Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates

Kangkang Niu[1,2], Lijun Xiang[1,2], Ying Jin[1,2], Yuling Peng[1,2], Feng Wu[1,2], Wenhuan Tang[1,2], Xiaojuan Zhang[1,2], Huimin Deng[1,2], Hui Xiang[1,2], Sheng Li[1,2], Jian Wang[3], Qisheng Song[4,*] and Qili Feng[1,2,*]

[1]Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China, [2]Guangzhou Key Laboratory of Insect Development Regulation and Application Research, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China, [3]Department of Entomology, University of Maryland, College Park, MD 20742, USA and [4]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

ABSTRACT

Double-stranded DNAs are usually present in the form of linear B-form double-helix with the base pairs of adenine (A) and thymine (T) or cytosine (C) and guanine (G), but G-rich DNA can form four-stranded G-quadruplex (G4) structures, which plays important roles in transcription, replication, translation and protection of telomeres. In this study, a RNA recognition motif (RRM)-containing protein, BmLARK, was identified and demonstrated to bind G4 structures in the promoters of a transcription factor BmPOUM2 and other three unidentified genes of *Bombyx mori*, as well as three well-defined G4 structures in the human genes. Homologous LARKs from *Bombyx mori*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* bound G4 structures in *BmPOUM2* and other genes in *B. mori* and *H. sapiens*. Upon binding, LARK facilitated the formation and stability of the G4 structure, enhancing the transcription of target genes. The G4 structure was visualized *in vivo* in cells and testis from invertebrate *B. mori* and vertebrate Chinese hamster ovary (CHO) cells. The results of this study strongly suggest that LARK is a novel and conserved G4-binding protein and that the G4 structure may have developed into an elaborate epigenetic mechanism of gene transcription regulation during evolution.

INTRODUCTION

The genetic DNA molecule is usually present in a double-helix structure with strict base pairings between adenine (A) and thymine (T) and between cytosine (C) and guanine (G) (1). However, when multiple Gs or Cs continuously and tandemly exist in double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA), the DNA molecule has been shown to form G-quadruplex (G4) or i-motif secondary structures, respectively (2,3). The G4 structure is formed by stacked G-tetrads, which are square co-planar arrays of four G bases stabilized by a monovalent cation, such as K^+ (4). Correspondingly, the complementary strand of the G4 structure can form an i-motif structure, which is formed by intercalated hemi-protonated C–C ($C–C^+$) base pairs under acidic conditions (5), at a neutral pH by molecular crowding of the cosolutes (6) or under high pressure (7). These non-linear and non-B-form double-helix secondary structures have been found in promoter regions, replication initiation sites, telomeres and mRNA molecules (8–11) and play important roles in the regulation of gene transcription (12), DNA replication (9), protein translation (11) and telomere protection (10). Recent studies in vertebrates and invertebrates suggest that changes in G4 or i-motif structures could lead to changes in gene transcription (13,14). Thus, the existence and function of secondary structures in DNA and RNA molecules provide a novel epigenetic mechanism of genetic activity regulation.

G4 structures have been identified and well characterized *in vitro* in chromosome telomeres and promoter regions of human genes, including *c-MYC*, *VEGF*, *c-kit* and *bcl-2* (12,15–17). Recently, G4s have also been identified in inver-

*To whom correspondence should be addressed. Tel: +86 20 85215291; Fax: +86 20 85215291; Email: qlfeng@scnu.edu.cn
Correspondence may also be addressed to Qisheng Song. Tel: +1 573 8829798; Fax: +1 573 8821469; Email: songq@missouri.edu

tebrates, including *Drosophila* and silkworm (14,18). Bioinformatic analyses predict the presence of numerous G4 structures in various genomes. The human genome might have over 700 000 distinct G4 motifs (19). In silkworm, 3174 G4 motifs have been predicted (14). However, most G4 motifs are predicted based on sequence analyses and have not been visually demonstrated *in vivo*.

In humans, over 40% of genes have one or more G4 structures in their promoter regions (20). By using G4-ChIP-seq with an engineered G4 antibody (BG4), ~10 000 G4 structures mainly located in regulatory and nucleosome-depleted regions were marked in human chromatin (21). The stabilization or destruction of G4 structures affects the *in vivo* gene transcription of some oncogenes, such as the human genes *c-MYC* (12), *KRAS* (22) and *c-kit* (23) and the zebrafish genes *col2al*, *fdz5*, and *nog3* (24).

To elucidate the possible functional mechanism of G4 structures in diverse biological processes, substantial efforts have been made to identify G4 structures and their binding proteins. hPARP-1, which was the first G4-binding protein identified, mediates transcriptional regulation and telomere end protection in humans (25). Nucleolin (26), Rif1 (27), Sub1 (28), CNBP (29) and SLIRP (30) have also been found to bind G4 structures in different genes, while hnRNP A1 and hnRNP A2 specifically bind unfolded G4 sequences (31,32). These findings suggest that G4 structures are regulated by different proteins. However, *in vivo* experimental data are still needed.

In our previous study, we found that there is a G4 structure in the reverse strand of the *BmPOUM2* promoter (14). Here, we report that the RNA recognition motif (RRM)- and CCHC-type zinc finger-containing protein BmLARK and its homologues specifically bind G4 structures in the promoter of the BmPOUM2 transcription factor and that different forms of G4s in genes of *Bombyx mori* and other species regulate the transcription of target genes. The *ex vivo* existence of G4 structures was visualized in invertebrate and vertebrate cells.

MATERIALS AND METHODS

Insect and cell lines

Silkworm *B. mori* strain Dazao was provided by the Research and Development Center of the Sericultural Research Institute of the Academy of Agricultural Sciences of Guangdong Province, China. Larvae were raised on fresh mulberry leaves at 26°C under a 14/10 h light/dark photoperiod. The *B. mori* cell line DZNU-Bm-12 (*Bm*12), which was originally developed from ovarian tissues (33), was maintained at 28°C in Grace medium (Gibco, New York, USA) supplemented with 10% FBS. The CHO (Chinese hamster ovary) cell line was maintained at 37°C in F-12K Nutrient Mixture medium (Gibco, New York, USA) supplemented with 10% FBS.

Protein pull-down and liquid chromatography–tandem mass spectrometry (LC–MS/MS) analyses

The protein pull-down assay was conducted as described in a previous study (14). Nuclear proteins from larval wing disc cells were extracted using NE-PER Nuclear and Cytoplasmic Extraction Reagents according to the manufacturer's instructions (Thermo Scientific, MA, USA). The protein concentration was determined using a BCA Protein Assay Reagent Kit (Thermo Scientific, MA, USA). DNA oligonucleotide sequences containing G4 structures were heated at 95°C for 10 min in 50 mM Tris buffer at pH 7.5 with 100 mM KCl and then slowly cooled to room temperature over 4 h. Single-stranded biotinylated DNA oligonucleotides (20 μg) were incubated with 100 μg streptavidin-coated Dynabeads (Life Technologies, California, USA) in 400 μl binding buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 1 M NaCl, and 0.003% NP40) for 30 min at room temperature under constant and slow rotation. After twice washed with binding buffer, the immobilized DNA was incubated for 30 min in 400 μl blocking buffer (2.5 mg/ml BSA, 10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA and 3.5% glycerol with 0.003% NP40 and 5 mg/ml polyvinyl pyrrolidone). Then, 100 μg of nuclear proteins was incubated with the immobilized DNA for 4 h at 4°C in 400 μl protein binding buffer (10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 80 mM KCl, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA and 3.5% glycerol with 0.001% NP40 and 1 μg non-specific DNA) under constant and slow rotation. After 4 h, the DNA/protein complexes were washed three times with 400 μl washing buffer (10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA, 3.5% glycerol, 0.5 mg/ml BSA and 0.05% NP40). The proteins bound to the DNA were eluted with 20 μl SDS-PAGE sample buffer (50 mM Tris, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, and 10% glycerol). The eluted proteins were subjected to 12% SDS-PAGE. The gels were stained with Coomassie Brilliant Blue R-250 for more than 4 h and destained with destaining solution (10% acetic acid, 5% ethanol and 85% water). The different protein bands were excised and analysed for protein identification by Huijun Biotechnology (Guangzhou, China) using liquid chromatography–tandem mass spectrometry (LC–MS/MS). A full report of LC-MS/MS data is provided as supplementary data.

Expression and purification of recombinant BmLARK protein

Total RNA extraction from the *Bm*12 cells and reverse transcription were conducted using a previously described method (34). The BmLark open reading frame (ORF) fragment was amplified using cDNA cloned from *Bm*12 cells with the 5′-GAATTCATGCCGGGCACCGGTACT-3′ forward primer and 5′-GCGGCCGCTTAATAAGGC GCATATCTTCTTGAC-3′ reverse primer (the restriction enzyme sites are underlined). The obtained DNA fragment was then sub-cloned into the pGEX-6P-1 vector between the EcoR I and Not I restriction enzyme sites to generate the BmLARK-pGEX-6P-1 recombinant expression vector.

The recombinant BmLARK protein was expressed in *Escherichia coli* cells (BL21) that were grown in Luria-Bertani (LB) medium containing 100 μg/ml ampicillin at 18°C with a 12 h induction by 0.1 mM isopropyl-β-D-thiogalactoside (IPTG). The cells were collected by centrifu-

gation and re-suspended in phosphate-buffered saline (PBS) (136 mM NaCl, 1.1 mM K_2HPO_4, 2.7 mM KCl and 8.0 mM Na_2HPO_4, pH 7.4). The suspension was centrifuged at 10 000 g at 4°C for 5 min after being lysed by sonication. The GST-LARK proteins were purified from the supernatant using BeyoGold™ GST-tag Purification Resin (Beyotime, Beijing, China) according to the manufacturer's recommended procedures. The other recombinant proteins (DmLARK, MmLARK, HsLARK, RRM1-2, RRM2 and ZnF regions of BmLARK) were expressed and purified using the same method. The ORF of RRM1 domain was subcloned into the pET32a vector and purified using a His-Bind Kit according to the manufacturer's instructions (Novagen, Darmstadt, Germany). To remove the glutathione and imidazole, the purified proteins were dialysed (25 kDa molar weight cut-off membrane) against 20 mM Tris–HCl buffer (pH 7.5) at 4 °C overnight. The protein concentration was determined using a BCA Protein Assay Reagent Kit (Thermo Scientific, MA, USA). The purified proteins were examined by using SDS-PAGE (Supplementary Figure S1).

