Alternative splicing modulation mediated by G-quadruplex structures in MALAT1 lncRNA

Arpita Ghosh[1,2], Satya Prakash Pandey[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, CSIR- Human Resource Development Centre (CSIR-HRDC) Campus, Sector 19, Kamla Nehru Nagar, Ghaziabad 201 002, Uttar Pradesh, India and [3]CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411 008, India

ABSTRACT

MALAT1, an abundant lncRNA specifically localized to nuclear speckles, regulates alternative-splicing (AS). The molecular basis of its role in AS remains poorly understood. Here, we report three conserved, thermodynamically stable, parallel RNA-G-quadruplexes (rG4s) present in the 3′ region of MALAT1 which regulates this function. Using rG4 domain-specific RNA-pull-down followed by mass-spectrometry, RNA-immuno-precipitation, and imaging, we demonstrate the rG4 dependent localization of Nucleolin (NCL) and Nucleophosmin (NPM) to nuclear speckles. Specific G-to-A mutations that abolish rG4 structures, result in the localization loss of both the proteins from speckles. Functionally, disruption of rG4 in MALAT1 phenocopies NCL knockdown resulting in altered pre-mRNA splicing of endogenous genes. These results reveal a central role of rG4s within the 3′ region of MALAT1 orchestrating AS.

INTRODUCTION

Long non-coding RNAs (lncRNAs) are a class of non-protein-coding transcripts that are usually >200 nucleotides in length (1,2). Contrary to the small non-coding RNAs which have defined functions within cells, lncRNAs carry out multiple functions by either acting as a scaffold, decoy, activator, repressor, miRNA sponge or even act as epigenetic modulator (3–7). Nuclear enriched lncRNAs interact with epigenetic regulators, and serve as scaffolds for protein complexes or guide proteins to a specific location (8). On the other hand, cytoplasmic lncRNAs mostly participate in translational and post-translational activities (9). LncRNAs can also function as decoys for binding proteins both in the nucleus and cytoplasm and also take part in the sequestration of miRNAs (10). Because of their varied regulatory function, lncRNAs are crucial for physiological and cellular functions like growth and development, cell-cycle progression, inflammation, differentiation, or even tumor formation, invasion and migration. Dysregulated lncRNAs are widely reported in different types of cancers and can have oncogenic or tumor-suppressive action (11,12).

RNA being single-stranded can fold into secondary structures which helps it accomplish diverse regulatory functions. G-rich RNA has the propensity to form the RNA G-quadruplex (rG4), where a set of four guanines, also known as a G-quartet, are stabilized by Hoogsteen hydrogen bonds in the presence of cations and form a square planar assembly to attain a non-canonical RNA secondary structure (13,14). The DNA G-quadruplex structures have been studied extensively in terms of function (15); however, detailed rG4 studies are lacking (16). Recently, rG4 domains have been shown to perform important functional roles both in mRNAs as well as non-coding RNAs (17,18). rG4s are associated with transcriptional and translational regulation, 3′ end processing, alternative splicing (AS), mRNA localization, protein binding, telomere RNA regulation and RNA stability (4,19–27). LncRNAs usually have low expression within cells and are known to be less conserved across species, yet, structural elements within such RNAs have high homology across species. Such conserved structural motifs could define the mode of action for that particular lncRNA in multiple regulatory processes. rG4 structures have been predicted to exist within a large number of lncRNAs (28,46–49). Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1) is an abundantly expressed, highly conserved nuclear-retained lncRNA. It is widely associated with cancer and is known to regulate alternative splicing and gene expression (6,29–33). Localization of MALAT1 to nuclear speckles (34) allows it to interact with several splicing factors like SRSF1,

---

*To whom correspondence should be addressed. Email: souvik@igib.res.in

SRSF2 (SC35), SRSF3 and depletion of MALAT1 results in a decreased nuclear speckle association of several exogenous or endogenous pre-mRNA splicing factors with no impact on overall nuclear speckle formation (29,35,36). The effect of secondary structures of this lncRNA in AS remains untested. MALAT1 being a long RNA of 8.7 kb can attain various dynamic conformations within cells which might aid in recruiting different proteins within the structural scaffold. Secondary and tertiary structures in MALAT1 are reported to regulate its stability and function through different molecular mechanisms. A recent study established that a 74nt region at the 3′ end of MALAT1 forms a unique, bipartite triple-helix where the U-rich stem-loop sequesters the A-rich tail, and protects the RNA from exonucleolytic degradation (37–40).

In this study, we provide evidence about the existence of three putative G-quadruplex forming motifs in the 3′ end of the lncRNA, that form stable rG4s and interact with proteins such as nucleolin (NCL) and nucleophosmin (NPM) in HeLa cells. Functionally, rG4s act as a scaffold for these proteins and translocate them into nuclear speckles. Our studies also show that rG4 mediated localization of NCL in the nuclear speckles is further responsible for the regulation of AS of RNA transcripts.

MATERIALS AND METHODS

Bioinformatics

To predict the putative rG4 forming sites in MALAT1 we used a bioinformatic 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. 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 PQS region, we used UCSC Genome Browser. It uses the PhyloP base-wise conservation scoring system which is derived from Multiz alignment of 46 vertebrate species.

In-vitro transcription (IVT) of RNAs

IVT of RNAs was carried out using Megascript T7 IVT kit (Ambion®) as per the manufacturer's instructions. Briefly, forward and reverse primers designed for each domain of MALAT1 harboring G-quadruplex namely Q1, Q2 and Q3 along with their mutant counterparts (Q1M, Q2M and Q3M) were heated and annealed to form a partial complementary double-stranded DNA. Taq polymerase (Geneaid) was used to extend the partial complementary double-stranded DNA to full complementary DNA containing T7 promoter. In vitro transcription was carried out overnight at 37°C in a PCR machine. Template DNA was digested using TURBO DNase (Ambion®) and RNAs were purified using NucAway columns (Ambion®). RNAs were further checked on poly-acrylamide gel electrophoresis (PAGE) for their integrity and purity using ethidium bromide staining. Pure RNA species were used for further experiments.

rG4 mutant library preparation

The plasmid containing full-length MALAT1 of 8.7 kb, under a pCMV promoter, was obtained as a gift from Dr K.V. Prasanth's lab, University of Illinois. Overlapping primer sets each for Q1, Q2, and Q3 regions were designed to perform site-directed mutagenesis (SDM) (Supplementary Table S1). Each primer identifying the rG4 forming region had G-A mutations introduced. Long PCR with the MALAT1 cloned pCMV plasmid (FL) as a template was performed using these primers to introduce the mutations into the plasmid as it was amplified. The entire 11.2 kb plasmid was amplified at each cycle using High Fidelity Pfu Polymerase (Agilent). 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-synthesised 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 identifying Q1, Q2 and Q3 separately. Once Q1m, Q2m and Q3m were verified by sequencing, the second mutation was introduced in their backbone. As, Q12m, Q23m and Q13m were sequence-verified, the third mutation was introduced in the backbone of double mutant plasmid and 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

In-vitro transcription of RNAs was carried out using forward and reverse primers designed containing the T7 promoter for each domain of MALAT1 rG4s (listed in Supplementary Table S3). Q5 High Fidelity Polymerase (New England Biolabs® Inc.) was used to extend the partial complementary forward and reverse primers spanning the rG4 regions of the plasmid double-stranded DNA to fully complementary DNA containing T7 promoter. RNAs were synthesized from Q1m, Q2m, Q3m, Q12m, Q13m, Q23m, Q123m and FL plasmids and PAGE purified. Thioflavin T (ThT) compound was obtained from Sigma. The experiment was carried out in a 96-well microplate from CORN-ING (Flat Bottom Black Polystyrol). The RNAs varying in concentration from 0, 2, 4, 6 and 8 μM were mixed with ThT at 2 μM final concentrations, in 10 mM Sodium Cacodylate, 100 mM KCl buffer, at a pH of 7.0. A similar setup for the experiment was also carried out in 10 mM sodium cacodylate, 100 mM LiCl buffer as control. The fluorescence emission was collected at 487 nm with excitation at 440 nm in a microplate reader (Tecan Microplate Reader Life Sciences) at 25°C. S.E.M. was plotted for three separate technical replicates.

Cell culture and transfections

All experiments were performed with the HeLa cervical cancer cell line (mycoplasma free, cell authenticated). MALAT1 null HeLa along with normal HeLa control cell lines was obtained from Dr Roderic Guigo and Dr Rory Johnson's lab in CRG Barcelona. MALAT1 null cells created by knockout of promoter region by dual sgRNA based CRISPR approach (41,42) (Supplementary Figure S2A), was verified by genotyping to check promoter deletion (Supplementary Figure S2B) and qRT-PCR to check MALAT1 expression in HeLa (wt) and MALAT1 knockout (ko) cells (Supplementary Figure S2C). The cell lines were maintained in DMEM (Dulbecco's Modified Eagle's Medium) with 10% Foetal Bovine Serum (FBS) without antibiotic or anti-mycotic and were incubated in 5% CO_2 in a humidified incubator at 37°C. 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 were incubated for 24 h. After 24 h, the cells were transfected with 1 μg (for 12-well plates) and 2 μg (for 6-well plates) of plasmid concentration 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™ (Modified Eagle's Media), which is 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 after which the cells were harvested for their respective experimental procedures.

qRT-PCR

RNA was isolated from HeLa cells after 48 h of transfection. TRIZOL® reagent (Ambion®) was used to isolate total RNA in the step-wise protocol provided with it. cDNA was prepared using Qiagen cDNA synthesis kit. After the cDNA was prepared, real-time qPCR was performed for all the samples and controls in triplicates. The reaction volume for each was 10 μl. qRT-PCR primers used in different experiments are listed in Supplementary Table S4. Transcripts were quantified using SYBR Green Master Mix: SYBR Premix Ex Taq II (Tli RNase H Plus) (from TaKaRa) in the instrument Light Cycler 480 (Roche). All the C_t 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 (43).

RNA-FISH

Stellaris® ShipReady RNA FISH probes for MALAT1 having Quasar® 570 and 670 dyes, was ordered from Biosearch Technologies (LGC). 22 mm square coverslips from Corning® were immersed in the 6-well plates before seeding the cells at a density of 8×10^4 cells per well. Post seeding, transfections were performed as mentioned in the protocol before. Forty-eight hours post-transfection, cells were harvested for RNA-FISH. RNA-FISH buffers were purchased from Stellaris®. Stellaris RNA FISH protocol was performed as prescribed by the manufacturer for adherent cells. All the steps were conducted in highly sterile,

RNase-free conditions. For Immuno-FISH experiments, after the entire FISH protocol was conducted, the coverslips were again fixed in 4% formaldehyde (ThermoFisher Scientific), and an Immuno-cyto-chemistry experiment was carried out in the same coverslips according to the protocol detailed in the next section. Coverslips were mounted on Corning® plain microscope slides, and ProLong® Diamond Antifade ThermoFisher Scientific was used. Slides were imaged under 60× objective in DeltaVision Microscope from GE Healthcare Life Sciences. Quantification of images was performed using ImageJ. Co-localization analysis for Immuno-FISH was carried out with Fiji using Colloq 2 plugin.

