

SCIENTIFIC REPORTS



OPEN

The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes

Matteo Scalabrin*, Ilaria Frasson*, Emanuela Ruggiero*, Rosalba Perrone, Elena Tosoni, Sara Lago, Martina Tassinari, Giorgio Palù & Sara N. Richter

Received: 04 January 2017

Accepted: 21 February 2017

Published: 24 March 2017

G-quadruplexes are four-stranded conformations of nucleic acids that act as cellular epigenetic regulators. A dynamic G-quadruplex forming region in the HIV-1 LTR promoter represses HIV-1 transcription when in the folded conformation. This activity is enhanced by nucleolin, which induces and stabilizes the HIV-1 LTR G-quadruplexes. In this work by a combined pull-down/mass spectrometry approach, we consistently found hnRNP A2/B1 as an additional LTR-G-quadruplex interacting protein. Surface plasmon resonance confirmed G-quadruplex specificity over linear sequences and fluorescence resonance energy transfer analysis indicated that hnRNP A2/B1 is able to efficiently unfold the LTR G-quadruplexes. Evaluation of the thermal stability of the LTR G-quadruplexes in different-length oligonucleotides showed that the protein is fit to be most active in the LTR full-length environment. When hnRNP A2/B1 was silenced in cells, LTR activity decreased, indicating that the protein acts as a HIV-1 transcription activator. Our data highlight a tightly regulated control of transcription based on G-quadruplex folding/unfolding, which depends on interacting cellular proteins. These findings provide a deeper understanding of the viral transcription mechanism and may pave the way to the development of drugs effective against the integrated HIV-1, present both in actively and latently infected cells.

G-quadruplexes (G4s) are unique four-stranded nucleic acid structures that may form in guanine-rich sequences. Based on the strand orientation, they can adopt three main topologies: parallel, antiparallel, and hybrid-type structures. G4s have been shown to be involved in key regulatory and pathological roles in eukaryotes, including transcriptional regulation of gene promoters and enhancers, translation, chromatin epigenetic regulation, DNA recombination^{1–3}. Formation of G4 *in vivo* has been consolidated by the development of G4 specific antibodies^{4,5}.

Recently, the presence of G4s in viruses and their involvement in key steps of viral infection has been provided⁶. G4s have been reported in the SARS coronavirus⁷, the human papilloma, hepatitis C, Zika and Ebola virus genomes^{8–11}. Among herpesviruses, RNA G4s have been implicated in the regulation of DNA replication and translation of the Epstein–Barr virus^{12,13}. We have shown that the herpes simplex virus 1 possesses several repeats of sequences forming stable G4s, which were visualized in infected cells by a G4-specific antibody¹⁴; stabilization of these tetraplex structures by a G4 ligand inhibited viral DNA replication¹⁵.

The largest body of evidence of G4-mediated regulation of viruses has been provided for the human immunodeficiency virus-1 (HIV-1), the etiologic agent of the acquired immune deficiency syndrome (AIDS). We and other groups have identified functionally significant G4s in the Nef coding region¹⁶ and in the unique long terminal repeat (LTR) promoter^{17–19} of HIV-1. When the HIV-1 G4s were stabilized by G4 ligands, antiviral effects, mainly dependent on inhibition of LTR-mediated transcription, were observed^{18,20,21}. Furthermore, the cellular protein nucleolin has been shown to stabilize the HIV-1 LTR G4s and induce potent inhibition of viral transcription²².

In general, several proteins that modulate G4s and/or serve as a bridge to recruit additional protein regulators have been reported²³. Besides the shelterin complex proteins that are involved in telomere homeostasis²⁴, G4 interacting proteins either stabilize (e.g. nucleolin, MAZ and nucleophosmin) or unfold (the helicase and

Department of Molecular Medicine, University of Padua, Padua, Italy. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.N.R. (email: sara.richter@unipd.it)

Exp#	Experimental monoisotopic mass	Expected monoisotopic mass	Protein match	Peptide match	Score*
1	1921.102	1921.079	Putative zinc finger protein 812	V145-R162	18
1, 2, 3	2188.904 2219.053	2188.898 2219.063	Heterogeneous nuclear ribonucleoprotein A2/B1	N326-R350 D130-R147	27 18
1	2429.146	2429.157	Heterogeneous nuclear ribonucleoprotein D0	I198-K218	31
2	1681.900	1681.907	Heterogeneous nuclear ribonucleoprotein C1/C2	M74-K89	39
2	863.487	863.487	Mediator of RNA polymerase II transcription subunit 11	V97-K103	22
2	1036.698	1036.676	Histone H1.2	A192-K201	20
2	1036.698	1036.687	Transcription factor AP-4	R294-R302	18
3	952.346	952.349	Cysteine/serine-rich nuclear protein 2	D249-R257	16
3	3450.685	3450.699	Non-POU domain-containing octamer-binding protein	G399-R434	28

Table 1. MS analysis of protein content of three independent experiments and collected from SDS

gels. *The score (assigned on probabilistic bases by mascot software, “http://www.matrixscience.com/help/interpretation_help.html#THRESHOLDS”)⁵¹ is the probability that the observed match is not a random event. The score is reported as $-10 \times \log_{10}(P)$ where P is the absolute probability. The mass of the putative peptides and of the 50 most intense fragment ions were used for the data base search. Significant Mascot hits were accepted as positive matches and their ion score reported. The highest score obtained in the three experiments is reported.

mutant LTR-II + III + IV M4 + 5 was bound with a lower affinity (K_D 35.5 ± 3.0 nM). χ^2 and U-values were <10 and <15 , respectively, indicating optimal data fitting^{28,29}. In contrast, binding to a random oligonucleotide of the same length as LTR-II + III + IV was so low that in these conditions it was not possible to obtain a meaningful K_D value (Fig. 2).

hnRNP A2 unfolds the LTR G4s. We next set out to investigate the effect of hnRNP A2 binding to the LTR G4s. A *Taq* polymerase stop assay was performed on the LTR-II + III + IV template. In the presence of 100 mM K^+ , stop sites corresponding to formation of LTR-III and LTR-II were visible (compare lanes 1 and 2, Fig. 3a). No stop corresponding to LTR-IV was detected, as expected, since LTR-IV has been previously reported to form upon induction by G4 ligands¹⁸. Upon addition of hnRNP A2, both LTR-III and LTR-II stop sites showed an insignificant decrease (compare lanes 3 and 2, Fig. 3a and b). Similarly, no effect was induced by addition of K^+ and hnRNP A2 in the control LTR-II + III + IV M4 + 5 template lacking the possibility to form G4 (lanes 5–7, Fig. 3a).