Antibody preparation of BmLARK and western blot

Anti-LARK antibodies were prepared by Biogot Technology (Nanjing, China). In short, the soluble GST-BmLARK was purified using GST-tag purification resin. The purified recombinant GST-BmLARK was mixed with Freund's adjuvant and injected into New Zealand White rabbits. Antisera were collected and purified after five booster injections. For Western blot, protein extracts were separated in 12% SDS-PAGE and transferred to a nitrocellulose membrane. The membrane was blocked with 3% bovine serum albumin (BSA) in Tris-buffered saline Tween (TBST) (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween 20) for 2 h at room temperature, and then incubated with the anti-BmLARK antibody in a dilution of 1:2000 at 4°C overnight, and then incubated 1 h at room temperature with the second antibody of horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG in a dilution of 1:5000 (Dingguo Biotechnology, Guangzhou, China). Immunoreactivity was detected by Enhanced Chemiluminescence (ECL) (Invitrogen, CA, USA). The membranes were washed with TBST three times, each for 5 min before color development.

Electrophoretic mobility shift assay (EMSA) of protein–DNA binding

An EMSA was conducted using a Light Shift Chemiluminescent EMSA Kit (Thermo Scientific, MA, USA). The wild-type and mutant oligonucleotides were labelled with biotin at the 5′-end, heated at 95°C for 10 min in 50 mM Tris buffer at pH 7.5 with or without KCl and slowly cooled to room temperature over 4 h. The oligonucleotides labelled with biotin at the 5′-end were synthesized by Invitrogen (CA, USA).

The binding reactions were performed according to the EMSA Kit instructions. Briefly, the reactions were conducted in a 20 μl mix [1 × binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT, pH 7.5), 2.5% glycerol, 0.05% NP-40, 5 mM $MgCl_2$, 4 mM EDTA, 780 nM (1 μg) purified recombinant LARK protein, and 20 fmol of an end-labeled biotinylated probe] at room temperature for 20 min. For the competition assay, cold probes (same DNA sequences but not biotinylated) were added to the binding reaction. For the pyridostatin (PDS) treatment, different concentrations (0.25, 0.5 and 1 μM) of PDS were incubated with 20 fmol end-labelled biotinylated probe for 30 min before the binding reaction. Then, the samples were separated on a 4% polyacrylamide gel on ice at 100 volts for 1.5 h. After electrophoresis, the gel was blotted onto a positively charged Nylon membrane (Amersham Biosciences, Boston, USA). Then, the membranes were developed using a Light Shift Chemiluminescent EMSA Kit according to the manufacturer's protocol. The oligonucleotide probes used in this study are shown in Supplementary Table S1.

Microscale thermophoresis (MST)

The DNA oligonucleotides were labelled with Cy5 and heated at 95°C for 10 min in 50 mM Tris buffer at pH 7.5 with 100 mM KCl and slowly cooled to room temperature over 4 h. The purified BmLARK with varying concentrations (from 2.93 to 6000 nM) was incubated with 25 nM of the labeled G4 structures at room temperature for 20 min in buffer containing 20 mM Tris (pH 7.5), 100 mM NaCl and 0.05% Tween-20. MST analysis was performed using standard Capillaries from Nanotemper and with LED 60% and 40% MST power, on the Monolith NT.115 (NanoTemper Technologies GmbH, Munich, Germany) at 24°C. All experiments were repeated three times for each measurement. Data analyses were carried out using NanoTemper analysis software.

Polymerase stop assay

The polymerase stop assay template was designed according to the POUM2 G4 sequence 5-GTGCGGGGCGCGAGGGGGCCGAGGGGCGGGGCAA-3. The template DNA was annealed in annealing buffer (50 mM Tris–HCl, pH 7.4 and 100 mM KCl) by heating at 95°C for 10 min, followed by slow cooling to room temperature. The 5′-TCGTTACTCGATTGCCCCG-3′ primer was 5′-labelled with FAM, and a PCR reaction was performed with different amounts of PDS. The PCR reaction contained (25 μl total volume) 2.5 μl PCR buffer, 0.25 μl of each primer (100 μM final concentration), 0.5 μl of the template DNA samples (5 μM), 2 μl dNTPs (2.5 mM), 2 U Taq polymerase, and different amounts of PDS. The mixtures were incubated at 95°C for 3 min, followed by 30 cycles at 95°C for 30 s, 60°C for 30 s and 72°C for 30 s. The products were detected by electrophoresis using a 20% denaturing polyacrylamide gel. The small molecule compound PDS (MCE, NJ, USA) used in the polymerase stop assay was dissolved in water and stored at −20°C.

Chromatin immunoprecipitation (ChIP) assay

The *Bm*12 cells were crosslinked with 1% formaldehyde for 10 min at room temperature after transfection with BmLARK-3× FLAG-EGFP (a terminator codon was

added before the EGFP coding sequence) or EGFP (control) plasmids for 48 h. Glycine was added to terminate the fixation, and the cells were washed twice with one media volume of ice-cold PBS. Cells in 1 ml of ice-cold PBS and 10 μl Halt Cocktail were collected by scraping. The cells were centrifuged at 3000 g for 5 min, and the cell pellet was broken up with extraction buffer containing protease/phosphatase inhibitors. The nuclei were collected by centrifugation at 9000 g for 3 min and digested by MNase. Digested chromatin was obtained by sonicating on ice with several pulses to break up the nuclear membrane with 20 s incubations on ice between the pulses and then centrifuged at 9000 g for 5 min. An immunoprecipitation experiment was performed with a Pierce™ Magnetic ChIP Kit (Thermo Scientific, Massachusetts, USA) according to the manufacturer's instruction. Ten micrograms of either rabbit anti-FLAG antibody (#14793, Cell Signaling Technology, MA, USA) or normal rabbit IgG (as a control) (Thermo Fisher Scientific, Massachusetts, USA) was used for the IP reactions that were incubated overnight at 4°C with constant mixing. The DNA/protein/antibody complexes were purified by incubation with ChIP Grade Protein A/G Magnetic Beads for 2 h at 4°C with mixing. The immunoprecipitated genomic DNA fragments were amplified using quantitative real-time PCR (qRT-PCR) with primers (Supplementary Table S3). The specificity of the primers was examined using Primer-BLAST (http://www.ncbi.nlm. nih.gov/tools/primer-blast/). For the qRT-PCR, the SYBR Green Kit was used according to the manufacturer's instructions (TaKaRa, Dalian, China). The qRT-PCR reaction contained 10 μl of 2× SYBR Premix Ex Taq, 0.4 μl of each primer at a final concentration of 10 μM, and 2 μl of the immunoprecipitated DNA samples in a 20 μl reaction volume. The mixtures were incubated at 95°C for 10 s, followed by 40 cycles at 95°C for 5 s and 60°C for 31 s using an ABI7300 fluorescence quantitative PCR system. The enrichment of the promoter sequence in the immuno-precipitated DNA samples was normalized to the DNA present in the 10% input material analysed using the $2^{-\triangle\triangle Ct}$ method (35). The PCR products of the enriched promoters were confirmed by sequencing at Tsing Ke Biotechnology (Guangzhou, China). The enriched DNA fragments were inserted in pUC118 vector using Blunting Kination Ligation (BKL) Kit (TaKaRa, Dalian, China). The recombinant vectors were used to transform *Escherichia coli* (DH5α). Ten colonies of transformed bacterial were randomly selected for sequencing to check the G4-sequence enrichment.

Circular dichroism (CD) analysis

The CD analyses were performed with a J-815 CD spectrometer (Jasco International, Tokyo, Japan). All spectra were collected at wavelengths ranging from 220 to 350 nm with a 1 nm step width and a 1 s response time. The CD spectra represent three averaged scans from the same sample obtained at room temperature and are baseline-corrected for the signal contributions due to buffer. The DNA oligonucleotide sequences were diluted to 5 μM in 50 mM Tris–HCl, pH 7.5. The unannealed ssDNA samples were incubated with either Tris–HCl, pH 7.5 buffer, protein dialysis buffer or specified amounts of recombinant Bm-

LARK protein at room temperature for 30 min to reach equilibrium prior to the CD spectroscopy. To fold the ss-DNAs into the G4 conformation, the oligonucleotide was heated at 95°C for 10 min in 50 mM Tris buffer (pH 7.5) with 100 mM KCl and then slowly cooled to room temperature over 4 h. The formed G4 structure was incubated with either protein dialysis buffer or recombinant protein at room temperature for 30 min prior to the CD spectroscopy.

RNA interference (RNAi) and overexpression of BmLARK and PDS treatment

For the BmLARK RNAi in the *Bm*12 cell line, a 458-bp unique fragment from nts 198–655 in the BmLARK ORF was used as a template for synthesizing the gene-specific dsRNA. The BmLARK dsRNA was synthesized from a linearized template using the T7 RiboMAXTM Express RNAi System (Promega, Madison, USA). DsRNA (4 μg) was transfected into *Bm*12 cells in Grace Medium with 6 μl Fugene HD transfection reagent in Opti-MEM Reduced Serum Medium. For the BmLARK overexpression in the *Bm*12 cell line, the BmLARK-3× Flag-EGFP recombinant plasmid was constructed, and a terminator codon was added before the EGFP coding sequence. Thus, the recombinant protein is BmLARK-3 × Flag. The BmLARK-3× Flag-EGFP plasmid (1 μg) was transfected into *Bm*12 cells in Grace Medium with 4 μl Fugene HD transfection reagent in Opti-MEM Reduced Serum Medium. The cells were collected 48 h after transfection. Total RNA extraction and reverse transcription were conducted as mentioned above (34). Western blot and qRT-PCR were performed to detect the overexpression and knock-down efficiencies. *RP49* was used as internal control in qRT-PCR assays. The Y-axis represents the relative expression levels of the genes after normalized with the *RP49* expression.

*Bm*12 cells were inoculated into culture media in 12-well culture plates (Corning, NY, USA) and cultured for 48 h with or without PDS. Then, the total RNA extraction and reverse transcription were conducted as mentioned above (34). qRT-PCR was performed to detect the mRNA levels of BGIBMGA007142-TA, BGIBMGA008159-TA and BGIBMGA006558-TA, which were predicted to have G4 structures in their promoters. The qRT-PCR reactions contained 10 μl of 2× SYBR Premix Ex Taq, 0.4 μl of each primer (10 μM final concentration) and 2 μl of the cDNA samples in a 20 μl reaction volume. The mixtures were incubated at 95°C for 10 s, followed by 40 cycles at 95°C for 5 s and 60°C for 31 s using an ABI7300 fluorescence quantitative PCR system.