Immuno-cyto-chemistry

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 square cover slips 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 as mentioned before. Forty-eight hours post-transfection, the cells were washed with 1× 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-immuno-cyto-chemistry 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-immuno-cyto-chemistry 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 60× objective. Quantification of images was performed using ImageJ. Co-localization analysis for Immuno-FISH was carried out with Fiji using Colloq 2 plugin.

RNA-Pulldown

In-vitro transcribed RNAs Q1, Q2, Q3 and their mutated counterparts were biotin-labeled using Pierce RNA 3′ end biotinylation kit. Labeling efficiency was determined with the Chemiluminescent detection kit module using the Manufacturer's instructions. 10×10^6 HeLa cells were lysed using commercially available RIPA buffer (Thermo Fisher Scientific). Briefly, cells were pelleted down and washed with filtered PBS followed by lysis in RIPA buffer on ice for 30 min. The supernatant was collected after centrifugation at 12 000 rpm for 15 min and stored at −80°C. The lysate was subjected to BCA protein assay to quantify the total protein content of the cell (kit used from ThermoFisher Scientific). 100 pmol of biotinylated RNAs (Q1, Q2 and Q3 and their mutant counterparts Q1m, Q2m, 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. Biotinylated RNA-protein complex was pulled down using magnetic Streptavidin beads (DynaBeads™ MyOne™ Streptavidin from Invitrogen) as per the manufacturer's instructions. Briefly, 50μl of beads were incubated with the binding reaction for 1 h at room temperature, washed thrice, boiled in SDS to break biotin–streptavidin interaction. Isolated proteins were identified using Mass spectrometry (QTOF 6600 (SCIEX) or used to perform 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 (44). 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 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/min for 10 min. Peptides were then separated on a reverse-phase C18 analytical column (ChromXP C18, 3 μm 120 Å, Eksigent) in a 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/min 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. 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 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). Paragon algorithm was used to get protein group identities. The search parameters were set as follows: sample type-identification, cysteine alkylation-iodoacetamide, 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. 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 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. 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

DelNucleolin (ΔNCL) having the N-terminal deleted from 1–283 amino acid residues was cloned into pET22b (+) vector between BamHI and XhoI sites and had a 6XHis-tag

at the C-terminal. The plasmid was transformed into *Escherichia coli* Rosetta cells for expression and protein production. Cells were grown in LB media containing ampicillin (100.0 μg/ml) and chloramphenicol (35 μg/ml) at 37.0°C till an O.D. of 0.6 at 600 nm. Cultures were induced with 0.2 mM IPTG and kept at 18°C with shaking at 200 rpm for 16 h. 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 with 5 s 'on' and 30 s 'off' pulse cycle for 10 min or till 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 were eluted with elution buffer (50.0 mM sodium phosphate pH 7.4, 150 mM NaCl, 10% glycerol, with a gradient of 250 mM and 500 mM imidazole) after adequate washing. Eluted protein fractions were analyzed on 12% SDS-PAGE gel. The fractions containing purified protein were pooled, dialyzed and further purified using size exclusion chromatography (SEC) and visualized on 12% SDS-PAGE for purity and homogeneity.

Circular dichroism (CD) spectroscopy

CD spectra were recorded for all the sequences using the Jasco 815 spectropolarimeter. The rG4s were prepared by slow cooling at 0.2°C/min after heating to 100°C for 5 min in 100 mM KCl,10 mM sodium cacodylate buffer (pH 7.4), and also in 100 mM LiCl and 10mM sodium cacodylate buffer (pH 7.4) at 5 μM strand concentrations of each. The represented spectrum is an average of three consecutive scans for each sample.

UV melting

Cary 100 (Varian) spectrophotometer, equipped with a thermoelectrically controlled cell holder, was used to perform UV melting experiments. rG4s were prepared as described above. An RNA duplex was also prepared with two complementary sequences annealed to each other (Forward Oligo strand: 5′ UCCAAAACAUGAAUUG 3′ and Reverse Oligo Strand: 5′ CAAUUCAUGUUUUGGA 3′). A temperature range of 10–90°C was used to monitor the changes in absorbance at 295 nm (also 260 nm for RNA duplex) at 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 h at 200V 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 with the BIAcore 3000 (BIAcore Inc.) system running with BIAcore 3000 control software version 4.1.2. 5′- biotinylated Q1, Q2 and Q3 sequences (Supplementary Table S6) were dissolved in filtered and degassed 10 mM HEPES buffer with 100 mM KCl with 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 chips. Flow cell 1 was left blank as a control to account for 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 a running buffer. Serial dilutions of the 10 μM ΔNCL stock were performed to make a concentration series in the running buffer. ΔNCL solution of different concentration ranges in between 10 nM to 500 nM was injected at 20 μl min^{-1} flow rate 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 two independent binding sites 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 χ^2 value which was either equal to or <1. All experiments were performed at 25°C. using running buffer.

Fluorescence titration for NCL-rG4 binding

ΔNCL-rG4 binding was monitored by monitoring tryptophan fluorescence in ΔNCL varying the rG4 concentrations. Fluorescence spectra of tryptophan were measured in a Fluoromax 4 (Spex) spectro-fluorophotometer equipped with a thermoelectrically temperature-controlled cell holder (quartz cuvette, 1 cm × 1 cm). Samples were excited at 290 nm, and the emission spectra were recorded from 320 to 500 nm. The experiments were performed in 10 mM sodium cacodylate buffer (pH 7.4) containing 100 mM KCl at 25°C. The protein concentration was fixed at 500 nM, and the concentration of preformed rG4 (prepared as described above) was varied from 0 to 1000 nM. Considering bound ΔNCL is directly proportional to the extent of fluorescence intensity change, $\Delta F/\Delta F_{max}$, where $\Delta F = F_0 - F$ and $\Delta F_{max} = F_0 - F_{final}$, $\Delta F/\Delta F_{max}$ was plotted as a function of rG4 concentrations. For data analysis, the observed fluorescence intensity was considered as the sum of the weighted contributions from an rG4 bound ΔNCL and an unbound ΔNCL form. Averaged data from three independent experiments were 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.

Splicing reporter (pRIP-flow cytometry) assay

pRIP plasmid was obtained as a gift from Dr Beena Pillai's lab, CSIR-IGIB. The plasmid has a pCMV promoter, downstream of which is an intronic element encoding the pre-miR-30b sequence, followed by an RFP reporter (57). MALAT+/+ cells, MALAT−/− cells, FL plasmid, and Q123m plasmid rescued in MALAT1−/− cells were transfected with pRIP plasmid as mentioned in the protocol before. Twenty-four hours post-transfection, the cells were trypsinized and the pellet was washed with $1\times$ PBS. After washing, the supernatant was discarded, and the pellet was re-suspended in 500 µl $1\times$ PBS. Data acquisition to check the percentage of RFP expression was done using BD FACS MELODY™.

Western blot

The treated cells were lysed using Cell Lytic™ (Sigma) buffer to prepare the protein lysate. 50 µl of the lysis buffer was added to each 12-well plate along with 5 µl of Proteinase Inhibitor Cocktail (Sigma). The cells were allowed to lyse for 1 h in a rocker at 4°C. Post incubation, protein lysate from each sample was collected. The concentration of protein was estimated using Pierce™ BCA Protein Assay Kit (ThermoFisher). For each sample, 40 µg of protein was loaded into the wells of 10% SDS gel and PAGE was performed. The proteins were then transferred from the gel to the PVDF membrane (GE Healthcare Life-Science) in Bio-Rad vertical gel Transfer Apparatus at 4°C for 3 h at 70 V. After the transfer was complete, 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 1:1000 dilution overnight at 4°C. After primary antibody incubation, the blots were washed three times for 15 min each with $1\times$ TBST. Post washing, they were incubated with a secondary antibody having HRP conjugate for 3 h at room temperature. Post incubation a similar washing step (as mentioned after primary antibody incubation) was performed. EMD Millipore™ Immobilon Western Chemiluminescent HRP Substrate (ECL) was used to develop the blots in the Syngene Gel doc instrument. ImageJ was used for the densitometry analysis of the blots.

RNA-sequencing and analysis

RNA isolation was performed from three independent biological replicates of MALAT1+/+, MALAT1−/−, FL, Q123m and NCL KD cells, as per the protocol mentioned earlier in Methods. A modified NEBNext RNA Ultra II directional protocol was used to prepare the libraries for total RNA sequencing. The first step for purifying the poly-A containing mRNA molecules using oligo-dT attached magnetic beads was omitted. Instead, the first step involved the removal of ribosomal RNA of cytoplasmic and mitochondrial origin using biotinylated, target-specific oligos combined with rRNA removal beads. Following purification, the RNA was fragmented using divalent cations under elevated temperatures. Next, the cDNA was synthesized using Reverse transcriptase and random hexamers in a first-strand synthesis reaction. Subsequently, the cDNA was converted to double-stranded cDNA where Uracil was added instead of Thymine. The strand specificity was preserved by a USER enzyme-based digestion of the second strand thereby leaving one functional strand mapped to the DNA strand from which it was transcribed. The USER digested single strand molecules were enriched and indexed in a limited cycle PCR followed by AMPure bead purification to create a final cDNA library for sequencing. Prepared libraries were sequenced on Illumina Novaseq to generate 80M, 2×150 bp reads/sample. Up to 75% of the sequenced bases were of Q30 value. Sequenced data were processed to generate FASTQ files. Raw reads were checked for quality using FastQC (v0.11.5). Trimmomatic (v0.39) was used to trim adaptor sequences and filter low-quality reads. Filtered reads were aligned to the Human Gencode (GRCh38.p13) reference genome and transcriptome using STAR (v2.7.9a) aligner (45) with RSEM (v1.3.1) (46) to estimate gene expression. Genes read counts from RSEM were normalized using TMM (weighted trimmed mean of M-values) followed by differential expression analysis by applying quasi-likelihood methods in the edgeR (v3.30.3) (47,48). Genes with differential expression from all conditions when compared to the MALAT+/+ condition were plotted using EnhancedVolcano (v1.6.0) (49). For the splicing analysis, reads were mapped to Human Gencode (GRCh38.p13) transcriptome by Salmon (v1.4.0) (50) in quasi-mapping-based mode. Further, SUPPA2 (51,52) was used to generate, calculate and compare the alternative splicing events (SE, A3, A5, MX, RI, AF, AL) across conditions. Differential splicing events with percentage or proportion spliced-in (ΔPSI) ≥ 0.10 and P-value < 0.05 are considered significant. To check the overlap of events across the conditions PSI ≥ 0.10 is considered and VennDiagram (v1.6.20) (53) was used to plot the Venn diagram. Pie charts are generated using the ggpubr and ggplot2 R package.

