetonitrile with 1% TFA. Peptide clean-up was performed using C18 Ziptip (Merck) as per the manufacturer's protocol before the LC–MS acquisition. Samples were acquired on a quadrupole-TOF hybrid mass spectrometer (TripleTOF 6600, SCIEX, USA) coupled to a nano-LC system (Eksigent NanoLC-425, SCIEX, USA). Protein identification was performed using ProteinPilot Software 5.0.1 (SCIEX, USA) using the Paragon algorithm. For each sample, peptides were loaded on a trap-column (ChromXP C18CL 5 μm 120 Å, Eksigent) where desalting was performed using 0.1% formic acid in water with a flow rate of 10 μl per minute for 10 min. Peptides were then separated on a reverse-phase C18 analytical column (ChromXP C18, 3 μm 120 Å, Eksigent) in 57 min gradient of buffer A (0.1% formic acid in water) and buffer B (0.1% formic in acetonitrile) at a flow rate of 5 μl/minute with the following gradient:

Time (min)	% A	% B
0	97	3
38	75	25
43	68	32
45	20	80
45.5	10	90
48	10	90
49	97	3
57	97	3

Data acquisition was performed using Analyst TF 1.7.1 Software (SCIEX) using optimized source parameters. The ion spray voltage was set to 5.5 kV, 25 psi for the curtain gas, 20 psi for the nebulizer gas, and 250°C as the source temperature. For DDA, a 1.8 s instrument cycle was repeated in high sensitivity mode throughout the whole gradient, consisting of a full scan MS spectrum (350–1250 m/z) with an accumulation time of 0.25 s, followed by 30 MS/MS experiments (100–1500 m/z) with 0.050 s accumulation time each, on MS precursors with charge state 2+ to 5+ exceeding a 120 cps threshold. The rolling collision energy was used and former target ions were excluded for 10 s. Protein identification was performed by searching the .wiff format DDA mode LC–MS/MS acquisition files against UniProtKB human FASTA database (Swissprot, 20394 proteins entries) using ProteinPilot™ Software 5.0.1 (SCIEX). The Paragon algorithm was used to get protein group identities. The search parameters were set as follows: sample type- identification, cysteine alkylation-iodoacetamide, and digestion-trypsin. A biological modification was enabled in ID focus. The search effort was set to 'Thorough ID' and the detected protein threshold [Unused ProtScore (Conf)] was set to >0.05 (10.0%). False discovery rate (FDR) analysis was enabled. Only proteins identified with 1% global FDR were considered.

RNA immunoprecipitation

100×10^6 cells were cultured, harvested, and crosslinked using 1% glutaraldehyde followed by quenching with the addition of 0.125 M Glycine. The pellet was washed twice with cold PBS and then lysed using commercially available RIPA buffer (Thermo Fisher Scientific) with the addition of SuperaseIN (Ambion®) to avoid RNA degradation. The cell lysate was quantified using the Pierce BCA protein assay kit. The cell lysate was pre-cleared with 5 μg of IgG antibody for 4 h at 4°C on a rotator. The IgG-cleared lysate was incubated with Nucleolin/Nucleophosmin antibody overnight at 4°C on a rotator. 50 μl of Protein A/G Dynabeads (Thermo Fisher Scientific) was used to pull down the antibody-RNA complex. Beads were washed twice with a RIPA buffer to remove non-specific binding and beads were then subjected to Proteinase K treatment at 55°C for 30 min with gentle shaking. RNA was isolated using Trizol (Ambion). qRT-PCR was carried out to detect MALAT1 using primers listed in Supplementary Table S4.

Protein expression and purification

The DelNucleolin (ΔNCL) protein, which had amino acids 1–283 deleted from the the N-terminus and a 6XHis-tag at the C-terminus, was cloned into the pET22b (+) vector between BamHI and XhoI sites. The plasmid was then transformed into *E. coli* Rosetta cells for expression and protein production. The cells were grown in LB media containing ampicillin (100.0 μg/ml) and chloramphenicol (35 μg/ml) at 37.0°C until an O.D. of 0.6 at 600 nm was reached. The cultures were induced with 0.2 mM IPTG and kept at 18°C with shaking at 200 rpm for 16 h. The cells were harvested by centrifugation at 6000 rpm for 15 min, and the cell pellet was resuspended in lysis buffer (50.0 mM sodium phosphate pH 7.4, 150 mM NaCl, 10% glycerol, 0.2% Tween-20, 1 mM DTT, 4 mM PMSF) and lysed by sonication using 5 s 'on' and 30 seconds 'off' pulse cycle for 10 min or until the lysate became clear. The supernatant was recovered by centrifugation at 12 000 rpm for 40 min at 4°C. C-terminal His-tagged proteins were purified using affinity chromatography with Ni-NTA agarose resin, and eluted with elution buffer (50.0 mM sodium phosphate pH 7.4, 150 mM NaCl, 10% glycerol, with a gradient of 250 and 500 mM imidazole) after adequate washing. Eluted protein fractions were analyzed on a 12% SDS-PAGE gel. The fractions containing purified protein were pooled, dialyzed, and further purified using size exclusion chromatography (SEC), and visualized on a 12% SDS-PAGE gel for purity and homogeneity.

Circular dichroism (CD) spectroscopy

The CD spectra of all sequences were recorded using the Jasco 815 spectropolarimeter. To prepare the rG4s, the

strands were heated to 100°C for 5 min in 100 mM KCl and 10 mM sodium cacodylate buffer (pH 7.4) or 100 mM LiCl and 10 mM sodium cacodylate buffer (pH 7.4) at a strand concentration of 5 μM, followed by slow cooling at 0.2°C/min. The presented spectrum is an average of three consecutive scans for each sample.

UV melting

The UV melting experiments were performed using a Cary 100 (Varian) spectrophotometer, which was equipped with a thermoelectrically controlled cell holder. The rG4s were prepared as mentioned previously. In addition, an RNA duplex was also prepared using two complementary sequences annealed to each other (Forward Oligo strand- 5′ UCCAAAACAUGAAUUG 3′ and Reverse Oligo Strand- 5′ CAAUUCAUGUUUUGGA 3′). The changes in absorbance at 295 nm (and 260 nm for RNA duplex) were monitored over a temperature range of 10°C to 90°C, with a heating/cooling rate of 0.2°C/min. The melting and annealing curves were analyzed using Origin 7.0.

Electrophoretic mobility shift assay

1 μM pre-formed rG4s (as described above) were taken and incubated with increasing concentrations of purified ΔNCL (0.1–20 μM) in a reaction volume of 15 μl at 37 °C for 30 min. Free rG4s and protein-RNA complexes were separated by electrophoresis through 10% w/v native polyacrylamide gels in 0.5× TBE, pH 8.0 (Tris–borate–EDTA buffer) for 1 hour at 200 V at room temperature (∼22°C), well below the melting temperature of the oligos used in the assay. Free RNA and protein–RNA complexes were stained using SYBR Gold stain for 30 min and detected by the Typhoon FLA phosphorimager).

Surface plasmon resonance (SPR)

SPR measurements were performed using the BIAcore 3000 system with BIAcore 3000 control software version 4.1.2. The 5′-biotinylated Q1, Q2, and Q3 sequences (listed in Supplementary Table S6) were dissolved in filtered and degassed 10 mM HEPES buffer with 100 mM KCl and 0.005% surfactant IGEPAL, pH 7.4. Solutions were heated to 100°C and annealed by slow cooling to form a quadruplex. The biotinylated sequences were immobilized in flow cells 2, 3 and 4 of streptavidin-coated sensor chips (Sensor chip SA, BIAcore Inc.) until an RU change of 300 was achieved. After immobilization, unbound RNA was removed by flowing an excess buffer over the chips. Flow cell 1 was left blank as a control to account for the non-specific background signal, which was subsequently subtracted from the signal obtained in flow cells 2, 3 and 4. Filtered and degassed 10 mM HEPES with 100 mM KCl and 0.005% surfactant IGEPAL, pH 7.4 was used as the running buffer. Serial dilutions of the 10 μM ΔNCL stock were performed to make a concentration series in the running buffer. ΔNCL solutions of different concentrations between 10 and 500 nM were injected at a flow rate of 20 μl/min for 300 s. Following this, dissociation from the surface was monitored for 300 s in the running buffer. Regeneration

was done for 60 s using a buffer containing 1 M NaCl and 50 mM NaOH. Analysis of the binding sensorgrams was carried out using a two-independent-binding-site model using BIA evaluation software version 4.1.1. The experiments were carried out in triplicates, and the standard error was calculated. For all binding studies, the goodness of the fitting was monitored by the χ^2 value, which was either ≤1. All experiments were performed at 25°C using the running buffer.

Fluorescence titration for NCL-rG4 binding

The binding of ΔNCL to rG4 was monitored by measuring tryptophan fluorescence of ΔNCL with varying concentrations of rG4. Fluorescence spectra of tryptophan were recorded in a Fluoromax 4 (Spex) spectrofluorophotometer with a thermoelectrically temperature-controlled cell holder (quartz cuvette, 1 cm × 1 cm). The samples were excited at 290 nm, and emission spectra were recorded from 320 to 500 nm. The experiments were carried out at 25°C in 10 mM sodium cacodylate buffer (pH 7.4) containing 100 mM KCl. The protein concentration was kept fixed at 500 nM, and the concentration of pre-formed rG4 (prepared as described above) was varied from 0 to 1000 nM. The extent of fluorescence intensity change, $\Delta F / \Delta F_{\max}$, where $\Delta F = F_0 - F$ and $\Delta F_{\max} = F_0 - F_{\text{final}}$, was plotted as a function of rG4 concentrations. As the bound ΔNCL is directly proportional to the extent of fluorescence intensity change, $\Delta F / \Delta F_{\max}$ was used to determine the binding. The observed fluorescence intensity was considered as the sum of the weighted contributions from an rG4-bound ΔNCL and an unbound ΔNCL form for data analysis. The data were averaged from three independent experiments and analyzed using nonlinear regression with the 'One Site-Specific Binding' model ($\Delta F / \Delta F_{\max}$ =[rG4]/(K_d + [rG4]) where [rG4] is the quadruplex concentration, and K_d is the dissociation constant) in Origin 7.0.