Immunostaining G4 structures in cell nuclei and chromatin

Cells grown on glass coverslips were fixed in 4% paraformaldehyde for 10 min, permeabilized with 0.5% Triton-X100 at room temperature, and then blocked in blocking solution (2% bovine serum and 5% goat serum in PBS) for 1 h at room temperature. After blocking, the cells were incubated with BmLARK-3× FLAG (10 μg/ml) in a 1:10 diluted blocking solution for 2 h at room temperature. The primary anti-FLAG antibody (#2368, Cell Signalling Technology, MA, USA) was incubated at a 1:100 dilution

in 1:10 diluted blocking solution overnight at 4°C. On the following day, the cells were incubated with anti-rabbit Alexa 594-conjugated (A11037, Invitrogen, CA, USA) secondary antibodies at a 1:50 dilution in a 1:10 diluted blocking solution for 1 h. The cells were washed with PBS three times before each procedure. Then, the coverslips were mounted with Prolong Gold/DAPI (4′,6-diamidino-2-phenylindole) (P36941, Invitrogen, California, USA), and confocal images were obtained under an Olympus Fluoview FV1000 confocal microscope.

The whole testis was isolated from the fourth instar larvae, and the fat body, weasand and other associated tissues were dissected in sodium citrate saline. The testis was fixed for 2 h with Carnot-type fixed solution after three hypotonic treatments of 0.075 mol//l KCl for 45 min at 37°C and historrhexis. Then, the testis was transferred to a poly-lysine-coated slide and dried before washing three times with 0.4% PBST for 10 min. The tissues were incubated with purified BmLARK protein for 3 h after the slides were blocked in 1% goat serum for 1 h. After the slides were washed three times with 0.4% PBST, the primary anti-BmLARK antibody was incubated at a 1:2000 dilution in 1% goat serum overnight at 4°C. Then, the anti-rabbit Alexa 594-conjugated (A11037, Invitrogen, California, USA) secondary antibody was added after washing three times with 0.4% PBST, followed by a 1 h incubation at room temperature. After washing three times with 0.4% PBST and staining with DAPI for 1 h, the slides were washed once with PBS and sealed with 50% glycerol before imaging under an Olympus Fluoview FV1000 confocal microscope.

RESULTS

Identification of a nuclear protein that binds the *BmPOUM2* G4 structure

To identify nuclear proteins that bind the *BmPOUM2* G4 structure, a DNA-protein pull-down assay was performed with a DNA probe containing the G4 structure and nuclear proteins isolated from larval wing disc cells (Figure 1Aa). Six proteins (Supplementary Table S2) were identified by LC-MS/MS based on: (i) their molecular mass, being ∼38.2 kDa, (ii) DNA binding domain and (iii) nuclear transcription factor. While the other five were found to not bind with G4 structure in subsequent binding assays, a 38.2 kDa protein bound to the G4 probe, but not the mutant ssDNA probe, was finally identified to be a homologue of LARK (GenBank accession no: NM_001043828). This protein has two RRM domains that are also present in the reported mammalian G4-binding proteins, such as nucleolin (26), SLIRP (30) and hnRNP A1 (31) (Supplementary Figure S2). LARK has one CCHC-type zinc finger domain, which is also present in the G4-binding protein CNBP (29). The sequence alignment shows that the CCHC-type zinc finger domains are similar in LARK and CNBP. These CCHC-type zinc finger domains have a low similarity with other three types of zinc finger domains, C2H2, A20-like and C2C2 (Supplementary Figure S3). The pre- and post-pull down and pull-down samples were tested by Western blot (Figure 1Ab). The result showed that the G4 pull-down product could be recognized by the anti-LARK antibody and the lysate of post-pull down and mutant G4 sequence

(cannot form G4) pull-down products could not be recognized by the antibody.

To determine whether this protein can bind the G4 structure in the *BmPOUM2* promoter, *B. mori* LARK cDNA was cloned and expressed as a recombinant protein for EMSA. BmLARK could bind the G4 sequence in the presence of 100 mM K$^+$ after a stable G4 structure was formed (Figure 1B). This binding was enhanced by increasing the protein concentration (Figure 1C) and reduced by the addition of unlabelled cold probe (Figure 1B). BmLARK did not bind the mutated G4 sequence that did not form the G4 structure (Figure 1B).

If BmLARK is truly a G4-binding protein, its binding should be affected by pyridostatin (PDS), a well-known G4 structure stabilizer (36,37). The polymerase stop assay (Figure 1D) showed that the intensity of the target PCR product band decreased as the PDS concentrations increased, and the PDS effect disappeared when the mutated PCR template, which cannot form a G4 structure, was used (Figure 1E). Furthermore, the EMSA results showed that the binding of BmLARK to the G4 structure was blocked by the increased PDS concentration (Figure 1F). These results indicate that PDS can interact with the *BmPOUM2* promoter G4 structure to inhibit the binding of BmLARK to the G4 structure.

BmLARK binds G4 structures in different genes

After demonstrating the binding of BmLARK to *BmPOUM2* G4 structures, to further determine whether BmLARK binds other G4 structures, three G4 sequences in the BGIBMGA007142-TA (142), BGIBMGA008159-TA (159) and BGIBMGA006558-TA (558) gene promoters were randomly selected from 147 predicted G4 sequences in the upstream regions of 322 genes in *B. mori* (14). The G4 sequences were first confirmed by performing CD analyses (Figure 2A–C), then the *in vitro* binding of BmLARK to these G4 structures was examined by EMSA. BmLARK bound all G4 structures, and this binding could be enhanced by 100 mM K$^+$ and competed off by unlabelled cold probes (Figure 2D–F). The protein could not bind the mutant probes, which cannot form G4 structures.

Furthermore, the binding of BmLARK to the well-defined G4 structures in the human genes *c-MYC* (38), *HIF-1a* (39) and *c-kit* (23) was examined by EMSA. BmLARK bound to the different types of reported G4 structures in these human genes (Figure 2G–I). This binding was enhanced by 100 mM K$^+$ and could be competed off by unlabelled cold probes. Furthermore, the concentration-dependent EMSAs (Figure 2D–I) showed that the binding was enhanced with the increase in the protein concentration. In addition, there were shifted bands in the absence of potassium. This may due to the BmLARK-induction of the G4 formation from ssDNA. The CD results further confirmed this speculation that in the presence of BmLARK, ssDNA of these G4 sequences could form G4 structures to some extent as shown by the increase in the feature absorption peak at 265 nm (Supplementary Figure S4). These results suggest that BmLARK binds G4 structures not only in *B. mori* genes but also human genes, suggesting that BmLARK is a broad-spectrum G4-binding protein. To deter-
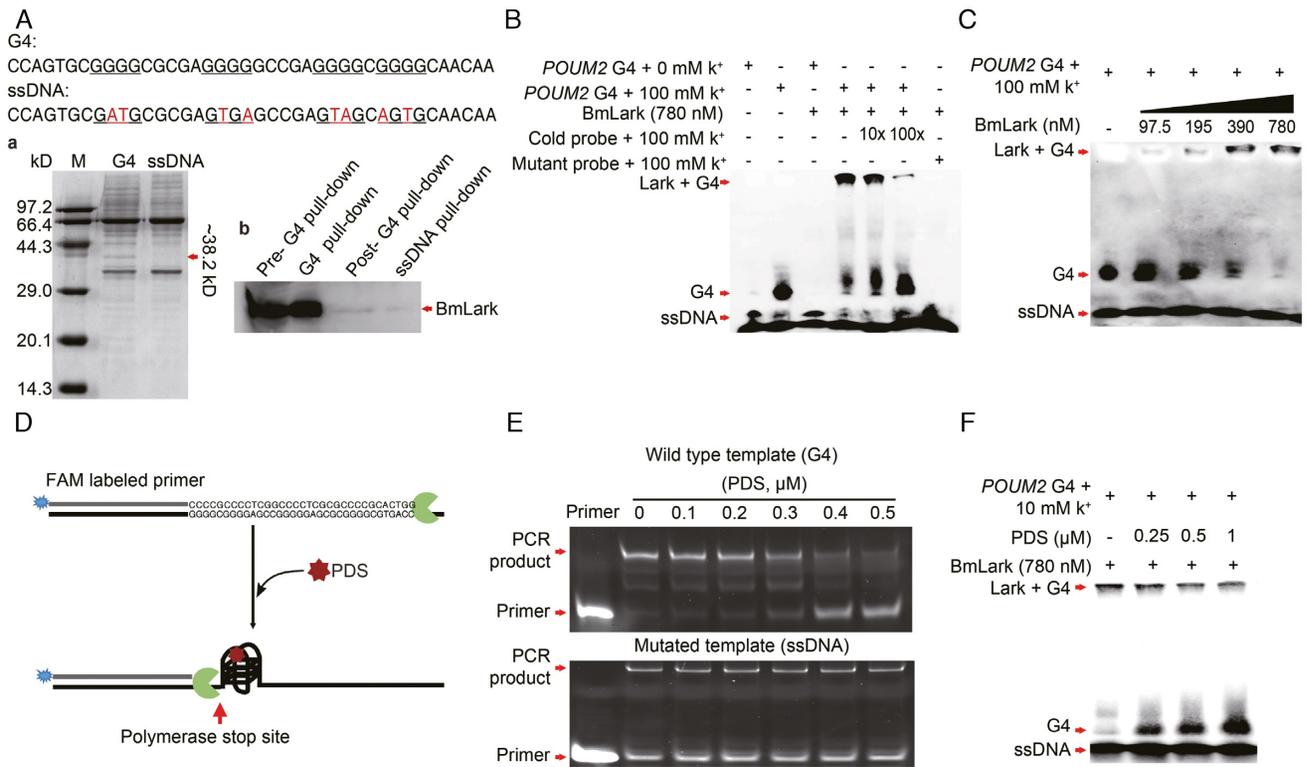
Figure 1. Identification and binding of the BmLARK protein to G4 structures in *BmPOUM2* promoter. (**Aa**) Identification of the G4-binding protein by a DNA pull-down experiment. (**Ab**) Western blot analysis of the lysates of pre- and post- pull down and pull-down samples. (**B** and **C**) EMSA analyses of recombinant BmLARK binding the G4 structure. (**D**) Descriptive diagram of the polymerase stop assay. In the absence of PDS, there were no stable G4 structures and the polymerase could goes through the transcription region. In the presence of PDS, the G4 structures were stabilized and the polymerase was prevented from passing through the transcription region. (**E**) Polymerase chain reaction (PCR) of polymerase stop assay using wild-type and mutated templates with or without PDS. (**F**) Effect of PDS at different concentrations on the binding of BmLARK to G4 structures by EMSA analysis. ssDNA: single-stranded DNA; G4s: G4 sequences; and PDS: small molecule ligand pyridostatin.

mine and compare the binding affinity of BmLARK to different G4 structures, MST analysis, which is a powerful method to quantify protein-DNA interactions (40,41), was conducted. As shown in Figure 3, the K_d values of BmLARK binding to the G4s in the promoters of *BmPOUM2*, *142*, *159*, *558*, *c-kit*, *HIF-1a* and *c-MYC* were calculated as 110.84, 107.35, 113.64, 120.32, 126.48, 130.84 and 269.23 nM, respectively.