Statistical analysis

The statistical analysis for experiments was performed using GraphPad Prism 8.0 to evaluate significance among experimental replicates. All data were presented as mean $\pm$ 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

A previous in-silico transcriptomic study from our group showed a significant number of putative rG4 forming sites residing in human lncRNAs (54). In addition, a transcriptome-wide rG4 profiling methods in other reports further established that PQS in lncRNAs, like MALAT1 and NEAT1, have the potential to form rG4s (55–58).

Similar observations were made in other studies, however, the functional significance of rG4s in lncRNA largely remains elusive (28,48,49). By using the QGRS mapper (59) (a bioinformatic G-quadruplex prediction tool), we identified three potential G-quadruplex forming motifs towards the 3′ end of the 8.7 kb lncRNA MALAT1. The three putative G-quadruplex motifs named Q1, Q2, and Q3 were positioned within the stretch of 6.3–7.7 kb region (Figure 1A). An earlier study from McCown *et al.* confirmed that MALAT1 shows 75% nucleotide conservation within 43 mammalian homologs (19). The region harboring rG4s was also seen to be conserved across species as extrapolated from the phyloP scoring in UCSC Genome Browser (Figure 1B, Supplementary Figure S1A). To investigate the functional role of these predicted structural domains, we performed biophysical assays to verify the formation of the G-quadruplex structures *in-vitro* using the Q1, Q2 and Q3 RNA sequences independently. Using circular dichroism (CD) spectroscopy, it was observed that all the three putative rG4 sequences could fold into stable G-quadruplex structures, as characterized by a positive peak at 266 nm and a negative peak at 236 nm in the CD spectra, a characteristic signature of the parallel G-quadruplex in the presence of K^+ ions (Figure 1C). The CD spectra for the three rG4 sequences were also measured in the presence of Li^+ ions as shown in Supplementary Figure S1D. Both the positive and negative peaks depicted a lower intensity as compared to when K^+ ions were present, which is a well-established property of rG4 structures. UV-thermal denaturation studies at 295 nm in sodium cacodylate buffer (pH 7.0) containing 100 mM KCl showed typical hypochromic sigmoidal transitions for all the three oligonucleotides with T_m values of 73 ($\pm$1), >90 and 75 ($\pm$1)°C for Q1, Q2 and Q3 respectively (Figure 1A, D). Melting and annealing curves were superimposable onto each other indicating the formation of a thermodynamically stable intramolecular quadruplex. Measuring T_ms over a 50-fold oligonucleotide concentration range (from 1 to 50 μM) showed no change in T_ms (data not shown), which is consistent with formation of the intramolecular rG4s. As a control, we performed similar UV Melting study for an RNA duplex at 260 nm and 295 nm (Supplementary Figure S1E), which did not display any curve at 295 nm, but, had a T_m of 49($\pm$1)°C as extrapolated from its melting curve at 260 nm. Considering the probability of formation of stable alternative Watson–Crick-based secondary structures that might compete with rG4 formation, which cannot be assessed by conventional *in vitro* studies using isolated putative sites, we wanted to verify the formation of all the three structures together in the longer sequence. For this purpose, we obtained the full-length 8.7 kb MALAT1 cloned in a pCMV vector plasmid (FL). Site directed mutagenesis was performed replacing G with A that impeded rG4 formation to obtain Q1m, Q2m, Q3m (single rG4 mutated), Q12m, Q23m, Q13m (double rG4 mutated) and Q123m (triple rG4 mutated) (Supplementary Figure S1B, C). Thioflavin T (ThT), a dye which upon binding to rG4 shows enhanced fluorescence signal as compared to single stranded RNA or RNA hairpin structure (60,61), was used. *In-vitro* transcription (IVT) was performed for the 1.4 kb region having all the three rG4/mutant

regions from the plasmids in the library. The RNAs were gel purified and then titrated from 0 to 8 μM against a fixed concentration of 2 μM ThT in 10 mM sodium cacodylate and 100 mM KCl buffer condition (Figure 1E). The sequence having all three G-quadruplexes mutated (Q123m) showed minimal fluorescence equivalent to the baseline, while the presence of even a single rG4 in Q12m, Q23m and Q13m, results in an increase in the fluorescence intensity. The intensity further increases when two rG4 sequences are present in Q1m, Q2m, Q3m and maximum in the FL RNA where all the three rG4 sequences are present. An exact similar experiment when performed in the presence of Li + ions as control, show similar pattern for fluorescence for all the different combination of rG4s/mutants, but with much lower intensity, signifying lower strength of the rG4s formed (Supplementary Figure S1F). Thus, these results show that all the three G-quadruplex structures are able to form in the longer 1.4 kb region of MALAT1 RNA as well by over-competing with other possible alternative canonical secondary structures.

The rG4 structures in MALAT1 do not affect its expression or localization within cells

rG4 structures are known to carry out a multitude of regulatory functions. The presence of rG4s in the human transcriptome is often shown to determine their stabilities and expression. To evaluate the functional implications of rG4s on the stability and expression of MALAT1, we measured the levels of MALAT1 in MALAT1−/− HeLa cells (see Methods for MALAT1−/− cell-line creation) upon transfection of plasmids containing full-length MALAT1 (FL) and different mutants, with disruption of all rG4s (Q123m), two rG4s (Q12m, Q13m, Q23m) and single rG4 (Q1m, Q2m, Q3m). qRT-PCR analysis showed that there was no change in the expression of MALAT1 across all the different conditions implying that rG4 depletion did not alter lncRNA expression within cells (Figure 2A). The presence of an rG4 consensus motif is reported to act as a localization element for dendritic mRNAs in mouse cortical neurons (26). MALAT1 is a nuclear enriched lncRNA, specifically known to localize in nuclear speckles (36). To determine whether the presence of rG4 structures in the lncRNA played any role in the localization, RNA-fluorescent in-situ-hybridization (FISH) assays were performed to analyze the sub-cellular distribution of MALAT1. It was observed that MALAT predominantly localized in the nuclear speckles of HeLa cells in MALAT1+/+ condition while no signal for the lncRNA was observed in MALAT1−/− cells (Figure 2B). We then checked the sub-cellular distribution of MALAT1 in the context of each rG4 using RNA-FISH assays by transfecting plasmids FL, Q123m, Q12m, Q13m, Q23m, Q1m, Q2m, and Q3m in the MALAT1−/− cells. FL was able to rescue the phenotype as MALAT1 was observed to be localized in the nuclear speckles (Figure 2B). Also, complementation with different rG4 mutants did not show any change in the phenotype for localization of this lncRNA (Figure 2B–E). Therefore, we inferred that the rG4 structures do not act as a localization cue for MALAT1. Earlier studies suggest MALAT1 is dispensable for the forma-
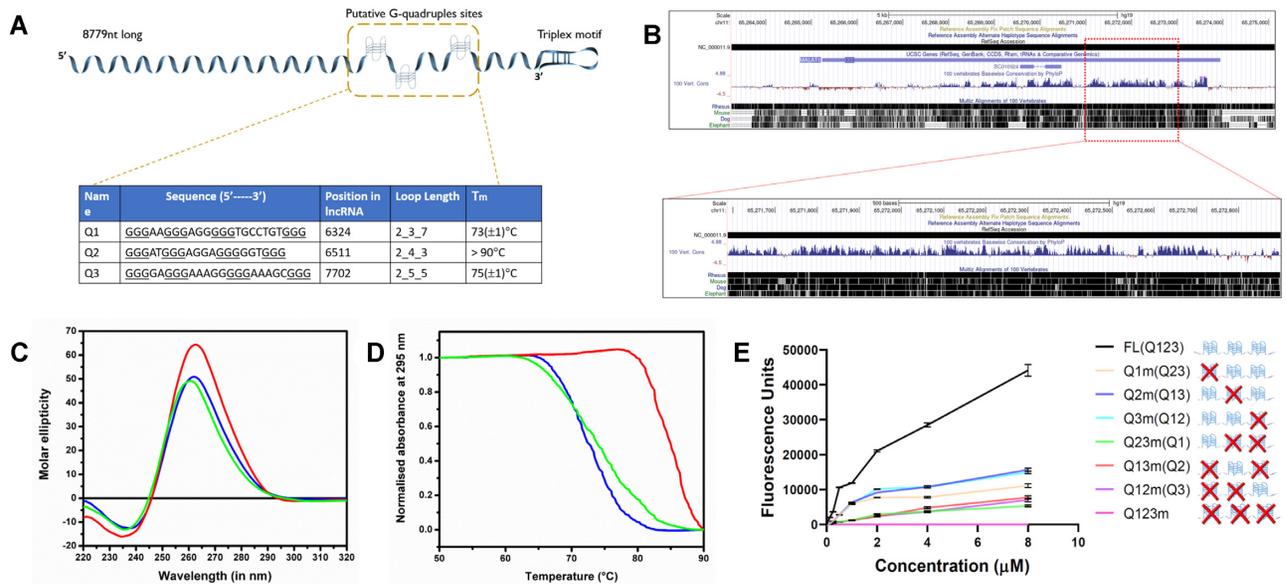
Figure 1. MALAT1 long non-coding RNA harbors three putative rG4 forming domains towards its 3′end. (**A**) The schematic for long non-coding RNA MALAT1 harboring three putative G-quadruplex motifs as predicted by QGRS mapper, its respective locus, and melting temperature (T_m) as observed in UV Melting studies. (**B**) MALAT1 conservation across vertebrates as depicted in UCSC. Highlighted part expanded below harbors the three rG4 domains and the region are 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.
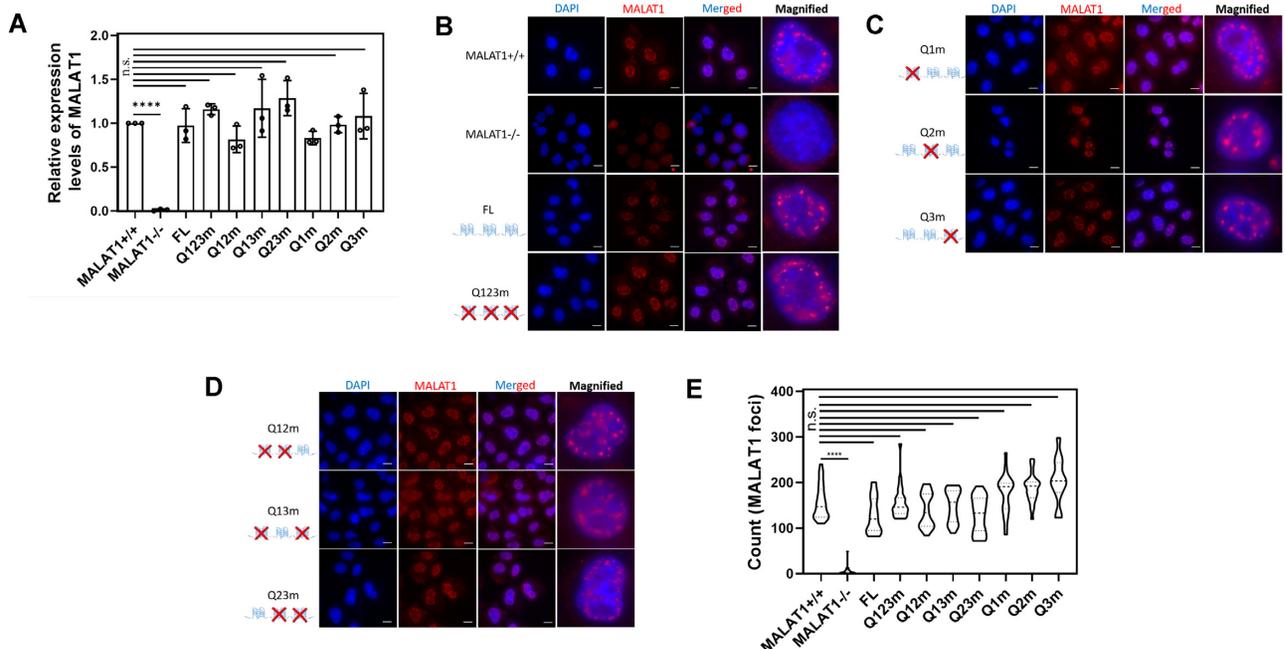
Figure 2. The rG4 structures in MALAT1 do not affect its expression or localization within cells (**A**) qRT-PCR to quantify the levels of MALAT1 across all the experimental conditions. Relative expression levels of MALAT1 in FL rescue as well as Q123m, Q12m, Q13m, Q23m, Q1m, Q2m and Q3m by transfecting 1 μg of plasmid was 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).