As suggested by Xodo³⁰, the apparent lack of effect in the presence of an unfolding protein could be due to the protein binding to the template sequence which would stimulate polymerase stop and mask G4 release at the same binding site. We thus switched to a dual-labelled system where oligonucleotide folding could be monitored by changes in fluorescence. Fluorescence resonance energy transfer (FRET) is a spectroscopic technique that provides information about structure and dynamics of nucleic acids folding. It involves a donor fluorophore in an excited state, the excitation energy of which can be transferred to a proximal acceptor chromophore. Given that the major determinant of FRET efficiency is the distance between the acceptor and the donor, fluorescence intensities depict nucleic acids folding states. Consequently, when annealed to its complementary sequence to form a double-stranded structure, the tested oligonucleotide would yield the maximum fluorescence intensity, while the G4 folded conformation would be the least fluorescent. In these conditions the measured fluorescence intensity allows to calculate the energy transfer (E) and the end-to-end distance (R) between the two fluorophores, and therefore the unfolding degree. The G4 folded LTR-II + III + IV sequence was characterized by R of $\sim 38 \text{ \AA}$, with $E \sim 0.84$. When the G4 structure was converted into the duplex conformation by addition of the complementary C-rich strand, the fluorophores were separated by $\sim 150 \text{ \AA}$, and FRET was approximately 0. Treatment of LTR-II + III + IV G4 with hnRNP A2 increased fluorescence by 3-folds with respect to the free G4, and E decreased to 0.51 (Fig. 4a and b, Supplementary Table S2). Considering that the complete unfolding of the G4 structure required $\Delta E = 0.72$, the $\Delta E = 0.30$ induced by hnRNP A2 reflected a 42% unfolding (Fig. 4c). The negative control protein BSA showed negligible unfolding activity in these conditions (Fig. 4c and Supplementary Table S2). To test the statistical significance of the effect observed in the presence of the protein, we applied a model comparison approach, which allows to test the statistical role of an independent variable (in our case the hnRNP A2 protein) through the comparison of two models differing only in that variable. The test yielded $F(2, 189) = 36.27$ and $P < 0.05$ ³¹, indicating that the difference observed in the presence and absence of the protein was statistically significant. Analysis of the activity of hnRNP A2 was extended to shorter LTR sequences, i.e. LTR-III + IV, LTR-III and LTR-IV G4s, folded in 100 mM K^+ . The unfolding of LTR-III + IV and LTR-IV was similar but lower than that on the full-length sequence (31%); unfolding of LTR-III was very low (8.5%) (Fig. 4c, Supplementary Table S1 and Supplementary Fig. S2–S4) When the G4-forming sequences were folded in lower K^+ concentrations, i.e. 50 and 25 mM, a distinct enhancement of unfolding was observed:

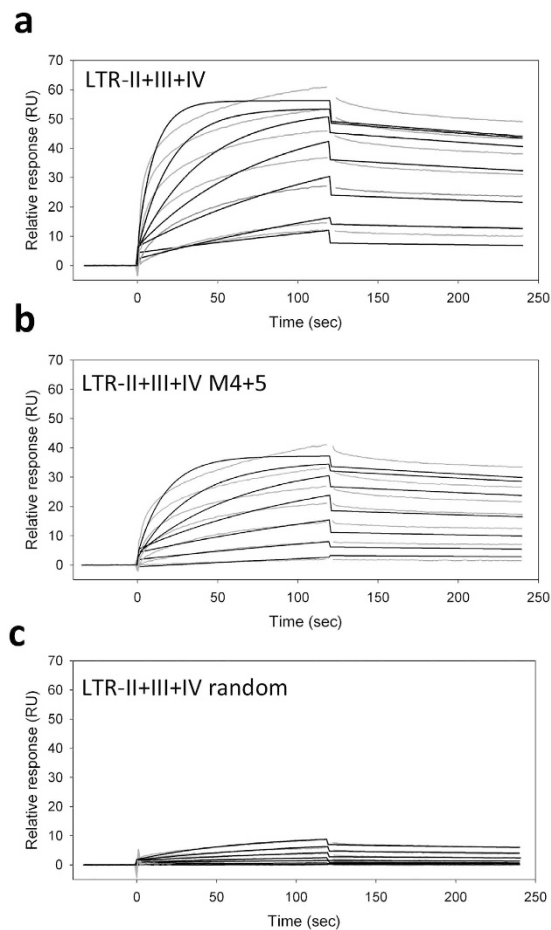


Figure 2. Binding affinity of hnRNP A2 for the LTR-II + III + IV G4 and mutant sequences measured by surface plasmon resonance (SPR) analysis. The human recombinant hnRNP A2 was immobilized on the SPR chip by amine coupling chemistry and the unlabelled oligonucleotides were bound at a flow rate of 25 $\mu\text{l}/\text{min}$, with contact time and dissociation time of 120 s in HEPES-KCl buffer. Sensograms were obtained in the oligonucleotide concentration range of 31.25 nM–2 μM . SPR sensograms of (a) LTR-II + III + IV with k_a $4.7 \times 10^4 \pm 5.2 \times 10^2 \text{ Ms}^{-1}$ and k_d $9.2 \times 10^{-4} \pm 7.5 \times 10^{-5} \text{ s}^{-1}$, (b) LTR-II + III + IV M4-5 with k_a $2.8 \times 10^4 \pm 3.4 \times 10^2 \text{ Ms}^{-1}$ and k_d $10.0 \times 10^{-4} \pm 8.4 \times 10^{-5} \text{ s}^{-1}$ and (c) LTR-II + III + IV Random, the k_a and k_d values of which are not reliably measurable. Sensograms are shown as grey lines and their respective fits as black lines.

unfolding up to 83% was obtained in the LTR-II + III + IV oligonucleotide (Fig. 4c, Supplementary Table S1 and Supplementary Fig. S1).

Because G4s are less stable at lower K^+ concentrations, we proposed that the diverse unfolding efficiency towards the different-length LTR G4s depended on the stability of the LTR G4s. To test this hypothesis, melting temperatures (T_m) of the dual-labelled oligonucleotides were thus measured by FRET-monitored thermal unfolding (Table 2). The most stable sequence was LTR-III (T_m $62.1 \pm 0.1^\circ\text{C}$), followed by LTR-IV and LTR-III + IV (T_m $59.1 \pm 0.1^\circ\text{C}$ and $55.2 \pm 0.1^\circ\text{C}$, respectively); LTR-II + III + IV was the least stable sequence (T_m 49.1 ± 0.1), confirming that the observed unfolding scale depended on the oligonucleotide stability in the FRET system. To note that in the unlabelled oligonucleotides measured by circular dichroism (CD), LTR-IV was the least stable sequence (Table 2), bias which likely derives by the absence of fluorophores at the oligonucleotides' 5'- and 3'-end, which affect T_m values, as previously noted³².

