

University of Groningen

Investigating the NRAS 5' UTR as a target for small molecules

Balaratnam, Sumirtha; Torrey, Zachary R; Calabrese, David R; Banco, Michael T; Yazdani, Kamyar; Liang, Xiao; Fullenkamp, Christopher R; Seshadri, Srinath; Holewinski, Ronald J; Andresson, Thorkell

Published in:
Cell Chemical Biology

DOI:
[10.1016/j.chembiol.2023.05.004](https://doi.org/10.1016/j.chembiol.2023.05.004)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Balaratnam, S., Torrey, Z. R., Calabrese, D. R., Banco, M. T., Yazdani, K., Liang, X., Fullenkamp, C. R., Seshadri, S., Holewinski, R. J., Andresson, T., Ferré-D'Amaré, A. R., Incarnato, D., & Schneekloth, J. S. (2023). Investigating the NRAS 5' UTR as a target for small molecules. *Cell Chemical Biology*, 30(6), 643–657. <https://doi.org/10.1016/j.chembiol.2023.05.004>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

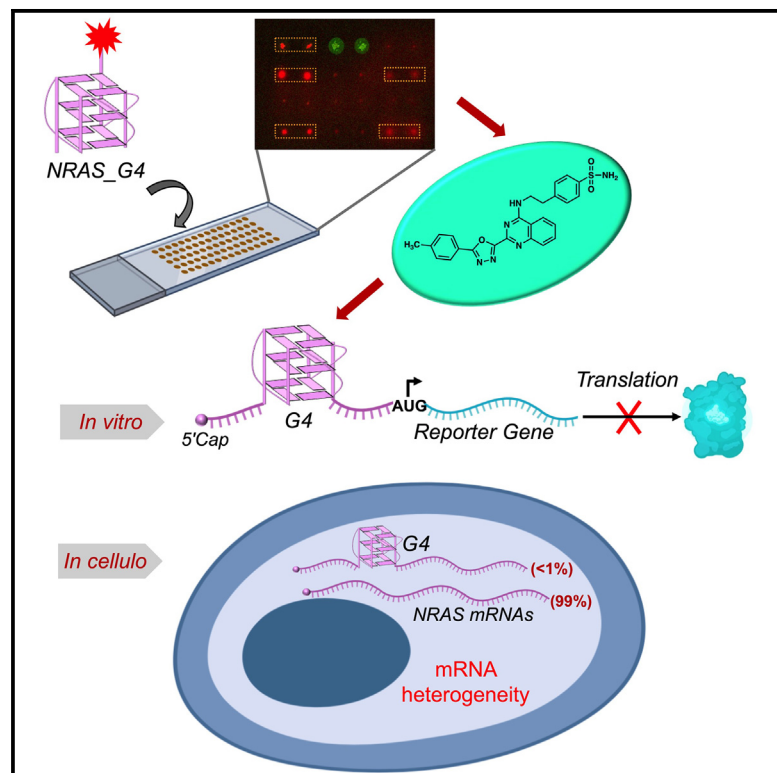
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Cell Chemical Biology

Investigating the *NRAS* 5' UTR as a target for small molecules

Graphical abstract



Authors

Sumirtha Balaratnam,
Zachary R. Torrey,
David R. Calabrese, ...,
Adrian R. Ferré-D'Amaré,
Danny Incarnato,
John S. Schneekloth, Jr.

Correspondence

schneeklothjs@mail.nih.gov

In brief

Balaratnam et al. used a high throughput small molecule microarray screen to identify novel small molecules that block translation by binding to a G-quadruplex structure from the *NRAS* mRNA. Detailed chemical, biophysical, structural, and biological evaluation provide insights into both the structure and targetability of this cancer-associated mRNA, revealing transcript heterogeneity as a challenge to targeting *NRAS* translation.

Highlights

- SMMs enable identification of a lead compound which binds to a G4 in the *NRAS* mRNA
- Chemical inhibition of translation by targeting G4 in the 5'-UTR of the *NRAS* mRNA
- Analysis indicates heterogeneity in the length of the *NRAS* mRNA 5'-UTR



Article

Investigating the *NRAS* 5' UTR as a target for small molecules

Sumirtha Balaratnam,¹ Zachary R. Torrey,¹ David R. Calabrese,¹ Michael T. Banco,² Kamyar Yazdani,¹ Xiao Liang,¹ Christopher R. Fullenkamp,¹ Srinath Seshadri,¹ Ronald J. Holewinski,³ Thorkell Andresson,³ Adrian R. Ferré-D'Amaré,² Danny Incarnato,⁴ and John S. Schneekloth, Jr.^{1,5,*}

¹Chemical Biology Laboratory, National Cancer Institute, Frederick, MD 21702, USA

²Biochemistry and Biophysics Center, National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA

³Protein Characterization Laboratory, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc, Frederick, MD 21702, USA

⁴Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Groningen, the Netherlands

⁵Lead contact

*Correspondence: schneeklothjs@mail.nih.gov

<https://doi.org/10.1016/j.chembiol.2023.05.004>

SUMMARY

Neuroblastoma RAS (*NRAS*) is an oncogene that is deregulated and highly mutated in cancers including melanomas and acute myeloid leukemias. The 5' untranslated region (UTR) (5' UTR) of the *NRAS* mRNA contains a G-quadruplex (G4) that regulates translation. Here we report a novel class of small molecule that binds to the G4 structure located in the 5' UTR of the *NRAS* mRNA. We used a small molecule microarray screen to identify molecules that selectively bind to the *NRAS*-G4 with submicromolar affinity. One compound inhibits the translation of *NRAS* *in vitro* but showed only moderate effects on the *NRAS* levels *in cellulo*. Rapid Amplification of cDNA Ends and RT-PCR analysis revealed that the predominant *NRAS* transcript does not possess the G4 structure. Thus, although *NRAS* transcripts lack a G4 in many cell lines the concept of targeting folded regions within 5' UTRs to control translation remains a highly attractive strategy.

INTRODUCTION

Messenger RNA (mRNA) transcripts exhibit highly complex and diverse structures through canonical and non-canonical base pairing interactions as well as interactions with proteins.^{1–4} Structured RNA elements impact biological functions such as RNA synthesis, metabolism, and regulatory pathways.^{1,5,6} The formation of folded structures in mRNA has been shown to play important roles in post-transcriptional regulation of gene expression including RNA maturation, translation, and degradation.^{7–9} Interestingly, folds in 5' untranslated regions (UTRs) of mRNAs have been recognized as a major feature that regulates the translation process.^{10,11} For example, ~60% of 5' UTRs in humans have structured regions near the 5' cap site and are believed to impact translation initiation.¹² One way these structures regulate cap-dependent translation initiation is through helicase-mediated remodeling of RNA structures and higher-order RNA interactions, as well as cap-independent translation initiation through internal ribosome entry sites.^{10,11,13} The formation of complex structures within 5' UTRs suggests opportunities for small molecule targeting and control of gene expression at the translational level.

NRAS is a proto-oncogene belonging to the RAS oncogene superfamily and was first identified in neuroblastoma.^{14–16} In many cancers, *NRAS* proteins are constitutively activated by

oncogenic mutations or overexpression. The mutation of *NRAS* accounts for about 15% of RAS-related human malignancies, notably in myeloid leukemias and cutaneous melanomas.^{17–19} In normal cells, *NRAS* proteins switch between active GTP-bound forms and inactive GDP-bound forms. The transition between the active and inactive state is mediated by GTPase-activating proteins.^{20,21} Hyperactivation of *NRAS* leads to the persistence of the GTP-bound state of *NRAS*, initiating constitutive MAPK signaling as well as AKT signaling, which drives malignancies by promoting cell growth, survival, and invasion.^{22–25} Targeting *NRAS* is therefore believed to be a promising strategy for developing anticancer therapies. However, the *NRAS* protein has been considered a highly challenging target due to a lack of drug binding pockets other than the nucleotide (GTP) binding pocket.^{19,25,26} Since the nucleotide binding pocket has picomolar binding affinity to GTP, developing GTP-competitive inhibitors is nontrivial. Still, the FDA approval of the covalent KRAS^{G12C} inhibitor AMG 510 validates Ras-family proteins as highly important anticancer targets²⁷ even though there are still no approved *NRAS* inhibitors yet. One intriguing strategy to expand the targetability of *NRAS* is to target the associated mRNA and inhibit translation. This approach could be accomplished by identifying a structured region that contains small molecule binding pockets in the *NRAS* mRNA and developing the small molecules that bind to these structures.



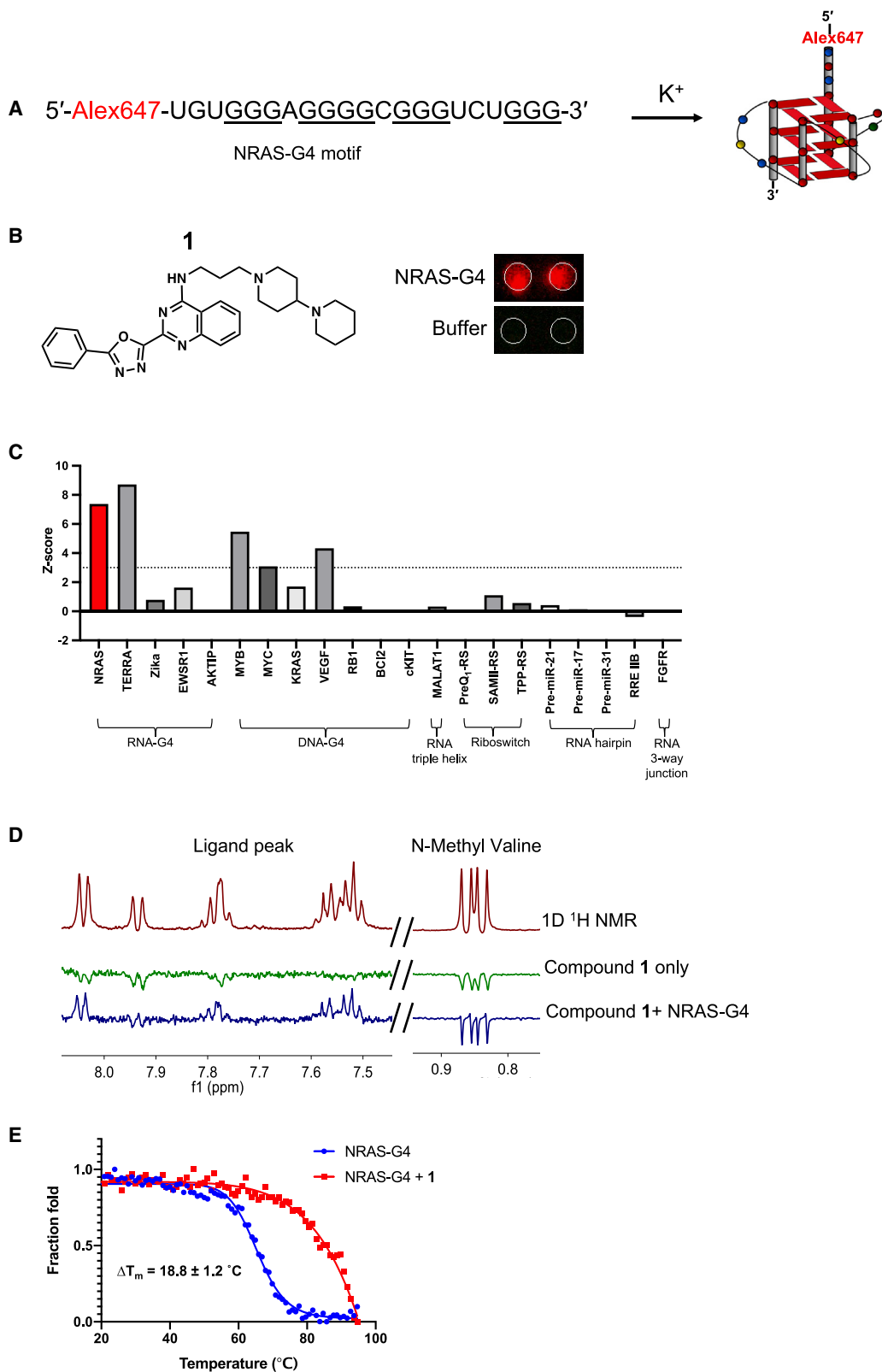


Figure 1. Small Molecule Microarray (SMM) screening using the NRAS-G4 and hit validation

(A) Sequence of the AlexaFluor647-labeled NRAS G-quadruplex forming sequence (5'-Alex647-NRAS-G4) used for SMM screening. Schematic drawing of parallel G4 structure formed by the folded sequence in the presence of K^+ ion.