tion of nuclear speckles. To validate whether the knockout of MALAT1, or the disruption of its rG4 structures hampered nuclear speckle formation, we performed an immuno-FISH for SC35 (a nuclear speckle protein) and MALAT1 in MALAT1+/+, MALAT1−/−, FL and Q123m rescue conditions and found that the nuclear speckle integrity was not lost in any (Supplementary Figure S2D–F). Therefore, from all the observations above, it can be concluded that the rG4 structures in MALAT1 does not affect its expression nor its localization within cells.

MALAT1 lncRNA anchors to specific proteins by its rG4 structures

An important mode of lncRNA function within the nucleus is by interacting with proteins. As we could not deduce any significance of the rG4s in MALAT1 in its expression and localization, we, therefore, hypothesized that these structures might function via RNA–protein interactions. To test this, biotinylated in-vitro transcribed Q1, Q2 and Q3 RNAs were incubated with HeLa cell lysate, and a streptavidin-based RNA pulldown followed by mass spectrometry was performed, to identify the protein partners, if any, for these motifs (as listed in Supplementary Table S8). rG4 mutants Q1m, Q2m and Q3m were used as controls (Figure 3A). Nucleolin (NCL) was one of the most enriched rG4 binders. Nucleophosmin (NPM) was another protein, which showed some rG4 specific binding. We also obtained hnRNP A2/B1, C and H, in our pull-down experiment, but these were equally enriched for both rG4 and mutated RNA sequences. We validated the interacting partners by performing western blot with specific antibodies to detect NCL and NPM and observed that both proteins specifically interact with Q1, Q2, Q3 and not with the mutated G-quadruplex motifs (Figure 3B, C). A similar experiment was performed to detect hnRNP A2/B1 as well and it was detected across all the rG4s and mutated sequences (Supplementary Figure S3). We further confirmed the binding of these proteins to the lncRNA by performing a reverse pulldown or RNA-immunoprecipitation (RIP) in HeLa cell lysate using NCL and NPM antibodies. We checked the enrichment by qRT-PCR and found MALAT1 5- to 7-fold enriched in both the proteins over IgG control (Figure 3D). Hence, these experiments validated that both NCL and NPM show interaction with MALAT1 lncRNA, specifically via the rG4 structures.

MALAT1 rG4s aid localization of the interacting protein partners

The rG4 structures present in MALAT1 do not participate in directing the lncRNA into the nuclear speckles but specifically binds to certain nuclear proteins. We hypothesized that the localization of these proteins is dependent on MALAT1. Thus, we performed immuno-cyto-chemistry (ICC) and imaging of NCL, NPM, hnRNP H, hnRNP C, and hnRNP A2/B1 in MALAT1+/+ and MALAT1−/− HeLa cells. As reported previously, both NCL and NPM localized to the nuclear speckles, apart from their major residence in the nucleolus in MALAT1+/+ cells (62–65). However, in MALAT1−/− cells, NCL was seen only in the

nucleolus, while the signal in the nuclear speckles disappeared (Supplementary Figure S4A). This implies that the localization of this protein to the nuclear speckles is dependent on the lncRNA. To validate whether the foci seen for NCL is a part of nuclear speckles or not, we performed co-localization experiments for both NCL and speckle marker SC35 in the MALAT1+/+ and MALAT1−/− cells. In the MALAT1+/+ condition, both proteins, NCL and SC35 colocalized, while in MALAT1−/− cells only SC35 was seen in the speckles, and NCL was found only in the nucleolus and not in speckles (Supplementary Figure S4B, S4C). ICC and imaging of NPM also gave similar results as that of NCL in both conditions (Supplementary Figure S4D). hnRNP H and hnRNP C in MALAT1+/+ and MALAT1−/− conditions showed the proteins to be dispersed across the entire nucleoplasm with no dependence on the lncRNA (Supplementary Figure S4E, S4F). Similarly, hnRNP A2/B1 puncta across the nucleus were also indifferent to the presence and absence of MALAT1 within the cells (Supplementary Figure S4G). These results indicate that the localization of both NCL and NPM to nuclear speckles is dependent on MALAT1. Next, we checked the dependence of the localization of NCL and NPM on rG4s in MALAT1. We performed a plasmid-dependent rescue of MALAT1, along with all the three rG4 mutated lncRNA in the MALAT1−/− cell line, followed by ICC to observe the localization of these proteins (schematic in Supplementary Figure S4H). First, we imaged NCL in MALAT1 rescue condition (FL) where we found it to be distributed to the nuclear speckles along with the nucleolus, a phenotype similar to the MALAT1+/+ condition. In the rescue of MALAT1 with all its rG4 mutated (Q123m), NCL did not appear in the nuclear speckles, and the phenotype mirrored the MALAT1−/− condition (Figure 4A). These results indicated that rG4s of MALAT1 are responsible for driving NCL into the nuclear speckles. We next asked whether all rG4s or any individual rG4 or any combination of the rG4s are a pre-requisite for this action. We performed rescue experiments of the full lncRNA with two rG4s mutated (Q12m, Q23m, Q13m) and single rG4 mutated (Q1m, Q2m, Q3m) in MALAT1−/− cells, followed by ICC and imaging of NCL. In all of these conditions, NCL was seen only in the nucleolus and not in the speckles, as observed in MALAT1−/− cells (Figure 4B, C). Quantification of the NCL foci in the nuclear speckles also asserts its presence in MALAT1+/+ or MALAT1 full length (FL) rescue, while none in MALAT1−/− and rescue conditions of MALAT1 with the different combination of rG4 mutants (Figure 4D). Hence, our data suggest that all the three rG4s of MALAT1 are essential for the localization of NCL into the nuclear speckles. Immuno-FISH experiment to check NCL and MALAT1 co-localization in MALAT1+/+, MALAT1−/−, rescue with FL and Q123m show the presence of MALAT1 in all the conditions except for MALAT1−/− cells, co-localizing with NCL in the nuclear speckles in MALAT1+/+ and FL rescue conditions only, as NCL was not present in the speckles in the other two conditions (Figure 4E, F). Similar to NCL, MALAT1 rescue could bring the NPM protein signal back in the nuclear speckles while lncRNA rescue with all the rG4s mutated could not (Figure 4G, H). Our results so far estab-
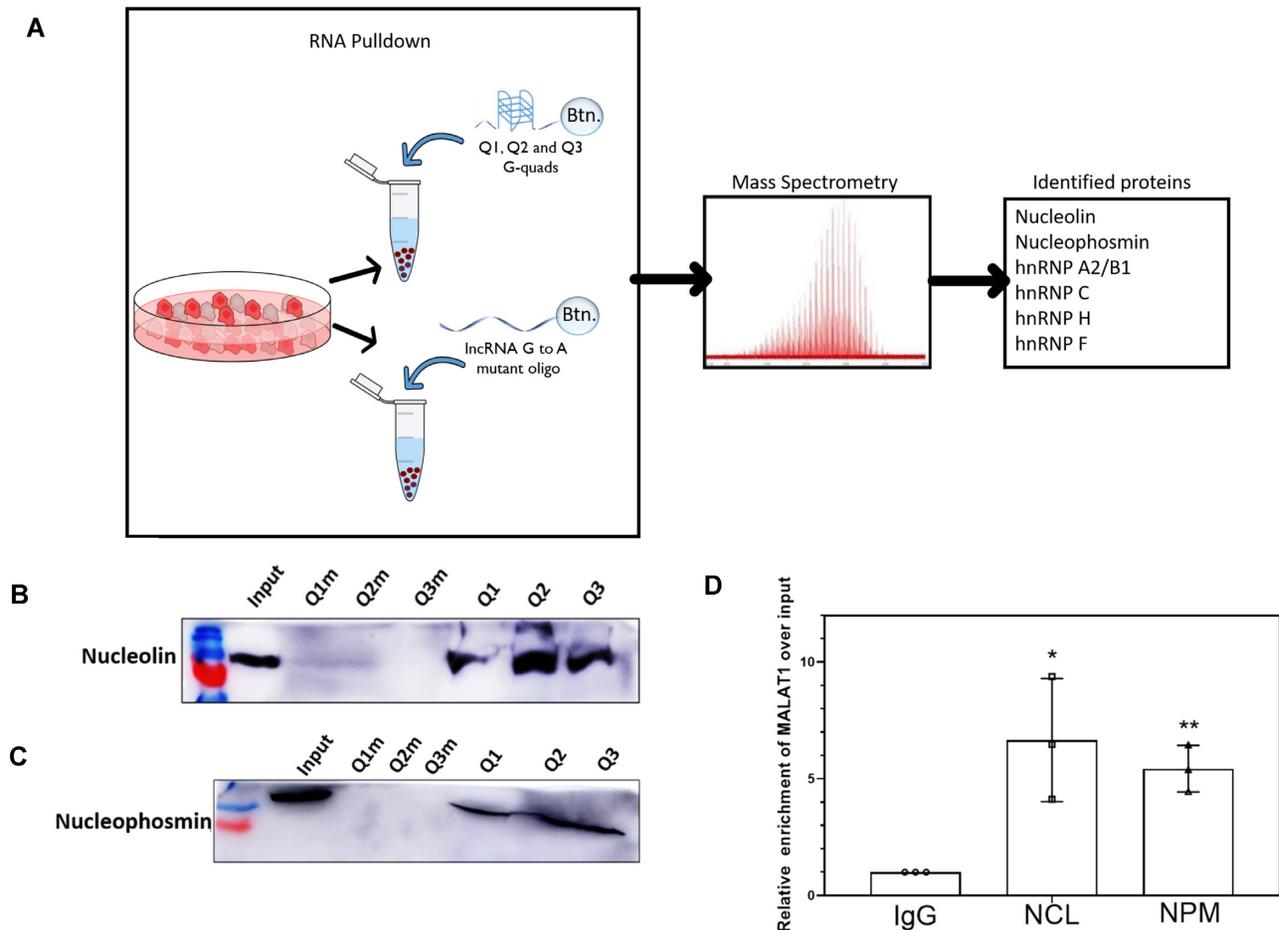
Figure 3. MALAT1 lncRNA anchors to specific proteins by its rG4 structures. (**A**) Mass spectrometry representation for RNA pulldown from HeLa cell lysate using Q1, Q2, and Q3 rG4 and their mutated counterparts. A list of enriched proteins mentioned showing interaction. (**B**) RNA pulldown of Q1, Q2, 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 RIP performed with NCL and NPM proteins. Enrichment of NCL is shown in comparison to IgG control. Error bars represent $\pm$ S.D. across three independent biological replicates. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$ (Student's *t*-test).