Western blot

The cells were lysed using Cell Lytic™ buffer (Sigma) to prepare protein lysates. To each 12-well plate, 50 μl of the lysis buffer and 5 μl of Proteinase Inhibitor Cocktail (Sigma) were added. The cells were allowed to lyse at 4°C on a rocker for 1 hour. After incubation, the protein lysate from each sample was collected, and the protein concentration was estimated using a Pierce™ BCA Protein Assay Kit (ThermoFisher). For each sample, 40 μg of protein was loaded into the wells of a 10% SDS gel, and PAGE was performed. The proteins were then transferred from the gel to a PVDF membrane (GE Healthcare Life-Science) in a Bio-Rad vertical gel Transfer Apparatus at 4°C for 3 h at 70 V. After transfer, the membrane was cut according to the required protein size and kept for blocking (with 5% BSA in TBST) on a rocker at room temperature for 5–6 h. After blocking, the blots were incubated with primary antibody (listed in Supplementary Table S5) at a 1:1000 dilution overnight at 4°C. After primary antibody incubation, the blots were washed three times for 15 min each with 1X TBST. Post-washing, they were incubated with a secondary antibody having HRP conjugate for 3 h at room temperature. After

incubation, a similar washing step (as mentioned after primary antibody incubation) was performed. The EMD Millipore™ Immobilon Western Chemiluminescent HRP Substrate (ECL) was used to develop the blots in the Syngene Gel doc instrument. The densitometry analysis of the blots was performed using ImageJ.

Statistical analysis

Statistical analysis was performed using GraphPad Prism 8.0 to evaluate the significance among experimental replicates. All data were presented as mean ± S.D. of three independent biological replicates except otherwise mentioned. A two-tailed unpaired Student's t-test was used to analyze the experimental data. The experimental results leading to a *P*-value <0.05 were considered statistically significant. One asterisk (*), two asterisks (**), three asterisks (***), and four asterisks (****) denote $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively. Biological replicates were similar to those employed in the field.

RESULTS

MALAT1 long non-coding RNA harbors three putative rG4 forming domains towards its 3′ end

In a previous transcriptomics study conducted by our team, in silico analysis was used to identify a significant number of potential rG4-forming sites within human lncRNAs (52). Subsequently, a transcriptome-wide profiling method confirmed the presence of putative rG4 structures in lncRNAs such as MALAT1 and NEAT1 (53,54). Although similar findings continue to be reported, the functional roles of rG4 structures in lncRNAs remain largely unknown (28,55,56). To investigate the potential functional roles of rG4 structures in MALAT1, we first employed the QGRS mapper (57), a bioinformatic tool for G-quadruplex prediction. This analysis revealed three potential G-quadruplex motifs, designated Q1, Q2, and Q3, that are located towards the 3′ end of the 8.7 kb MALAT1 lncRNA (Figure 1A). Nucleotide conservation analysis demonstrated that the region harboring the rG4s exhibited conservation across species, as evidenced by the phyloP scoring in the UCSC Genome Browser (Figure 1B, Supplementary Figure S1A).

To validate the formation of G-quadruplex structures, we performed biophysical assays using the Q1, Q2 and Q3 RNA sequences individually. Circular dichroism (CD) spectroscopy revealed the characteristic signature of stable G-quadruplex structures, represented by a positive peak at 266 nm and a negative peak at 236 nm in the CD spectra. These features are indicative of parallel G-quadruplex formation in the presence of K^+ ions (Figure 1C). CD spectra were also obtained for the three rG4 sequences in the presence of Li^+ ions (Supplementary Figure S1B). The intensity of both positive and negative peaks was reduced compared to the presence of K^+ ions, which is a well-known property of rG4 structures. Notably, Q2 displayed greater stability in the presence of Li^+ ions compared to Q1 and Q3, suggesting that its stability is not solely dependent on K^+ ions. Furthermore, UV-thermal denaturation studies in sodium cacodylate buffer (pH 7.0) containing 100 mM KCl revealed

hypochromic sigmoidal transitions for all three oligonucleotides, with T_m values of 73 (±1), >90, and 75 (±1)°C for Q1, Q2 and Q3, respectively (Figure 1A, D). These results provide experimental evidence supporting the formation of stable G-quadruplex structures within the identified rG4 motifs in MALAT1 lncRNA. The melting and annealing curves of the rG4 structures were found to overlap, indicating the formation of thermodynamically stable intramolecular G quadruplexes. Analysis of T_m values across a 50-fold concentration range (from 1 to 50 μM) revealed no changes, supporting the intramolecular nature of the rG4s (data not shown). For comparison, we performed a UV melting study for a control RNA duplex at 260 nm and 295 nm (Supplementary Figure S1C). The curve at 295 nm did not show any significant changes, while the melting curve at 260 nm yielded a T_m of 49 (±1)°C. This control experiment confirmed that the observed stabilities and melting behaviors for Q1, Q2 and Q3 are specific to the rG4 structures.

Conventional *in vitro* studies typically focus on isolated putative rG4 sites and often overlook the possibility of alternative Watson-Crick base-paired secondary structures that could compete with rG4 formation. To investigate this possibility, we generated a full-length 8.7 kb MALAT1 construct in a pCMV vector plasmid (referred to as FL). Site-directed mutagenesis was then performed by replacing G with A to prevent rG4 formation, resulting in Q1m, Q2m, Q3m (single rG4 mutated), Q12m, Q23m, Q13m (double rG4 mutated), and Q123m (triple rG4 mutated) constructs (Supplementary Figure S1D, E). Thioflavin T (ThT), a dye that exhibits enhanced fluorescence upon binding to rG4 structures compared to single-stranded RNA or RNA hairpin structures, was employed to validate the formation of the rG4 structures (58,59). In vitro transcription (IVT) was performed for the 1.4 kb region containing all three rG4/mutant regions from the plasmid library. The resulting RNAs were gel purified and induced to form rG4 structures in the long strand. Subsequently, they were titrated from 0 to 8 μM against a fixed concentration of 2 μM ThT in a buffer containing 10 mM sodium cacodylate and 100 mM KCl (Figure 1E). The RNA sequence with mutations in all three G-quadruplexes (Q123m) exhibited minimal fluorescence, comparable to the baseline, indicating the absence of rG4 structures. Conversely, the presence of even a single rG4 in Q12m, Q23m and Q13m led to an increase in fluorescence intensity. The intensity further increased when two rG4 sequences were present, as observed in Q1m, Q2m, Q3m and reached its maximum in the FL RNA, where all three rG4 sequences were present. These results provide evidence that the fluorescence signal of ThT is enhanced in the presence of rG4 structures and supports the formation of rG4 structures within the identified sequences. Upon closer examination of Figure 1E, it becomes apparent that the fluorescence intensities of ThT exhibit a significant increase up to a concentration of 0.75 μM for the FL RNA, which contains all three rG4 sequences. Beyond this point, the intensities show a tendency towards saturation. Similarly, for the RNA containing two rG4 sequences, the fluorescence intensities increase up to a concentration of 2 μM, while for the RNA with one rG4 sequence, the intensities increase up to 2 μM before reaching saturation. Subsequently, the fluorescence intensities of ThT continue to rise steadily throughout the
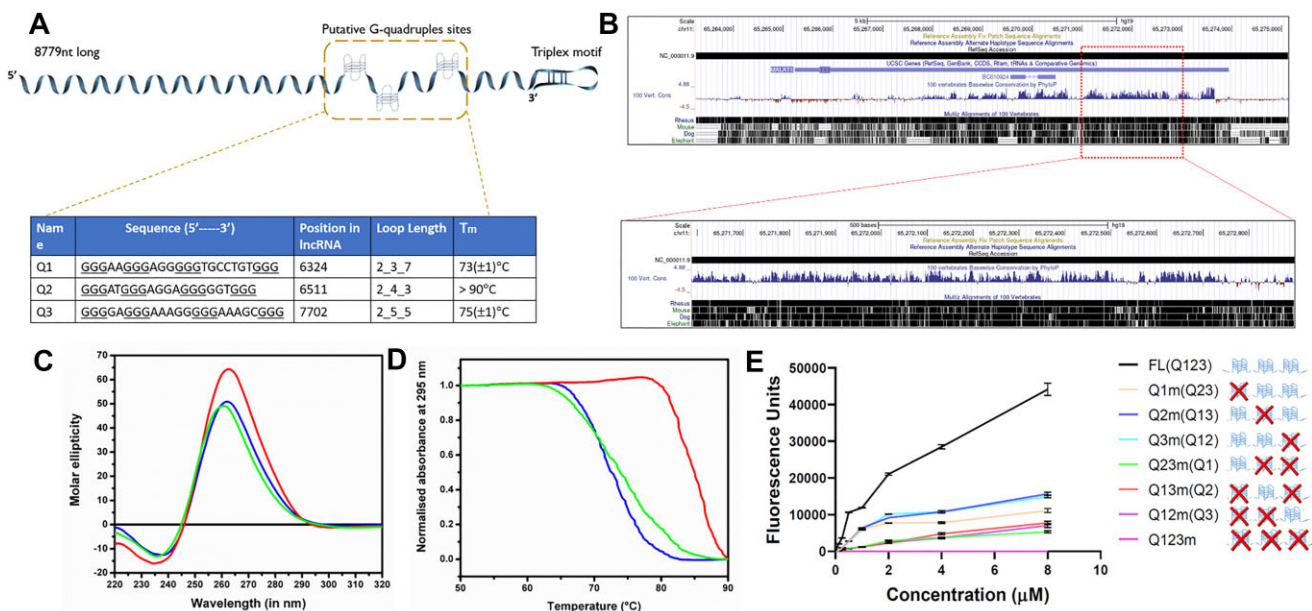
Figure 1. MALAT1 long non-coding RNA harbors three putative rG4 forming domains towards its 3′ end. (**A**) The schematic for long non-coding RNA MALAT1 harboring three putative G-quadruplex motifs as predicted by QGRS mapper, its respective locus, and melting temperature (T_m) as observed in UV Melting studies. (**B**) MALAT1 conservation across vertebrates as depicted in UCSC. Highlighted part expanded below harbors the three rG4 domains and the region is conserved with high PhyloP scores. (**C**) CD spectra for Q1, Q2 and Q3 of MALAT1 depicting characteristic rG4. (**D**) UV Melting of Q1, Q2 and Q3 rG4 of MALAT1 to determine thermal stability. For (C) and (D), Blue, Red and Green lines denote Q1, Q2 and Q3 respectively. (**E**) Fluorescence titration assay of long RNA sequence from MALAT1 containing Q1, Q2 and Q3 together (Q123), the single (Q23m, Q13m, Q12m), double (Q1m, Q2m, Q3m), and triple (Q123m) G-quadruplex mutation (0–8 μM) against ThT (2 μM). S.E.M. plotted for two consecutive readings.