In addition, the longest sequences, i.e. LTR-II + III + IV and LTR-III + IV, where multiple G4s may form, display T_m values that represent the average stability, which depends on the effective formation of the possible G4s. We thus performed a *Taq* polymerase stop assay to assess the G4 mutual formation and stability at increasing K^+ concentration in the full-length LTR. LTR-III was used as a control sequence where only one G4 could form. We observed that in the full-length LTR-II + III + IV sequence, both LTR-II and LTR-III G4s formed upon addition of K^+ (lanes 2–3, Fig. 5a); of these, LTR-III G4 induced a slightly more intense stop than LTR-II, suggesting a preferred formation of LTR-III in the full-length sequence (Fig. 5b). LTR-III was also the only G4 forming in the LTR-III + IV template (lanes 5–7, Fig. 5a). Interestingly, the LTR-III that formed in the full-length sequence was less intense than LTR-III forming in the shortest oligonucleotides (compare lanes 2–3, with 6–7 and 9–10, Fig. 5a and b). These data indicate that i) LTR-III is the most prominent G4 among the mutually exclusive LTR G4s and ii)

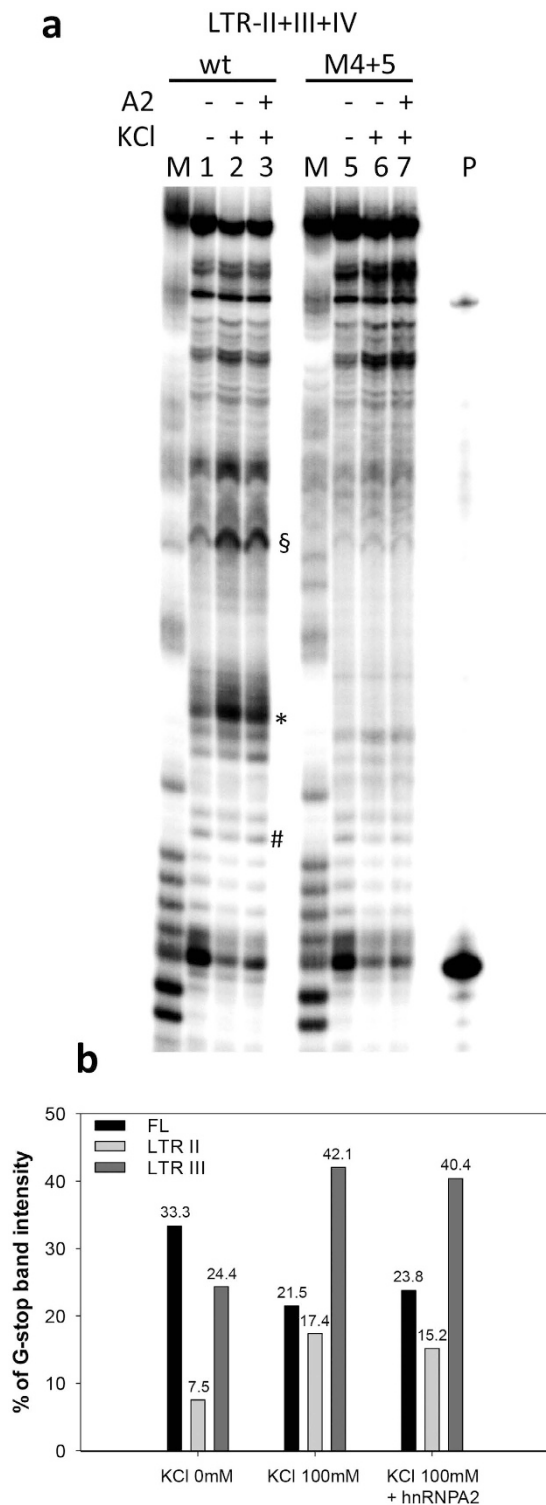


Figure 3. *Taq* polymerase stop assay in the presence of hnRNP A2. (a) *Taq* polymerization was performed in the presence/absence of K^+ and hnRNP A2, as indicated, on the wt and the mutant LTR-II + III + IV M4 + 5 sequences. Amplification of the wt template was performed at 37 °C for 30 min. Stop regions corresponding to LTR-III are indicated by the * symbol, to LTR-II by § and to LTR-IV by#. M is a marker lane obtained with the Maxam & Gilbert sequencing protocol. P indicates the band of the labelled primer. (b) Quantification of lanes 1–3 in panel (a). Quantification of stop bands corresponding to formation of LTR-II and LTR-III G4s and of the full-length amplification product (FL) is shown.

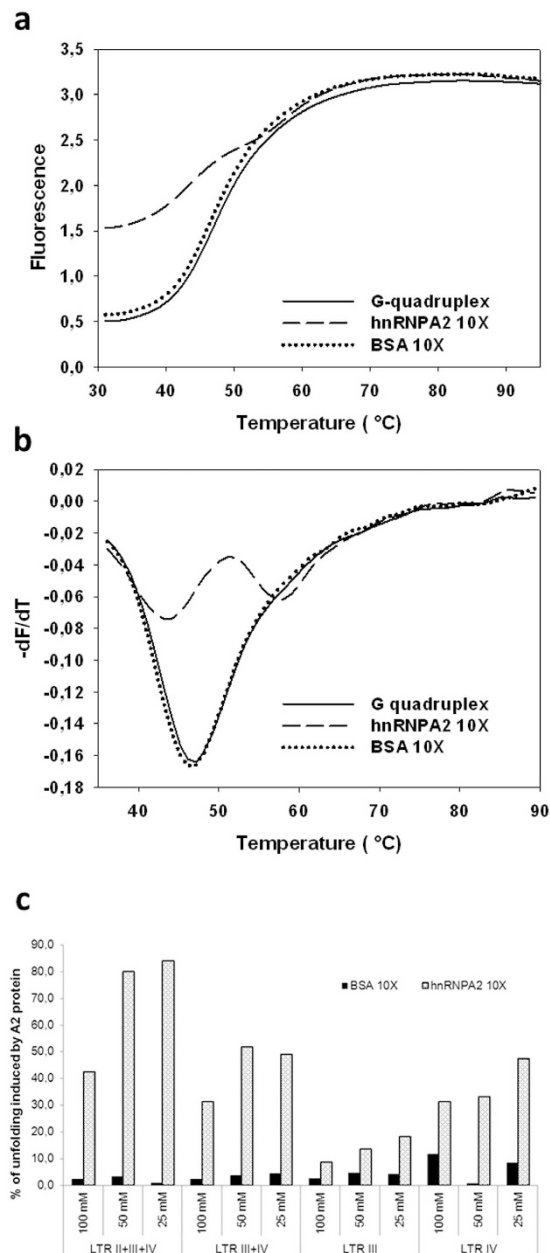


Figure 4. FRET analysis of hnRNP A2 unfolding. (a) FRET-melting curves of LTR F-II + III + IV-T, in the absence or presence of hnRNP A2 1:10 or of BSA as control. (b) Corresponding first derivative curves, dF/dT versus T. Proteins were incubated with the LTR G4s for 30 min at 37 °C prior to analysis. (c) Unfolding % of all tested sequences at 100, 50, 25 mM K^+ , incubated in the presence of hnRNP A2 or BSA at 37 °C for 30 min. Unfolding values are referred to unfolding in the absence of proteins (100%).

when LTR-III forms in the full-length LTR promoter, its stability maintains a level that is fully susceptible to hnRNP A2 processing.