(legend continued on next page)

G-quadruplexes (G4s) are a class of secondary structure formed by G-rich sequences in DNA or RNA.²⁸ The general formula for canonical G4-forming sequences is $G_{2-5} N_{1-7} G_{2-5} N_{1-7} G_{2-5}$ where N can be any nucleotide located in the loops of the G4.²⁹ One G from each of the four G-tracts is bonded to form a co-planar G-quartet by Hoogsteen base pairings. Two or more G-quartets can stack on top of each other to form G4 structures. G4s are stabilized by monovalent cations, most commonly K^+ .³⁰ Although initial genome-wide studies revealed the existence of DNA-G4 structures in genomes and their enrichments in telomeric regions, the occurrence of G4s in RNA has been more controversial.³¹ Studies have shown the presence of RNA-G4 structures in mRNAs, viral genomic RNA, pre-miRNAs, precursor piRNA transcripts, mature piRNAs and long noncoding RNAs (lncRNAs),³² while other reports also indicate that RNA G-quadruplexes (rG4s) are globally unfolded in the steady state (presumably by helicases or other protein factors).^{33–35} Still, rG4s are thought to play important roles in diverse of biological processes and related to human diseases.^{36,37} One bioinformatic search for “regular” rG4s in 5′ UTRs of the human transcriptome indicated that 9,979 5′ UTRs contain at least one potential G4-forming sequence.³⁸ Further studies reported that formation of RNA-G4 structures in 5′ UTRs inhibited mRNA translation by interfering with the recruitment of the pre-initiation complex.³⁹ In the human *NRAS* gene, Kumari and co-workers reported the presence of an RNA-G4 structure in the 5′ UTRs and that this G4 motif is conserved in both its sequence and its positions relative to the translation start site among different organisms such as chimpanzee, macaque, mouse, rat, and dog.⁴⁰ They used a luciferase reporter assay to demonstrate that formation of RNA-G4 structure in the 5′ UTRs of *NRAS* mRNA inhibits translation by >80% in rabbit reticulocyte lysates. This result suggests that native RNA-G4 structures in 5′ UTRs could act as regulatory elements of translation and become an attractive secondary structure for therapeutic targeting to control *NRAS* mRNA at the post-transcriptional level. Thus, stabilization of the *NRAS*-G4 represents an opportunity for pharmacological suppression of *NRAS* expression with small molecule ligands. The development of a few small molecule stabilizers for *NRAS*-G4 has been reported in recent years, though selectivity remains a challenge.^{41,42}

In this study, we report the discovery of a series of small molecules that bind directly to the *NRAS*-G4 structure using an SMM screening strategy. The lead compound **18** showed selective binding to the *NRAS*-G4 structure over other RNA structures. Biophysical and biochemical experiments confirmed reversible submicromolar binding affinity of the lead compound. Additionally, luciferase-based reporter assays indicated that compound **18** inhibits translation in rabbit reticulocyte lysates via stabilizing the *NRAS*-G4 structure *in vitro*. Structure probing and X-ray crystallographic studies support the formation of a parallel-stranded G4 in the context of the luciferase reporter construct.

Treatment of **18** in SK-MEL-2 and MCF-7 cells has an unexpectedly moderate effect on *NRAS* translation (~20%). Gene expression profiling using RNA sequencing and global proteomics assays confirmed that G4-associated genes were not generally perturbed by **18**. In-depth analysis of *NRAS* transcripts in various cell lines using RT-PCR and 5′ Rapid Amplification of cDNA Ends (RACE) confirmed that majority of *NRAS* transcript isoforms are shorter and lack a G4-forming sequence in the 5′ UTR. This result was further supported by analysis of publicly available rG4 sequencing and CAGE-seq data and aligns with the recently re-annotated *NRAS* transcription start site (TSS) in the UCSC genome database.^{43–47} Importantly, this work does not rule out the existence of an *NRAS* rG4, but rather demonstrates that the predominant *NRAS* transcript lacks it in all the cell lines evaluated. Finally, this work demonstrates that the strategy of controlling translation by targeting structured regions of 5′ UTRs is a viable approach.

RESULTS

Discovery of an Neuroblastoma RAS G4-binding small molecule using SMM

To identify small molecules that bind to the *NRAS*-G4, we used an SMM screening strategy.⁴⁸ SMM is a convenient method to rapidly screen tens of thousands of compounds to identify selective RNA-binding small molecules and has successfully been employed by our lab and others to identify RNA-binding ligands.⁴⁸ In the iteration of this approach used by our laboratory, a library of 26,227 compounds is spatially arrayed and covalently linked to the glass surface. An AlexaFluor 647-labeled *NRAS*-G4 oligonucleotide was designed as a screening construct, folded into a G4 structure in K^+ buffer, and incubated with printed arrays (Figure 1A). The fluorescence intensity is measured at each location on the array to identify binding interactions between the *NRAS*-G4 and the compound printed at each location. A composite Z score is then calculated for each compound (printed in duplicate) in the library. Next, Z scores for the *NRAS*-G4 incubated dataset are compared to a control (buffer incubated) dataset. Compounds are considered hits if (1) they had a composite Z score greater than three (2) Coefficient of Variation (compensation voltage (CV)) of the two compound spots was lower than 200% and (3) compounds showed an increase in fluorescence in the presence of the RNA compared to the buffer incubated slides.⁴⁸ We identified 235 hits from a collection of 26,227 compounds screened in the SMM, for an overall hit rate of 0.89%. To further investigate selectivity, Z scores for each hit compound were compared to 20 different SMM screens that were performed with various RNA and DNA structures, including 11 other DNA or RNA G4s (Table S1). After selectivity analysis, 30 unique hit compounds were identified as having high Z score and good selectivity (Table S2). As a representative example, direct binding on SMM slides (Figure 1B) and selectivity data (Figure 1C)

(B) SMM images and chemical structure of hit compound **1**. Note that the compounds are printed in duplicate.

(C) Selectivity of the hit compound **1** across 20 different oligonucleotides screened using the SMM platform as measured by composite Z score.

(D) NMR validation of hit compound **1**. The ¹H NMR of **1** (100 μM) and N-methyl-L-valine (100 μM) (non-binding control) (Top, red spectrum), WaterLOGSY NMR of **1** and N-methyl-L-valine in the absence (middle, green spectrum) and presence (bottom, blue spectrum) of *NRAS*-G4 RNA (5 μM).

(E) Representative thermal melting curve of *NRAS*-G4 (5 μM) in the absence (blue) and presence (red) of compound **1** (10 μM). Melting was performed using 1 mM K^+ . Error bars indicate the standard deviation determined from three independent measurements.

are shown for compound **1**. From the set of *NRAS*-G4 selective hit molecules, 14 compounds were selected for further analysis based on their availability and chemotype (Table S3). To validate the 14 hit compounds as *NRAS*-G4 binders, binding assays were performed by surface plasmon resonance (SPR)^{49,50} (Table S3). Based on the measured equilibrium constant (K_D), compound **1** was chosen for further analysis.

Biophysical analysis confirmed the binding of compound **1** to *NRAS*-G4

We first evaluated the binding interaction between the initial hit compound **1** and the *NRAS*-G4 using ligand-observed NMR experiments. To determine the solubility of the compound in the aqueous buffer, compound **1** was observed in a standard ¹H NMR experiment (Figure 1D upper panel). WaterLOGSY is widely used for detecting macromolecule (RNA, DNA and Protein)-ligand interactions. In WaterLOGSY, peaks from ligands that bind to the RNA exhibit phase opposite (positive) to the non-binding control and peaks which are not engaged in binding interaction phase down (negative), providing a straightforward readout of the binding interaction.⁵¹ WaterLOGSY was performed on **1** both in the presence and absence of the *NRAS*-G4. As shown in the Figure 1D, in the absence of RNA, all peaks phased negatively, confirming the compound does not aggregate in aqueous buffer (aggregating compounds phase positively). Upon the addition of *NRAS*-G4 RNA, peaks corresponding to **1** phased positively, while peaks for N-methyl-L-valine (used as an internal, non-binding control) remain unchanged in all cases confirming the direct binding of **1** to the *NRAS*-G4.

To measure the binding affinity of compound **1** with *NRAS*-G4, we used SPR experiments. A biotinylated *NRAS*-G4 oligonucleotide was immobilized to a streptavidin-coated SPR chip, and binding was measured as a function of compound concentration. This experiment demonstrated that compound **1** bound to *NRAS*-G4 with an equilibrium dissociation constant (K_D) of $0.45 \pm 0.11 \mu\text{M}$ (Figures S1 A and S1B). The pan-G4 binder BRACO 19 was used as a positive control in the SPR. It bound to the *NRAS*-G4 with an equilibrium dissociation constant (K_D) of $79 \pm 1.2 \text{ nM}$ (Figure S1C) which is roughly within 2- to 3-fold of values previously reported for G4s.⁵² Thus, the SPR assay is a valid method to measure compound affinities. To validate binding further using an orthogonal biophysical assay, a fluorescence titration assay (FIA) was also performed to measure the affinity of **1** to *NRAS*-G4. In this experiment, compound **1** was titrated into a solution containing AlexaFluor 647-labeled *NRAS*-G4 (the same construct used in the SMM screen) and changes in fluorescence intensity were monitored as a function of compound concentration (Figure S1D). An apparent dissociation constant was measured by fitting the binding curve between the normalized fluorescence intensity and compound concentration. In this assay, compound **1** had a K_D of $1.2 \pm 0.5 \mu\text{M}$. In sum, biophysical experiments confirmed the direct binding of **1** to the *NRAS*-G4 structure and good solubility in aqueous buffer.

Next, we evaluated the effect of **1** on *NRAS*-G4 thermal stability by performing a circular dichroism (CD)-based thermal melt assay, in which molar ellipticity was measured as a function of increasing temperature. In the absence of **1**, the CD spectrum of the *NRAS*-G4 exhibited a maximum at 263 nm and a minimum at 240 nm, confirming proper folding of a parallel G4 structure.⁴⁰

The melting temperature (T_M) of the RNA in 1 mM K^+ was measured to be $65.4 \pm 1.9^\circ\text{C}$, which was consistent with previously reported values.⁴⁰ Due to the unusually high stability of this G4, complete unfolding was not observed at high potassium levels. Therefore, unfolding was performed using 1 mM K^+ buffer so that full unfolding of the G4 can be observed, and accurate measurements collected. In the presence of $2.5 \mu\text{M}$ of compound **1**, the T_M was also increased by $6.9 \pm 0.6^\circ\text{C}$, while at $10 \mu\text{M}$ of compound **1** increased the T_M by $18.8 \pm 1.2^\circ\text{C}$ (Figures 1E and S1E).

Structure-activity relationship study of **1**

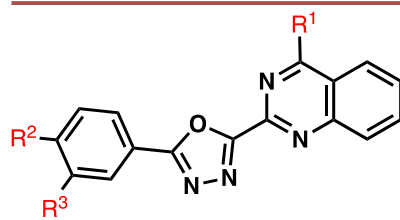
To identify an improved binder, we performed a preliminary structure-activity relationship (SAR) study by using a focused series of commercially available analogs of **1** (Table 1). Based on the chemical structure of compound **1**, 13 analogs of compound **1** were purchased from the commercially available focused library. These analogs have altered R^1 side chain groups, while R^2 and R^3 were also altered in several analogs. Each compound was evaluated for binding affinity toward the *NRAS*-G4 structure by fluorescence titration assay (Figures 2A and S2A). Most of the analogs showed strong binding behavior toward the *NRAS*-G4 structure, and 12 analogs out of 13 (except compound **8**) were identified as stronger binders than the parent compound (**1**). Among them, compounds **2**, **14**, and **18** showed the tightest binding affinities, and the calculated binding affinity was less than 300 nM. To further validate the binding affinities of these compounds, SPR experiments were performed (Figures 2B, 2C, S2B, and S2C). Here, analogs **2** and **14** showed weaker binding affinity with $4.5 \pm 1.2 \mu\text{M}$ and $>41 \mu\text{M}$ respectively. However, compound **18** bound to *NRAS*-G4 with an equilibrium dissociation constant (K_D) of $932 \pm 38 \text{ nM}$ in SPR experiments, compared to a 250 nM K_D in fluorescence titrations. Based on these results, we selected **18** as a lead compound for further analysis. To examine whether compound **18** influenced the stability of the *NRAS*-G4 structure, we performed a CD-based thermal melting assay. In the presence of compound **18**, the melting temperature of *NRAS*-G4 increased by $6.2 \pm 0.4^\circ\text{C}$ (Figure 2D).