lished that the rG4s present in the 3′ region of MALAT1 is important for subcellular localization of NCL and NPM in the nuclear speckles.

NCL directly binds to G-quadruplex structures in-vitro and is a bonafide partner

NCL, a well-known RNA-binding protein, has been reported to function via interactions with rG4 structure in multiple studies. From our observation, NCL is the prime protein of interest due to its high peptide score with the MALAT1 rG4s in the mass spectrometry analysis and our immuno-FISH experiments also signify that the sub-cellular localization of NCL is dependent on the G-quadruplex structures of MALAT1. We next performed detailed investigations on its interaction with the rG4s of MALAT1. To establish that NCL directly binds to the rG4s, we performed in-vitro binding studies. We tried to purify NCL, but due to the self-cleaving nature of the N-terminal end of the protein, only a truncated version could be pu-

rified , which had all of the four RNA Recognition Motifs (RRMs), and the GAR domain intact (Supplementary Figure S5A, S5B). Using ΔNCL and pre-formed Q1, Q2 and Q3 quadruplex individually, electrophoretic mobility shift assays (EMSA) were performed to establish binding of the protein to these rG4s (Figure 5A–C). Keeping the concentration of the pre-formed rG4s constant while increasing the ΔNCL concentration, a shift was seen in all the EMSAs done for Q1, Q2 and Q3, indicating the binding of the protein to the rG4s. To analyze the binding parameters, surface plasmon resonance (SPR) and fluorescence titration experiments (Figure 5D) were performed to obtain binding affinities between all the rG4s and ΔNCL. All three structures showed high binding affinity towards ΔNCL with Q2 having the maximum affinity with a K_d value of 4 nM (Table 1). Further, we confirmed the binding of the proteins to rG4 structures of the lncRNA by performing RIP using NCL antibody from HeLa cell lysates of MALAT+/+, MALAT−/−, FL and Q123m transfected MALAT−/− cells. We checked the enrichment by
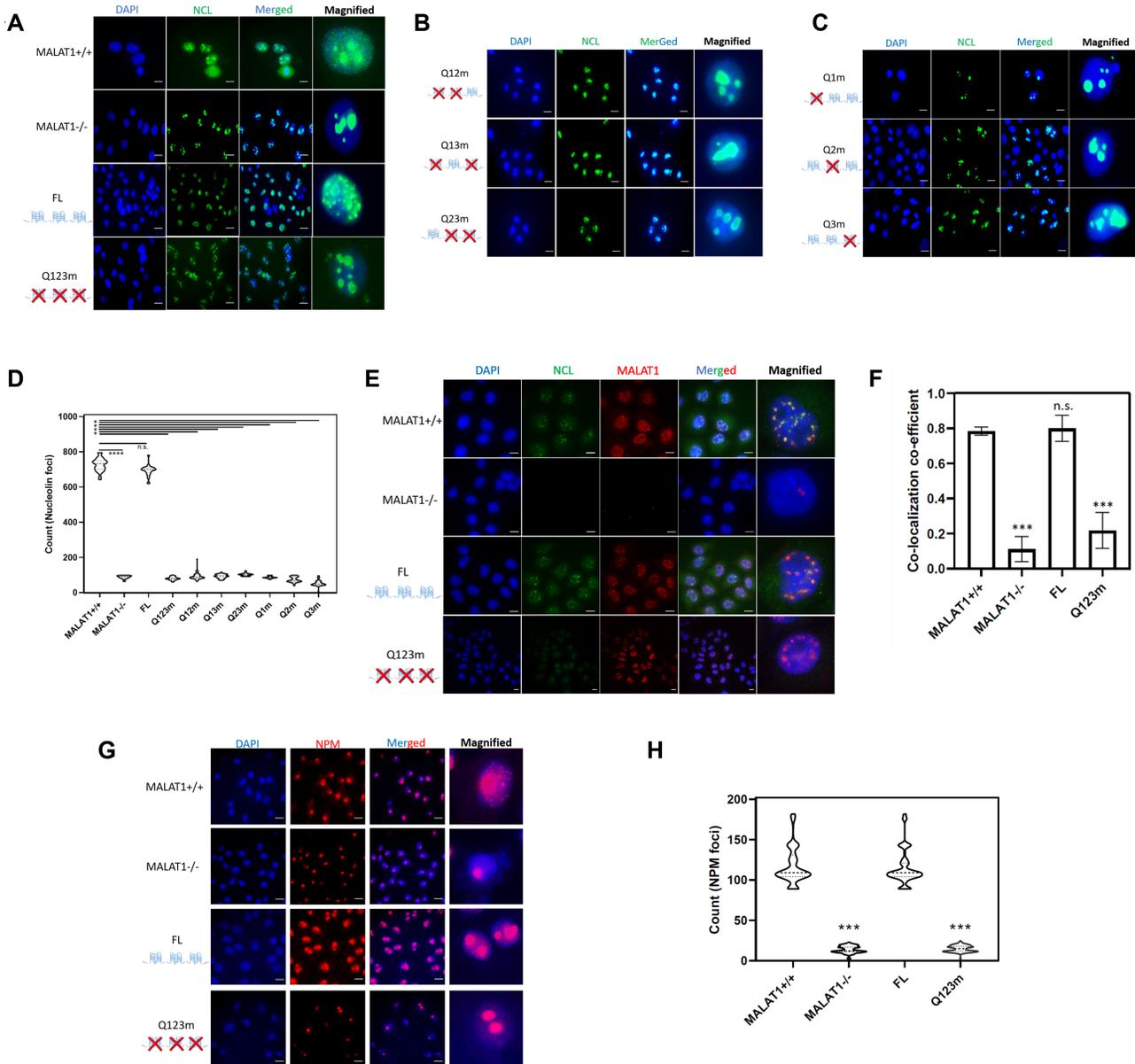
Figure 4. MALAT1 rG4s aid localization of the interacting protein partners. (**A**) ICC to observe the localization of NCL across different experimental conditions: MALAT1+/+, MALAT1−/−, FL and Q123m. (**B**) ICC to observe the localization of NCL across different experimental conditions: Q12m, Q13m, 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 immuno-FISH. 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 Co-efficient 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).

qRT-PCR and found MALAT1 to be 7-fold and 5-fold enriched in the MALAT1+/+ and MALAT1 FL rescue conditions respectively, while the enrichment in MALAT1 rescue with the Q123m triple G-quadruplex mutant lncRNA was equivalent to that as seen in the MALAT1−/− condition. Therefore, we could confirm from all these experiments that the three G-quadruplex structures of MALAT1 Q1, Q2 and Q3 have NCL as its bonafide partner.

MALAT1 rG4 structures together with NCL orchestrate alternative splicing of specific endogenous pre-mRNAs