concentration range for the FL RNA with all three rG4 sequences. In contrast, only minimal increases are observed for the other RNAs with two or one rG4 sequence. The observed increases in fluorescence intensities after the saturation point, which occurs at 1 μM for the FL RNA containing all three rG4 sequences and at 2 μM for the other two RNAs (with two or one rG4 sequence) are likely attributable to the formation of micro-aggregates. As the FL RNA with all three rG4 sequences possesses the highest propensity for micro-aggregate formation, it exhibits more pronounced increases in fluorescence intensities at higher concentrations compared to the other RNAs. Recent studies have demonstrated that RNA sequences containing extended stretches of Gs, such as in (G3A1-2)4, are particularly susceptible to the formation of such aggregates under conditions resembling those found in eukaryotic cells (60). When Li⁺ ions were used as a control, a comparable fluorescence pattern was observed for all the various combinations of rG4s/mutants, albeit with slightly lower intensity. This reduction in intensity can be attributed to the relatively low stability of the rG4 structures in the presence of Li + ions (Supplementary Figure S1F). These findings provide evidence that all three G-quadruplex structures can form within the 1.4 kb region of MALAT1 RNA, effectively outcompeting other potential alternative canonical secondary structures.
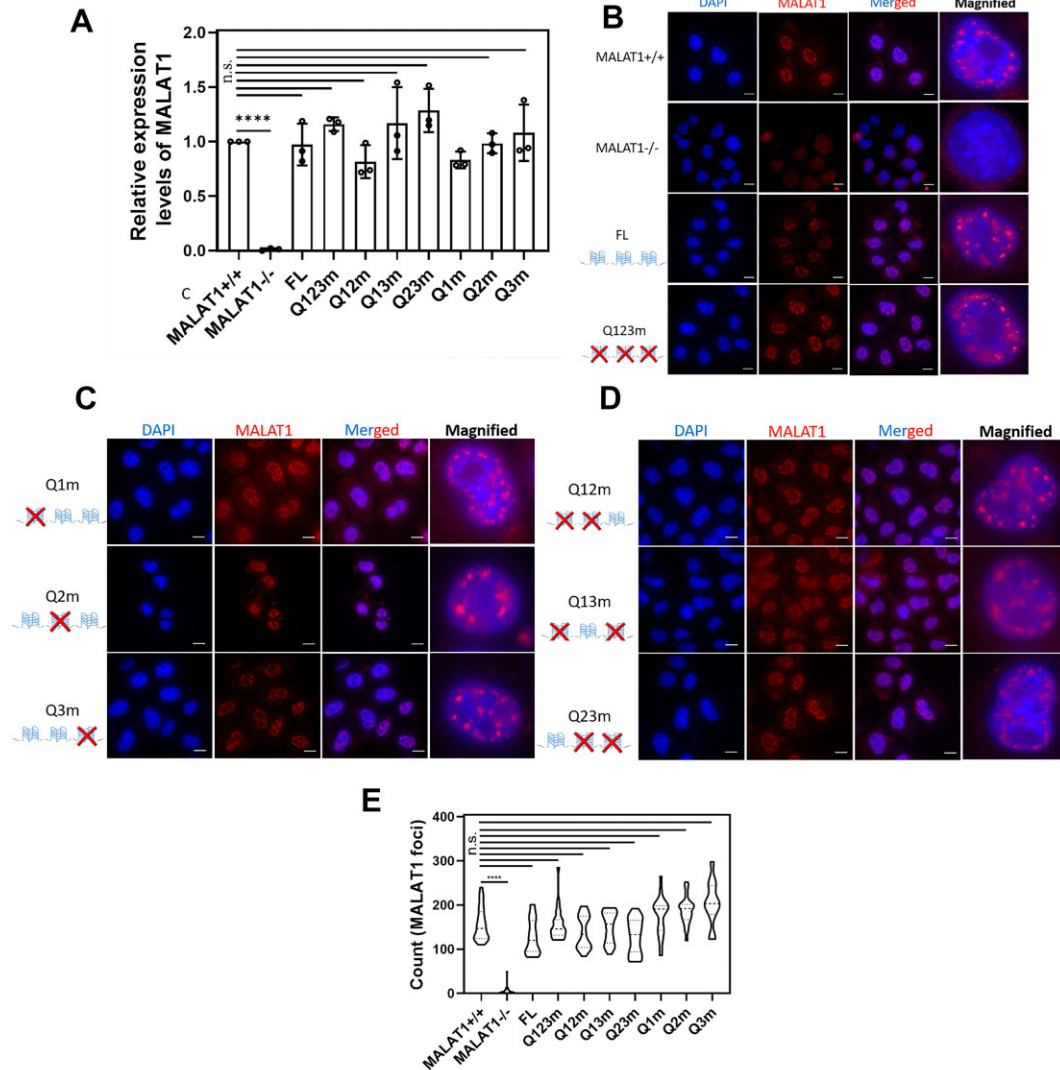
The rG4 structures in MALAT1 do not affect its abundance or localization within cells

The rG4 structures found in other RNAs are known to carry out a variety of structural and regulatory functions.

To assess the impact of the rG4 structures inMALAT1, we transfected MALAT1−/− HeLa cells (see Methods for MALAT1−/− cell-line creation) with plasmids containing full-length MALAT1 (FL) or various mutants, which perturbed rG4s individually. The mutant constructs included disruptions of all rG4s (Q123m), two rG4s (Q12m, Q13m, Q23m), or a single rG4 (Q1m, Q2m, Q3m). We performed qRT-PCR analysis to determine the abundance of the various MALAT1 RNAs. Surprisingly, the results indicated that there was no discernible change in the abundance of MALAT1 across all tested conditions (Figure 2A). This implies that disrupting the rG4 structures did not have an impact on the stability of this long non-coding RNA (lncRNA) inside cells.

Previous studies have reported that the presence of an rG4 consensus motif acts as a localization element for dendritic mRNAs in mouse cortical neurons (26). In the case of MALAT1, a nuclear-enriched lncRNA known to localize in nuclear speckles (36), we sought to investigate whether the presence of rG4 structures in the lncRNA was necessary for its localization. To analyze the subcellular distribution of MALAT1, we performed RNA-fluorescence *in situ* hybridization (FISH) assays. The results revealed that MALAT1 predominantly localized in the nuclear speckles of HeLa cells harboring the MALAT1+/+ plasmid. In contrast, no signal for the lncRNA was observed in MALAT1−/− cells (Figure 2B). Subsequently, we examined the effect of each rG4 on the subcellular distribution of MALAT1 by conducting RNA-FISH assays in MALAT1−/− cells transfected with plasmids containing FL, Q123m, Q12m, Q13m, Q23m, Q1m, Q2m, and Q3m. The rescue experiment with FL MALAT1 demonstrated that it regained localization

Figure 2. The rG4 structures in MALAT1 do not affect its abundance or localization within cells. (**A**) qRT-PCR to quantify the levels of MALAT1 across all the experimental conditions. Relative expression levels of MALAT1 in FL rescue, as well as Q123m, Q12m, Q13m, Q23m, Q1m, Q2m and Q3m by transfection 1 μg of plasmid, were comparable to that of MALAT1+/+ condition. β-Actin house-keeping gene is used as an internal normalization control. Error bars represent ± S.D. across three independent biological replicates. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$ (Student's *t*-test). (**B**) RNA FISH to observe the localization of MALAT1 across different experimental conditions: MALAT1+/+, MALAT1−/−, FL and Q123m. (**C**) RNA FISH to observe the localization of MALAT1 across different experimental conditions: Q12m, Q13m and Q23m. (**D**) RNA FISH to observe the localization of MALAT1 across different experimental conditions: Q1m, Q2m, and Q3m, Q12m. For (B–D), Cells are counterstained with DAPI to mark the nucleus and the scale bar corresponds to 10 μm. (**E**) Image quantification for RNA FISH; MALAT1 foci calculated using ImageJ across different fields for each condition. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$ (Student's *t*-test).

in nuclear speckles (Figure 2B). Notably, complementation with different rG4 mutants did not lead to any change in the localization pattern of this lncRNA (Figure 2B–E). Thus, it can be inferred that the rG4 structures do not serve as a localization cue for MALAT1. Previous studies have indicated that MALAT1 is dispensable for the formation of nuclear speckles. To validate whether the knockout of MALAT1 or the disruption of its rG4 structures affected nuclear speckle formation, we conducted an immunoFISH assay targeting SC35, a nuclear speckle protein, along with MALAT1. Our findings demonstrated that the integrity of nuclear speckles was preserved in all tested conditions, including MALAT1+/+, MALAT1−/−, FL and Q123m rescue conditions (Supplementary Figure S2D–F). Therefore,

based on the collective observations, it can be concluded that the rG4 structures present in MALAT1 do not impact its expression or its localization within cells.