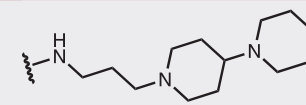


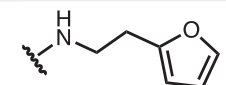


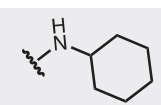


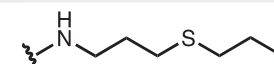


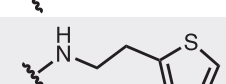


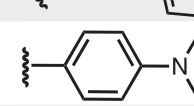


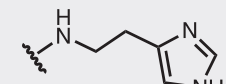


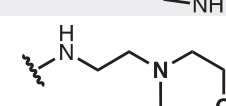


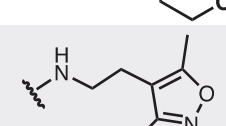


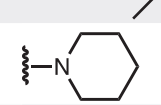


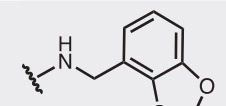


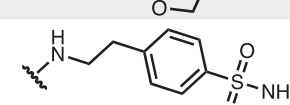


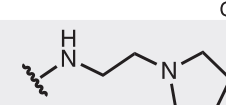


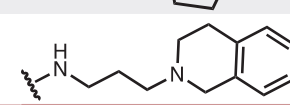


The binding selectivity of compound **18** to *NRAS*-G4 over other G4 structures was also evaluated by FIA. For this study, nine different G4 structures from cancer-relevant genes were selected. This panel includes 6 DNA G4s from the promoter region of oncogenes (*BCL2*, *KRAS*, *mTOR*, Telomeric DNA G4, *VEGF*, and *MYCN*) and 3 RNA-G4s from mRNA 5' UTRs or ncRNAs (*AKTIP*, *TERRA* and *EWSR1*) (Table S1). Compound **18** was titrated into the various 5'-Cy5- or AlexaFluor647-labeled DNA/RNA G4s, and apparent K_D values were calculated. We observed that compound **18** exhibited weak binding toward the RNA-G4 structure formed by *AKTIP* and DNA-G4 from *BCL2*, (Figure S3A). For all other G4s, the binding affinity (K_D) could not be calculated because **18** showed no significant quenching. These results indicate that **18** binds to the *NRAS*-G4 via its unique structure.

Enzymatic structure mapping indicates a binding site of **18** on *NRAS*-G4

To biochemically investigate the binding mode of compound **18** on the *NRAS*-G4 structure, we performed Ribonuclease A (RNase A) structure mapping.⁵³ 5'-(Alexa Fluor 647)-labeled

Table 1. Focused library of analogs of compound 1 and affinity for the NRAS G4



Name	R ¹	R ²	R ³	FIA K _D (μM)
1				1.2 ± 0.5
2				0.26 ± 0.07
3				0.43 ± 0.12
6				0.74 ± 0.03
7				0.36 ± 0.06
8				3.0 ± 1.7
10				0.48 ± 0.08
12				0.30 ± 0.06
14				0.28 ± 0.04
16				0.35 ± 0.05
17				0.56 ± 0.17
18				0.25 ± 0.06
24				0.50 ± 0.14
25				0.95 ± 0.45

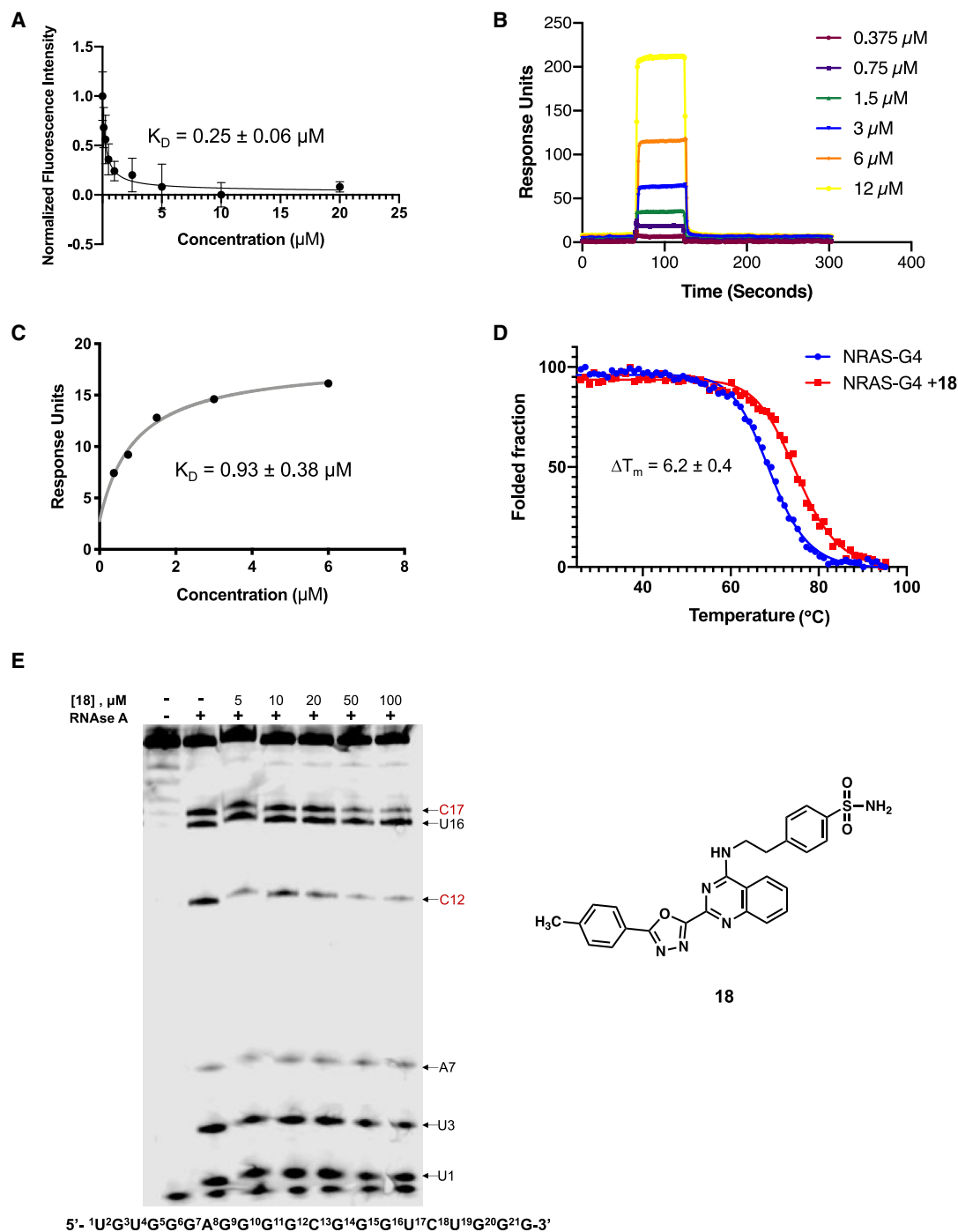


Figure 2. Affinity of compound 18 to the NRAS-G4

(A) Fluorescence intensity assay of 5'-(Alexa Fluor 647)-labeled NRAS-G4 RNA (50 nM) in the presence of **18**. Error bars indicate the standard deviation determined from three independent measurements.

(B) Sensorgrams and (C) binding curve corresponding to **18** interacting with a 5' biotin labeled NRAS-G4 RNA (5 μM) in SPR.

(D) CD-based thermal melting of NRAS-G4 (5 μM) in the absence (blue) and presence (red) of 10 μM compound **18** in 1 mM K^+ buffer. Error bars indicate the standard deviation determined from three independent measurements.

(E) RNase A structure probing of NRAS-G4 (5 μM) with increasing concentration of compound **18** (5 μM –100 μM). Compound **18** protected nucleotides (C17 and C12) in the loops (labeled in red). Chemical structure of compound **18**.

NRAS-G4 RNA (5 μ M) was folded into a G4 structure and incubated with increasing concentrations of compound **18** and RNase A. As shown in Figure 2E, in the absence of **18**, RNase A predominantly cleaved the U and C bases in the loop regions of G4 structure. Upon incubation with increasing concentrations of **18** however, the resulting RNase A cleavage pattern was markedly different at C12 and C17, which are flexible bases anticipated to be on neighboring loops (Figures 2E and S3B). The C12 nucleotide was protected from RNase A cleavage at lower concentrations of **18** (5 μ M), while C17 protection was only observed at higher concentrations of **18** (50 μ M and 100 μ M). These effects indicate that compound **18** binding impacts the flexibility or accessibility of the loop regions, and presumably elucidates the binding site of **18** on the *NRAS*-G4 structure. To further evaluate the binding mode of **18** on the *NRAS*-G4, we designed two 2-aminopurine-labeled (2-AP) *NRAS*-G4 constructs by replacing C12 and U16 with 2-AP. For both 2-AP constructs (12C-2AP-*NRAS*-G4 and 16U-2AP-*NRAS*-G4), titration with **18** resulted in a decrease in fluorescence, and compound-induced changes in fluorescence were used to derive a K_D value (Figures S3C and S3D). Observation of fluorescence quenching in both cases indicates that both C12 and U16 become less solvent exposed upon the binding of **18**, indicating they are near the binding site of **18**, which is consistent with protection in footprinting assays. Finally, molecular modeling/docking studies (Figures S4A–S4C) revealed several poses for **18** consistent with these experiments whereby the compound interacts with the tetrads and groove formed by loops containing C12 and C17.

Compound **18** inhibits the translation of a Neuroblastoma RAS reporter gene *in vitro*

To evaluate the effect of compound **18** on the efficiency of *NRAS* translation, we utilized a reporter system developed by the Balasubramanian group.⁴⁰ In these constructs, 254 bp encoding the human *NRAS* 5' UTR were cloned downstream of a T7 promoter (pSKC11), which possesses the transcript containing the native G4 structure (*NRAS*-G4-FL) (Figure 3A). As a control, another plasmid was generated by deleting the first 29 bp of the 5' UTR to produce a transcript that lacks the G4-forming region (*NRAS*-G4-Del-FL). Each of the transcripts was generated by *in vitro* transcription using T7 RNA polymerase, and the transcripts were subjected to *in vitro* translation in rabbit reticulocyte lysates in the presence of compound **18**. The translation efficiency was measured by quantifying the resulting luminescence and normalizing to controls without compound. As shown in Figure 3B, dose-dependent translational inhibition was observed with *NRAS*-G4-FL in the presence of compound **18**. However, no translational inhibition was observed for **18** with the control *NRAS*-G4-Del-FL transcript. Thus, compound **18** only inhibits translation of the reporter gene containing the *NRAS* 5' UTR when the G4 is present.

To validate proper formation of the *NRAS*-G4 structure in the chimeric reporter construct, and the observed *in vitro* translation effect was due to the G4 binding of **18**, we performed Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension-Mutational Profiling (SHAPE-MaP) experiments.⁵⁴ SHAPE-MaP is an RNA structure probing technique that combines chemical probing of unpaired nucleotides with next-generation sequencing to mea-

sure the flexibility of individual nucleotides in long RNAs. We treated the pre-folded *NRAS*-G4-FL mRNA construct with the SHAPE reagent, 2-methylnicotinic acid imidazole (NAI), and SHAPE-MaP was performed to generate a reactivity profile for the entire chimeric luciferase reporter construct (*NRAS*-G4-FL) (Figure S4D). SHAPE reactivity indicates the relative flexibility of a nucleotide, which correlates with occurrence of base pairing and base stacking (Figure S4D). SHAPE reactivities above 0.8 likely indicate unpaired bases, while reactivities below 0.4 likely indicate basepaired nucleotides. As shown in Figure 3C, the median window SHAPE reactivity was plotted across the transcript using a sliding window (median window-median transcript). This analysis demonstrated that there are specific regions in the transcript that have low median SHAPE reactivity. Most importantly, the SHAPE reactivities of the G4-forming region (highlighted in red) revealed that this region is structurally constrained and are a part of a folded structure.