Consistent with earlier reports, we observed co-localization of NCL and MALAT1 in nuclear speckles. We demonstrate that this behavior requires the presence of rG4s in MALAT1. Nuclear speckles are dynamic, self-organizing nuclear domains that are hubs of pre-mRNA processing
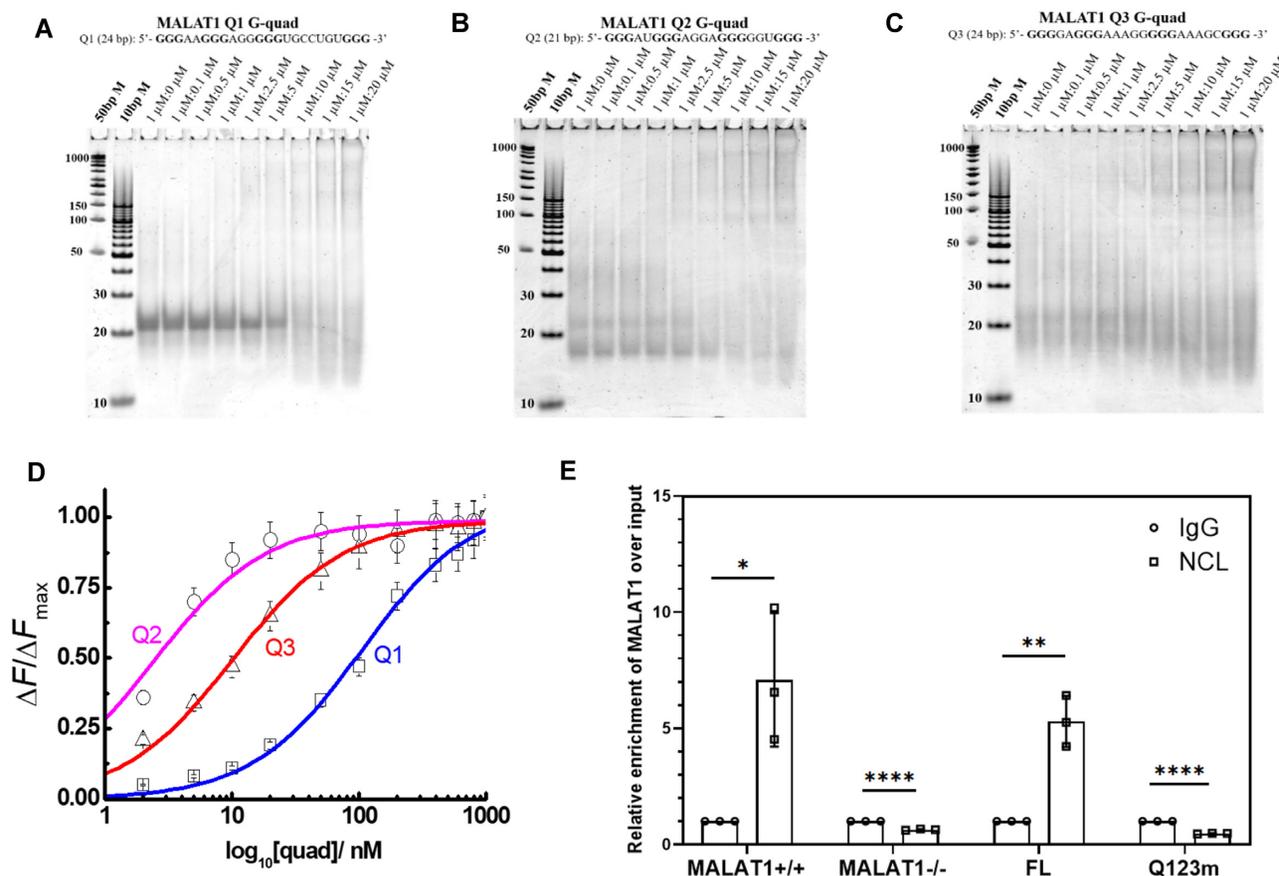
Figure 5. NCL directly binds to MALAT1 G-quadruplex structures in-vitro and is a bonafide partner. (**A–C**) EMSA of Q1, Q2 and Q3 MALAT1 RNA G-quadruplex with ΔNCL protein, with the RNA kept constant and protein concentrations varied from 0 to 20 μM. (**D**) Fluorescence titrations of ΔNCL protein in 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)

and thus are rich in splicing protein machinery. As both our lncRNA and protein of interest localize to nuclear speckles, we hypothesized that the interaction of MALAT1 and NCL via the rG4s might play a role in splicing. To validate our hypothesis, we used a splicing reporter plasmid (pRIP) and transfected it in the following conditions: MALAT1+/+, MALAT−/−, rescue of FL and Q123m in MALAT1−/− cells. The plasmid construct has an RFP reporter, which fluoresces only if its intronic sequence has been spliced, or else is cleaved off by Dicer as the intronic sites also encode a pre-miR sequence (66). Flow cytometry analysis 24 ho post-transfection of the reporter showed that RFP expression was maximum under MALAT1−/− and Q123m rescue conditions compared to MALAT1+/+ and FL res-

cue conditions (Supplementary Figure S6A). This hinted at the possibility that MALAT1 rG4s might participate in splicing regulation. A previous study confirms the role played by MALAT1 in AS by modulating the phosphorylation level of SRSF1 protein, which binds to the 5′ end of the lncRNA (29). To understand if endogenous transcripts are similarly affected, we selected a few candidate transcripts from the microarray data generated in the previous study upon MALAT1 down-regulation and followed their expression. It was shown that downregulation of MALAT1 modulated alternative splicing patterns of MGEA6, CAMK2B, PAX2, B-MYB and ARHGEF1 (29). Importantly these genes have been implicated in cancer and tumor progression (30). We designed primers for qRT-PCR to detect the

specific exon inclusion event of these transcripts (Supplementary Figure S6B, Table S7). qRT-PCR analysis to detect these transcripts in MALAT1+/+, MALAT1−/−, full MALAT1 (FL) and all the different combinations of rG4 mutated MALAT1 rescue conditions, was performed. We observed that in the presence of MALAT1 or its intact three rG4s (MALAT1+/+ and FL rescue), exon inclusion happens in MGEA6, CAMK2B and PAX2, while in B-MYB and ARHGEF1 transcript, exon exclusion could be seen (Figure 6A, B). This effect was lost in the absence of MALAT1 or absence of any of the rG4s (MALAT1−/− and Q123m, Q12m, Q23m, Q13m, Q1m, Q2m, Q3m rescues). A mitochondrial transcript ND1 was chosen as a control whose levels remained unaltered across all of the different conditions (Figure 6C). Thus, our experiments established that MALAT1 rG4s play a significant role in modulating the levels of specific isoforms of certain RNA transcripts. NCL on the other hand majorly participates in ribosomal biogenesis as almost 90% of the protein resides in the nucleolus although studies show a small pool of NCL translocating to the nuclear speckles. Also, it is reported that NCL co-localizes and interacts with the pre-catalytic spliceosome complex and participates in the alternative splicing of fibronectin mRNA. R-DeeP, a tool used to investigate protein-protein interactions in the presence and absence of RNase (67,68), further indicated that NCL interacts with splicing proteins SON, SRSF2, SRSF3, SRSF5, and SRSF11 in an RNA dependent manner as this interaction is hampered in the absence of RNA (upon RNase treatment). Hence, to understand whether NCL also affects AS of the same transcripts similarly as modulated by MALAT1 rG4s, we first performed a siRNA-mediated down-regulation of the protein. NCL downregulation was confirmed by western blot (Figure 6D), which was followed by ICC to check its localization in the nuclear speckles (Figure 6E). Around 50% knock-down of the protein was obtained and NCL was also dislodged from the nuclear speckles in the siRNA transfected cells. qRT-PCR analysis depicted around 70% NCL knock-down, while MALAT1 and ND1 levels remained unaffected (Figure 6F, G). MGEA6, PAX2, and CAMK2B showed exon inclusion in the presence of NCL, and their levels dropped as NCL was down-regulated. On the contrary, B-MYB and ARHGEF1 exon exclusion events occurred in the presence of NCL and upon its down-regulation, their levels increased for that specific isoform. Therefore, the AS pattern of transcripts in NCL down-regulated condition mimicked MALAT1−/− or rG4 mutated lncRNA rescue conditions. This strongly indicates that NCL interacts with rG4s of MALAT1 that facilitate NCL co-localization in nuclear speckles to form pre-catalytic spliceosome complexes that modulate the alternative spliced variants of specific transcripts.