MALAT1 lncRNA binds to specific proteins via its rG4 structures

To investigate whether the rG4s in MALAT1 RNA interact with proteins, we performed streptavidin-based RNA pulldowns with biotinylated Q1, Q2 and Q3 RNAs synthesized *in vitro* and incubated with HeLa cell lysate, followed by mass spectrometry. We included rG4 mutants Q1m, Q2m and Q3m as controls (Figure 3A). Among the enriched proteins identified, nucleolin (NCL) exhibited significant
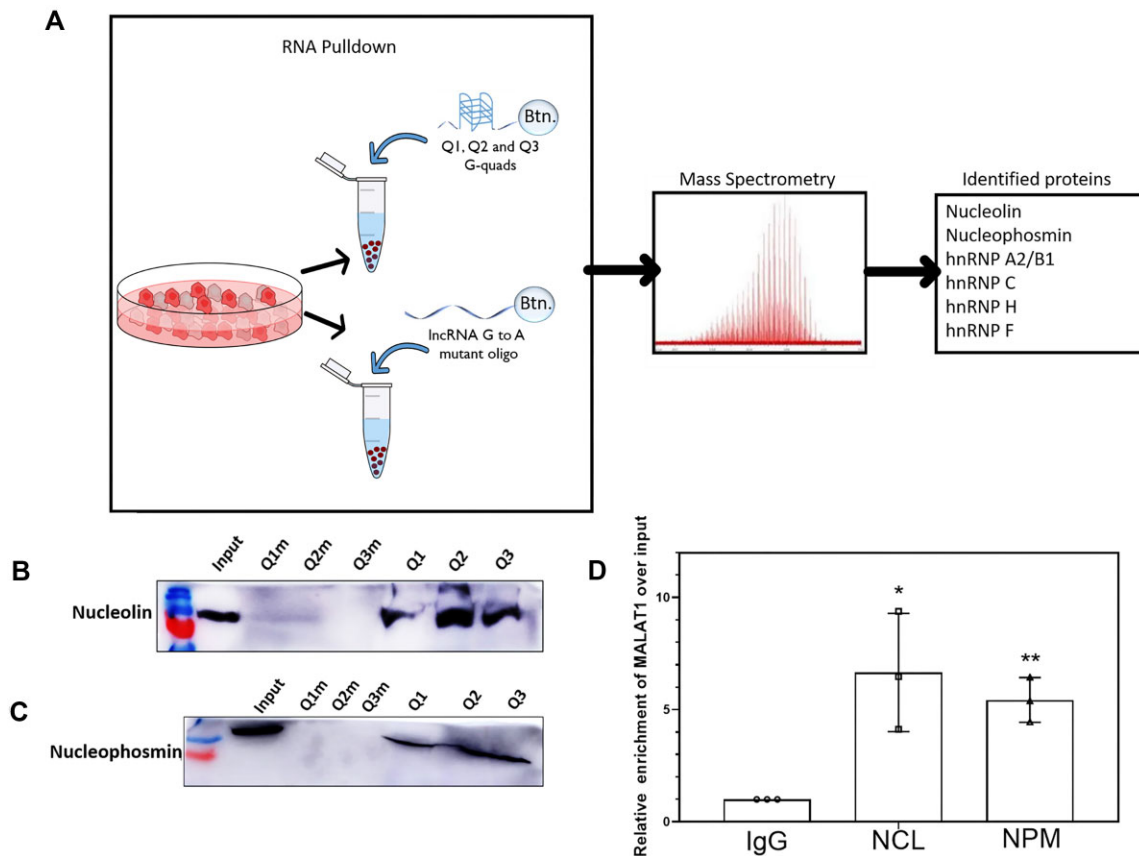
Figure 3. MALAT1 lncRNA binds to specific proteins via its rG4 structures. (**A**) Mass spectrometry representation for RNA pulldown from HeLa cell lysate using Q1, Q2 and Q3 rG4 and their mutated counterparts. A list of enriched proteins is mentioned showing interaction. (**B**) RNA pulldown of Q1, Q2, Q3 G-quadruplex, and their mutated counterparts followed by western blot to detect NCL protein. (**C**) RNA pulldown of Q1, Q2, Q3 G-quadruplex, and their mutated counterparts followed by western blot to detect NPM protein. (**D**) qRT-PCR to quantify the levels of MALAT1 enrichment in reverse pulldown or RNA-immunoprecipitation performed with NCL and NPM proteins. Enrichment of NCL is shown in comparison to IgG control. Error bars represent ± S.D. across three independent biological replicates. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$ (Student's *t*-test).

binding to the rG4 motifs. Another protein, nucleophosmin (NPM), displayed some specific binding to rG4 structures as well. hnRNP A2/B1, C and H were also detected in the pulldown, but their enrichment was comparable between wild-type and mutated rG4s. To validate the RNA-protein interactions, we performed western blot analysis using specific antibodies against NCL and NPM. These experiments confirmed that both proteins specifically bind Q1, Q2 and Q3, but not the mutated G-quadruplex motifs (Figure 3B, C). A reverse pulldown or RNA-immunoprecipitation experiment using NCL and NPM antibodies further confirmed their binding to MALAT1 lncRNA. Quantitative real-time PCR analysis showed a 5- to 7-fold enrichment of MALAT1 in both NCL and NPM samples compared to the IgG control (Figure 3D).

MALAT1 rG4s aid localization of the interacting protein partners

The rG4 structures present in MALAT1 do not directly contribute to its localization within the nuclear speckles. Instead, they bind specifically to certain nuclear proteins. We hypothesized that the localization of these proteins might be dependent on MALAT1. To investigate this possibility, we performed immunocytochemistry (ICC) and imaging of NCL, NPM, hnRNP H, hnRNP C, and hnRNP A2/B1 in both MALAT1+/+ and MALAT1−/− HeLa cells. In MALAT1+/+ cells, NCL and NPM were observed in the nucleolus as well as in the nuclear speckles, consistent with previous reports (61–64). However, in MALAT1−/− cells, NCL was detected only in the nucleolus, not in nuclear speckles (Supplementary Figure S4A), suggesting that the localization of NCL to speckles depends on MALAT1. To confirm that the NCL foci were indeed part of the nuclear speckles, we performed co-localization experiments between NCL and the speckle marker SC35 in both MALAT1+/+ and MALAT1−/− cells. In MALAT1+/+ cells, NCL and SC35 exhibited colocalization, while in MALAT1−/− cells, only SC35 was observed in the speckles, with NCL confined to the nucleolus and absent from the speckles (Supplementary Figure S4B, C). Similar results were obtained for NPM in both conditions (Supplementary Figure S4D). In contrast, hnRNP H and hnRNP C appeared dispersed throughout the nucleoplasm in the MALAT1+/+ and MALAT1−/− cells, (Supplementary Figure S4E, F). Similarly, the distribution of hnRNP A2/B1

puncta across the nucleus did not exhibit any difference in the presence or absence of MALAT1 (Supplementary Figure S4G).

Having shown that the localization of NCL and NPM to the nuclear speckles is dependent on the presence of MALAT1, we went on to examine the dependence of NCL and NPM localization on the rG4s within MALAT1. We performed a plasmid-dependent rescue of MALAT1, along with the expression of all three rG4-mutated versions of the lncRNA, in the MALAT1–/– cell line. Subsequently, we conducted ICC to observe the localization of these proteins (as depicted in the schematic in Supplementary Figure S4H). With the plasmid carrying FL MALAT1, NCL was observed in the nuclear speckles as well as in the nucleolus, as in MALAT1+/+ cells. In contrast, NCL did not localize to the nuclear speckles in cells carrying MALAT1 with mutant rG4s (Q123m), as in MALAT1–/– cells (Figure 4A). These findings strongly support a crucial role for the rG4 structures within MALAT1 in localizing NCL to the nuclear speckles. To further investigate the requirement of individual or combined rG4s for proper localization, we conducted complementation experiments in MALAT1–/– cells using the full-length lncRNA with mutations in two rG4s (Q12m, Q23m, Q13m) or a single rG4 (Q1m, Q2m, Q3m). Subsequent ICC and imaging of NCL revealed that in all cases, NCL was exclusively observed in the nucleolus and not in the speckles, as in MALAT1–/– cells (Figure 4B, C). Quantitative analysis of NCL foci within the nuclear speckles further confirmed their presence in MALAT1+/+ or full-length (FL) rescue conditions, and absence in MALAT1–/– or any of the rG4 mutant combinations (Figure 4D). Thus, our data strongly suggest that each of the three rG4s within MALAT1 is essential for localizing NCL to the nuclear speckles.