Comparison of the normalized SHAPE reactivities of the guanine nucleotides involved in the formation of the G4 structure with all other guanine nucleotides in the chimeric luciferase construct demonstrated that guanines involved in G4 formation have low SHAPE reactivity (Figure 3D). The observed low SHAPE reactivity values indicate a high degree of ordered structure in the transcript, presumably corresponding to a folded G4. Interestingly, G12 showed higher SHAPE reactivity compared to the other guanines in the *NRAS*-G4-forming sequence. Thus, the dominant *NRAS*-G4 structure is likely to consist of G12 located in a loop region connecting the G-quartets, rather than facilitating the formation of the G-quartet itself. Efforts to map compound binding to this mRNA did not meet with success, potentially because the G4 is already highly folded in the absence of compound.

In order to gain molecular insights into the structure of the G4, we determined the X-ray crystal structure of an *NRAS*-G4 variant at 2.9 Å resolution (Figures 3E and 3F and Table S4). To prevent conformational heterogeneity of *NRAS*-G4 structural isomers, residue G8 was mutated to ensure formation of a single conformer. A single molecule of *NRAS*-G8U was observed in the asymmetric unit. The *NRAS*-G8U structure adopts a canonical parallel quadruplex that is comprised of three G-quartets coordinated by potassium ions. The guanines of the G4 are in the anti-conformation but consist of a mixture C2'- and C3'-endo sugar puckers (Figure S4E). In the *NRAS*-G8U structure, the G-tracts are connected by propeller-type loops. The loop nucleotides corresponding to A7 and U8 lacked electron density, were presumed to be disordered, and thus were not modeled. The other loops of the G4 possessed electron density, comprised of C12 and C17. In the crystal structure, the 3' G-quartet exhibits substantial buckling, whereas the other two G-quartets form largely planar tetrads (Figure 3G). The buckled guanines of the 3' G-quartet adopt a C2'-endo pucker, which deviates from the preferred C3'-endo pucker observed in most RNA G4s.²⁹ This buckled quartet participates in crystal packing; thus, its conformation in solution may differ.

Compound **18** has moderate effects on Neuroblastoma RAS levels *in cellulo*

To examine the effect of compound **18** on cellular *NRAS* levels, we selected a breast cancer cell line (MCF-7) and a melanoma cell line (SK-MEL-2). Both cancer cell lines have a high level of

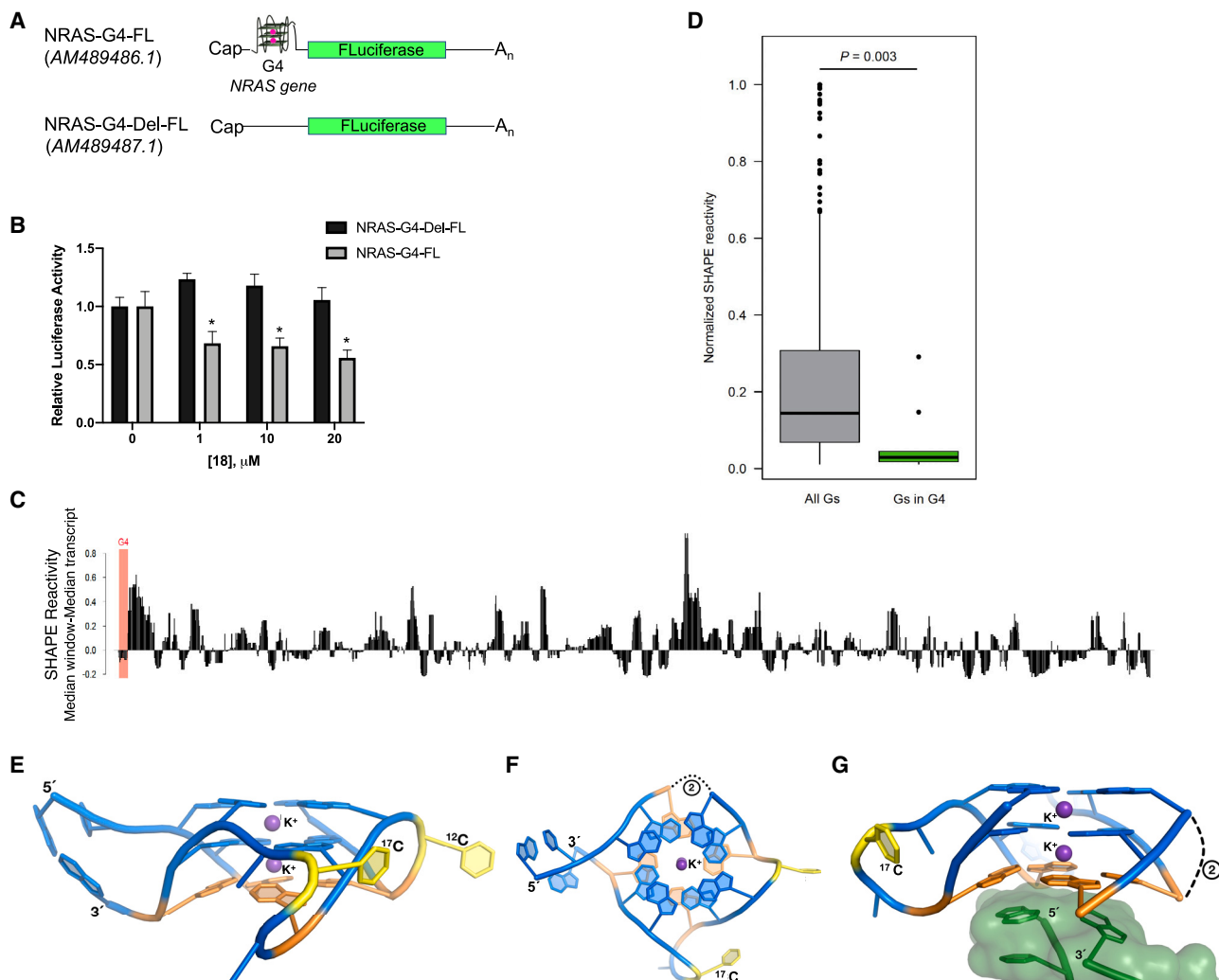


Figure 3. Effect of 18 on *NRAS*-FL reporter gene translation and structural analysis of the *NRAS* G4

(A) Schematic representation of firefly luciferase reporter constructs that contain the *NRAS* 5' UTR: *NRAS*-G4-FL (contains G4, top) and *NRAS*-G4-Del-FL (lacks G4, bottom).

(B) Relative translation efficiency of constructs (500 ng) in the presence of increasing concentrations of **18** (1 μ M, 10 μ M and 20 μ M), measured by quantitation of luciferase enzyme activity. Results were normalized to data for constructs with DMSO. Error bars represent the standard deviation of three independent experiments ($n = 3$, $*p < 0.05$).

(C) Median SHAPE reactivities across the chimeric luciferase construct (*NRAS*-G4-FL). Regions below the X axis indicate more structure than average, while regions above the X axis indicate less structure than average.

(D) Bar graph representing normalized SHAPE reactivity of all Gs, and Gs involved in G4 formation. Note that nucleotide numbering is relative to the mRNA transcript and is therefore different than numbering in Figures 1A and 2E. The statistical significance was calculated by t-test analysis ($n=3$).

(E and F) Cartoon representation of the X-ray crystal structure of the *NRAS* G4 viewed on the G tetrads plane (E) and rotated 90° (F). The RNA is colored blue, except the buckled 3' quartet (orange) and loop nucleotides (yellow). Potassium ions are depicted as purple spheres. Black dashes denote two disordered nucleotides not observed in the structure.

(G) Interface between an *NRAS*-G8U molecule and an adjacent crystallographic asymmetric unit (ASU). A portion of the molecule from an adjacent ASU is highlighted by the translucent green molecular surface. The crystal contacts may contribute to bulking of the 3' G-quartet (colored orange).

NRAS expression, either by oncogenic mutation (SK-MEL-2) or overexpression (MCF-7). Compound **18** was evaluated for its capacity to reduce cell viability in both cell lines. We observed that **18** decreased MCF-7 and SK-MEL-2 cell viability with an IC_{50} of $1.1 \pm 0.3 \mu$ M and $0.9 \pm 0.4 \mu$ M respectively (Figure 4A). To examine the effect of compound **18** on *NRAS* mRNA stability, we performed a quantitative PCR assay (qPCR) to quantify the *NRAS* mRNA levels after treatment with compound **18**. Treatment with **18** up to 25 μ M

did not drastically change the *NRAS* mRNA levels in either MCF-7 or SK-MEL-2 cells (Figure 4B). To assess broad changes in the transcriptome following treatment with compound **18**, we performed gene expression profiling by RNA-seq. Both cell lines were treated with 1 μ M of compound **18** and total RNA was isolated from cells after 48 h. RNA sequencing was performed, and differential gene expression analysis was carried out to compare DMSO and **18**-treated sample groups. This analysis identified

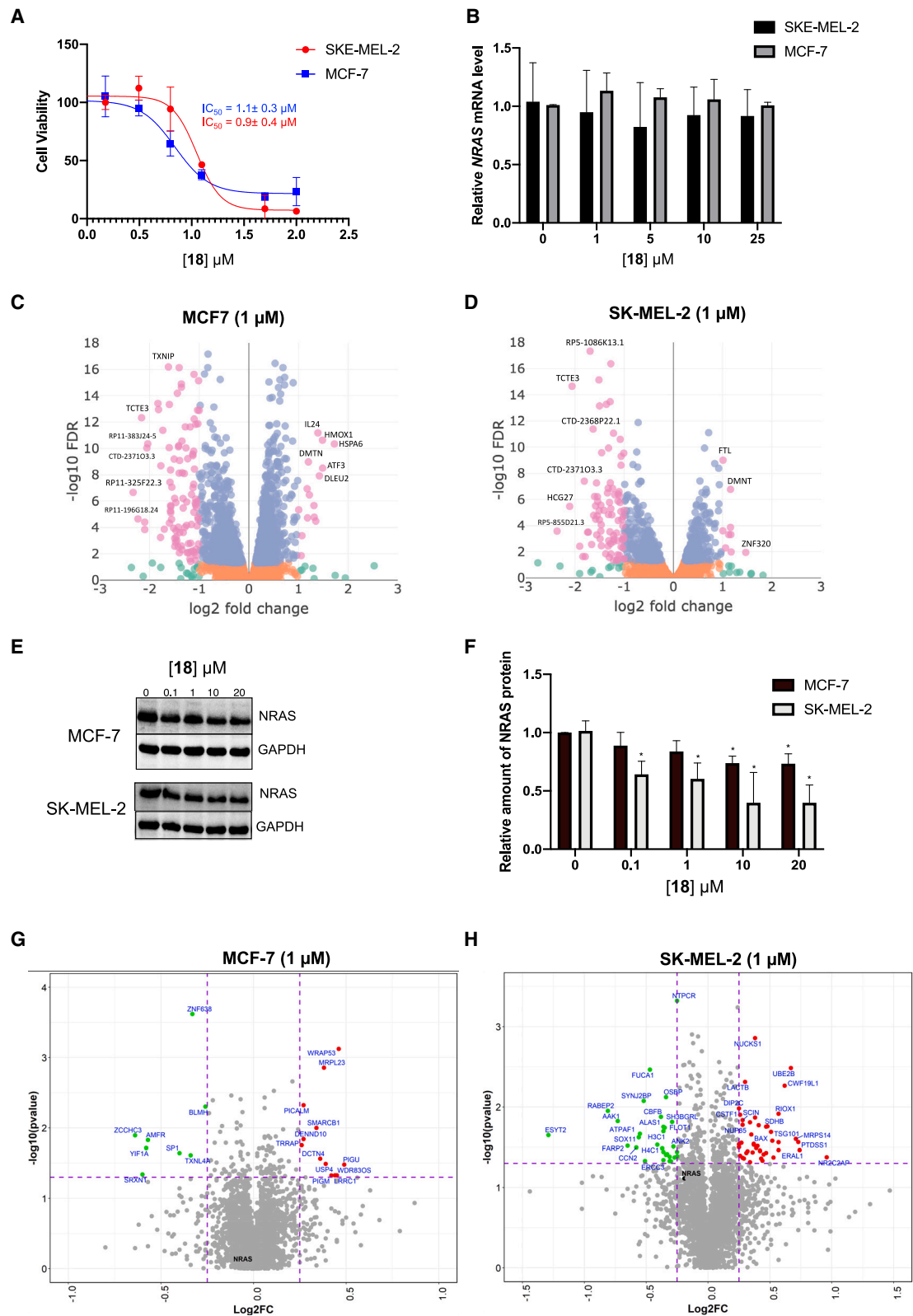


Figure 4. Effect of 18 on NRAS expression in cellulo

(A) Inhibition of SK-MEL-2 cell, and MCF-7 cell proliferation with treatment of 18. The results are presented as the mean \pm SEM (n = 9) of three independent experiments with triplicate in each.