Depletion of hnRNP A2/B1 decreases viral transcription. We have shown that LTR G4 folding inhibits viral transcription, whereas when point mutations that disrupt G4s are introduced in the LTR promoter, an increase in promoter activity is evidenced¹⁸. Silencing of the LTR G4 folding/stabilizing protein nucleolin significantly increased promoter activity, indicating the inhibitory effect of nucleolin on LTR-driven transcription²². To assess the effect of hnRNP A2/B1 binding/unfolding of the full-length LTR G4s, a set of anti-hnRNP A2/B1 siRNAs was used, which effectively silenced hnRNP A2/B1 expression in 293T cells (Fig. 6a). In these conditions, LTR-driven transcription decreased by 45% at the highest siRNA concentration in the wt LTR sequence (Fig. 6b). In contrast, the LTR sequence with two point mutations that disrupt G4 folding was only marginally affected by hnRNP A2/B1 depletion (Fig. 6b), indicating that the effect was G4-specific. In these conditions, hnRNP A2/B1-silenced 293T cells showed no reduction in cell viability at 48 h with respect to the controls, thus confirming

LTR	K ⁺ (mM)	T _m (°C)	
		FRET	CD
II + III + IV	100	49.1 ± 0.1	61.1 ± 0.2
	50	45.1 ± 0.1	nd
	25	43.2 ± 0.1	nd
III + IV	100	55.2 ± 0.1	64.6 ± 0.1
	50	49.1 ± 0.1	nd
	25	46.2 ± 0.1	nd
III	100	62.1 ± 0.1	68.7 ± 0.5
	50	56.0 ± 0.1	nd
	25	52.1 ± 0.1	nd
IV	100	59.1 ± 0.1	50.6 ± 0.8
	50	57.1 ± 0.1	nd
	25	51.1 ± 0.1	nd

Table 2. Melting temperatures (T_m) of LTR G4s measured on dual-labelled oligonucleotides by FRET and on unlabelled oligonucleotides by CD. T_m values are reported with standard deviation. Nd stands for “not-determined”.

the consistency of the inhibitory effects on LTR promoter activity. Silencing of hnRNP A2/B1 was also performed in the TZM-bl reporter cell line, which contains stably integrated copies of the luciferase gene under control of the LTR wt promoter and thus closely resembles the integrated provirus. The LTR-driven reporter transcription decreased up to 33% in hnRNP A2/B1-silenced cells (Supplementary Fig. S5). These data further confirm the unfolding activity of hnRNP A2/B1 on the LTR G4s.

Discussion

We have shown that hnRNP A2/B1 selectively binds the HIV-1 LTR G4s and unfolds them. While hnRNPs are RNA/protein complexes that bind newly synthesized RNA (pre-mRNA) in the cell nucleus during gene transcription and post-transcriptional modifications³³, two of the most abundant of them, hnRNP A1 and hnRNP A2/B1, have been reported to bind also DNA in the quadruplex conformation. Initially hnRNP A1 was found to unfold the telomeric G4s and promote telomerase activity^{34–36}, to unfold the G4 in the promoter of the KRAS oncogene and regulate transcription³⁰ and to destabilize the tetraplex forms of the fragile X expanded sequence d(CGG)_n³⁷. hnRNP A2 has been shown to unfold the fragile X repeats³⁸ and the shorter hnRNP A2 splice variant to promote telomere extension in mammalian cells³⁹.

Our finding that hnRNP A2 binds and unfolds the G4s in the promoter of HIV-1 not only confirms the DNA G4 unfolding activity of this protein, but also indicates for the first time that this type of activity is exploited by a virus.

In HIV-1, hnRNP A2/B1 has been reported to play with Rev an important accessory role in promoting nuclear viral RNA retention and nucleocytoplasmic viral RNA transport⁴⁰. In this context, the lower virus production observed upon siRNA-mediated depletion of hnRNP A2/B1 in infected cells was ascribed to accumulation of viral genomic RNA in cytoplasmic compartments or in the nucleus⁴¹. Our results showing decreased transcriptional activity upon depletion of hnRNP A2/B1 indicate a possible additional mechanism of inhibition of virus production mediated by increased G4 folding in the HIV-1 LTR promoter in the absence of hnRNP A2/B1. Surprisingly, overexpression of this protein and other hnRNP members induced similar effects, i.e. inhibition of transcription and reduction of virus production⁴². We suggest that upon stimuli that boost viral transcription (i.e. integral release of LTR G4s upon hnRNP A2/B1 massive overexpression) the virus (or the cell) activates counteracting mechanisms to avoid excessive exploitation of the cell by the virus that would lead to fast cell death, in the end impairing virus production. Therefore, the reported transcription inhibition might be mediated by other factors triggered by hnRNP A2/B1 overexpression.

The activity of hnRNP A2/B1 is fit to unfold the G4s that actually form in the HIV-1 LTR region. In fact, G4s folded in short oligonucleotides, such as LTR-III and LTR-IV, which displayed higher thermal stability, were only partially unfolded by the protein. In contrast, the same G4 structures folding in a longer context (LTR-II + III + IV) were less stable and were effectively processed by hnRNP A2/B1. Obviously, the longer oligonucleotide better mimics the actual G4 condition in the full-length LTR, as also proved by transcription inhibition obtained by hnRNP A2/B1 depletion in cells that contained the entire HIV-1 LTR.

The LTR G4 system displays regulating features similar to those described for the *c-myc*^{43,44}, *HRAS*²⁶ and *KRAS* oncogene-promoters^{30,45–47}. As in these eukaryotic G4-modulated promoters, the HIV-1 LTR promoter is processed by G4 stabilizing (nucleolin)²² and destabilizing proteins (hnRNP A2/B1). Because the LTR promoter has been suggested to be the region where viral latency is regulated^{48,49}, the G4 switch may play a role not only in activation of effective viral transcription, but also in its shift to latency. Currently, only the actively transcribed virus is targeted by antiviral drugs, while eradication of the HIV-1 infection has been made impossible by the existence of reservoirs of the latent virus⁵⁰. Therefore, our findings not only advance our understanding on the mechanism of viral transcription but may also constitute a progress from a therapeutic point of view.