To assess the co-localization of NCL and MALAT1, we performed immunoFISH experiments with MALAT1+/+, MALAT1–/–, FL rescue and Q123m cells. These experiments revealed the presence of MALAT1 in all except the MALAT1–/– cells. Additionally, MALAT1 co-localized with NCL in the nuclear speckles only in MALAT1+/+ and FL rescue conditions and was absent from speckles in the other two conditions (Figure 4E, F). Similarly, the NPM protein signal within nuclear speckles was restored in the presence of MALAT1 but not with any of the rG4 mutants (Figure 4G, H). In summary, our results strongly support a crucial role for the rG4 structures present in the 3′ region of MALAT1 in the subcellular localization of both NCL and NPM within nuclear speckles.

NCL directly binds to G-quadruplex structures *in vitro* and is a bona fide partner

NCL, a well-known RNA-binding protein, has been previously reported to interact with rG4 structures in various studies (24,65,66). Given its high peptide score with the MALAT1 rG4s in the mass spectrometry analysis and the results of our immunoFISH experiments showing that NCL's subcellular localization is dependent on the G-quadruplex structures of MALAT1, we conducted further experiments to determine whether it directly interacts with the rG4s of MALAT1 in vitro. Since the N-terminus of NCL

undergoes self-cleavage (67,68), we purified a truncated version (ΔNCL) that retained all four RNA Recognition Motifs (RRMs) and the GAR domain (Supplementary Figure S5A, B). Electrophoretic mobility shift assays (EMSA) were performed using ΔNCL and pre-formed Q1, Q2 and Q3 quadruplexes individually to confirm binding of the protein to these rG4s (Figure 5A–C). Increasing the concentration of ΔNCL while keeping the pre-formed rG4 concentration constant resulted in a shift in all EMSAs performed for Q1, Q2 and Q3, indicating the binding of the protein to the rG4s.

To analyze the binding parameters, we conducted surface plasmon resonance (SPR) and fluorescence titration experiments (Figure 5D) to determine the binding affinities between the rG4s and ΔNCL. All three structures exhibited a high binding affinity towards ΔNCL, with Q2 showing the highest affinity with a Kd value of 4 nM (Table 1). Furthermore, we confirmed the binding of NCL to the rG4 structures of MALAT1 by conducting RNA-immunoprecipitation using NCL antibody from HeLa cell lysates of MALAT1+/+, MALAT1–/–, FL, and Q123m transfected MALAT1–/– cells. Enrichment analysis by qRT-PCR revealed a 7-fold and 5-fold enrichment of MALAT1 in the MALAT1+/+ cells and cells carrying the MALAT1 FL plasmids, respectively, while there was no enrichment in cells carrying the MALAT1 Q123m triple G-quadruplex mutant, with a signal equivalent to that observed in the MALAT1–/– cells (Figure 5E).

To validate the specificity of the NCL antibody, we performed siRNA-mediated down-regulation of NCL and examined the knockdown efficiency through qRT-PCR assays and western blotting (Supplementary Figures S6A, B). The knockdown resulted in an approximately 50% reduction of NCL protein levels, and subsequent ICC in the NCL down-regulated cells (Supplementary Figure S6C) demonstrated a significant decrease in NCL signal within the nuclear speckles of the siRNA-transfected cells. This finding confirms the specificity of the antibody for NCL detection. Consequently, based on these experiments, we can confidently conclude that NCL is an authentic interacting partner of the three G-quadruplex structures of MALAT1 (Q1, Q2 and Q3). NCL primarily resides in the nucleolus, where it plays a critical role in ribosomal biogenesis, constituting around 90% of the protein population. However, a minor portion of NCL is capable of translocating to the nuclear speckles. Prior research has indicated that NCL exhibits co-localization and interacts with the pre-catalytic spliceosome complex, actively contributing to the alternative splicing of fibronectin mRNA (69). Using R-DeeP, a tool designed for exploring protein-protein interactions in the presence or absence of RNAs (70,71), it was observed that NCL engages with splicing proteins SON, SRSF2, SRSF3, SRSF5 and SRSF11 in an RNA-dependent manner. The interaction between NCL and these splicing proteins was disrupted when RNA was absent, as demonstrated by the loss of interaction following RNase treatment (Supplementary Figure S6D, Supplementary File S1). Based on these findings, we propose that the rG4-mediated localization of NCL to nuclear speckles in MALAT1 may be a dynamic and transient process, facilitating its participation in the splicing mechanism through interaction with SR proteins, as supported by the R-DeeP analysis.
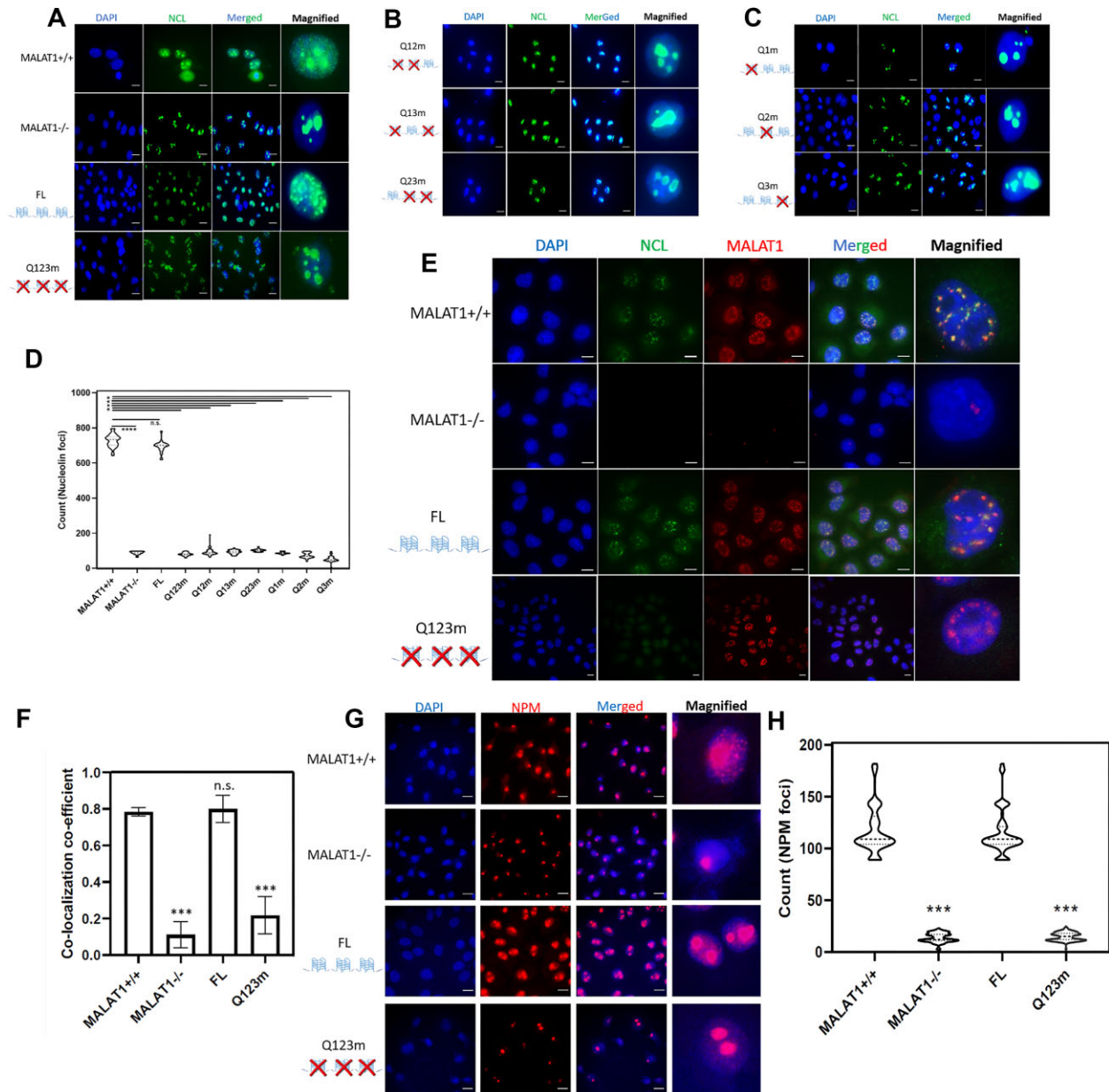
Figure 4. MALAT1 rG4s aid localization of the interacting protein partners. (**A**) ICC to observe the localization of NCL across different experimental conditions: MALAT1+/+, MALAT1−/−, FL, and Q123m. (**B**) ICC to observe the localization of NCL across different experimental conditions: Q12m, Q13m and Q23m. (**C**) ICC to observe the localization of NCL across different experimental conditions: Q1m, Q2m and Q3m. (**D**) Image quantification for ICC; NCL foci calculated using ImageJ across different fields under each condition. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ (Student's *t*-test). (**E**) Co-localization experiment to confirm MALAT1 and NCL localization together in the nuclear speckles by immunoFISH. Antibody used for acetylated (K88) NCL to mark protein specifically in nuclear speckles and not nucleolus. Co-localization shows a yellow signal. (**F**) Quantification for Pearson Coefficient of Co-localization for NCL and MALAT1 performed by ImageJ. Error bars represent ± S.D. across three independent biological replicates. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ (Student's *t*-test). (**G**) ICC to observe the localization of NPM across different experimental conditions: MALAT1+/+, MALAT1−/−, FL and Q123m (where all the three G-quadruplexes are mutated). For (A–C), (E) and (G), cells are counterstained with DAPI to mark the nucleus, and the scale bar corresponds to 10 μM. (**H**) Image quantification for NPM foci calculated using ImageJ across different fields for each condition. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ (Student's *t*-test).

DISCUSSION

As a multi-functional lncRNA, MALAT1 is involved in multivalent interactions with different RNA binding proteins (RBPs), and structural motifs in MALAT1 may serve as a scaffold to facilitate such interactions. In this study, we establish for the first time that three rG4 structures are present in MALAT1, and that they provide a structural motif to interact with NCL and NPM.