BmLARK binds *B. mori* G4 structures *ex vivo*

After demonstrating the binding of BmLARK to *BmPOUM2, 142, 159* and *558* G4 structures *in vitro*, whether BmLARK binds the G4 structure *ex vivo* was examined using a ChIP assay with a 3× FLAG tag fused to BmLARK and an anti-FLAG tag antibody. The cells were transfected with the recombinant plasmid BmLARK-3× FLAG or a control vector and then collected for a ChIP assay 48 h after transfection. BmLARK-3× FLAG expression was confirmed by Western blot (Figure 4A). The anti-FLAG antibodies, but not control IgG, precipitated and enriched the G4 fragment from *BmPOUM2, 142, 159* and *558* promoters in the cells transfected with the BmLARK-3× FLAG-expressing plasmid (Figure 4B–E). The enriched G4 sequences amplified by PCR were confirmed by DNA sequencing (Figure 4F–I). These experiments demonstrate

that BmLARK can bind not only the *BmPOUM2* promoter G4 structure *ex vivo* but also G4s in the promoters of other *B. mori* genes. Additionally, the non-G4 containing DNA fragments were also examined and the percentages of the non-G4 sequences were 0.4% over the input (Supplementary Figure S5). The percentages of these non-G4 DNA are much lower than those of the G4 sequences immunoprecipitated by LARK in Figure 4. These results further demonstrated that LARK is a specific G4 binding protein.

Function of BmLARK-G4 binding in transcriptional regulation

Subsequently, whether the binding of BmLARK to the G4 structures affects gene transcription was determined. To address this question, BmLARK overexpression and RNA interference (RNAi) were conducted in cell lines. The qRT-PCR and Western blot results showed that when BmLARK was up-regulated by overexpression (Figure 5A), the transcription levels of *BmPOUM2* and genes *142, 159* and *558* were also up-regulated (Figure 5B–E). In contrast, when BmLARK was down-regulated by RNAi (Figure 5F), the transcription levels of *BmPOUM2* and genes *142, 159* and *558* were consequently down-regulated (Figure 5G–J). When the cells were treated with the inhibitor PDS, a decrease in the transcription levels of *BmPOUM2* and genes
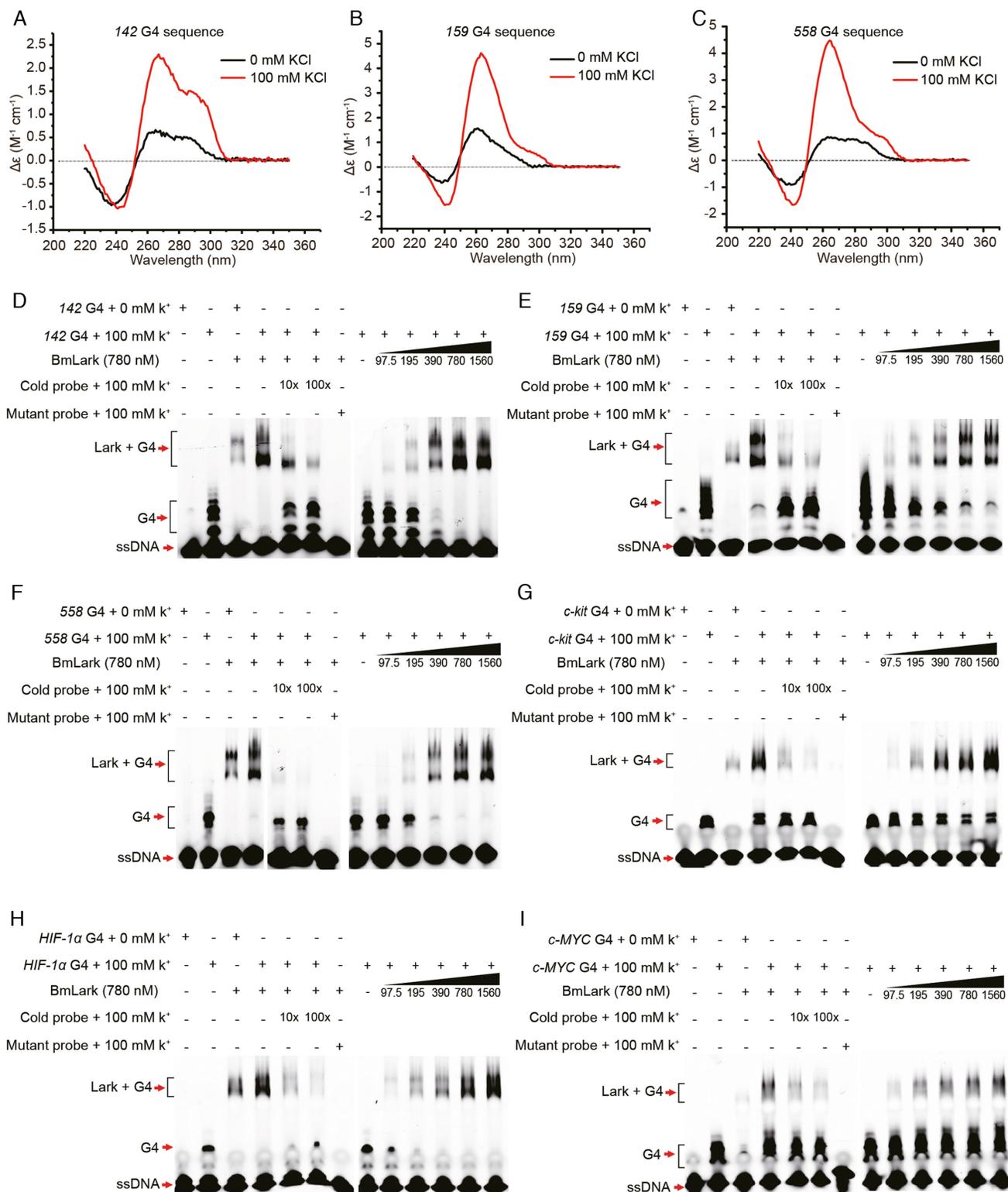
Figure 2. Circular dichroism (CD) analyses of BmLARK binding with other G4 structures in genes of *B. mori* and humans. (**A–C**) CD analyses of G4 structures in the promoters of genes *142*, *159* and *558*. (**D–F**) Binding of BmLARK to G4 structures in the promoters of genes *142*, *159* and *558*. (**G–I**) Binding of BmLARK to well-defined G4 structures in human *c-kit, HIF-1a* and *c-MYC* genes.
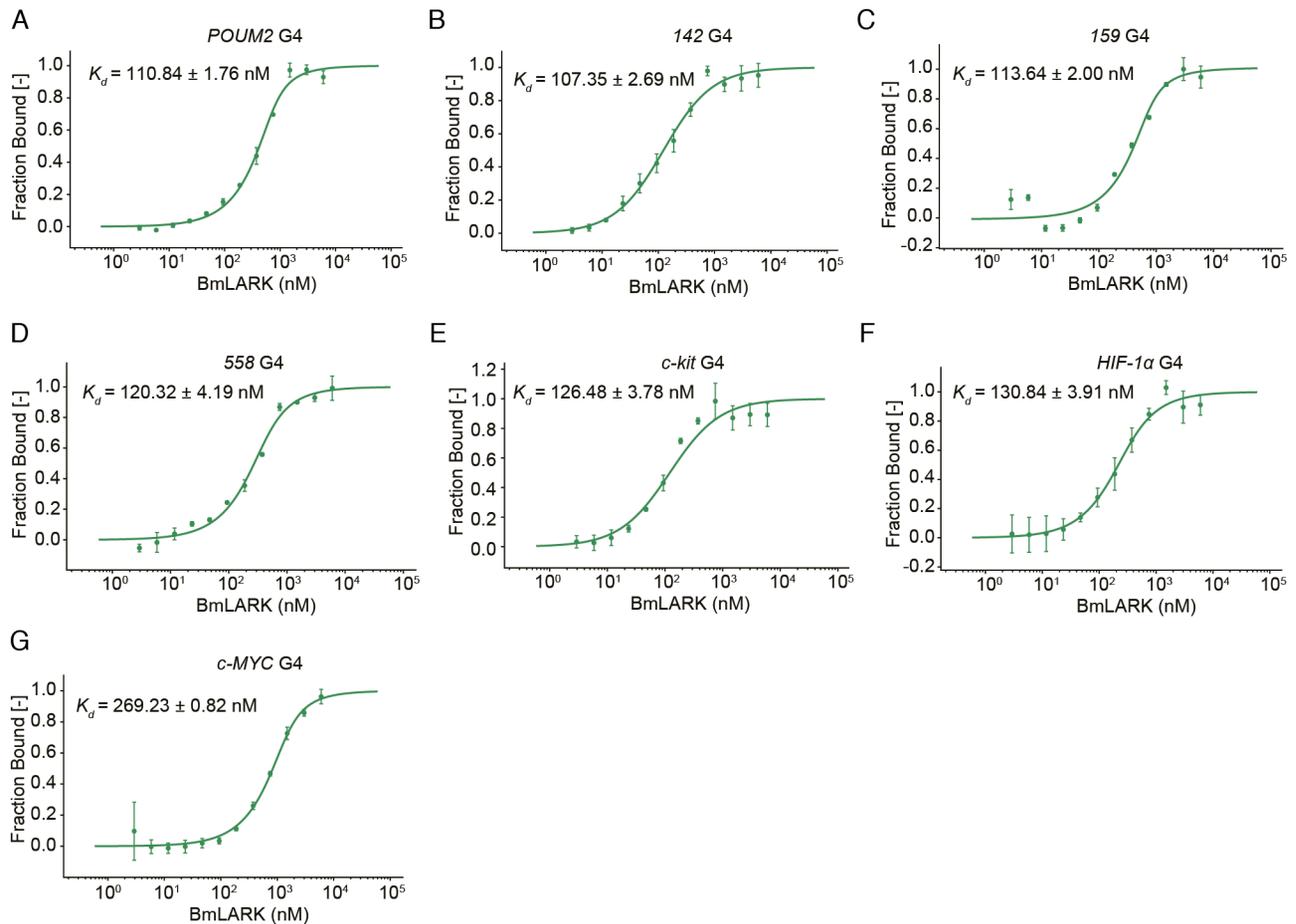
Figure 3. MST measurements of the binding of the G4 structures with BmLARK at varying concentrations. (**A–G**) The quantitative analyses of the binding of BmLARK to the different G4s in the promoters of *BmPOUM2*, *142*, *159*, *558*, *c-kit*, *HIF-1a* and *c-MYC*, respectively, showing the binding curves and K_d values. Error bars represent standard deviation with $n = 3$.