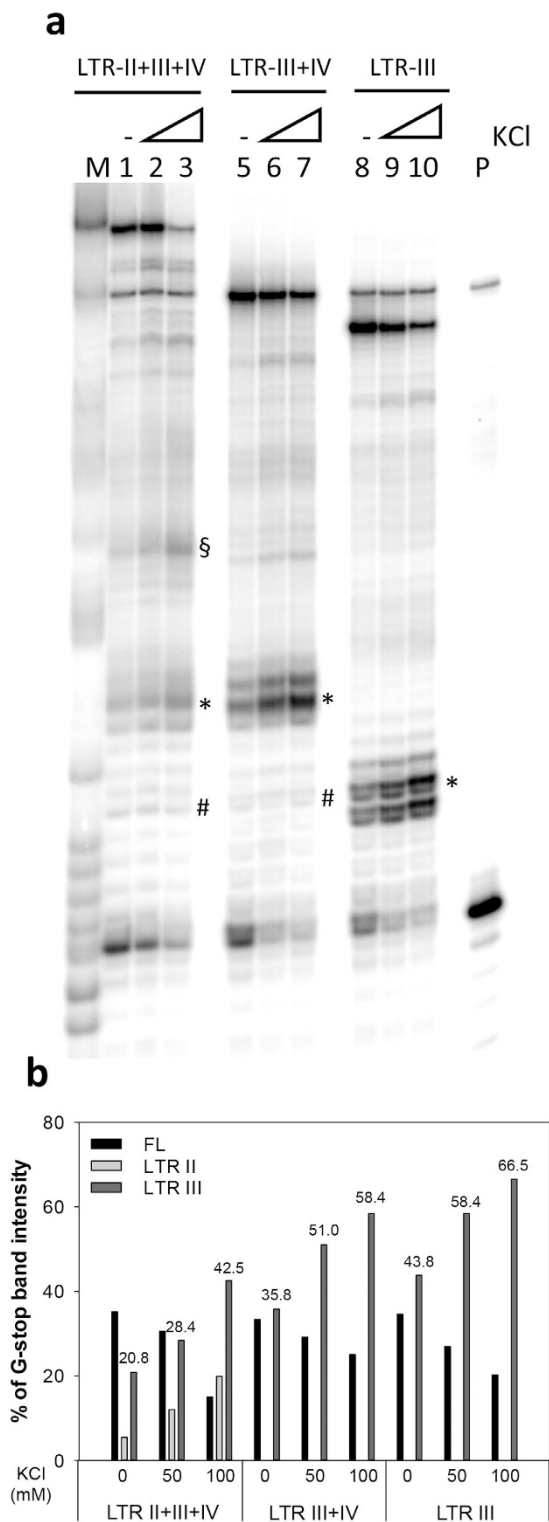


Figure 5. Taq polymerase stop assay at increasing K^+ concentrations. (a) Oligonucleotides were folded in the absence or presence of K^+ (50–100 mM). Oligonucleotides were used as templates in a Taq polymerase reaction at 37 °C. Stop regions corresponding to LTR-III are indicated by the * symbol, to LTR-II by § and to LTR-IV by #. P indicates the band of the labelled primer. M is a marker lane obtained with the Maxam & Gilbert sequencing protocol. (b) Quantification of lanes 1–10 of panel (a). Quantification of stop bands corresponding to formation of LTR-II and LTR-III G4s and of the full-length amplification product (FL) is shown.

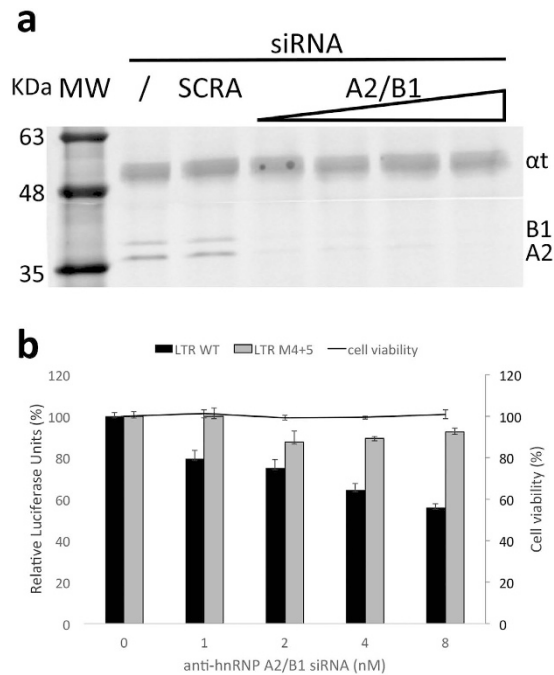


Figure 6. Effect of hnRNP A2/B1 on the HIV-1 LTR promoter in cells. (a) hnRNP A2/B1 depletion in 293T cells by siRNAs analysed by western blot with anti hnRNP A2/B1 antibody. Scra indicates scrambled siRNAs. Detection of α -tubulin (α t) was used as control. The symbol “/” indicates untreated cells. (b) Analysis of the luciferase activity of the wt and M4 + 5 LTR promoters in 293T cells treated with hnRNP A2/B1 siRNAs, normalized to protein content.

Methods

Oligonucleotides and cells. All desalted oligonucleotides were purchased from Sigma-Aldrich, Milan, Italy (Supplementary Table S1). TZM-bl (obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. J.C. Kappes, Dr. X. Wu, and Tranzyme Inc.) is a HeLa cell line stably expressing large amounts of CD4 and CCR5 and containing integrated copies of the luciferase and β -galactosidase genes under control of the HIV-1 promoter. Human embryonic kidney (HEK) 293T (ATCC # CRL-3216) and TZM-bl cells were grown in DMEM (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Cells were grown in a humidified incubator maintained at 37°C with 5% CO₂.