Although a previous study disputed the presence of rG4 structures in mouse Malat1 (28), multiple reports have supported the existence of rG4s in other lncRNAs and in different cellular contexts (20,25,53–56,72). rG4 structures are highly dynamic and do not always form in living cells. Their
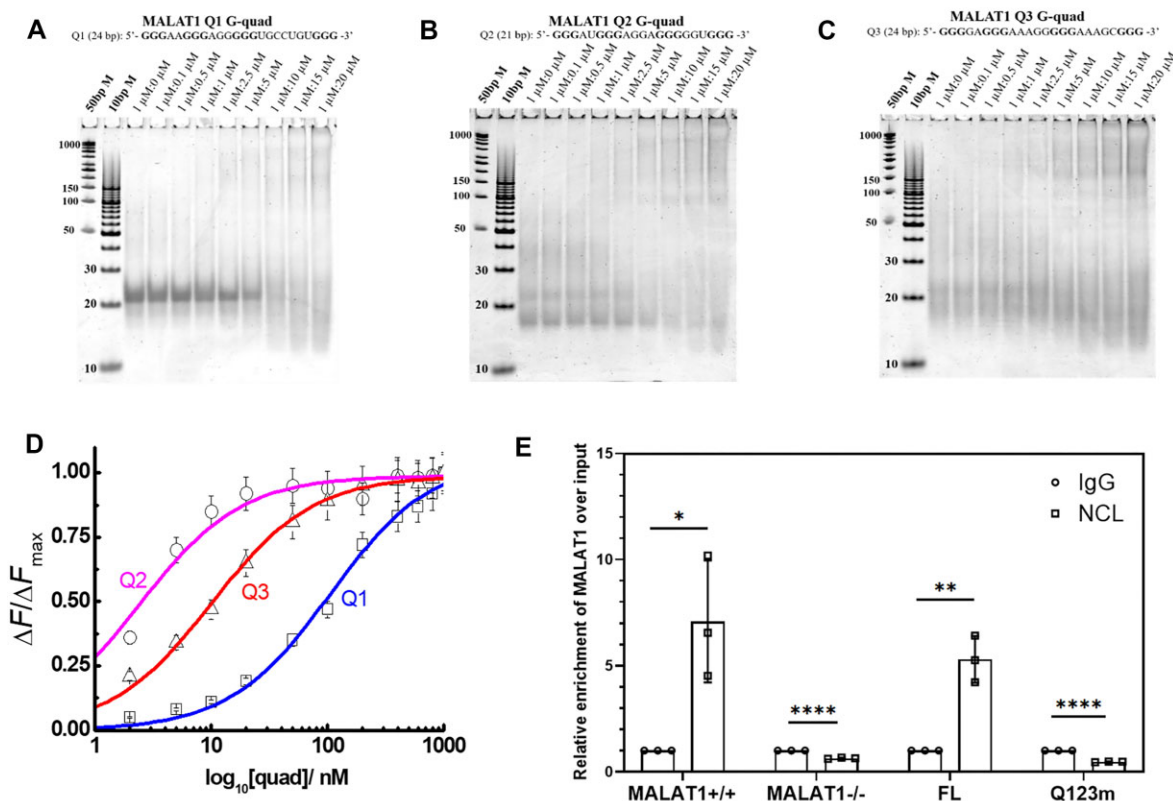
Figure 5. NCL directly binds to G-quadruplex structures *in vitro* and is a bona fide partner. (**A–C**) EMSA of Q1, Q2 and Q3 MALAT1 RNA G-quadruplex with ΔNCL protein, with the RNA kept constant and protein concentrations varied from 0 to 20 μM. (**D**) Fluorescence titrations of ΔNCL protein in the presence of an increasing concentration of Q1 (blue), Q2 (magenta) and Q3 (red) rG4s. Data points are the average of three technical replications. Solid lines represent fits of the experimental data points with the binding equation mentioned in the methods section. (**E**) RNA immunoprecipitation of NCL to observe an enrichment of MALAT1 compared to IgG control. The experiment was performed across MALAT1+/+, MALAT1–/–, FL and Q123m conditions. Error bars represent ± S.D. across three independent biological replicates. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$ (Student's *t*-test).

Table 1. Binding parameters for MALAT1 rG4s and ΔNCL by surface plasmon resonance (SPR) and fluorescence titration of rG4s and ΔNCL

	SPR			Fluorescence
rG4	K_a (M^{-1} s^{-1})	k_d (s^{-1})	K_d (nM)	K_d (nM)
Q1	$0.89 (\pm 0.10) \times 10^4$	$7.01 (\pm 0.11) \times 10^{-4}$	79	103 ($\pm$ 7.5)
Q2	$1.66 (\pm 0.13) \times 10^4$	$0.61 (\pm 0.10) \times 10^{-4}$	4	3 ($\pm$ 0.2)
Q3	$2.72 (\pm 0.06) \times 10^4$	$4.07 (\pm 0.24) \times 10^{-4}$	15	10 ($\pm$ 0.9)

formation may require specific cues and could be resolved once their biological function is fulfilled. These structures might arise during specific cell cycle stages or in particular disease conditions where the cellular environment supports their formation. Furthermore, certain RNA quadruplexes may be specific to developmental stages or particular tissues. Recent research utilizing three experimental datasets proposed a potential secondary structural model for human MALAT1 spanning 8425 nucleotides (19). According to this model, human MALAT1 exhibits extensive structure, including 194 helices, 13 pseudoknots, five structured tetraloops, as well as numerous internal loops and

long-range intramolecular interactions. Some local structures within MALAT1 have also been characterized in detail. One extensively studied local structure is the triple helix known as the 3′ terminal stability element for nuclear expression (ENE), which protects MALAT1 from exonucleolytic degradation (37,39,40). It has been demonstrated that the methyltransferase-like protein 16 (METTL16), an abundant nuclear protein, interacts with the MALAT1 triple helix both *in vitro* and *in vivo* (73). Additionally, two well-structured hairpins at positions 2509–2537 nucleotide and 2556–2586 nucleotide have been identified, facilitating the binding of heterogeneous nuclear ribonucleoprotein G and ribonucleoprotein C (hnRNP G and hnRNP C), respectively. These proteins are involved in pre-mRNA processing (74,75). Bioinformatics analyses have revealed the presence of three conserved rG4s in the 3′ end of MALAT1, and we have experimentally confirmed that these motifs can individually form stable intramolecular parallel G-quadruplex structures *in vitro*. When testing the longer transcript of MALAT1, fluorescence emission using the ThT rG4-sensing fluorescent molecule increased with the introduction of each additional quadruplex site, reaching its maximum when all three rG4 motifs were intact. Thus, the rG4s within this lncRNA represent a novel structural

module that fine-tunes the biological regulation mediated by MALAT1.

Although MALAT1 is specifically localized in nuclear speckles, depleting it does not affect the formation of these structures (29,35). We have determined that the rG4 structures within the MALAT1 transcript do not influence its stability, expression, or act as localization cues for the lncRNA itself. However, we have discovered that several proteins, including NCL and NPM, specifically associate with these rG4 structures. Previous studies have also identified proteins interacting with MALAT1. Chen *et al.* identified 127 potential MALAT1-interacting proteins using an RNA pulldown followed by quantitative proteomics using the Stable Isotope Labeling of Amino acids In Cell culture (SILAC) method on a fragment of human MALAT1 (32). Similarly, Scherer *et al.* performed mass spectrometry studies on 14 non-overlapping fragments of full-length mouse MALAT1 to identify potential nuclear-interacting proteins (76). Among the 35 binding proteins they identified, 14 overlapped with our study. While some proteins, such as hnRNPs, were common to both studies, our work revealed NCL and NPM as interacting protein partners for MALAT1 for the first time. Differences in the observed MALAT1 interacting protein partners could be attributed to the use of different cell lysate sources in previous experiments. We used HeLa cell lysates, whereas HepG2 and NSC-34 cell lysates were used in earlier studies. Furthermore, RNA secondary structures are highly dependent on the cellular environment, including factors like salt and buffer conditions. For our experiments, we employed high monovalent salt concentrations (100 mM KCl) to promote rG4 formation, which differs from the salt concentrations used in previous studies (120 and 10 mM NaCl). As a result, the RNAs used in the pull-down experiments may adopt different structures and interact with distinct sets of proteins. Indeed, the proteins we identified, such as NCL, are known to specifically interact with rG4 structures (23,24). Interestingly, we observed that both NCL and NPM bound specifically to the rG4s of MALAT1 and not to its mutant counterpart, while some hnRNP proteins interacted with both structured and unstructured RNA. We also found that the subcellular distribution of hnRNPs remained unaffected regardless of the presence or absence of the MALAT1 transcript. Recent studies have highlighted that certain hnRNPs, like hnRNP H1, can interact with both unstructured G-tracts and rG4 structures *in vitro*, which supports our findings (77). Importantly, we demonstrated that the localization of NCL and NPM to nuclear speckles was dependent on the rG4 structures of MALAT1. We further established that all three rG4s were responsible for localizing these proteins to the nuclear speckles. To the best of our knowledge, this is the first evidence showing that rG4 structures in a lncRNA play a role in the subcellular localization of specific proteins. In our study, we measured the *in vitro* binding affinity of NCL to the three rG4s present in MALAT1 and found that NCL strongly binds to all three structures. Our experiments validate that NCL can recognize and bind to each rG4 of MALAT1. However, in HeLa cells, disruption of any one rG4 results in the loss of NCL localization to nuclear speckles.