142, *159* and *558* was detected (Figure 5K-N). These results indicate that changes in the BmLARK levels affect the transcription of target genes through its binding the G4 structures in the promoters of these genes.

BmLARK homologues bind G4 structures

To detect whether BmLARK homologues from other organisms also bind G4 structures, homologous LARK cDNAs from *D. melanogaster* (GenBank accession number: NM_079233.6), *M. musculus* (GenBank accession number: NM_001290122.1) and *H. sapiens* (GenBank accession number: NM_002896.3) were cloned, and recombinant proteins (DmLARK, MmLARK and HsLARK) were generated. The EMSA results showed that DmLARK (Figure 6A), MmLARK (Figure 6B) and HsLARK (Figure 6C) bound the G4 structure in the *BmPOMU2* promoter. Additionally, MmLARK (Figure 6D and F) and HsLARK (Figure 6E and G) bound to the well-studied mammalian G4 structures in c-*kit* (23) and *c-MYC* (38). These results reveal that LARK is a structurally and functionally conserved G4-binding protein from arthropods to vertebrates.

BmLARK binding domains that involved in the interaction with G4 structure

BmLARK has three main domains: two RRM motifs and a Zn finger structure. To examine which domain or region of the BmLARK protein is involved in the binding with the G4 structure of the *BmPOUM2* promoter, EMSA was conducted with the truncated RRM1, RRM2, RRM1-2 and ZnF proteins (Figure 7A) and the G4 structure of the *Bm-POUM2* promoter. The results revealed that when both RRM motifs were present, the interaction between Bm-LARK and the G4 structure occurred, which could be competed off by the unlabeled cold probe (Figure 7B). While RRM1, RRM2 or zinc finger region alone could not bind with the G4 structure (Figure 7C–E). These results suggest that the simultaneous presence of RRM1 and RRM2 is required for the binding with the G4 structure.

Effect of BmLARK on the formation of the G4 structure

To investigate whether or not BmLARK changes the G4 structure of the *BmPOMU2* promoter when it binds the structure, CD experiment was conducted with BmLARK incubated with the G4 or ssDNA (non-heat treated G4 DNA) sequence. The results showed that incubation of Bm-
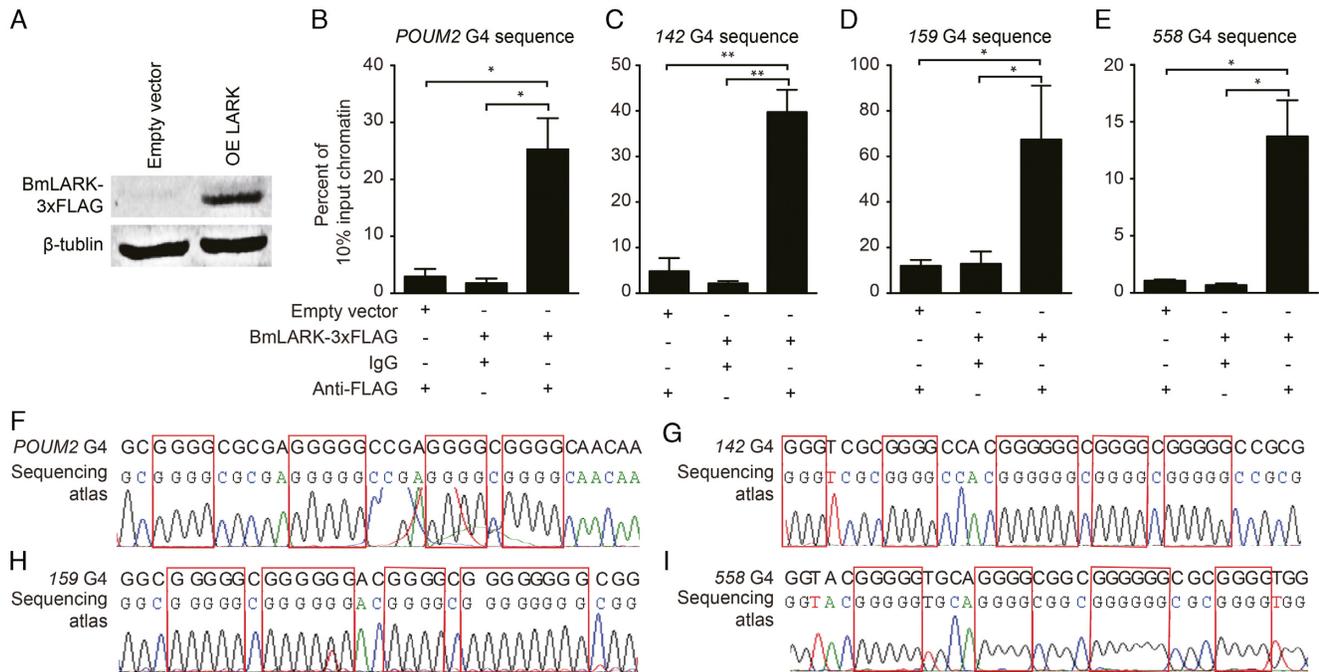
Figure 4. Chromatin immunoprecipitation (ChIP) analysis of the BmLARK binding with G4 structures in *Bm*12 cells. (**A**) Western blot analysis confirming BmLARK-3 × FLAG expression in *Bm*12 cells. OE LARK: BmLARK overexpression. (**B–E**) qRT-PCR detecting the enrichment of the G4 sequences from the *BmPOUM2* and gene *142*, *159* and *558* promoters. Data represent the mean±SEM ($n = 3$). **$P < 0.01$ (Student's *t*-test). (**F–I**) G4 sequences and the sequencing atlas of the enriched reverse transcription PCR products of the *BmPOUM2* and genes *142*, *159* and *558* promoters. Data are the mean ± SEM ($n = 3$). * indicates that the differences are significant at $P < 0.05$, and ** indicates that the differences are significant at $P < 0.01$ (Student's *t*-test).

LARK with the folded G4 did not result in any change in the CD spectra of the G4 structures, suggesting that Bm-LARK could bind to, but did not destruct or destabilize the structure (Figure 7F). However, the absorption peak at 265 nm was significantly changed when BmLARK was incubated with ssDNA G4 (non-heat treated G4 DNA) sequence (Figure 7G), but not with the mutant G4 sequence (Figure 7H), implying that BmLARK could enhance the formation of G4 structure from ssDNA, instead unfold the structure.

Visualization of G4 structures *in vivo*

Many computational analyses and *in vitro* experimental data have predicted that G4 structures widely exist in eukaryote (42), plant (43) and mammalian (44) genomes. To visualize the G4 structures in *Bm*12 and Chinese hamster ovary (CHO) cells, BmLARK-3× FLAG protein and an anti-FLAG antibody were used in an immunohistochemistry assay. The binding of BmLARK to the G4 structures in the cell nuclei was visualized by fluorescence signals generated using a fluorescence-labelled secondary antibody (Figure 8A). Punctate nuclear staining was detected in *Bm*12 (Figure 8B and C) and CHO cells (Figure 8D and E) in the presence, but not the absence, of BmLARK. The mammalian CHO cells appeared to show more G4 structures than the invertebrate *Bm*12 cells. When the cells were treated with PDS, which blocks protein binding to G4 structures, an obvious decrease in nuclear staining was observed (Figure 8B,vii–ix). The *in vivo* presence of G4 structures was also detected in the chromosomes of testis cells from *B. mori* (Fig-

ure 8F). These results suggest that G4 structures can form *in vivo* in the nuclei and chromosomes of cells and tissues.

DISCUSSION

Many computational analyses and *in vitro* experimental data indicate that G4 structures widely exist in the genomes of many organisms (19,42–45), but direct and *in vivo* evidence is generally lacking. Furthermore, the identification and characterization of the nuclear proteins that interact with G4 structures is extremely helpful for better understanding the biological functions of G4 structures in gene regulation, DNA replication and mRNA translation. In this study, several important findings are germane. First, Bm-LARK was identified to be a novel G4-binding protein capable of binding multiple gene G4 structures; similarly, the LARK protein in several species ranging from invertebrates to humans was able to bind G4s from *B. mori* and *H. sapiens*. Second, the G4 structures and LARK binding were visualized in invertebrate and vertebrate cells and/or organ.