Deletion of MALAT1 rG4 structures leading to genome-wide AS pattern changes

Our study till now established that MALAT1 lncRNA interacts with NCL protein via its rG4 structures. Absence of rG4s causes loss of NCL localization from the nuclear speckles. This in turn causes AS changes in specific transcripts as the NCL protein might be interacting with other splicing proteins within the splicing hub: nuclear speckles. An earlier study reported the change in AS in around 200 mRNAs via microarray performed in MALAT1 downregulated HeLa cells from which the candidate transcripts for AS read-out were chosen for the cells lacking MALAT1 rG4s (29). To characterize the AS patterns as well as the mRNA expression in the different experimental conditions used in our study, we performed genome-wide RNA sequencing in MALAT+/+, MALAT1−/−, FL, Q123m and NCL knock-down (KD) conditions. The total number of AS reads occurring in each condition were analyzed, which was further investigated to assess the number of events under each AS type, namely Alternative 3′ splice site (A3), Alternative first exon (AF), Mutually exclusive exon (MX), Skipped exons (SE), Alternative 5′ splice site (A5), Alternative last exon (AL) and Retained intron (RI) as shown in Figure 7A (Supplementary Information 1). It was observed that the maximum AS events was obtained for NCL KD (136834 events) followed by MALAT−/− cells supplemented with rG4 deleted MALAT1, i.e., Q123m (129013 events). There was an overall increase in the number of events under the different types of AS in these two conditions and no inclination towards a particular AS type was observed in the absence of MALAT1 or due to the deletion of its rG4s. We further performed analysis to check the differential AS events in MALAT1−/−, FL, Q123m and NCL KD condition in comparison to MALAT1+/+ cells (Figure 7B, Supplementary Information 1). As expected, the FL condition which had rescued MALAT1 levels in the MALAT1−/− cell line, showed the minimal number of differential AS events (261) across each AS subtype, which was followed by the MALAT1−/− (313 events) and NCL KD (354 events) conditions respectively. Interestingly, our data indicated an increase in differential AS events in the Q123m condition when compared to MALAT1+/+ cells (796 events). This rise in differential AS reads due to the loss of rG4 structures in MALAT1 indicates its importance for this function. To understand the extent of splicing regulation that might be happening due to the interaction of these MALAT1 rG4s and NCL, we further analyzed the number of common events for AS in the four conditions MALAT1+/+, MALAT1−/−, Q123m and NCK KD (Figure 7C). The Venn diagram for the AS reads occurring in each condition represents a total of 96570 events that are common among all. Almost 5111 AS events that occurred due to the loss of MALAT1 lncRNA overlaps with NCL KD as well as Q123m condition. As NCL is a crucial protein for cell survival, its downregulation would affect cells and lead to co-lateral AS events (which is also reflected in the maximum number of AS events for NCL KD). Still, the unavailability of NCL in NCL KD cells would also affect its interaction with the rG4s of MALAT1 resulting in some specific AS changes in transcripts. Therefore, the common 4799 AS reads of NCL KD and Q123m which do not appear in either MALAT1+/+ or MALAT1−/− cells might be occurring due to the interaction of MALAT1 rG4s and NCL within cells. Apart from AS events, to understand the global mRNA profile of HeLa cells in the different experimental conditions, the differential mRNA expression was analyzed from the RNA-sequencing study (Figure 7D, Supplementary Information 2). Loss of MALAT1 lncRNA caused a
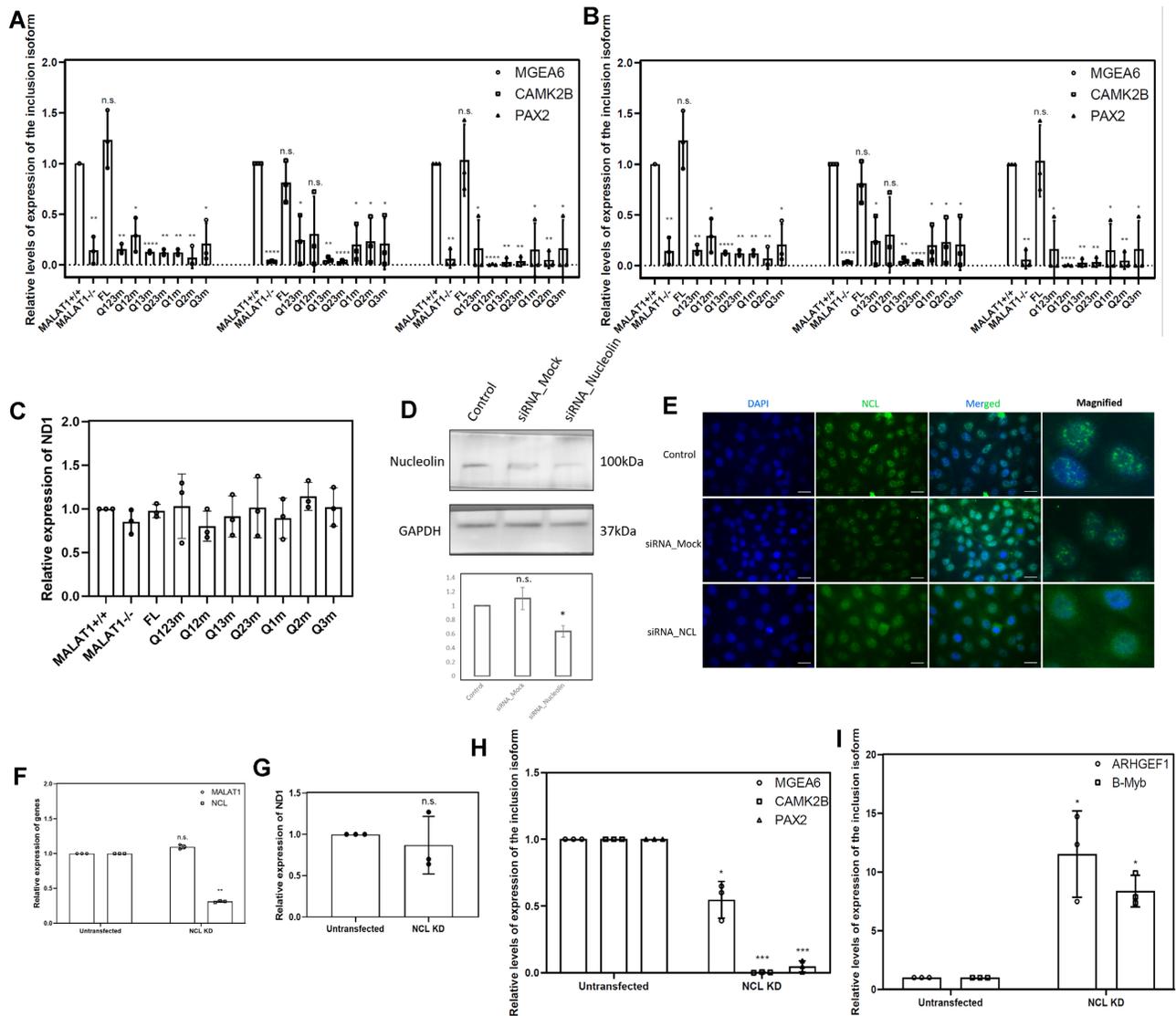
Figure 6. MALAT1 rG4 structures and NCL together orchestrate alternative splicing of specific transcripts in cervical cancer (HeLa) cells. (**A, B**) qRT-PCR to check relative levels of expression of the inclusion isoforms upon MALAT1 and rG4 depletion. (**C**) qRT-PCR to check relative levels of ND1 in mitochondria as a control transcript unaffected across the different conditions. (**D**) Western blot to validate the down-regulation of NCL by siRNA mediated knockdown compared to mock siRNA and untransfected controls. The quantification for the blot as shown is performed by ImageJ. ±S.D. across three independent biological replicates. (**E**) ICC for NCL to check its localization specifically in the nuclear speckles after siRNA mediated knockdown compared to mock and untransfected HeLa cells. Antibody against NCL is used to specifically capture NCL in nuclear speckles. The cells were counterstained with DAPI to mark the nucleus and the scale bars shown corresponds to 10 μM. (**F**) qRT-PCR to quantify down-regulation of NCL by siRNA mediated knockdown, showing ∼70% effective downregulation at the RNA level. MALAT1 levels, shown in normal and NCL down-regulated conditions, remain unchanged. (**G**) qRT-PCR to quantify the levels of ND1 mitochondrial transcript as a control unaffected by siRNA mediated NCL knockdown compared to wild type (WT) or untransfected HeLa cells. (**H, I**) qRT-PCR to check relative levels of expression of the inclusion isoforms spliced alternatively upon NCL knockdown compared to untreated HeLa cells. For (A–C, F–I) Beta-Actin housekeeping gene was 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).

change in the expression of 742 genes when compared to MALAT1+/+ cells. As the MALAT1 expression is restored in the FL rescue, only 38 genes had their expression different than the MALAT1+/+ cells. NCL KD itself would affect multiple genes, yet, it was interesting to note that in both NCL KD as well as the MALAT1 rescue with deleted rG4s (Q123m) differential expression of genes was skewed towards up-regulation when compared to MALAT1+/+ cells.

A further extrapolation of the data indicated 1357 common genes which were differentially regulated in NCL KD and Q123m conditions (Figure 7E). Therefore, a holistic picture of the global landscape of mRNA expression and AS regulation due to deletion of the rG4s in MALAT1 was obtained and most of them were in sync with NCL KD, which makes it evident that rG4 and NCL interaction plays a crucial role in orchestrating these events within cells.
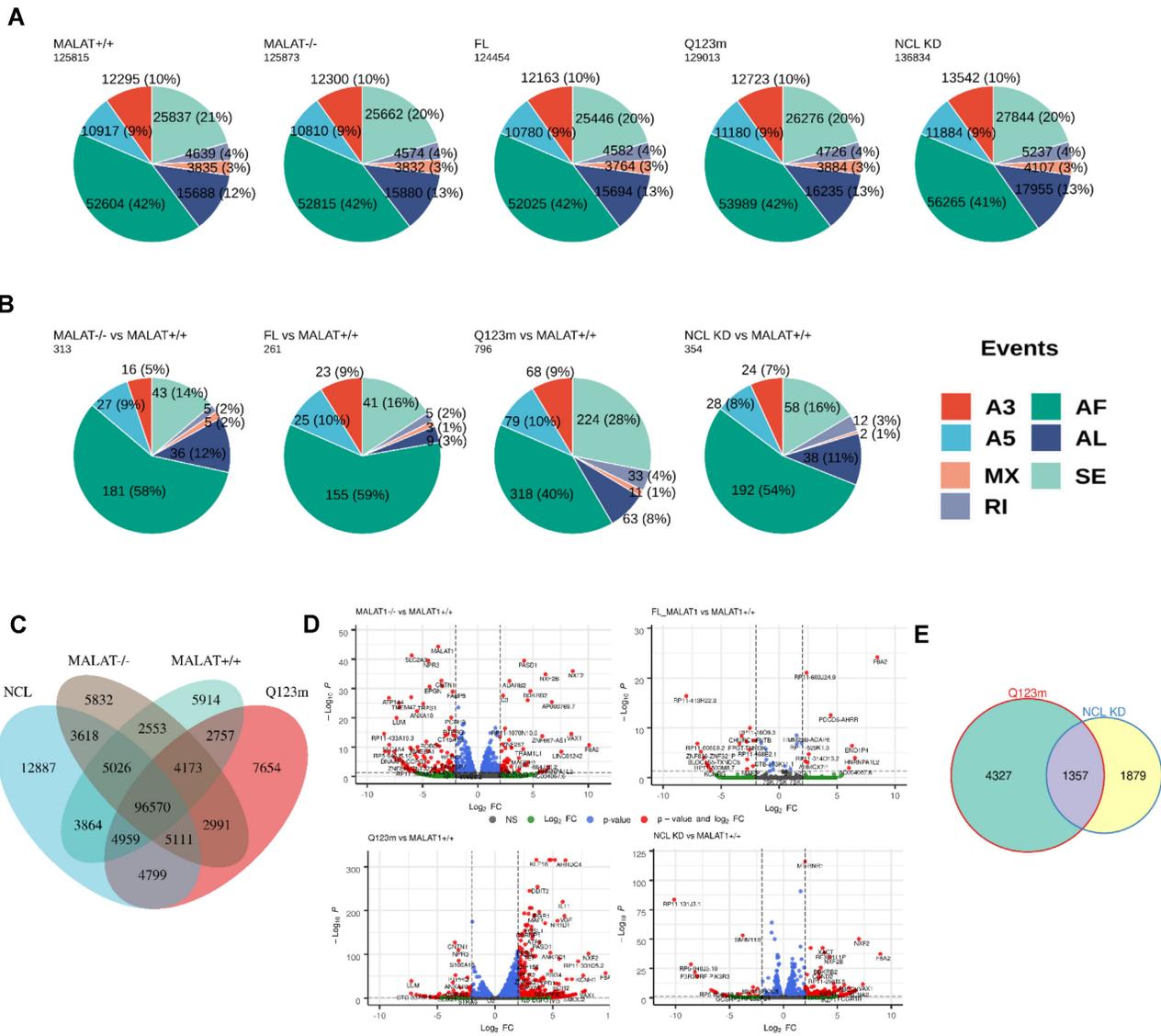
Figure 7. Deletion of MALAT1 rG4 structures leading to genome-wide AS pattern changes (**A**) AS reads with specific events occurring under each AS sub-type for the different experimental conditions used in this studyMALAT1+/+, MALAT1−/−, FL, Q123m and NCL KD. The AS subtypes as denoted in the figure are Alternative 3′ splice site (A3), Alternative first exon (AF), Mutually exclusive exon (MX), Skipped exons (SE), Alternative 5′ splice site (A5), Alternative last exon (AL) and Retained intron (RI). (**B**) Differential AS events captured under each AS sub-type for MALAT1−/−, Q123m and NCL KD with comparison to MALAT1+/+. (**C**) Over-all common AS reads under the MALAT1+/+, MALAT1−/−, Q123m and NCL KD conditions represented via Venn-diagram. (**D**) Differential mRNA expression occurring in MALAT1−/−, FL, Q123m and NCL KD with comparison to MALAT1+/+ conditions. (**E**) Common mRNAs differentially expressed with respect to MALAT1+/+ in Q123m and NCL KD conditions as represented via Venn-diagram.

DISCUSSION

So far, two alternative modes of action have been proposed for MALAT1: regulation of gene expression or AS. MALAT1 is specifically localized in nuclear speckles but its depletion does not affect the formation of nuclear speckles (29,35). Being a multi-functional lncRNA, MALAT1 is involved in multivalent interactions with different RNA binding proteins (RBPs) and structural motifs in MALAT1 may serve as an ideal scaffold to facilitate such interactions. In this study, for the first time, we establish that three rG4s in MALAT1 provide a unique platform to interact with NCL and regulate AS.