Due to their extended single-stranded nature, lncRNAs often undergo intramolecular interactions, leading to the formation of loops that facilitate various interactions with other molecules or within the lncRNA molecule itself. We propose the presence of such loops in MALAT1 where the rG4s form, which may interconnect or depend on each other. Consequently, disrupting one rG4 structure in the cellular context could potentially result in the loss of the other two, ultimately leading to the loss of NCL recognition and re-localization. Additionally, it is possible that the NCL protein, while interacting with each rG4, also interacts with other proteins, particularly splicing proteins, as evidenced by the R-DeeP analysis. It is plausible that all three rG4 structures play a crucial role in maintaining the integrity and function of the entire complex. Therefore, the loss of any one rG4 may destabilize the entire complex, ultimately resulting in the removal of NCL from nuclear speckles.

MALAT1 has been extensively studied due to its involvement in alternative splicing regulation in various cancer cells (29,78,79). Conversely, NCL primarily functions in ribosomal synthesis and is predominantly present in the nucleolus. The smaller pool of NCL localized in nuclear speckles does not contribute to rRNA production, as evidenced by rDNA ChiP studies (64). Instead, NCL is strongly implicated in splicing processes. It colocalizes with SC35 in nuclear speckles, interacts with the pre-catalytic spliceosome complex (69), and regulates alternative splicing, particularly of fibronectin. NCL also plays a crucial role in splice site selection (80). Our study sheds light on the rG4-mediated localization of NCL and NPM in nuclear speckles within the context of MALAT1. The R-DeeP analysis presented here supports the hypothesis that the rG4-mediated localization of NCL to nuclear speckles through MALAT1 may explain its role in splicing. These findings lay the groundwork for a novel approach that targets the structural scaffold of MALAT1 lncRNA rG4s, providing a platform for designing ligands that selectively target specific binding partners. Small molecules targeting the rG4 structure could potentially disrupt interactions with trans-acting factors and structural elements within MALAT1, opening new avenues for therapeutic intervention.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD026386.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Quinn,J.J. and Chang,H.Y. (2015) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.*, **17**, 47–62.
2. van Bakel,H., Nislow,C., Blencowe,B.J. and Hughes,T.R. (2010) Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.*, **8**, e1000371.
3. Kapranov,P., Cawley,S.E., Drenkow,J., Bekiranov,S., Strausberg,R.L., Fodor,S.P.A. and Gingeras,T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 916–919.
4. Xue,Z., Hennelly,S., Doyle,B., Gulati,A.A., Novikova,I.V., Sanbonmatsu,K.Y. and Boyer,L.A. (2016) A G-rich motif in the lncRNA Braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell*, **64**, 37–50.
5. Guttman,M., Amit,I., Garber,M., French,C., Lin,M.F., Feldser,D., Huarte,M., Zuk,O., Carey,B.W., Cassady,J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding rnas in mammals. *Nature*, **458**, 223–227.
6. Pruszko,M., Milano,E., Forcato,M., Donzelli,S., Ganci,F., Di Agostino,S., De Panfilis,S., Fazi,F., Bates,D.O., Bicciato,S. *et al.* (2017) The mutant p53-ID4 complex controls VEGFA isoforms by recruiting lncRNA MALAT1. *EMBO Rep.*, **18**, 1331–1351.
7. Marchese,F.P., Raimondi,I. and Huarte,M. (2017) The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.*, **18**, 206.
8. Sun,Q., Hao,Q. and Prasanth,K.V. (2018) Nuclear long noncoding rnas: key regulators of gene expression. *Trends Genet.*, **34**, 142–157.
9. Aillaud,M. and Schulte,L.N. (2020) Emerging roles of long noncoding rnas in the cytoplasmic milieu. *Non-coding RNA*, **6**, 44.
10. Statello,L., Guo,C.J., Chen,L.L. and Huarte,M. (2021) Gene regulation by long non-coding rnas and its biological functions. *Nat. Rev. Mol. Cell Biol.*, **22**, 96–118.
11. Huarte,M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.*, **21**, 1253–1261.
12. Huarte,M. and Rinn,J.L. (2010) Large non-coding rnas: missing links in cancer? *Hum. Mol. Genet.*, **19**, 152–161.
13. Agarwala,P., Pandey,S. and Maiti,S. (2015) The tale of RNA G-quadruplex. *Org. Biomol. Chem.*, **13**, 5570–5585.
14. Fay,M.M., Lyons,S.M. and Ivanov,P. (2017) RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.*, **429**, 2127–2147.
15. Hänsel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
16. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
17. Herdy,B., Mayer,C., Varshney,D., Marsico,G., Murat,P., Taylor,C., D'Santos,C., Tannahill,D. and Balasubramanian,S. (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.*, **46**, 11592–11604.
18. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
19. McCown,P.J., Wang,M.C., Jaeger,L. and Brown,J.A. (2019) Secondary structural model of human MALAT1 reveals multiple structure–function relationships. *Int. J. Mol. Sci.*, **20**, 5610.
20. Matsumura,K., Kawasaki,Y., Miyamoto,M., Kamoshida,Y., Nakamura,J., Negishi,L., Suda,S. and Akiyama,T. (2017) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, **36**, 1191–1199.
21. Murat,P., Zhong,J., Lekieffre,L., Cowieson,N.P., Clancy,J.L., Preiss,T., Balasubramanian,S., Khanna,R. and Tellam,J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.*, **10**, 358–364.
22. Yamazaki,T., Souquere,S., Chujo,T., Kobelke,S., Chong,Y.S., Fox,A.H., Bond,C.S., Nakagawa,S., Pierron,G. and Hirose,T. (2018) Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell*, **70**, 1038–1053.
23. Wu,R., Li,L., Bai,Y., Yu,B., Xie,C., Wu,H., Zhang,Y., Huang,L., Yan,Y., Li,X. *et al.* (2020) The long noncoding RNA LUCAT1 promotes colorectal cancer cell proliferation by antagonizing nucleolin to regulate MYC expression. *Cell Death. Dis.*, **11**, 908.
24. Bian,W.X., Xie,Y., Wang,X.N., Xu,G.H., Fu,B.S., Li,S., Long,G., Zhou,X. and Zhang,X.L. (2019) Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res.*, **47**, 56–68.
25. Simko,E.A.J., Liu,H., Zhang,T., Velasquez,A., Teli,S., Haeusler,A.R. and Wang,J. (2020) G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Res.*, **48**, 7421–7438.
26. Subramanian,M., Rage,F., Tabet,R., Flatter,E., Mandel,J.-L. and Moine,H. (2011) G–quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.*, **12**, 697–704.
27. Pietras,Z., Wojcik,M.A., Borowski,L.S., Szewczyk,M., Kulinski,T.M., Cysewski,D., Stepien,P.P., Dziembowski,A. and Szczesny,R.J. (2018) Dedicated surveillance mechanism controls G-quadruplex forming non-coding rnas in human mitochondria. *Nat. Commun.*, **9**, 2558.
28. Guo,J.U. and Bartel,D.P. (2017) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, aaf5371.
29. Tripathi,V., Ellis,J.D., Shen,Z., Song,D.Y., Pan,Q., Andrew,T., Freier,S.M., Bennett,C.F., Sharma,A., Bubulya,P.A. *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, **39**, 925–938.
30. Tripathi,V., Shen,Z., Chakraborty,A., Giri,S., Freier,S.M., Wu,X., Zhang,Y., Gorospe,M., Prasanth,S.G., Lal,A. *et al.* (2013) Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLos Genet.*, **9**, e1003368.
31. West,J.A., Davis,C.P., Sunwoo,H., Simon,M.D., Sadreyev,R.I., Wang,P.I., Tolstorukov,M.Y. and Kingston,R.E. (2014) The long noncoding rnas NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell*, **55**, 791–802.
32. Chen,R., Liu,Y., Zhuang,H., Yang,B., Hei,K., Xiao,M., Hou,C., Gao,H., Zhang,X., Jia,C. *et al.* (2017) Quantitative proteomics reveals that long non-coding RNA MALAT1 interacts with DBC1 to regulate p53 acetylation. *Nucleic Acids Res.*, **45**, 9947–9959.
33. Zhang,B., Arun,G., Mao,Y.S., Lazar,Z., Hung,G., Bhattacharjee,G., Xiao,X., Booth,C.J., Wu,J., Zhang,C. *et al.* (2013) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.*, **2**, 111–123.
34. Lamond,A.I. and Spector,D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.*, **4**, 605–612.
35. Nakagawa,S., Ip,J.Y., Shioi,G., Tripathi,V., Zong,X., Hirose,T. and Prasanth,K.V. (2012) Malat1 is not an essential component of nuclear speckles in mice. *RNA*, **18**, 1487–1499.
36. Miyagawa,R., Tano,K., Mizuno,R., Nakamura,Y., Ijiri,K., Rakwal,R., Shibato,J., Masuo,Y., Mayeda,A., Hirose,T. *et al.* (2012) Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA*, **18**, 738–751.
37. Steitz,T.A. and Steitz,J.A. (2015) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.*, **21**, 633–640.
38. Brown,J.A., Kinzig,C.G., Degregorio,S.J. and Steitz,J.A. (2016) Hoogsteen-position pyrimidines promote the stability and function of the MALAT1 RNA triple helix. *RNA*, **22**, 743–749.

39. Brown,J.A., Valenstein,M.L., Yario,T.A., Tycowski,K.T. and Steitz,J.A. (2012) Formation of triple-helical structures by the 3′-end sequences of MALAT1 and menβ noncoding rnas. *Proc. Natl. Acad. Sci. USA*, **109**, 19202–19207.