(legend continued on next page)

384 differentially expressed genes (290 downregulated and 94 upregulated genes) in MCF-7 cells and 292 genes (226 downregulated and 66 upregulated genes) in SK-MEL-2 cells with a false discovery rate (FDR) < 0.05 (Figures 4C and 4D). As with qPCR analysis, compound **18** treatment did not change levels of *NRAS* mRNA in either cell line (Figures 4C, 4D and S5–S5C). We also observed that **18** had no effect on the levels of other RAS-family genes (*KRAS* and *HRAS*). Additionally, ontology analysis (Table S6) did not indicate perturbations to RAS-related signaling pathways in either cell line. To further evaluate the effect of compound **18** on other mRNAs which contain G4 structures in their UTR regions (3' UTR and 5' UTR) or CDS, we analyzed the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values in 29 different G4-containing mRNAs, including 24 mRNAs with a G4 in the 5' UTR, 3 with a G4 in the 3' UTR and 2 containing G4s in the coding sequence. No significant change in expression was observed with compound **18** treatment in mRNAs which contain G4 structures either in their UTR regions (3' UTR and 5' UTR) or CDS (Figures S5D–S5F). Overall, the RNA seq analysis indicates that compound **18** does not broadly affect the expression of genes with G4 structures in mRNA or promoters (DNA).

Next, to investigate the functional inhibition of **18** on *NRAS* translation, MCF-7 and SK-MEL-2 cells were treated with increasing concentrations of compound **18** (0.1–20 μ M). As shown in Figures 4E and 4F, 20 μ M of **18** caused a modest 20–30% decrease in *NRAS* levels. To more broadly evaluate the effects of **18** on protein levels, we performed global proteomics in both MCF-7 and SK-MEL-2 cells after treating with **18** for 48 h at 1 μ M (Figures 4G and 4H). Among the 3,214 proteins detected in SK-MEL-2 cells, 74 proteins (2.3%) were significantly affected (Abs 2Fc > 0.25 and p < 0.05) by compound **18** relative to the DMSO control. In MCF-7 cells, 3,266 proteins were detected by proteomic analysis and 20 proteins (0.6%) were affected by compound **18** treatment using the same metrics. Interestingly, only 16% of significantly affected proteins in SK-MEL-2 and 15% in MCF-7 cell lines possess stable G4 structures in their mRNAs when analyzed using QGRS mapper, indicated by a QGRS⁵⁵ score \geq 40. A similar analysis was performed using a more advanced algorithm G4Hunter.^{56,57} When applying a threshold of 1.7 (recommended to remove false positives), none of the 20 proteins identified from MCF-7 cells or the 74 proteins identified in SK-MEL-2 cells were predicted to contain stable G4s in their mRNAs. Moreover, when we analyze functionally known G4-driven mRNA-related proteins, we did not observe significant effects on their expression at the protein level with compound **18**. Taken together, evaluation of expression levels using global proteomics combined with RNA-seq indicate that **18** does not broadly modulate the expression of G4-associated genes *in celulo*. Since we observed a weaker effect of compound **18** on

NRAS levels in cells than in the luciferase reporter *in vitro*, we performed an in-depth analysis of the 5' UTR structure of the endogenous *NRAS* transcript.

Analysis of the architecture of 5' UTR of Neuroblastoma RAS mRNA

Based on NCBI annotation, two different transcripts for the *NRAS* mRNA exist: an older version (NM_002524.4) and a revised, newer version (NM_002524.5). The original *NRAS* transcript annotation contained a 4454 nt long sequence including a 254 nt 5' UTR which contains the G4 (between 15–32 nt from the 5'-end). However, a revised annotation of the *NRAS* transcript is 4326 nt long and has a 131 nt long 5' UTR region that lacks 123 nt from the 5' end of the older transcript. Importantly, the revised annotation of the *NRAS* transcript is devoid of a G4 structure in its 5' UTR. Transcripts with differences in 5'-UTR length are not uncommon due to the presence of multiple promoters, alternative transcription start sites, alternative splicing mechanisms within UTRs, or inaccurate mapping.⁵⁸

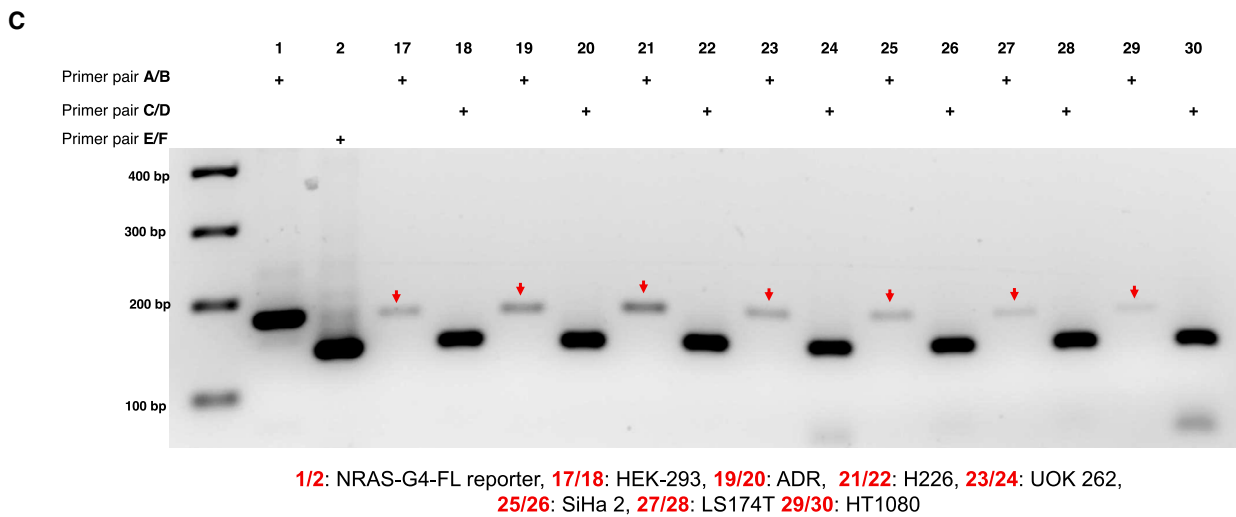
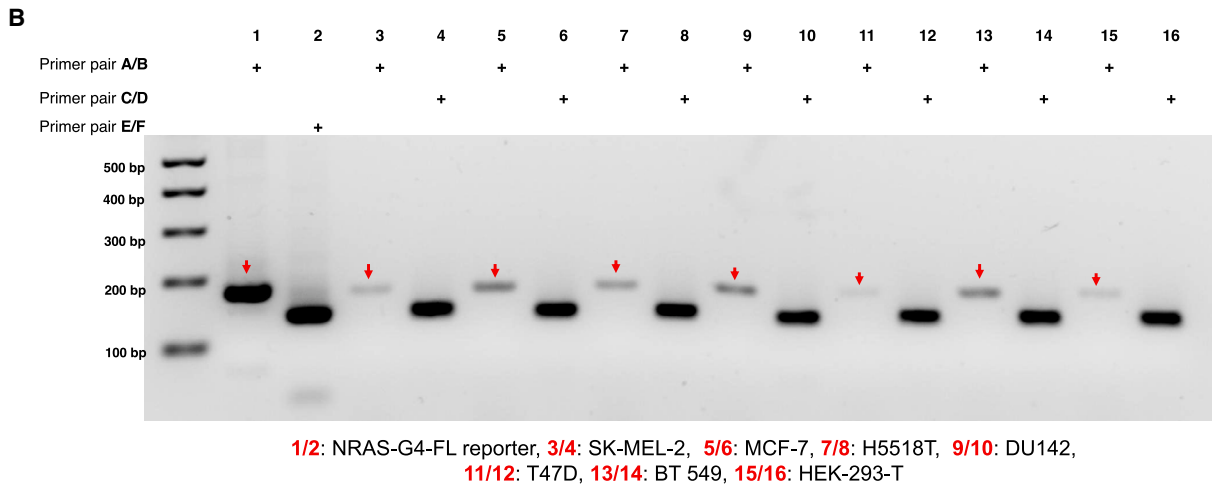
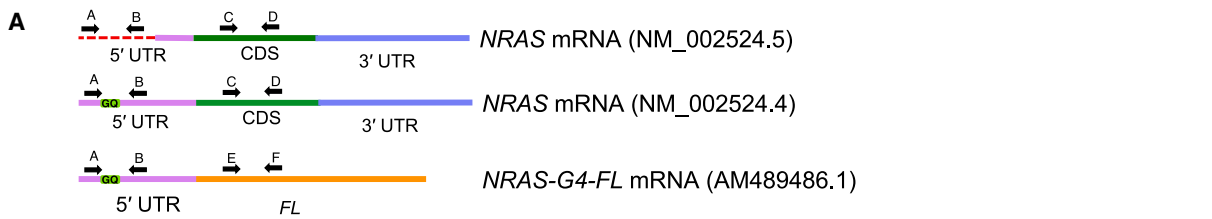
To experimentally define the *NRAS* mRNA 5' UTR, we performed a series of RT-PCR experiments. We designed primer pairs A/B that specifically bind and amplify the G4-containing 5' UTR of the *NRAS* mRNA and C/D pair that bind and amplify the part of the coding region of the *NRAS* mRNA (Figure 5A). The chimeric luciferase construct transcript (*NRAS*-G4-FL) acted as a positive control for this experiment as it contains the G4 in its 5' UTR. For this control, E/F primers were also designed to amplify the coding region of the luciferase mRNA. As shown in Figure 5B and 5C, 14 different cell lines were analyzed, all of which contained measurable levels of *NRAS* mRNA. In the control system (*NRAS*-G4-FL), both the G4-containing and coding sequence primers were amplified equally well. However, all the cell lines examined showed only trace amplification of the G4-containing 5' UTR (Figure 5B and 5C, indicated with red arrows), while in all cases the coding sequence primers were amplified with similar intensity to the control *NRAS*-G4-FL. In addition, we performed RT-qPCR assays on these amplicons from all 14 cell lines. We observed that less than 1% of the transcripts contain the G4 in all cases investigated (Figure S6A). These results indicate that the majority of *NRAS* transcripts are devoid of a G4 structure in the 5' UTR. To further confirm this observation, we performed 5' RACE experiments to map the 5' end of the *NRAS* mRNA in HEK-293 cells.⁵⁹ RACE is a powerful PCR-based technique for the rapid characterization of the 5' end of mRNAs and the start of transcription. As shown in Figure 5D, RACE experiments confirm that the predominant *NRAS* transcript in HEK-293 cells lacks a G4 structure and the G4-forming regions are upstream from the TSS. To further validate these observations, we analyzed publicly available Cap Analysis of Gene Expression Sequencing (CAGE-Seq) datasets of the *NRAS*

(B) Histogram representing the level of *NRAS* mRNA relative to *GAPDH* after **18** treatments in SK-MEL-2 and MCF-7 cells. The data were generated by qPCR analysis. The results are presented as the mean \pm SEM (n = 3) of three independent experiments.