Protein nuclear extraction and pull-down assay. Protein nuclear extracts of HEK 293T cells were obtained by using NXTRACT kit (Sigma-Aldrich, Milan, Italy). HEK 293T protein nuclear extracts (150 μ g) were incubated with biotinylated LTR-II + III + IV G4, mutated M4 + 5 and random R oligonucleotides folded (600 nM) in 250 μ l of reaction containing Tris-HCl 20 mM, pH 8, KCl 30 mM, MgCl₂ 1.5 mM, protease inhibitor cocktail (Sigma-Aldrich, Milan, Italy) 1%, NaF 5 mM, Na₃VO₄ (Sigma-Aldrich, Milan, Italy) 1 mM, poly [dI-dC] (Sigma-Aldrich, Milan, Italy) 1.25 ng/ μ l for 2 h at 37°C. The binding reaction was followed by incubation (2 h at 37°C) with 30 μ l of streptavidin-agarose beads (Sigma-Aldrich, Milan, Italy). After PBS and NaCl washes (0.2 and 1 M), beads were collected by brief centrifugation, resuspended in 50 μ l of Laemmli buffer, and finally incubated at 95°C for 5 min. Supernatants were separated in 12% SDS-PAGE and, after coomassie staining, the gel lanes of LTR G4 and LTR M4 + 5 were directly in-gel digested and analysed for protein identification by mass spectrometry (MS) (see MS protein identification in ref. 22). Briefly, gel lanes cut in ~0.5 cm pieces were first washed with 50% CH₃OH and 2.5% acetic acid, dehydrated with CH₃CN, then reduced with 30 μ l of DTT (10 mM in 100 mM NH₄HCO₃) for 30 min at room temperature; the excess of DTT was next neutralized by alkylation with 30 μ l of iodoacetamide (50 mM in 100 mM NH₄HCO₃) for 30 min at room temperature. Bands were washed with 100 mM NH₄HCO₃, dehydrated with CH₃CN twice, and then digested overnight with 1 μ g of MS-grade trypsin (ThermoFisher Scientific, Waltham, MA, USA) in 50 μ l of 50 mM NH₄HCO₃. Peptides were extracted twice with 5% formic acid and twice with 50% CH₃CN/5% formic acid. The peptide mixture was further desalted in a silica nanocolumn (Polymicro Technologies, Phoenix, AZ, USA) packed in house with pinnacle C18 pack material (Thermo Fisher Scientific, Waltham, MA, USA). The desalted mixture was finally analyzed by direct infusion electrospray ionization (ESI) on a Thermo Fisher Scientific (Waltham, MA, USA) LTQ-Orbitrap Velos mass spectrometer utilizing quartz emitters produced in house. A stainless steel wire was inserted through the back-end of the emitter to supply an ionizing voltage that ranged between 0.8 and 1.2 kV. Putative peptides present in wt LTR-II + III + IV (but not in M4 + 5) samples were submitted to tandem mass spectrometric (MS/MS) analysis. The masses of the 50 most intense fragment ions were employed to perform a Mascot Database Search⁵¹ to identify their parent protein. Significant Mascot hits were accepted as positive matches and their ion score reported

(see MS protein identification in ref. 22 for more details). Supernatants were further analysed by Western blot with an anti-hnRNP A2/B1 antibody (mouse monoclonal DP3B3; Santa Cruz Biotechnology, Dallas, TX, USA).

Surface plasmon resonance (SPR) analysis. SPR was performed on the Biacore T100 platform (GE Healthcare, Life Science, Milan, Italy). About 1000 RU of human recombinant hnRNP A2 (Origene Technologies, Rockville, USA) were immobilized on Serie S sensor chip CM5 by amine coupling chemistry. Immobilization was performed in HEPES-NaCl running buffer (HEPES pH 7.4 10 mM, NaCl 150 mM, EDTA 3 mM). Flow cell 1 was blank immobilized to permit reference subtraction. Binding analysis of LTR-II + III + IV, LTR-II + III + IV M4 + 5 and LTR-II + III + IV Random sequences was performed at a flow rate of 25 μ l/min, with contact time and dissociation time of 120 s in HEPES-KCl buffer (HEPES pH 7.4 10 mM, KCl 150 mM, EDTA 3 mM). Oligonucleotides folding into G4 structures was performed in HEPES-KCl buffer after heat denaturation at 95 °C for 5 min and gradual cooling at room temperature. Sensograms were obtained in the concentration range of 31.25 nM–2 μ M. After each oligonucleotide injection the chip surface was regenerated with NaCl 2.5 M solution. All sensograms were corrected by double reference subtraction of blank flow cell response and buffer injection response. Data were fitted to a global 1:1 binding model using BIAevaluation software (GE Healthcare).

FRET-melting assay. Oligonucleotides were diluted to 0.1 μ M in lithium cacodylate buffer (10 mM, pH 7.4) and KCl (indicated in figure legends) heat denatured for 5 min at 95 °C, and folded in G-quadruplex structure at room temperature for 16 h. Samples were incubated alone, with hnRNP A2 or bovine serum albumin (BSA, negative control) and subsequently fluorescence intensity was monitored from 30 °C to 95 °C (1 °C/min) in a LightCycler II (Roche, Milan, Italy) by observing 6-carboxyfluorescein (6FAM) emission. FRET efficiency (E) was calculated as the total fluorescence intensity of the donor in the absence (I_D) and presence of the acceptor (I_{DA}), by using $E = 1 - (I_{DA}/I_D)$. I_D and I_{DA} values were obtained on a LS 55 Fluorescence spectrometer (Perkin Elmer, Waltham, MA, USA) setting the excitation wavelength at 480 nm and recording the emission from 500 to 650 nm. The dual-labelled oligonucleotide was converted into the corresponding duplex, in which the fluorophores are at a distance (R) for which FRET is ~0. R was calculated from FRET efficiency values as $R = R_0^6 \sqrt{1/E} - 1$, where R_0 (Förster distance) is the distance at which energy transfer is 50% of the maximum value. Between FAM and TAMRA fluorophores, R_0 is assumed to be 50 Å²⁶. F-test (F) and the probability (P) values were calculated using R statistical environment (v. 3.3.2)³¹. A conventional alpha = 0.05 was considered to evaluate the test significance.

Taq Polymerase Stop Assay. Taq polymerase stop assay was carried out as previously described^{18,22}. Briefly, the 5'-end labelled primer was annealed to its template (Table S1) in lithium cacodylate buffer in the presence or absence of KCl (50–100 mM) and by heating at 95 °C for 5 min and gradually cooling to room temperature. Where specified, samples were incubated with 29 ng of human recombinant hnRNP A2 (Origene Technologies, Rockville, USA) at 37 °C for 15 min. Primer extension was conducted with 2 U of AmpliTaq Gold DNA polymerase (Applied Biosystem, Carlsbad, California, USA) at 37 °C for 30 min. Reactions were stopped by ethanol precipitation, primer extension products were separated on a 15% denaturing gel, and finally visualized by phosphorimaging (Typhoon FLA 9000).

Circular Dichroism Spectroscopy. DNA oligonucleotides were diluted to a final concentration (4 μ M) in lithium cacodylate buffer (10 mM, pH 7.4) and KCl 100 mM. Samples were annealed by heating at 95 °C for 5 min and gradually cooled to room temperature overnight. CD spectra were recorded on a Chirascan-Plus (Applied Photophysics, Leatherhead, UK) equipped with a Peltier temperature controller using a quartz cell of 5 mm optical path length, over a wavelength range of 230–320 nm. For the determination of T_m , spectra were recorded over a temperature range of 20–90 °C, with temperature increase of 5 °C. The reported spectra are baseline-corrected for signal contributions due to the buffer. Observed ellipticities were converted to mean residue ellipticity (θ) = deg \times cm² \times dmol⁻¹ (mol ellip). T_m values were calculated according to the van't Hoff equation, applied for a two-state transition from a folded to unfolded state, assuming that the heat capacity of the folded and unfolded states are equal⁵².