In an effort to identify G4-binding proteins, BmLARK was found to be able to bind G4s in the promoters of *BmPOUM2* and other genes in *B. mori* and *H. sapiens* (Figures 1–4). Additionally, LARK proteins from different species, including *B. mori, D. melanogaster, M. musculus* and *H. sapiens,* could also bind G4 structures in different genes from *B. mori* and *H. sapiens* (Figure 6). These results suggest that LARK is a novel and conserved G4-binding protein with a broad spectrum. Investigating how LARK binds different forms of G4 structures and which amino acid residues in the protein and basic G4 structures or groups are required
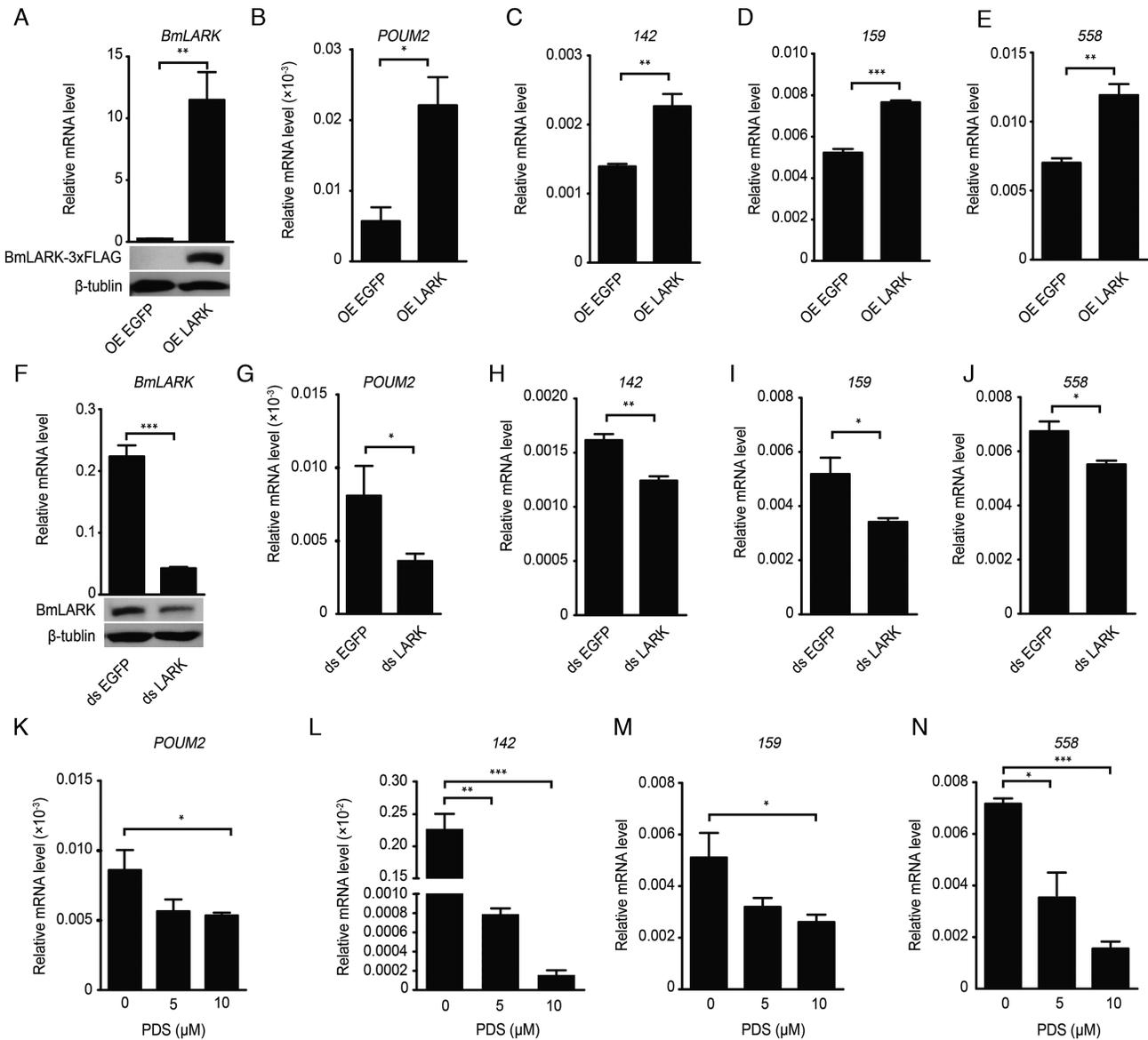
Figure 5. Function of BmLARK binding to G4 structures in the transcriptional regulation of genes. (**A**) Confirming BmLARK overexpression by qRT-PCR (top panel) and Western blot (bottom panel) in *Bm*12 cells. (**B–E**) mRNA expression of *BmPOUM2* and genes *142*, *159* and *558* after BmLARK overexpression in *Bm*12 cells. (**F**) Confirming the efficiency of BmLARK RNAi by qRT-PCR (top panel) and Western blot (bottom panel) in *Bm*12 cells. (**G-J**) mRNA expression of *BmPOUM2* and genes *142*, *159* and *558* after applying BmLARK dsRNA in *Bm*12 cells. (**K–N**) mRNA expression of *BmPOUM2* and genes *142*, *159* and *558* after treatment with PDS in *Bm*12 cells. Data are the mean±SEM ($n = 3$). * indicates that the differences are significant at $P < 0.05$; ** indicates that the differences are significant at $P < 0.01$; and *** indicates that the differences are significant at $P < 0.001$ (Student's *t*-test). OE LARK: BmLARK overexpression; OE EGFP: EGFP overexpression.

for the binding could be interesting. LARK contains two RRM domains (locations: 8–73 and 87–152) and a CCHC-type zinc finger (location: 169–185). Our results indicate that the two RRM domains were required for the binding (Figure 7A–E). The number of G-tetrads and the length of loops that link the G-tetrads are various in different G4 structures, however, the G-tetrad is unique in different G4s. LARK probably can recognize the common G-tetrad in different G4 structures through its RRM domains. This seems reflected by the K_d values, which almost are in the same order of magnitude, although they are somehow different (Figure 3). This hypothesis needs to be further investigated.

LARK is an RRM and CCHC-type zinc finger-containing protein that is evolutionarily conserved from fruit fly to humans (46) (Supplementary Figure S6). LARK is a component of the circadian pacemaker output pathway (47) and was initially found to be involved in rhythmic behaviours in *Drosophila* (48) by regulating the circadian eclosion rhythm via interactions with distinct RNA targets, such as E74. Mutated LARK causes rhythm-specific circadian defects (49). The *Drosophila* DBT, which is required for circadian clock function, is directly regulated by LARK (50). LARK also interacts with dFMRP to regulate eye development and circadian behaviours (51). In addition, LARK is criti-
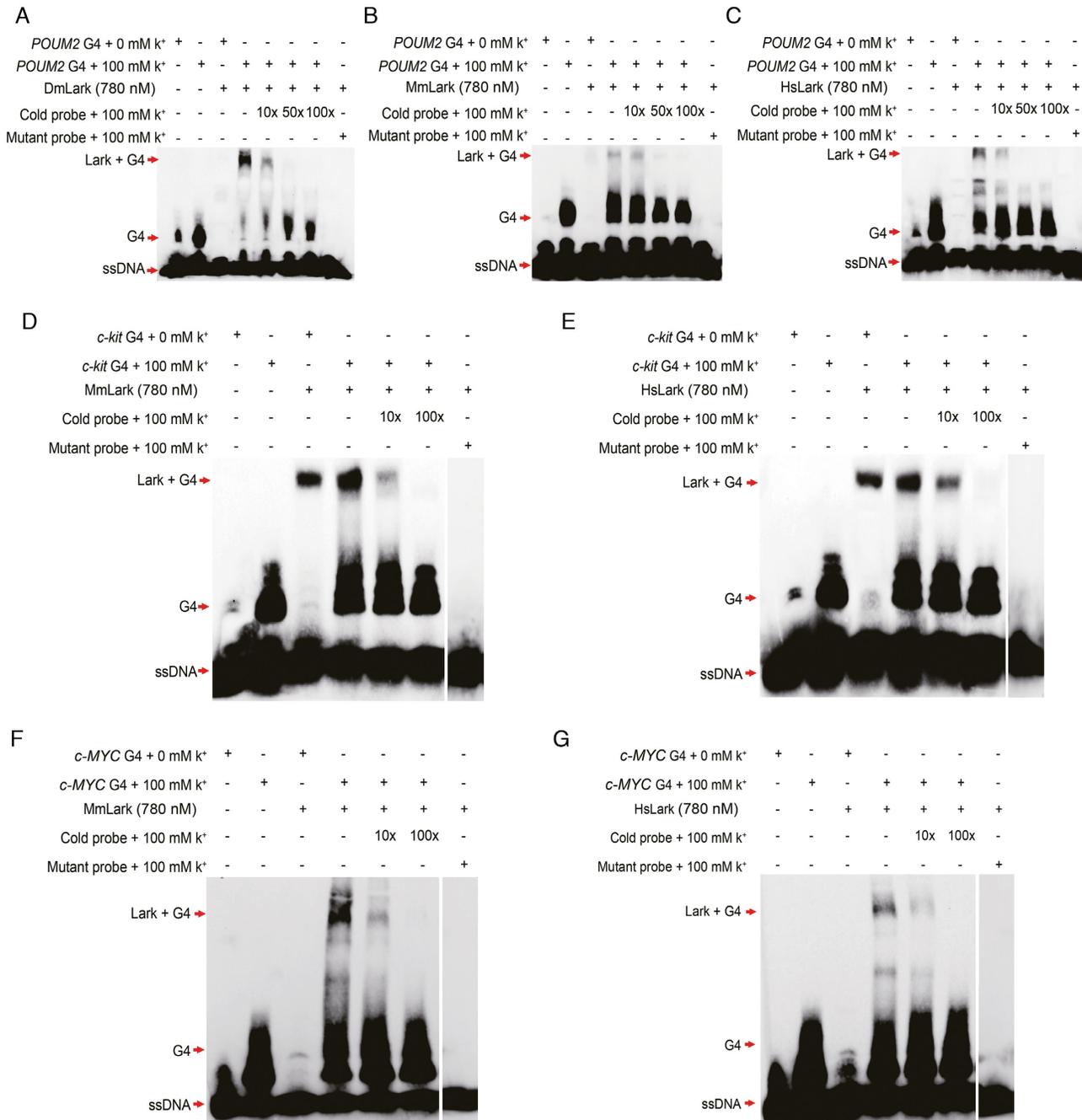
Figure 6. Analyses of the binding of various LARK proteins to G4 structures by EMSA. (**A–C**) Binding of *D. melanogaster* DmLARK, *M. musculus* MmLARK and *H. sapiens* HsLARK to the *BmPOUM2* G4 structure. (**D** and **F**) Binding of MmLARK to the G4 structures of human genes *c-kit* and *c-MYC*. (**E** and **G**) Binding of HsLARK to the G4 structures of human genes *c-kit* and *c-MYC*.

cal for neuronal development and physiology (52). Reports regarding the function of LARK in *B. mori* are lacking. Bm-LARK exhibits a ubiquitous and constitutive expression pattern throughout the developmental stages in many different tissues in *B. mori* (53), which is similar to the expression pattern of LARK in *Drosophila* and humans (46,49). The constitutive expression of LARK suggests that it plays an important role in developmental regulation in diverse organisms. In consideration of the highly conservation of Bm-LARK with DmLARK (Supplementary Figure S6), Bm-