(C and D) Volcano plot of differential gene expression analysis (DeSeq2) with 1 μ M compound **18** treatment in MCF-7 cells and SK-MEL-2 cells. Cells treated with DMSO act as a control. All analyses performed in 3 independent replicate samples.

(E) Western blot of *NRAS* protein levels after compound **18** treatment in SK-MEL-2 cells and MCF-7 cells.

(F) Relative level of *NRAS* protein levels compared to *GAPDH* in SK-MEL-2 cells and MCF-7 cells after **18** treatment, quantified by densitometry analysis. The statistical significance was calculated by t-test analysis. (n = 3, *p < 0.05). Volcano plots indicating differentially expressed proteins from global proteomics assays performed on (G) MCF-7 cells and (H) SK-MEL-2 cells treated with **18** (1 μ M) for 48 h.



(legend on next page)

transcript from several cell lines. CAGE-Seq is used to accurately annotate the 5' end of RNAs carrying a cap site and utilizes a "cap-trapping" technology. CAGE-Seq data also confirm that the major *NRAS* transcript lacks a G4 structure and the TSS is located downstream of the G4 (Figure S6B).

To further investigate the existence of G4s in *NRAS* transcripts in biological contexts, we analyzed genome-wide rG4-seq (G4s in RNA) data reported by the Balasubramanian group.⁴³ We observed that the G4-forming region in the *NRAS* transcript has lower mismatch rate in both samples K⁺ and pyridostatin (PDS)-stabilized datasets compared to Li⁺ dataset, indicating the absence of a G4 in the *NRAS* 5'-UTR (Figure S6C). For comparison, we analyzed the G4-seq data (G4s in DNA) also reported by Balasubramanian.⁴⁷ In this case, a higher mismatch rate was observed in both K⁺ and PDS-stabilized datasets, indicating a DNA G4 in the *NRAS* promoter region, upstream of the predominant TSS (Figure S6D).

DISCUSSION

Direct targeting of *NRAS* at the protein level has proven to be highly challenging due to a lack of suitable small molecule binding pockets. Here, we evaluate an alternative approach to control *NRAS* expression via targeting a structured region in the 5' UTR of *NRAS* mRNA with small molecules. The development of such molecules is an attractive approach to control *NRAS* expression at the post-transcriptional level. We utilized an SMM screening method to identify small molecules that bind to an rG4 structure reported to be within the *NRAS* 5' UTR. The best compound, **18**, identified through SAR studies showed reversible binding to the *NRAS*-G4 structure with submicromolar equilibrium dissociation constants in multiple orthogonal biophysical assays. Further, weaker or no binding was observed in a variety of other RNA and DNA G4 structures. Structural analysis using RNase A probing indicated that the lead compound binds to a site near C12 and C17, which was shown to be on neighboring loops by X-ray crystallography.

The effects of **18** on the translational efficiency of a reporter gene containing the *NRAS* 5' UTR were confirmed by *in vitro* translation assays. Compound **18** showed dose-dependent translation inhibition in a wild type luciferase construct (*NRAS*-G4-FL) but not in a G4-deletion control (*NRAS*-G4-Del-FL). Using SHAPE-MAP, we demonstrated that the G4 structure folds in the context of this reporter construct (*NRAS*-G4-FL). Compound **18** caused an increase in the melting temperature of the G4 in thermal unfolding experiments, confirming that it stabilizes the *NRAS*-G4 structure upon binding. Taken together, these observations indicate that compound **18** inhibits translation via interactions that stabilize the G4 structure in the *NRAS* 5'UTR. Thus, this study provides a proof of principle that small molecules that bind to structured elements within the 5' UTR of

mRNAs can block translation, consistent with other studies on the *NRAS* mRNA.

More in-depth studies aimed at controlling wild type *NRAS* translation revealed that the compound had only modest effects in cells. RNA-seq and qPCR analysis demonstrated no changes in levels of *NRAS* mRNA, and no major changes in a variety of other G4-associated genes. Additionally, the compound had minimal effects on *NRAS* protein levels in cells, prompting us to investigate the structure of the wild type *NRAS* transcript. We used a series of qPCR experiments with amplicons designed to cover both the G4 region and coding sequence of the *NRAS* mRNA. Remarkably, we observe that in 14 different cell lines, the predominant *NRAS* transcript lacks a G4 in the mRNA, and the majority TSS is downstream of the G4-forming sequence. These qPCR experiments were supported further by 5' RACE analysis, confirming that the predominant *NRAS* transcript lacks a G4 in the 5'-UTR in HEK 293 cells. In addition, analysis of multiple public datasets, including CAGE-seq data, rG4-seq, and G4-seq (for DNA G4s) indicate that the G4-forming sequence is mostly not transcribed and may fold into a G4 in the DNA. Finally, it is worth noting that the updated reference sequence for *NRAS* mRNA (NM_002524.5) also lacks a G4 in the 5' UTR. Thus, efforts to target *NRAS* translation with G4-binding small molecules should focus on the identification of cellular contexts in which the G4 is actively transcribed in this gene, or on other regions of the *NRAS* transcript.

The observations reported here demonstrate the importance of rigorously validating any target in an mRNA 5' UTR when evaluating such sequences as a target for small molecules. We report that in the case of *NRAS*, transcript heterogeneity impacts the ability to pharmacologically target this mRNA, with most transcripts in the cell lines evaluated lacking a G4 within the 5' UTR. Thus, compounds targeting the G4 are not effective modulators of *NRAS* translation in cells since they are unable to target the shorter transcripts. A critical aspect of this work is that it does not rule out the existence of the *NRAS* G4. For example, it is conceivable that under different biological contexts such as different cell lines, tissues, or stimuli, cells may predominantly express longer transcripts that contain the G4. Interestingly, there appear to be other G4-forming sequences in the *NRAS* transcript, which could explain some of the results seen in other literature reports studying G4 structures in the *NRAS* mRNA. More broadly, it has been shown that both UTR length and sequence can be modulated in diseases, highlighting the need for accurate annotation, and understanding, of the impact of UTR sequence/structure in gene regulation. Finally, in a broader sense, the strategy of targeting rG4 sequences or other highly structured regions in 5' UTRs with small molecules remains an attractive strategy to pharmacologically control mRNA translation.

Figure 5. Analysis of the G4-forming region in the 5' UTR of *NRAS*

(A) Schematic representation of *NRAS* mRNAs (NM_002524.4), (NM_002524.5), and chimeric luciferase construct mRNA (*NRAS*-G4-FL). The primer binding regions were marked in black arrow (top). The A/B primer pair aligned in the G4 region in the 5' UTR of *NRAS* mRNA and C/D primer pair aligned in the CDS of *NRAS*. The E/F primer pair specific for CDS of firefly luciferase mRNA.

(B and C) The agarose gel electrophoresis of RT-PCR products of *NRAS* mRNA from 14 different cell lines. Red arrows represent the amplicon of the G4 spanning region in the 5' UTR of *NRAS* mRNA.

(D) 5' RACE analysis of the *NRAS* TSS in HEK-293 cells. The G4 regions marked with red block and are in the promoter region.

Limitations of the study

This study reports that many cell lines contain shorter transcripts that lack a G4 in the 5' UTR of the *NRAS* mRNA, limiting the ability to target *NRAS* translation through this structure. However, further studies could reveal contexts where cells express longer, G4-containing *NRAS* transcripts. Further, although this work represents the first crystal structure of the *NRAS* G4, a structure in complex with the ligand has not yet been solved.

SIGNIFICANCE

Therapeutic targeting of RNA with small molecules is an emerging field, and structured regions of the mRNAs that encode “undruggable” proteins are attractive targets. Among the Ras-family proteins, NRAS remains an unsolved challenge and there are no small molecules capable of targeting the protein directly. Here we report a new class of small molecules that bind to a G4 structure within the 5' UTR of the NRAS mRNA that blocks translation in an *in vitro* system. However, within many different cell lines the predominant NRAS transcript is shorter and lacks this G4-containing sequence. Thus, the strategy of controlling mRNA translation by targeting structured regions within the 5' UTR of “undruggable” proteins remains highly attractive. However, in the specific case of NRAS efforts should focus on targeting regions of the mRNA that are more consistently expressed.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell lines
- METHOD DETAILS
 - SMM screening of a NRAS-G4
 - Water ligand observed gradient spectroscopy (waterLOGSY)
 - Thermal melt assays
 - Fluorescence intensity titration
 - Surface plasmon resonance
 - 2-Aminopurine (2-AP) fluorescence titration
 - RNase A footprinting
 - *In vitro* transcription
 - *In vitro* translation and luciferase assay
 - *In vitro* RNA structure probing
 - Crystallization and structural determination
 - Cell viability assay
 - qPCR
 - Western blotting
 - RNA seq
 - Proteomic analysis
 - RT-PCR

- RT-qPCR
- Molecular docking
- 5'-RACE
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.chembiol.2023.05.004>.

ACKNOWLEDGMENTS

The authors thank Dr. S. Tarasov and M. Dyba (Biophysics Resource, SBL, NCI at Frederick) for assistance with CD-based experiments. M.T.B. is a Lenfant Postdoctoral Fellow of the National Heart, Lung, and Blood Institute (NHLBI). This research was supported (in part) by the Intramural Research Programs of the NIH, National Cancer Institute, Center for Cancer Research (grant 1-BC011970) and the NHLBI, NIH. We thank the Balasubramanian group for the generous gift of luciferase promoters used in *in vitro* translation assays in Figure 3B.

AUTHOR CONTRIBUTIONS

S.B., Z.R.T., D.R.C., K.Y., X.L., C.R.F., and S.S. performed screening, biochemical, biological, and synthetic chemistry experiments described in the manuscript and helped write/edit the manuscript. M.B. performed X-ray crystallographic experiments and helped edit the manuscript. R.J.H. and T.A. performed and analyzed proteomics assays. A.R.F. edited the manuscript. D.I. performed SHAPE and CAGE experiments and edited the manuscript. J.S.S. conceived experiments and wrote/edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 30, 2022

Revised: February 24, 2023

Accepted: May 10, 2023

Published: May 30, 2023

REFERENCES

1. Mortimer, S.A., Kidwell, M.A., and Doudna, J.A. (2014). Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* *15*, 469–479.
2. Cruz, J.A., and Westhof, E. (2009). The dynamic landscapes of RNA architecture. *Cell* *136*, 604–609.
3. Tinoco, I., Jr., and Bustamante, C. (1999). How RNA folds. *J. Mol. Biol.* *293*, 271–281.
4. Lescoute, A., and Westhof, E. (2006). The interaction networks of structured RNAs. *Nucleic Acids Res.* *34*, 6587–6604.
5. Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* *35*, 169–178.
6. McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* *21*, 373–379.
7. Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* *367*, 13–37.
8. Silverman, I.M., Li, F., and Gregory, B.D. (2013). Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. *Plant Sci.* *205–206*, 55–62.
9. Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M., and Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* *485*, 264–268.

- Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **83**, 2850–2854.
- Leppke, K., Das, R., and Barna, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19**, 158–174.
- Davuluri, R.V., Suzuki, Y., Sugano, S., and Zhang, M.Q. (2000). CART classification of human 5' UTR sequences. *Genome Res.* **10**, 1807–1816.
- Pelletier, J., and Sonenberg, N. (1987). The involvement of mRNA secondary structure in protein synthesis. *Biochem. Cell. Biol.* **65**, 576–581.
- Barbacid, M. (1987). Ras GENES. *Annu. Rev. Biochem.* **56**, 779–827.
- Shimizu, K., Goldfarb, M., Perucho, M., and Wigler, M. (1983). Isolation and preliminary characterization of the transforming gene of a human neuroblastoma cell line. *Proc. Natl. Acad. Sci. USA* **80**, 383–387.
- Hall, A., Marshall, C.J., Spurr, N.K., and Weiss, R.A. (1983). Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1. *Nature* **303**, 396–400.
- Fedorenko, I.V., Gibney, G.T., and Smalley, K.S.M. (2013). NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. *Oncogene* **32**, 3009–3018.
- Bos, J.L. (1989). Ras oncogenes in human cancer: a review. *Cancer Res.* **49**, 4682–4689.
- Cox, A.D., Fesik, S.W., Kimmelman, A.C., Luo, J., and Der, C.J. (2014). Drugging the undruggable RAS: mission possible? *Nat. Rev. Drug Discov.* **13**, 828–851.
- Milburn, M.V., Tong, L., deVos, A.M., Brünger, A., Yamaizumi, Z., Nishimura, S., and Kim, S.H. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939–945.
- Pylayeva-Gupta, Y., Grabocka, E., and Bar-Sagi, D. (2011). RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774.
- Samatar, A.A., and Poulikakos, P.I. (2014). Targeting RAS-ERK signaling in cancer: promises and challenges. *Nat. Rev. Drug Discov.* **13**, 928–942.
- Mendoza, M.C., Er, E.E., and Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36**, 320–328.
- Prior, I.A., Lewis, P.D., and Mattos, C. (2012). A comprehensive survey of ras mutations in cancer. *Cancer Res.* **72**, 2457–2467.
- Vu, H.L., and Aplin, A.E. (2016). Targeting mutant NRAS signaling pathways in melanoma. *Pharmacol. Res.* **107**, 111–116.
- Johnson, D.B., and Puzanov, I. (2015). Treatment of NRAS-mutant melanoma. *Curr. Treat. Options Oncol.* **16**, 15.
- Canon, J., Rex, K., Saiki, A.Y., Mohr, C., Cooke, K., Bagal, D., Gaida, K., Holt, T., Knutson, C.G., Koppada, N., et al. (2019). The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223.
- 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.
- Banco, M.T., and Ferré-D'Amaré, A.R. (2021). The emerging structural complexity of G-quadruplex RNAs. *RNA* **27**, 390–402.
- Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Front. Chem.* **4**, 38.
- Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474.
- Mou, X., Liew, S.W., and Kwok, C.K. (2022). Identification and targeting of G-quadruplex structures in MALAT1 long non-coding RNA. *Nucleic Acids Res.* **50**, 397–410.
- Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637.
- Fay, M.M., Lyons, S.M., and Ivanov, P. (2017). RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.* **429**, 2127–2147.
- Millevoi, S., Moine, H., and Vagner, S. (2012). G-quadruplexes in RNA biology. *WIREs RNA* **3**, 495–507.
- Kharel, P., Balaratnam, S., Beals, N., and Basu, S. (2020). The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *WIREs RNA* **11**, e1568.
- Cammass, A., and Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* **45**, 1584–1595.
- Beaudoin, J.D., and Perreault, J.P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **38**, 7022–7036.
- Song, J., Perreault, J.-P., Topisirovic, I., and Richard, S. (2016). RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **4**, e1244031.
- Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221.
- Peng, W., Sun, Z.Y., Zhang, Q., Cheng, S.Q., Wang, S.K., Wang, X.N., Kuang, G.T., Su, X.X., Tan, J.H., Huang, Z.S., and Ou, T.M. (2018). Design, synthesis, and evaluation of novel p-(methylthio)styryl substituted quindoline derivatives as neuroblastoma RAS (NRAS) repressors via specific stabilizing the RNA G-quadruplex. *J. Med. Chem.* **61**, 6629–6646.
- Kawauchi, K., Sugimoto, W., Yasui, T., Murata, K., Itoh, K., Takagi, K., Tsuruoka, T., Akamatsu, K., Tateishi-Karimata, H., Sugimoto, N., and Miyoshi, D. (2018). An anionic phthalocyanine decreases NRAS expression by breaking down its RNA G-quadruplex. *Nat. Commun.* **9**, 2271.
- Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635.
- Morioka, M.S., Kawaji, H., Nishiyori-Sueki, H., Murata, M., Kojima-Ishiyama, M., Carninci, P., and Itoh, M. (2020). Cap analysis of gene expression (CAGE): a quantitative and genome-wide assay of transcription start sites. *Methods Mol. Biol.* **2120**, 277–301.
- Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. USA* **100**, 15776–15781.
- 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.
- Connelly, C.M., Abulwerdi, F.A., and Schneekloth, J.S., Jr. (2017). Discovery of RNA binding small molecules using small molecule microarrays. *Methods Mol. Biol.* **1518**, 157–175.
- White, E.W., Tanius, F., Ismail, M.A., Reszka, A.P., Neidle, S., Boykin, D.W., and Wilson, W.D. (2007). Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophys. Chem.* **126**, 140–153.
- Prado, E., Bonnat, L., Bonnet, H., Lavergne, T., Van der Heyden, A., Pratviel, G., Dejeu, J., and Defrancq, E. (2018). Influence of the SPR experimental conditions on the G-quadruplex DNA recognition by porphyrin derivatives. *Langmuir* **34**, 13057–13064.
- Calabrese, D.R., Connelly, C.M., and Schneekloth, J.S., Jr. (2019). Ligand-observed NMR techniques to probe RNA-small molecule interactions. *Methods Enzymol.* **623**, 131–149.
- Moore, M.J.B., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanius, F.A., Wilson, W.D., and Neidle, S. (2006). Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the

- 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J. Med. Chem.* **49**, 582–599.
53. Cuchillo, C.M., Nogués, M.V., and Raines, R.T. (2011). Bovine pancreatic ribonuclease: fifty years of the first enzymatic reaction mechanism. *Biochemistry* **50**, 7835–7841.
 54. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616.
 55. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682.
 56. Bedrat, A., Lacroix, L., and Mergny, J.L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759.
 57. Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Štátný, J., and Mergny, J.-L. (2019). G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* **35**, 3493–3495.
 58. Steri, M., Idda, M.L., Whalen, M.B., and Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdiscip. Rev. RNA* **9**, e1474.
 59. Schramm, G., Bruchhaus, I., and Roeder, T. (2000). A simple and reliable 5'-RACE approach. *Nucleic Acids Res.* **28**, E96.
 60. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675.
 61. Incarnato, D., Morandi, E., Simon, L.M., and Oliviero, S. (2018). RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res.* **46**, e97.
 62. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E., and Weeks, K.M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965.
 63. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
 64. Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J.M., and Incarnato, D. (2020). Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **48**, 12436–12452.
 65. Kabsch, W. (2010). *Xds. Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
 66. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
 67. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
 68. Trajkovski, M., Endoh, T., Tateishi-Karimata, H., Ohya, T., Tanaka, S., Plavec, J., and Sugimoto, N. (2018). Pursuing origins of (poly)ethylene glycol-induced G-quadruplex structural modulations. *Nucleic Acids Res.* **46**, 4301–4315.
 69. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
 70. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367.
 71. 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.
 72. Camacho Londoño, J., and Philipp, S.E. (2016). A reliable method for quantification of splice variants using RT-qPCR. *BMC Mol. Biol.* **17**, 8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-NRAS antibody-N-terminal	abcam	RRID: AB_188369
GAPDH monoclonal antibody	ThermoFisher	Cat# MA5-15738; RRID: AB_10977387
rabbit anti-mouse IgG	abcam	RRID: AB_6728
Bacterial and virus strains		
DH5 α competent cells	In-house	N/A
Chemicals, peptides, and recombinant proteins		
Chemical compounds (Tables S2 and S3)	This paper	ChemDiv and ChemBridge
DNase I	Omega	Cat# E1091-02
Perfecta SYBR Green Super-Mix	Quanta Biosciences	Cat# 95054
NAI	In-house	N/A
RIPA buffer	Thermo Fisher	Cat# 89900
oligo-dT coated magnetic beads	New England BioLabs	Cat# S1419S
Platinum SuperFi II DNA polymerase	Thermo Fisher	Cat# 12361010
Cap-Clip Acid Pyrophosphatase	Fisher Scientific	Cat# C-CC15011H
Trypsin/LysC	Thermo Fisher	Cat# A40007
TURBO DNase	Thermo Fisher	Cat# AM2239
SuperaseIN RNase inhibitor (Ambion)	Thermo Fisher	Cat# AM2694
rSAP	New England BioLabs	Cat# M0371
Western Blotting Luminol Reagent	Cell Signaling Technology	Cat# 7003
EasyPep lysis buffer	Thermo Fisher	Cat# A40006
N-methyl-L-valine	In-house	N/A
RNase A (Ambion)	Thermo Fisher	Cat# AM2270
DMEM	Thermo Fisher	Cat# 12491023
MTT	Thermo Fisher	Cat# M6494
Critical commercial assays		
mMessage mMachine T7	ThermoFisher Scientific	Cat# AM1344
Cell-free translation system of rabbit reticulocyte lysates	Promega	Cat# L4960
NEBNext® Ultra™ II DNA Library Prep Kit	New England Biolabs	Cat# E7103
E.Z.N.A.® Micro-Elute Total RNA Kit	Omega	Cat# R6831-01
qScript cDNA Super-Mix kit	Quanta Biosciences	Cat# 95048-025
Superscript first-strand synthesis kit	Invitrogen	Cat#11904018
Agilent RNA 6000 Nano kit.	Agilent	Cat# 5067-1511
Illumina TruSeq Stranded mRNA Library Kit	Illumina	Cat# 20020595
NEBNext® Multiplex Small RNA Library Prep Kit	New England Biolabs	Cat# E7300S
Deposited data		
Raw and analyzed RNA seq data and SHAPE data	This paper	GEO: GSE191144
Raw and analyzed proteomic data	This paper	Data: MassIVE database and accession https://doi.org/10.25345/C5VX0673N
NRAS-G4 (apo) structure	This paper	PDB: 7SXP
Experimental models: Cell lines		
MCF-7	ATCC	Cat# HTB-22
SK-MEL-2	ATCC	Cat# HTB-68

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Continued		
Oligonucleotides		
Oligonucleotides sequences for SMM (Table S1)	This paper	N/A
RNA and DNA G4 sequences for SMM (Table S1)	This paper	N/A
Primers for qPCR (Table S5)	This paper	N/A
Primers for RT-PCR and RT-qPCR (Table S5)	This paper	N/A
Recombinant DNA		
pSKC11-UTRQ	Kumari et al. ⁴⁰	Addgene; Cat#110490
pSKC12-DelQ	Kumari et al. ⁴⁰	Addgene; Cat#110493
Software and algorithms		
ImageJ	Schneider et al. ⁶⁰	https://imagej.net/ij/index.html
GraphPad Prism 7.0	GraphPad Software, Inc.	https://www.graphpad.com/scientificsoftware/prism/
BIAevaluation 4.0 software	GE Healthcare	N/A
RNA-Seq pipeline	NIH (https://hpc.nih.gov)	https://github.com/CCBR/Pipelinier
Proteome Discoverer 2.4	ThermoFisher	https://thermo.flexnetoperations.com/control/thmo/login
Enrichr	maayanlab.cloud/Enrichr/	https://maayanlab.cloud/Enrichr/
ICM Molsoft version 3.9-2d	icm-browser64-3.9-3a.msi	https://molsoft.com/getbrowser.cgi?product=icm&act=list
RNA Framework	Incarnato et al. ⁶¹ Siegfried et al. ⁶²	https://github.com/dincarnato/RNAFramework
Phaser crystallographic software	McCoy et al. ⁶³	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents used in this study should be directed to the Lead Contact, John Schneekloth (schneeklothjs@mail.nih.gov)

Materials availability

All cell lines, plasmids, and other stable reagents generated in this study are available from the [lead contact](#) with a completed Materials Transfer Agreement.

Data and code availability

- Raw sequencing data including RNA seq, invitro structure map and 5'RACE have been deposited at GEO and are publicly available as of the date of publication. Accession number (GSE191144) is listed in the [key resources table](#). The proteomics data generated in this study available at MassIVE database (massive.ucsd.edu) and accession <https://doi.org/10.25345/C5VX0673N> included in the [key resources table](#). The NRAS-G8U crystal structure was deposited in the PDB possessing the accession code 7SXP and is listed in the [key resources table](#).
- All data reported in this paper will be shared by the [lead contact](#) upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

MCF-7 cells (derived from 69-year-old white, female with Adenocarcinoma) and SK-MEL-2 cells (derived from 60-year-old white, male with malignant melanoma) were grown in Dulbecco's modified Eagle's medium (DMEM) with 1% glutamine, 10% fetal bovine serum and 1% antibiotics (streptomycin and penicillin) at 37°C in 5% CO₂ in a humidified incubator according to ATCC's recommendations. Cells were grown in 96 well plates (for MTS assay) or 6-well plates for (RT-PCR, qPCR and western blotting).