As MALAT1 is a multi-functional lncRNA, much attention has been paid to understand its secondary structure extensively. Although an earlier study refutes the formation of mouse Malat1 rG4 structures (28), multiple reports suggest its presence within cells (20,25,46–49,60,69). Undoubtedly, rG4 structures are very dynamic and are not always formed in-cellulo. It may require specific cues to form and might resolve as soon as their biological action is performed. They may arise at specific cell cycle stages, or in particular disease conditions where the cellular niche provides the suitable environment for the formation of such structures. Therefore, this kind of occurrence for a particular biomolecule might be developmental stage-specific and tissue-specific as well.

Using three experimental datasets that probed RNA structures, a recent report suggested a probable secondary structural model for 8425 nucleotides of human MALAT1 (19). According to this model, human MALAT1 is highly structured, forming 194 helices, 13 pseudoknots, five structured tetraloops along with many internal loops as well as intramolecular long-range interactions. Few high-resolution local structures have also been reported. An extensively characterized local structure in MALAT1 is the formation of a triple helix, the 3′ terminal stability element for nuclear expression (ENE) that result in the protection of MALAT1 from exonucleolytic degradation (37,39,40). It has been shown that the methyltransferase-like protein 16 (METTL16), an abundant nuclear protein, interacts with the MALAT1 triple helix *in-vitro* and *in-vivo* (70). Two well-structured hairpins at 2509–2537nt and 2556–2586nt have also been characterized that facilitate binding of the heterogeneous nuclear ribonucleoprotein G and ribonucleoprotein C (hnRNP G and hnRNP C) respectively, and abundant nuclear RNA binding proteins responsible for pre-mRNA processing (71,72). Bioinformatic analyses suggested that MALAT1 3′ end harbors three conserved rG4s and we have confirmed that these motifs individually form stable intramolecular parallel G-quadruplex structures in vitro. When the longer transcript of MALAT1 was tested for the formation of the three rG4 structures together with the help of ThT rG4 sensing fluorescent molecule, fluorescence emission was increased each time an additional quadruplex site was introduced in the longer transcript and was maximum when all the three rG4 motifs were intact. Thus, the rG4s in this lncRNA introduced a new structural module to fine-tune MALAT1 mediated biological regulation.

We show that rG4s present in the MALAT1 transcript neither impact its stability and expression nor act as a localization cue. However, we found that many proteins, including NCL and NPM, are associated with these rG4s specifically. Other studies of MALAT1 interacting proteins have also been reported. Chen et. al. identified 127 potential MALAT1-interacting proteins by performing an RNA pulldown followed by quantitative proteomics using stable isotope labeling of amino acids in cell culture (SILAC) method on a fragment of human MALAT1 (32). On the other hand, Scherer *et al.* performed mass spectrometry studies on 14 non-overlapping fragments covering the full-length mouse MALAT1 to identify possible nuclear interacting proteins (73). They identified 35 binding proteins of MALAT1, 14 of which were already identified from this study. Few proteins like hnRNPs overlapped with protein partners found in our study; however, NCL and NPM were identified as interacting protein partners for MALAT1 for the first time in our work. The difference in observations of MALAT1's interacting protein partners could be due to the use of different cell lysate sources in previous experiments. We used cell lysates from HeLa cells, whereas HepG2 and NSC-34 cell lysates were used in the earlier two studies. Moreover, RNA secondary structures, being dynamic, are very much environment-dependent (salt, buffer conditions, etc.). For instance, we have used high monovalent salt concentrations (100 mM KCl) to promote rG4 formation which is different from the salt concentrations used in earlier stud-

ies (120 and 10 mM NaCl) due to which RNAs used in the pull-down experiments may adopt different structures and thereby interact with different set of proteins. Indeed, the identified proteins in our study such as NCL are known to interact with rG4s specifically (23,24). In particular, we observed that both NCL and NPM bound specifically to the MALAT1 rG4s and not its mutant counterpart, whereas, some hnRNP proteins were found to interact with structured as well as unstructured RNA. We also saw that the sub-cellular distribution of hnRNPs remained unperturbed irrespective of the presence and absence of the MALAT1 transcript. Moreover, recent studies highlight that certain hnRNPs such as hnRNP H1 can interact with both rG1 unstructured G-tracts and rG4s *in-vitro* that corroborate with our findings (74). Interestingly, we observed that the localization of NCL and NPM to the nuclear speckles was dependent on the rG4s of MALAT1. We also established that all three rG4s were responsible for localizing the proteins to the nuclear speckles. To the best of our knowledge, this is the first demonstration that rG4s in lncRNA are responsible for the sub-cellular localization of certain proteins. We measured the in-vitro binding affinity of NCL to these three rG4s that were present in MALAT1 and found that NCL binds strongly with all three rG4s.

MALAT1 has been reported widely to take part in AS regulation in various cancer cells (29,75,76). NCL on the other hand participates in ribosomal synthesis as it is most abundant in the nucleolus although the smaller pool of NCL that localizes to the nuclear speckles plays no role in rRNA production as confirmed from the rDNA chip studies (65,77–79). Therefore, its function is highly implicated in AS. NCL in the nuclear speckles colocalizes with SC35 (as we also observe in our study), and is reported to interact with the pre-catalytic spliceosome complex and regulate AS of fibronectin (80). NCL is also known to play a major role in splice site selection (81). qRT-PCR performed to detect exon inclusion events for selected transcripts established the role of rG4s in AS. Interestingly, a similar pattern of exon inclusion was also observed for the transcript isoforms in the NCL knockdown condition. This indicated that loss of any one of the rG4s of MALAT1, or loss of NCL affect the AS pattern of certain transcripts similarly as in MALAT1−/− cells or MALAT1−/− cells rescued with MALAT1 having all rG4s mutated. From the genome-wide landscape of AS occurring, events in the rG4 mutated MALAT1 supplemented cells overlapped with NCL downregulated condition, and so does the expression of certain genes. Overall, our study provides a glimpse into the molecular mechanism of MALAT1 rG4 mediated NCL localization into speckles that modulate the fate of certain transcripts and globally affect AS splicing events. This observation also provides a foundation for a novel approach to target this rG4's structural scaffold present in the MALAT1 lncRNA and offers a highly unique platform for the future design of ligands for specific targets.

Since its discovery, MALAT1 has been linked to cancer and is associated with many clinical parameters that influence tumor cell proliferation, apoptosis, migration, invasion, or the metastatic spread of tumor cells (11,12,29). However, the specific involvement of MALAT1 in each of these diverse processes is not yet fully established. One of

the ways of MALAT1 action is by participating in splicing by serving as a scaffold for several splicing and chromatin remodeling factors (19,30,31,35). Thus, it can be hypothesized that some functions of MALAT1 might be attributed to its binding with such factors that regulate disease-specific AS. Either the canonical or non-canonical structures or the sequence of MALAT1 may serve as a scaffold for recruiting different RBPs to achieve appropriate temporal and spatial gene regulation. Here, in this study, we show that G-rich sequences within 3′regions of MALAT1 form rG4 structures and serve as a framework for recruiting and re-localizing a few RBPs including NCL and NPM to nuclear speckles. Analysis of qRT-PCR data on five selected transcripts under deficient rG4 conditions depicted specific transcripts isoform levels to be significantly altered, which is analogous to the observation in MALAT null or NCL knockdown conditions confirming the functional role of rG4s of MALAT1 in AS. RNA-sequencing studies also reveal the global AS events exclusively occurring due to lack of the MALAT1 rG4 motifs coincide with 4799 events occurring in the NCL knockdown condition confirming the regulatory role of these rG4s within the lncRNA. We speculate that targeting rG4 by selective small molecules may interfere with interactions between such trans-factors and structural elements in MALAT1 and thus help develop novel routes towards therapeutic modulation of disease states.

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. The RNA sequencing files are submitted in the Sequence Read Archive (SRA) database with the Accession code: PRJNA764507 and Submission-ID: SUB10379484.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

Conflict of interest statement. None declared.

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.* (2014) 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. U.S.A.*, **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 9 ends of long noncoding RNAs that lack poly (A) tails. *Genes Dev.*, **26**, 2392–2407.

41. Pulido-Quetglas,C., Aparicio-Prat,E., Arnan,C., Polidori,T., Hermoso,T., Palumbo,E., Ponomarenko,J., Guigo,R. and Johnson,R.

42. 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.

43. 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.

44. 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.

45. 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.

46. Parrish,N., Hormozdiari,F. and Eskin,E. (2011) Assembly of non-unique insertion content using next-generation sequencing. *BMC Bioinformatics*, **12**, S3.

47. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

48. McCarthy,D.J., Chen,Y. and Smyth,G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.*, **40**, 4288–4297.

49. Blighe,K., Rana,S. and Lewis,M. (2021) EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. R package version, 1.10.0.

50. Patro,R., Duggal,G., Love,M.I., Irizarry,R.A. and KIngsford,C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, **14**, 417–419.

51. Alamancos,G.P., Pagès,A., Trincado,J.L., Bellora,N. and Eyras,E. (2015) Leveraging transcript quantification for fast computation of alternative splicing profiles. *Rna*, **21**, 1521–1531.

52. Trincado,J.L., Entizne,J.C., Hysenaj,G., Singh,B., Skalic,M., Elliott,D.J. and Eyras,E. (2018) SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.*, **19**, 40.

53. Schwenk,A.J. (1984) Venn diagram for five sets. *Math. Mag.*, **57**, 297.

54. 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.

55. 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.

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

57. 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.

58. Renard,I., Grandmougin,M., Roux,A., Yang,S.Y., Lejault,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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. Qiu,L., Wang,H., Xia,X., Zhou,H. and Xu,Z. (2008) A construct with fluorescent indicators for conditional expression of miRNA. *BMC Biotechnol.*, **8**, 77.

67. 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.

68. 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.

69. 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.

70. 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. U.S.A.*, **113**, 14013–14018.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. Ginisty,H., Sicard,H., Roger,B. and Bouvet,P. (1999) Structure and functions of nucleolin. *J. Cell Sci.*, **11**, 761–772.

78. Kumar,S., Gomez,E.C., Chalabi-Dchar,M., Rong,C., Das,S., Ugrinova,I., Gaume,X., Monier,K., Mongelard,F. and Bouvet,P. (2017) Integrated analysis of mRNA and miRNA expression in HeLa cells expressing low levels of Nucleolin. *Sci. Rep.*, **7**, 9017.

79. Berger,C.M., Gaume,X. and Bouvet,P. (2015) The roles of nucleolin subcellular localization in cancer. *Biochimie*, **113**, 78–85.

80. 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.

81. Shefer,K., Boulos,A., Gotea,V., Chaim,Y.B., Sperling,J., Elnitski,L. and Sperling,R. (2020) A novel role for nucleolin in splice site selection. bioRxiv doi: https://doi.org/10.1101/2020.12.02.402354, 03 December 2020, preprint: not peer reviewed.