LARK may have similar biological function in *B. mori*. Nucleolin is a reported G4-binding protein (26) and is involved in ribosome biogenesis (54), chromatin remodeling (55,56), transcription (57–61) and apoptosis (62). BmLARK is a novel G4-binding protein and contains two RRM domains similar with nucleolin (Supplementary Figure S2). Thus, like nucleolin, BmLARK may participate in similar biological process. This study suggests a novel function of LARK targeting at G4 structure. The RNA G4 binding activity of BmLARK was also tested in this study and BmLARK
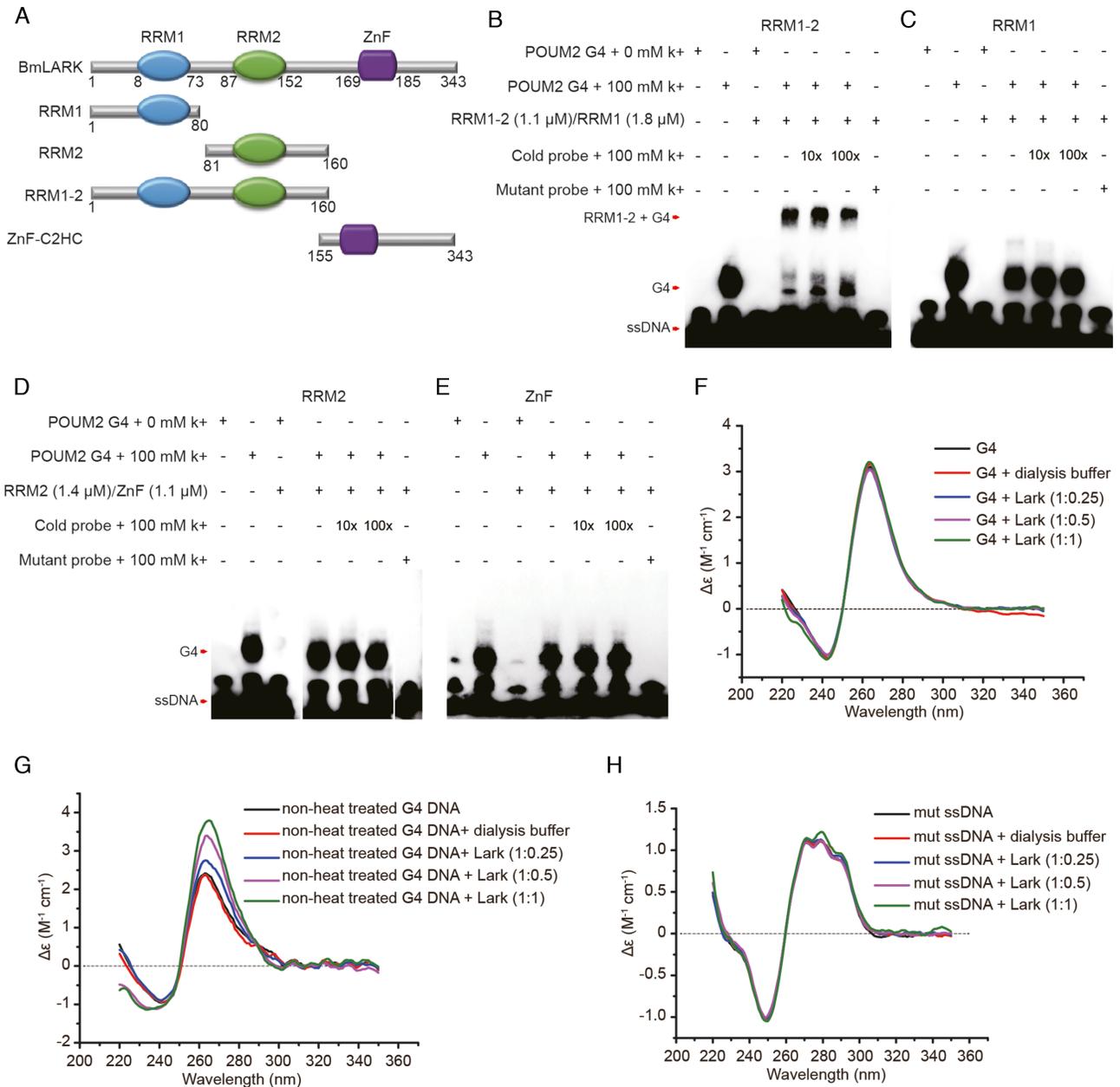
Figure 7. Binding regions identification and effect of BmLARK on the G4 structure of *BmPOUM2* promoter. (**A**) Sequences of BmLARK with three different domains and the constructs containing RRM1, RRM2, RRM1-2 and ZnF. (**B–E**) The binding of the RRM1-2, RRM1, RRM2 and ZnF fragments with the *BmPOUM2* G4 structure. (**F**) CD analysis of the effect of BmLARK on *BmPOUM2* G4 structure unwinding. (**G**) CD analysis of the effect of BmLARK on *BmPOUM2* G4 structure formation. (**H**) CD analysis of the effect of BmLARK on mutated *BmPOUM2* G4 sequence.

bound two well-defined RNA G4s, *BCL2* RNA G4 (63) and *ERS1* RNA G4 (64) (Supplementary Figure S7). This result implies that LARK may not only participate in the gene expression regulation at the transcription level but also in translational regulation or RNA alternative splicing post transcription. This should be further studied.

What is the function of BmLARK binding to gene G4 structures? The overexpression and RNAi experiments indicated that the binding of BmLARK to G4 structures enhanced the transcription of target genes (Figure 5). Our findings suggest that the interaction of LARK and G4 structure may have developed into an elaborate epigenetic

mechanism of gene transcription regulation during evolution. In our previous study, we found that BmILF bound the i-motif at the same position as the G4 structure on the reverse strand of the *BmPOUM2* promoter and enhanced gene transcription (14). Knowing how these two proteins, i.e., BmLARK and BmILF, synergistically and cooperatively interact with G4 and i-motif structures, respectively, to regulate *BmPOUM2* expression could be interesting.

There has been a long-standing debate regarding whether i-motif and G4 DNA structures exist inside cells. To date, researchers have observed G4 structures in human cells using engineered antibodies against G4 structures (65). The
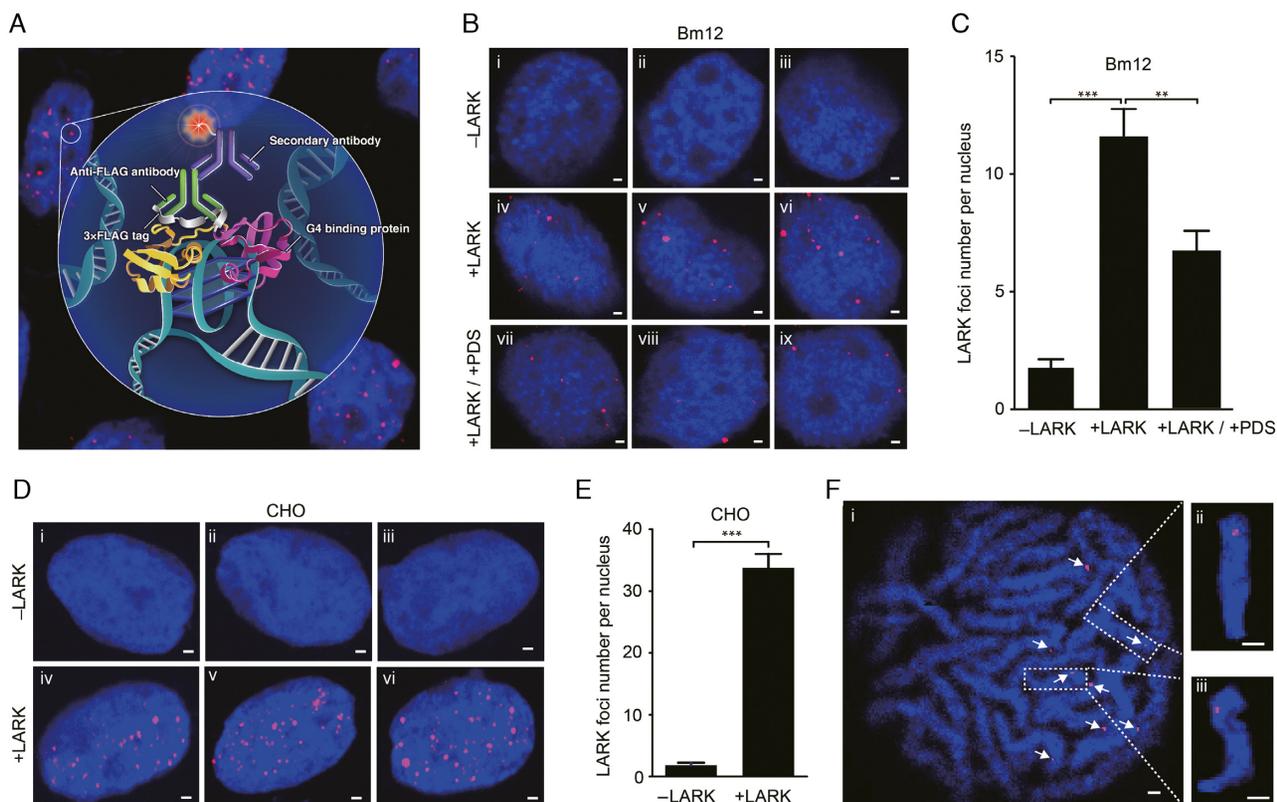
Figure 8. Visualization of G4 structures in cell nuclei and chromatin. (**A**) Diagram of G4 structure visualization method. Fixed cells were incubated with BmLARK-3 × FLAG protein, followed by incubation with an anti-FLAG primary antibody. The cells were then incubated with an anti-rabbit Alexa 594-conjugated secondary antibody, which shows red fluorescence signal. (**B** and **D**) Immunofluorescence showing G4 structures (red stains) in *Bm*12 and CHO cell nuclei, respectively. (**C** and **E**) Graph showing the quantification of the number of G4 structures per nucleus in *Bm*12 and CHO cells, respectively. Sixty nuclei from a set of 3 replicates were counted in each of conditions. (**F**) Immunofluorescence showing G4 structures (red stains) in chromatin from *B. mori* testis cells. All scale bars, 1 μm.

i-motif structure has also been recently visualized *ex vivo* using an engineered antibody against i-motif structures (66). Here, we demonstrate the existence of G4 structures *ex vivo* and *in vivo* by visualizing LARK-G4 interactions in both invertebrate and vertebrate cells and/or tissues (Figure 8). This observation demonstrates the presence of G4 structures in the nuclei and chromosomes of cells in invertebrates (*B. mori* cells and testis) for the first time and confirms the presence of G4 structures in vertebrate cells (CHO cells). These results provide solid evidence that G4 and i-motif DNA secondary structures form in the genomes of living cells, thus guaranteeing their possible *in vivo* functions.

In summary, this study strongly indicates that LARK is a novel and conserved G4-binding protein in *B. mori* and other organisms and is involved in the regulation of gene transcription. Using LARK, G4 structures were *ex vivo* and *in vivo* detected in invertebrate and vertebrate living cells or organs. LARK proteins and their interactions with G4 structures may provide novel targets and approaches for regulating gene expression, controlling cell function and pests, and applying gene therapy for disease prevention and control without transgenic manipulation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Watson,J.D. and Crick,F.H. (1953) Molecular Structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
2. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
3. Guéron,M. and Leroy,J.L. (2000) The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.*, **10**, 326–331.
4. Dingley,A.J., Peterson,R.D., Grzesiek,S. and Feigon,J. (2005) Characterization of the cation and temperature dependence of DNA

quadruplex hydrogen bond properties using high-resolution NMR. *J. Am. Chem. Soc.*, **127**, 14466–14472.