40. Wilusz,J.E., Jnbaptiste,C.K., Lu,L.Y., Marzluff,W.F., Kuhn,C., Joshua-tor,L. and Sharp,P.A. (2012) A triple helix stabilizes the 3' ends of long noncoding rnas that lack poly (A) tails. *Genes & Dev.*, **26**, 2392–2407.

41. Sanchez de Groot,N., Armaos,A., Grana-Montes,R., Alriquet,M., Calloni,G., Vabulas,R.M. and Tartaglia,G.G. (2019) RNA structure drives interaction with proteins. *Nat. Commun.*, **10**, 3246.

42. Corley,M., Burns,M.C. and Yeo,G.W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell*, **78**, 9–29.

43. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.

44. Schaeffer,C., Bardoni,B., Mandel,J.L., Ehresmann,B., Ehresmann,C. and Moine,H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBOJ*, **20**, 4803–4813.

45. Sexton,A.N. and Collins,K. (2011) The 5′ guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol. Cell. Biol.*, **31**, 736–743.

46. Booy,E.P., Meier,M., Okun,N., Novakowski,S.K., Xiong,S., Stetefeld,J. and McKenna,S.A. (2012) The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.*, **40**, 4110–4124.

47. Huang,H., Zhang,J., Harvey,S.E., Hu,X. and Cheng,C. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.*, **31**, 2296–2309

48. Pulido-Quetglas,C., Aparicio-Prat,E., Arnan,C., Polidori,T., Hermoso,T., Palumbo,E., Ponomarenko,J., Guigo,R. and Johnson,R. (2017) Scalable design of paired CRISPR guide rnas for genomic deletion. *PLoS Comput. Biol.*, **13**, e1005341.

49. Aparicio-Prat,E., Arnan,C., Sala,I., Bosch,N., Guigó,R. and Johnson,R. (2015) DECKO: single-oligo, dual-CRISPR deletion of genomic elements including long non-coding rnas. *BMC Genomics*, **16**, 846.

50. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2−\Delta\Delta CT$ method. *Methods*, **25**, 402–408.

51. Yadav,A.K., Bhardwaj,G., Basak,T., Kumar,D., Ahmad,S., Priyadarshini,R., Singh,A.K., Dash,D. and Sengupta,S. (2011) A systematic analysis of eluted fraction of plasma post immunoaffinity depletion: implications in biomarker discovery. *PLoS One*, **6**, e24442.

52. Jayaraj,G.G., Pandey,S., Scaria,V. and Maiti,S. (2012) Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol.*, **9**, 81–89.

53. Kwok,C.K., Marsico,G., Sahakyan,A.B., Chambers,V.S. and Balasubramanian,S. (2016) RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods*, **13**, 841–844.

54. Kwok,C.K., Marsico,G. and Balasubramanian,S. (2018) Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harb. Perspect. Biol.*, **10**, a032284.

55. Yang,S.Y., Lejault,P., Chevrier,S., Boidot,R., Robertson,A.G., Wong,J.M.Y. and Monchaud,D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.*, **9**, 4730.

56. Renard,I., Grandmougin,M., Roux,A., Yang,S.Y., Legault,P., Pirrotta,M., Wong,J.M.Y. and Monchaud,D. (2019) Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. *Nucleic Acids Res.*, **47**, 502–510.

57. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, 676–682.

58. Lages,A., Proud,C.G., Holloway,J.W. and Vorechovsky,I. (2019) Thioflavin T monitoring of Guanine quadruplex formation in the rs689-dependent INS Intron 1. *Mol. Ther. - Nucleic Acids*, **16**, 770–777.

59. Xu,S., Li,Q., Xiang,J., Yang,Q., Sun,H., Guan,A., Wang,L., Liu,Y., Yu,L., Shi,Y. *et al.* (2016) Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes. *Sci. Rep.*, **6**, 24793.

60. Williams,A.M., Poudyal,R.R. and Bevilacqua,P.C. (2021) Long tracts of guanines drive aggregation of RNA G-quadruplexes in the presence of spermine. *Biochemistry*, **60**, 2715–2726.

61. Liu,X., Liu,Z., Jang,S.W., Ma,Z., Shinmura,K., Kang,S., Dong,S., Chen,J., Fukasawa,K. and Ye,K. (2007) Sumoylation of nucleophosmin/B23 regulates its subcellular localization, mediating cell proliferation and survival. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 9679–9684.

62. Lam,L., Aktary,Z., Bishay,M., Werkman,C., Kuo,C.Y., Heacock,M., Srivastava,N., MacKey,J.R. and Pasdar,M. (2012) Regulation of subcellular distribution and oncogenic potential of nucleophosmin by plakoglobin. *Oncogenesis*, **1**, e4.

63. Tarapore,P., Shinmura,K., Suzuki,H., Tokuyama,Y., Kim,S.H., Mayeda,A. and Fukasawa,K. (2006) Thr199 phosphorylation targets nucleophosmin to nuclear speckles and represses pre-mRNA processing. *FEBS Lett.*, **580**, 399–409.

64. Das,S., Cong,R., Shandilya,J., Senapati,P., Moindrot,B., Monier,K., Delage,H., Mongelard,F., Kumar,S., Kundu,T.K. *et al.* (2013) Characterization of nucleolin K88 acetylation defines a new pool of nucleolin colocalizing with pre-mRNA splicing factors. *FEBS Lett.*, **587**, 417–424.

65. Santos,T., Miranda,A., Campello,M.P.C., Paulo,A., Salgado,G., Cabrita,E.J. and Cruz,C. (2021) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.*, **189**, 114208.

66. Figueiredo,J., Miranda,A., Lopes-Nunes,J., Carvalho,J., Alexandre,D., Valente,S., Mergny,J.L. and Cruz,C. (2021) Targeting nucleolin by RNA G-quadruplex-forming motif. *Biochem. Pharmacol.*, **189**, 114418.

67. Fang,S.H. and Yeh,N.H. (1993) The self-cleaving activity of nucleolin determines its molecular dynamics in relation to cell proliferation. *Exp. Cell. Res.*, **208**, 48–53.

68. Chen,C.M., Chiang,S.Y. and Yeh,N.H. (1991) Increased stability of nucleolin in proliferating cells by inhibition of its self-cleaving activity. *J. Biol. Chem.*, **266**, 7754–7758.

69. Ugrinova,I., Chalabi-Dchar,M., Monier,K. and Bouvet,P. (2019) Nucleolin interacts and Co-localizes with components of pre-catalytic spliceosome complexes. *Sci.*, **1**, 33.

70. Caudron-Herger,M., Rusin,S.F., Adamo,M.E., Seiler,J., Schmid,V.K., Barreau,E., Kettenbach,A.N. and Diederichs,S. (2019) R-DeeP: proteome-wide and quantitative identification of RNA-dependent proteins by density gradient ultracentrifugation. *Mol. Cell*, **75**, 184–199.

71. Caudron-Herger,M., Wassmer,E., Nasa,I., Schultz,A.S., Seiler,J., Kettenbach,A.N. and Diederichs,S. (2020) Identification, quantification and bioinformatic analysis of RNA-dependent proteins by rnase treatment and density gradient ultracentrifugation using R-DeeP. *Nat. Protoc.*, **15**, 1338–1370.

72. Martone,J., Mariani,D., Santini,T., Setti,A., Shamloo,S., Colantoni,A., Capparelli,F., Paiardini,A., Dimartino,D., Morlando,M. *et al.* (2020) SMaRT lncRNA controls translation of a G-quadruplex-containing mRNA antagonizing the DHX36 helicase. *EMBO Rep.*, **21**, e49942.

73. Brown,J.A., Kinzig,C.G., Degregorio,S.J. and Steitz,J.A. (2016) Methyltransferase-like protein 16 binds the 3′-terminal triple helix of MALAT1 long noncoding RNA. *Proc. Natl. Acad. Sci.*, **113**, 14013–14018.

74. Liu,N., Zhou,K.I., Parisien,M., Dai,Q., Diatchenko,L. and Pan,T. (2017) N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.*, **45**, 6051–6063.

75. Liu,N., Dai,Q., Zheng,G., He,C., Parisien,M. and Pan,T. (2015) N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, **518**, 560–564.

76. Scherer,M., Levin,M., Butter,F. and Scheibe,M. (2020) Quantitative proteomics to identify nuclear rna- binding proteins of malat1. *Int. J. Mol. Sci.*, **21**, 1166.

77. Neckles,C., Boer,R.E., Aboreden,N., Cross,A.M., Walker,R.L., Kim,B.H., Kim,S., Schneekloth,J.S. and Caplen,N.J. (2019) HNRNPH1-dependent splicing of a fusion oncogene reveals a targetable RNA G-quadruplex interaction. *RNA*, **25**, 1731–1750.

78. Meseure,D., Vacher,S., Lallemand,F., Alsibai,K.D., Hatem,R., Chemlali,W., Nicolas,A., De Koning,L., Pasmant,E., Callens,C. *et al.* (2016) Prognostic value of a newly identified MALAT1 alternatively spliced transcript in breast cancer. *Br. J. Cancer*, **114**, 1395–1404.

79. Jadaliha,M., Zong,X., Malakar,P., Ray,T., Singh,D.K., Freier,S.M., Jensen,T., Prasanth,S.G., Karni,R., Ray,P.S. *et al.* (2016) Functional and prognostic significance of long non-coding RNA MALAT1 as a metastasis driver in ER negative lymph node negative breast cancer. *Oncotarget*, **7**, 40418–40436.

80. Shefer,K., Boulos,A., Gotea,V., BenChaim,Y., Muharram,A., Isaac,S., Eden,A., Sperling,J., Elnitski,L. and Sperling,R. (2022) A novel role for nucleolin in splice site selection. *RNA Biology*, **19**, 333–352.