Immunoblot analysis. Immunoblot analysis was performed on cell protein extracts obtained in RIPA Buffer (20 mM Tris-HCl pH 7.5; 150 mM NaCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate, 1 mM Na₃VO₄, 1x protease inhibitors). Protein concentrations were quantified using the Pierce[®] BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). Each sample was electrophoresed on 10% SDS-PAGE and transferred to a nitrocellulose blotting membrane (Amersham TM Protan TM, GE Healthcare Life science, Milan, Italy) by using trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Milan, Italy). The membranes were blocked with 2.5% skim milk in PBST (0.05% Tween 20 in PBS). Membranes were incubated with the respective primary antibody directed against hnRNP A2/B1 (mouse monoclonal; Santa Cruz Biotechnology, Dallas, TX, USA), alpha-tubulin (mouse monoclonal; Sigma-Aldrich, Milan, Italy). After three washes in PBST, membranes were incubated with ECL Plex Goat- α -Mouse IgG-Cy5 (GE Healthcare Life sciences, Milan, Italy). Images were captured on the Typhoon FLA 9000.

siRNA and luciferase reporter assay. Gene-specific pooled siRNA trilencer targeting human hnRNP A2/B1 and a scrambled negative control duplex were purchased from Origene (human hnRNP A2/B1 Trilencer-27 Human siRNA, OriGene Technologies, Rockville, MD, USA). 293T cells or TZM-bl reporter cells were transfected with increasing concentrations of siRNA (1, 2, 4 and 8 nM for 293T and 1, 5, 10 nM for TZM-bl cells) and scrambled siRNA by using Lipofectamine RNAiMAX (Invitrogen, Thermo Fisher Scientific, Waltham, MA,

USA) following the manufacturer's instructions. Only for experiments in 293T cells, pLTR luciferase plasmids (pGL4.10-LTRwt or pGL4.10-LTR-M4 + 5) were transfected 24 h later by using Lipofectamine 3000 (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Luciferase activity was measured using the britelite plus Reporter Gene Assay System (PerkinElmer Inc., Milan, Italy) at a Victor X2 multilabel plate reader (PerkinElmer Inc., Milan, Italy), according to the manufacturer's instructions. Cells were lysed in 1% Triton-X100-PBS and protein concentration was determined by BCA assay (Thermo Scientific Pierce, Monza, Italy). Luciferase signals were subsequently normalized to total protein content, according to the manufacturer's protocol (<http://ita.promega.com/~pdf/resources/pubhub/cellnotes/normalizing-genetic-reporter-assays/>). Each assay was performed in duplicate and each set of experiments was repeated at least three times.

Cell Viability Assay. The MTT (3-(4,5-dimethylthiazol 2-yl)-2,5-diphenyltetrazolium bromide, Sigma Aldrich, Milan Italy) assay was performed to assess cell viability on silenced cells. 293T cells (6×10^3) were plated in 96-well plates and incubated for 24 h. Following siRNA transfection (from 1 nM to 8 nM), cells were incubated for an additional 48 h. Control cells (transfected in absence of siRNA) were treated in the exact same conditions. Cell survival was evaluated by MTT assay: 10 μ L of freshly dissolved solution of MTT (5 mg/mL in PBS) were added to each well, and after 4 h of incubation, MTT crystals were solubilized in solubilization solution (10% sodium dodecyl sulphate (SDS) and 0.01 M HCl). After overnight incubation at 37 °C, absorbance was read at 540 nm. Data were expressed as mean values of at least two individual experiments conducted in triplicate. The percentage of cell survival was calculated as follows: cell survival = $(A_{\text{well}} - A_{\text{blank}})/(A_{\text{control}} - A_{\text{blank}}) \times 100$, where blank denotes the medium without cells.

References

- Maizels, N. G4-associated human diseases. *EMBO reports* **16**, 910–922, doi: 10.15252/embr.201540607 (2015).
- Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic acids research* **43**, 8627–8637, doi: 10.1093/nar/gkv862 (2015).
- Zhou, B., Liu, C., Geng, Y. & Zhu, G. Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD. *Scientific reports* **5**, 16673, doi: 10.1038/srep16673 (2015).
- Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5**, 182–186, doi: 10.1038/nchem.1548 (2013).
- Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research* **42**, 860–869, doi: 10.1093/nar/gkt957 (2014).
- Metifiot, M., Amrane, S., Litvak, S. & Andreola, M. L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic acids research* **42**, 12352–12366, doi: 10.1093/nar/gku999 (2014).
- Tan, J. *et al.* The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS pathogens* **5**, e1000428, doi: 10.1371/journal.ppat.1000428 (2009).
- Thluckova, K. *et al.* Human papillomavirus G-quadruplexes. *Biochemistry* **52**, 7207–7216, doi: 10.1021/bi400897g (2013).
- Wang, S. R. *et al.* A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Science advances* **2**, e1501535, doi: 10.1126/sciadv.1501535 (2016).
- Fleming, A. M., Ding, Y., Alenko, A. & Burrows, C. J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS infectious diseases* **2**, 674–681, doi: 10.1021/acsinfedc.6b00109 (2016).
- Wang, S. R. *et al.* Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell chemical biology* **23**, 1113–1122, doi: 10.1016/j.chembiol.2016.07.019 (2016).
- Murat, P. *et al.* G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nature chemical biology* **10**, 358–364, doi: 10.1038/nchembio.1479 (2014).
- Norseen, J., Johnson, F. B. & Lieberman, P. M. Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *Journal of virology* **83**, 10336–10346, doi: 10.1128/JVI.00747-09 (2009).
- Artusi, S. *et al.* Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic acids research* **44**, 10343–10353, doi: 10.1093/nar/gkw968 (2016).
- Artusi, S. *et al.* The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral research* **118**, 123–131, doi: 10.1016/j.antiviral.2015.03.016 (2015).
- Perrone, R. *et al.* Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PLoS one* **8**, e73121, doi: 10.1371/journal.pone.0073121 (2013).
- Amrane, S. *et al.* Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *Journal of the American Chemical Society* **136**, 5249–5252, doi: 10.1021/ja501500c (2014).
- Perrone, R. *et al.* A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *Journal of medicinal chemistry* **56**, 6521–6530, doi: 10.1021/jm400914r (2013).
- Piekna-Przybylska, D., Sullivan, M. A., Sharma, G. & Bambara, R. A. U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* **53**, 2581–2593, doi: 10.1021/bi4016692 (2014).
- Perrone, R. *et al.* Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *The Journal of antimicrobial chemotherapy* **69**, 3248–3258, doi: 10.1093/jac/dku280 (2014).
- Perrone, R. *et al.* Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *Journal of medicinal chemistry* **58**, 9639–9652, doi: 10.1021/acs.jmedchem.5b01283 (2015).
- Tosoni, E. *et al.* Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic acids research* **43**, 8884–8897, doi: 10.1093/nar/gkv897 (2015).
- Sissi, C., Gatto, B. & Palumbo, M. The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie* **93**, 1219–1230, doi: 10.1016/j.biochi.2011.04.018 (2011).
- Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annual review of genetics* **42**, 301–334, doi: 10.1146/annurev.genet.41.110306.130350 (2008).
- Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K. & Xodo, L. E. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *The Journal of biological chemistry* **285**, 22003–22016, doi: 10.1074/jbc.M110.101923 (2010).
- Cogoi, S., Shchekotikhin, A. E. & Xodo, L. E. HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic acids research* **42**, 8379–8388, doi: 10.1093/nar/gku574 (2014).
- He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cellular and molecular life sciences: CMLS* **66**, 1239–1256, doi: 10.1007/s00018-008-8532-1 (2009).