5. Gehring,K., Leroy,J.L. and Guéron,M.A. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561–565.

6. Rajendran,A., Nakano,S. and Sugimoto,N. (2010) Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun. (Camb.)*, **46**, 1299–1301.

7. Takahashi,S. and Sugimoto,N. (2015) Pressure-dependent formation of i-motif and G-quadruplex DNA structures. *Phys. Chem. Chem. Phys.*, **17**, 31004–31010.

8. Brooks,T.A., Kendrick,S. and Hurley,L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.

9. Besnard,E., Babled,A., Lapasset,L., Milhavet,O., Parrinello,H., Dantec,C., Marin,J.M. and Lemaitre,J.M. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.*, **19**, 837–844.

10. Lin,W.Q., Sampathi,S., Dai,H., Liu,C., Zhou,M., Hu,J., Huang,Q., Campbell,J., Shin-Ya,K., Zheng,L. *et al.* (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J.*, **32**, 1425–1439.

11. Bugaut,A. and Balasubramanian,S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.

12. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.

13. Moruno-Manchon,J.F., Koellhoffer,E.C., Gopakumar,J., Hambarde,S., Kim,N., McCullough,L.D. and Tsvetkov,A.S. (2017) The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and down regulates transcription of BRCA1 in neurons. *Aging (Albany, NY)*, **9**, 1957–1970.

14. Niu,K.K., Zhang,X.J., Deng,H.M., Wu,F., Ren,Y.D., Xiang,H., Zheng,S.C., Liu,L., Huang,L.H., Zeng,B.J. *et al.* (2018) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyxmori*. *Nucleic Acids Res.*, **46**, 1710–1723.

15. Kendrick,S., Kang,H.J., Alam,M.P., Madathil,M.M., Agrawal,P., Gokhale,V., Yang,D., Hecht,S.M. and Hurley,L.H. (2014) The dynamic character of the BCL2 promoter i-Motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.*, **136**, 4161–4171.

16. Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.

17. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.

18. Comoglio,F., Schlumpf,T., Schmid,V., Rohs,R., Beisel,C. and Paro,R. (2015) High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep.*, **11**, 821–834.

19. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.

20. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.

21. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.

22. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.

23. Fernando,H., Reszka,A.P., Huppert,J., Ladame,S., Rankin,S., Venkitaraman,A.R., Neidle,S. and Balasubramanian,S. (2006) A

24. David,A.P., Margarit,E., Domizi,P., Banchio,C., Armas,P. and Calcaterra,N.B. (2016) G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res.*, **44**, 4163–4173.

25. Soldatenkov,V.A., Vetcher,A.A., Duka,T. and Ladame,S. (2008) First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem. Biol.*, **3**, 214–219.

26. González,V., Guo,K., Hurley,L. and Sun,D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.

27. Kanoh,Y., Matsumoto,S., Fukatsu,R., Kakusho,N., Kono,N., Renard-Guillet,C., Masuda,K., Iida,K., Nagasawa,K., Shirahige,K. *et al.* (2015) Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.*, **22**, 889–897.

28. Gao,J., Zybailov,B.L., Byrd,A.K., Griffin,W.C., Chib,S., Mackintosh,S.G., Tackett,A.J. and Raney,K.D. (2015) Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA. *Chem. Commun. (Camb.)*, **51**, 7242–7244.

29. Qiu,J., Liu,J., Chen,S., Ou,T.M., Tan,J.H., Gu,L.Q., Huang,Z.S. and Li,D. (2015) Role of hairpin-quadruplex DNA secondary structural conversion in the promoter of hnRNP K in gene transcriptional regulation. *Org. Lett.*, **17**, 4584–4587.

30. Williams,P., Li,L., Dong,X. and Wang,Y. (2017) Identification of SLIRP as a G quadruplex-binding protein. *J. Am. Chem. Soc.*, **139**, 12426–12429.

31. Paramasivam,M., Membrino,A., Cogoi,S., Fukuda,H., Nakagama,H. and Xodo,L.E. (2009) Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.*, **37**, 2841–2853.

32. Wang,F., Tang,M.L., Zeng,Z.X., Wu,R.Y., Xue,Y., Hao,Y.H., Pang,D.W., Zhao,Y. and Tan,Z. (2012) Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 20413–20418.

33. Khurad,A.M., Zhang,M.J., Deshmukh,C.G., Bahekar,R.S., Tiple,A.D. and Zhang,C.X. (2009) A new continuous cell line from larval ovaries of silkworm, *Bombyxmori*. *In Vitro Cell Dev. Biol. Anim.*, **45**, 414–419.

34. Deng,H.M., Zheng,S.C., Yang,X.H., Liu,L. and Feng,Q.L. (2011) Transcription factors BmPOUM2 and BmbFTZ-F1 are involved in regulation of the expression of the wing cuticle protein gene BmWCP4 in the silkworm, *Bombyxmori*. *Insect Mol. Biol.*, **20**, 45–60.

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

36. Rodriguez,R., Müller,S., Yeoman,J.A., Trentesaux,C., Riou,J.F. and Balasubramanian,S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Che. Soc.*, **130**, 15758–15759.

37. Koirala,D., Dhakal,S., Ashbridge,B., Sannohe,Y., Rodriguez,R., Sugiyama,H., Balasubramanian,S. and Mao,H. (2011) A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nat. Chem.*, **3**, 782–787.

38. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-myc promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.

39. DeArmond,R., Wood,S., Sun,D., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, **44**, 16341–16350.

40. Zhang,W., Duhr,S., Baaske,P. and Laue,E. (2014) Microscale thermophoresis for the assessment of nuclear protein-binding affinities. *Methods Mol. Biol.*, **1094**, 269–276.

41. Kazemier,H.G., Paeschke,K. and Lansdorp,P.M. (2017) Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. *Nucleic Acids Res.*, **45**, 5913–5919.

42. Capra,J.A., Paeschke,K., Singh,M. and Zakian,V.A. (2010) G-Quadruplex DNA sequences are evolutionarily conserved and

associated with distinct genomic features in Saccharomyces cerevisiae. *PLOS Comput. Biol.*, **6**, e1000861.

43. Garg,R., Aggarwal,J. and Thakkar,B. (2016) Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. *Sci. Rep.*, **6**, 28211.

44. Zhao,Y., Du,Z. and Li,N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett.*, **581**, 1951–1956.

45. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

46. Jackson,F.R., Banfi,S., Guffanti,A. and Rossi,E. (1997) A novel zinc finger-containing RNA-binding protein conserved from fruitflies to humans. *Genomics*, **41**, 444–452.

47. Schroeder,A.J., Genova,G.K., Roberts,M.A., Kleyner,Y., Suh,J. and Jackson,F.R. (2003) Cell- specific expression of the lark RNA-binding protein in Drosophila results in morphological and circadian behavioral phenotypes. *J. Neurogenet.*, **17**, 139–169.

48. Newby,L.M. and Jackson,F.R. (1993) A new biological rhythm mutant of Drosophila melanogaster that identifies a gene with an essential embryonic function. *Genetics*, **135**, 1077–1090.

49. Newby,L.M. and Jackson,F.R. (1996) Regulation of a specific circadian clock output pathway by lark, a putative RNA-binding protein with repressor activity. *J. Neurobiol.*, **31**, 117–128.

50. Huang,Y., McNeil,G.P. and Jackson,F.R. (2014) Translational regulation of the DOUBLETIME/CKIδ/ε kinase by LARK contributes to circadian period modulation. *PLOS Genet.*, **10**, e1004536.

51. Sofola,O., Sundram,V., Ng,F., Kleyner,Y., Morales,J., Botas,J., Jackson,F.R. and Nelson,D.L. (2008) The Drosophila FMRP and LARK RNA-binding proteins function together to regulate eye development and circadian behavior. *J. Neurosci.*, **28**, 10200–10205.

52. Huang,Y., Howlett,E., Stern,M. and Jackson,F.R. (2009) Altered LARK expression perturbs development and physiology of the Drosophila PDF clock neurons. *Mol. Cell. Neurosci.*, **41**, 196–205.

53. Wang,Z.L., Li,J., Xia,Q.Y., Zhao,P., Duan,J., Zha,X.F. and Xiang,Z.H. (2005) Identification and expression pattern of Bmlark, a homolog of the Drosophila gene lark in Bombyx mori. *DNA Seq.*, **16**, 224–229.

54. Ginisty,H., Sicard,H., Roger,B. and Bouvet,P. (1999) Structure and functions of nucleolin. *J. Cell Sci.*, **112**, 761–772.

55. Angelov,D., Bondarenko,V.A., Almagro,S., Menoni,H., Mongélard,F., Hans,F., Mietton,F., Studitsky,V.M., Hamiche,A., Dimitrov,S. *et al.* (2006) Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. *EMBO J.*, **25**, 1669–1679.

56. Mongelard,F. and Bouvet,P. (2007) Nucleolin: a multiFACeTed protein. *Trends Cell Biol.*, **17**, 80–86.

57. Yang,T.H., Tsai,W.H., Lee,Y.M., Lei,H.Y., Lai,M.Y., Chen,D.S., Yeh,N.H. and Lee,S.C. (1994) Purification and characterization of nucleolin and its identification as a transcription repressor. *Mol. Cell Biol.*, **14**, 6068–6074.

58. Grinstein,E., Wernet,P., Snijders,P.J., Rösl,F., Weinert,I., Jia,W., Kraft,R., Schewe,C., Schwabe,M., Hauptmann,S. *et al.* (2002) Nucleolin as activator of human papillomavirus type 18 oncogene transcription in cervical cancer. *J. Exp. Med.*, **196**, 1067–1078.

59. Grinstein,E., Du,Y., Santourlidis,S., Christ,J., Uhrberg,M. and Wernet,P. (2007) Nucleolin regulates gene expression in CD34-positive hematopoietic cells. *J. Biol. Chem.*, **282**, 12439–12449.

60. Brys,A. and Maizels,N. (1994) LR1 regulates c-myc transcription in B-cell lymphomas. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 4915–4919.

61. Dempsey,L.A., Sun,H., Hanakahi,L.A. and Maizels,N. (1999) G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.*, **274**, 1066–1071.

62. He,T.C., Sparks,A.B., Rago,C., Hermeking,H., Zawel,L., da Costa,L.T., Morin,P.J., Vogelstein,B. and Kinzler,K.W. (1998) Identification of c-MYC as a target of the APC pathway. *Science*, **281**, 1509–1512.

63. Shahid,R., Bugaut,A. and Balasubramanian,S. (2010) The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, **49**, 8300–8306.

64. Balkwill,G.D., Derecka,K., Garner,T.P., Hodgman,C., Flint,A.P. and Searle,M.S. (2009) Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry*, **48**, 11487–11495.

65. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

66. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.