28. Sinha-Datta, U., Khan, S. & Wadgaonkar, D. Label-free interaction analysis as a tool to demonstrate biosimilarity of therapeutic monoclonal antibodies. *Biosimilars* **5**, 83–91, doi: 10.2147/BS.S85537 (2015).
29. Zhang, X. & Oglesbee, M. Use of surface plasmon resonance for the measurement of low affinity binding interactions between HSP72 and measles virus nucleocapsid protein. *Biological procedures online* **5**, 170–181, doi: 10.1251/bpo59 (2003).
30. Paramasivam, M. *et al.* Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic acids research* **37**, 2841–2853, doi: 10.1093/nar/gkp138 (2009).
31. Team, R. C. (ed Vienna R Foundation for Statistical Computing, Austria) (2016).
32. Gray, R. D. & Chaires, J. B. Analysis of multidimensional G-quadruplex melting curves. *Current protocols in nucleic acid chemistry*/edited by Serge L. Beaucage ... [et al.] Chapter 17, Unit17 14, doi: 10.1002/0471142700.nc1704s45 (2011).
33. Mayeda, A., Munroe, S. H., Caceres, J. F. & Krainer, A. R. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *The EMBO journal* **13**, 5483–5495 (1994).
34. Hudson, J. S., Ding, L., Le, V., Lewis, E. & Graves, D. Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1. *Biochemistry* **53**, 3347–3356, doi: 10.1021/bi500351u (2014).
35. Myers, J. C., Moore, S. A. & Shamoo, Y. Structure-based incorporation of 6-methyl-8-(2-deoxy-beta-ribofuranosyl) isoxanthopteridine into the human telomeric repeat DNA as a probe for UP1 binding and destabilization of G-tetrad structures. *The Journal of biological chemistry* **278**, 42300–42306, doi: 10.1074/jbc.M306147200 (2003).
36. Zhang, Q. S., Manche, L., Xu, R. M. & Krainer, A. R. Hnrnp A1 associates with telomere ends and stimulates telomerase activity. *Rna* **12**, 1116–1128, doi: 10.1261/rna.58806 (2006).
37. Weisman-Shomer, P., Cohen, E. & Fry, M. Distinct domains in the CArG-box binding factor A destabilize tetraplex forms of the fragile X expanded sequence d(CGG)n. *Nucleic acids research* **30**, 3672–3681 (2002).
38. Khateb, S., Weisman-Shomer, P., Hershco, I., Loeb, L. A. & Fry, M. Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic acids research* **32**, 4145–4154, doi: 10.1093/nar/gkh745 (2004).
39. Wang, F. *et al.* Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20413–20418, doi: 10.1073/pnas.1200232109 (2012).
40. Gordon, H. *et al.* Depletion of hnRNP A2/B1 overrides the nuclear retention of the HIV-1 genomic RNA. *RNA biology* **10**, 1714–1725, doi: 10.4161/rna.26542 (2013).
41. Levesque, K. *et al.* Trafficking of HIV-1 RNA is mediated by heterogeneous nuclear ribonucleoprotein A2 expression and impacts on viral assembly. *Traffic* **7**, 1177–1193 (2006).
42. Jablonski, J. A. & Caputi, M. Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *Journal of virology* **83**, 981–992, doi: 10.1128/JVI.01801-08 (2009).
43. Brooks, T. A. & Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes & cancer* **1**, 641–649, doi: 10.1177/1947601910377493 (2010).
44. Dexheimer, T. S. *et al.* NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III(1). *Molecular cancer therapeutics* **8**, 1363–1377, doi: 10.1158/1535-7163.MCT-08-1093 (2009).
45. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *The FEBS journal* **277**, 3459–3469, doi: 10.1111/j.1742-4658.2010.07759.x (2010).
46. Chu, P. C. *et al.* Regulation of oncogenic KRAS signaling via a novel KRAS-integrin-linked kinase-hnRNPA1 regulatory loop in human pancreatic cancer cells. *Oncogene*, doi: 10.1038/onc.2015.458 (2015).
47. Cogo, S. & Xodo, L. E. G4 DNA in ras genes and its potential in cancer therapy. *Biochimica et biophysica acta* **1859**, 663–674, doi: 10.1016/j.bbagr.2016.02.002 (2016).
48. Duverger, A. *et al.* An AP-1 binding site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. *Journal of virology* **87**, 2264–2277, doi: 10.1128/JVI.01594-12 (2013).
49. van der Sluis, R. M., Jeeninga, R. E. & Berkhout, B. Establishment and molecular mechanisms of HIV-1 latency in T cells. *Current opinion in virology* **3**, 700–706, doi: 10.1016/j.coviro.2013.07.006 (2013).
50. Finzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
51. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567, doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 (1999).
52. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols* **1**, 2527–2535, doi: 10.1038/nprot.2006.204 (2006).
53. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* **11**, 601, doi: 10.1002/bms.1200111109 (1984).

Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation (GCE grant numbers OPP1035881, OPP1097238); the European Research Council (ERC Consolidator grant 615879); and the Italian Ministry of University and Research (FIRB-Ideas RBID082ATK) to SNR. Funding for open access charge: Bill and Melinda Gates Foundation.

Author Contributions

M.S. performed the mass spectrometry analysis and protein identification; I.F. provided silencing, reporter and viability data in cells, E.R. performed the FRET analysis, R.P. performed the Taq polymerase stop assays and silencing in TZM-bl cells, E.T. performed the pull-down assay and western blot, S.L. provided the SPR data, M.T. performed CD analysis, G.P. commented on the manuscript, S.N.R. conceived of the work and wrote the manuscript. All authors analyzed the data and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Scalabrin, M. *et al.* The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **7**, 45244; doi: 10.1038/srep45244 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017