METHOD DETAILS

SMM screening of a NRAS-G4

Small molecule microarray screening was carried out as previously described.⁴⁸ Briefly, γ -aminopropyl silane (GAPS) microscope slides (Corning) were functionalized with a short Fmoc-protected amino polyethylene glycol spacer. After deprotection using piperidine, 1,6-diisocyanatohexane was coupled to the surface by urea bond formation to provide functionalized isocyanate-coated microarray slides that can react with primary alcohols and amines to form immobilized chemical screening libraries. A total of 26,227 unique small molecule stock solutions (10 mM in DMSO) from MIPE and NCI Diversity set V screening collections, in addition to dyes and controls, were printed in duplicate onto one slide and exposed to pyridine vapor in a vacuum desiccator overnight to facilitate covalent attachment to the slide surface. After drying, slides were incubated with a 1:20 polyethylene glycol:DMF (v/v) solution to quench unreacted isocyanate surface. The 5'-AlexaFluor647-NRAS-G4 RNA was dissolved in 10 mM Tris buffer (pH 6.9) with 100 mM KCl, diluted to 5 μ M. Then the sample was annealed by heating to 95°C for 3 min, followed by slowly cooling to room temperature for 1 hour. The annealed RNA was then further diluted to 1 μ M in 10 mM Tris buffer (pH 6.9) with 100 mM KCl for screening. Next, printed slides were incubated with the RNA at a concentration of 1 μ M for 2 hours at room temperature. Following incubation, slides were gently washed three times for 2 min in 10 mM Tris buffer (pH 6.9) with 100 mM KCl, once in deionized water, and dried by centrifugation for 2 min at 4000 rpm. Fluorescence intensity was measured (650 nm excitation, 670 nm emission) on an Innopsys Innoscan 1100 AL Microarray Scanner with a resolution of 5 μ m. The scanned image was aligned with the corresponding GenePix Array List (GAL) file to identify individual features. Hits were identified based on signal-to-noise ratio (SNR), defined as (mean foreground – mean background)/standard deviation of background. The Z-score is defined as: $Z = (\text{mean SNR}_{635}(\text{compound}) - \text{mean SNR}_{635}(\text{library})) / (\text{SD SNR}_{635}(\text{library}))$ with the following criteria: (i) SNR > 0, (ii) Z score > 3, (iii) coefficient of variance (CV) of replicate spots < 100, (iv) SNR of negative control slide < 1 and (v) visual comparison of intensity with other nucleic acid structures screened. To further measure selectivity, the Z-score for each selected compound is compared across many different SMM screens (in this case 20 different oligonucleotides, (Table S1)).

Water ligand observed gradient spectroscopy (waterLOGSY)

A reference 1D-¹H and 1D WaterLOGSY spectrum of 100 μ M N-methyl-L-valine and 100 μ M compound **1** was collected, followed by a separate sample containing 5 μ M NRAS-G4 RNA oligo, 100 μ M N-methyl-L-valine, and 100 μ M compound **1**. NRAS-G4 RNA (UGUGGGAGGGCGGGUCUGGG) was buffer exchanged into 10 mM Tris-d11 buffer (pH 6.9, containing 100 mM KCl) using centrifugal filtration (3 kDa MWCO, EMD Millipore) and were annealed by heating to 95 °C for 3 min, followed by slowly cooling to room temperature for 1 hour. A sample of compound **1** and N-methyl-L-valine, each at 100 μ M, was prepared in 10 mM Tris-d11 buffer (pH 6.9, containing 100 mM KCl and 5% DMSO-d₆), and 1D reference proton and WaterLOGSY spectra with and without oligonucleotide were recorded. These spectra were recorded at 20°C on a Bruker AVANCE III 500 MHz spectrometer equipped with TCI cryogenically cooled probe. The “zgesgp” excitation sculpting water suppression pulse sequence from Bruker was used for data acquisition with 128 scans. All data were processed and visualized with MestReNova software (Version 8.1.2–11880).

Thermal melt assays

The 5 μ M NRAS-G4 RNA was folded in 10 mM Tris (pH=6.9) and 1 mM KCl by heated at 95°C for 3 min and cooled to room temperature over 1 hour. Then compound **1** (2.5 μ M, 5 μ M and 10 μ M) or compound **18** (10 μ M) was added into the folded RNA samples and incubated for 15 min in room temperature. For the control sample 5% DMSO was added into the folded RNA sample. Then, thermal stability of the NRAS-G4 oligonucleotide with and without compounds was determined by heating from 25 to 97 °C at 1 °C/min in a 0.1 cm quartz cuvette using an Aviv Biomedical Model 420 Circular Dichroism (CD) Spectrometer equipped with a ThermoCube temperature regulator. To calculate the T_m of each sample, ellipticity (folded fraction) was plotted as a function of temperature and fit in GraphPad Prism 7 software using a nonlinear sigmoidal dose-response model with a variable slope. Each condition was performed in triplicate, with ΔT_m values calculated using $T_m(\text{+compound}) - T_m(\text{apo})$ and then averaged to yield the final value.

Fluorescence intensity titration

AlexaFluor 647-labeled NRAS-G4 RNA and other appropriate oligonucleotides were heated at 95°C for 3 min, allowed to cool to room temperature for 1 hour, and diluted to 50 nM in 10 mM Tris buffer (pH 6.9, containing 100 mM KCl). Compound was added as a solution in buffer containing 5% DMSO, and the sample was allowed to equilibrate for 15 min. Fluorescence intensity spectra were recorded at room temperature using a Photon Technology International, Inc. QuantaMaster 600TM Spectrofluorometer equipped with Felix GX 4.2.2 software. Fluorescence intensity was recorded at an excitation wavelength of 645 nm, with the resulting emission spectrum recorded from 650 to 800 nm. Total area under the peak from 655–700 nm was quantified and was then normalized to the values obtained for RNA incubated with a DMSO control. Normalized fluorescence for three independent replicates were averaged and plotted against small molecule concentration. K_D values were determined using a single site-binding model.

Surface plasmon resonance

SPR was conducted using a Biacore 3000 (Biacore, Inc) instrument. A CM5 SPR biochip was loaded into the system and primed with running buffer (10 mM Tris, pH 6.9, 100 mM KCl, 0.005% Tween 20, 5% DMSO). To make the surface, the flow rate was set at 5 μ L/min. Then both Flow Cell (Fc) 1 and 2 were activated by EDC/NHS (0.4 M/0.1 M) aqueous solution for 15 min, followed

with an injection of streptavidin (SA) solution (0.2 mg/mL in 10 mM sodium acetate buffer, pH4.5) for 30 min. After the immobilization amount of SA reached 8,000~10,000 RU, the surface was deactivated by flowing 1 M ethanolamine aqueous solution (pH 8.5) for 10 min and regenerated with 10 mM NaOH for 2 min to remove the unbound SA. Then, 5 μ M folded 5'-biotin *NRAS*-G4 RNA in 10 mM Tris, 100 mM KCl, pH 6.9 was immobilized on Fc 2 of the SPR chip to a density of 1400 RU. Each of the compound solutions were prepared in non-DMSO running buffer by dilution, resulting in a final concentration of 5% DMSO. Then, 50 μ L of compound solution was injected at a flow rate of 30 mL/min in Fc 1-2 flow path for 120 sec. for association, followed by 200 sec. of buffer, for dissociation. The final binding curve was obtained by reference subtraction. To determine the binding affinity (K_D), a series of diluted compound solutions were injected, and K_D was calculated by BIAevaluation 4.0 software (GE Healthcare) using Langmuir 1:1 binding model.

2-Aminopurine (2-AP) fluorescence titration

Fluorescence titrations were performed according to previously reported protocols. Deprotected *NRAS*-G4 RNA containing a 2-AP label substituted with either C12 (12C-2AP-*NRAS*-G4: UGUGGGAGGGG(2AP)GGGUCUGGG) or U16 (16U-2AP-*NRAS*-G4: UGUGGGAGGGGCGGG(2AP)CUGGG) in the loop was purchased from Integrated DNA technology, and dissolved in 10 mM Tris, 100 mM KCl, pH 6.9 to a concentration of 100 μ M. The RNA was annealed by briefly heating to 95 °C for 3 min, followed by cooling to RT for 1 h, and was subsequently diluted to 10 μ M in 10 mM Tris, 100 mM KCl, pH 6.9. Compound **18** solutions were prepared as serial dilutions in DMSO. In a black 96-well plate, **18** was diluted to final concentrations ranging from 0.05–50 μ M in triplicate in 10 mM Tris, 100 mM KCl, pH 6.9 (5% final DMSO concentration), and the 12C-2AP-*NRAS*-G4 or 16U-2AP-*NRAS*-G4 RNA were added at a final concentration of 2 μ M. Background fluorescence of **18** was assessed in wells containing compound at each final concentration in 10 mM Tris, 100 mM KCl, pH 6.9 in the absence of RNA. After delivery of the RNA, the plate was centrifuged (1000 rpm, 2 min) and allowed to incubate for 30 min at RT with shaking. Fluorescence was then measured on a Synergy Mx microplate reader (BioTek) with an excitation wavelength of 310 nm and an emission wavelength of 365 nm. Background fluorescence of **18** in the absence of RNA was subtracted from wells containing RNA, and the apparent dissociation constants were determined by fitting a sigmoidal dose-response curve to the mean fluorescence of background subtracted triplicate measurements.

RNase A footprinting

The 5'-end-AlexaFluor-647-labeled *NRAS*-G4 RNA (5 μ M) was folded in 10 mM Tris pH 6.9, 100 mM KCl buffer by heating at 95 °C for 3 min and was cooled to room temperature over 1 hour. Then, increasing concentrations of compound **18** (5 μ M to 100 μ M) were added to the folded RNA samples and incubated for 15 min at room temperature. The folded RNAs were digested with 0.0001 μ g of RNase A (Ambion) for 3 min at room temperature. The reactions were terminated by heating at 95 °C for 5 min with an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl (pH 7.5), and 0.1 mM EDTA). Samples were electrophoresed on a 17% denaturing PAGE. The gel images were obtained by scanning the PAGE gel on a Typhoon Imager (Amersham) and bands were quantified using Image J software.

In vitro transcription

The plasmids pSKC11 and pSKC12, which encode the transcripts *NRAS*-G4-FL and *NRAS*-G4-Del-FL, respectively, were received from Balasubramanian group⁴⁰. The plasmids were linearized at the 3' end using EcoRI, and linearized plasmids were used as a template for *in vitro* transcription. The 5'-capped transcripts were synthesized *in vitro* using mMessage mMachine T7 (ThermoFisher Scientific), followed by incubation with DNase for 15 minutes at 37 °C to remove residual template DNA. All the transcripts were purified using Monarch RNA Cleanup Kit (NEB) and the concentration was determined using a Nanodrop. Integrity and size of each transcript was confirmed using a 1% agarose gel.

In vitro translation and luciferase assay

In vitro translation of *NRAS*-G4-FL and *NRAS*-G4-Del-FL transcripts in the presence of compound **18** were carried out in a cell-free translation system of rabbit reticulocyte lysates (Promega) following the manufacturer's protocol. Briefly, 500 ng of *NRAS*-G4-FL and *NRAS*-G4-Del-FL transcripts were folded at 95 °C for 5 min and slowly cooled down to room temperature for 1 hour. Then compound **18** (1 μ M, 10 μ M and 20 μ M) or 5% DMSO were added to the folded RNA and incubated for 15 minutes at room temperature. *In vitro* translation was carried out by adding the reticulocyte lysate into the samples and incubating at 30 °C for 90 min. Firefly luciferase activity was measured using luciferase assay reagent (Promega) according to the manufacturer's protocol on a Synergy Mx microplate reader (BioTek).

In vitro RNA structure probing

In vitro RNA folding and probing were carried out as previously described.⁶⁴ Briefly, 2 μ g of *NRAS*-G4-FL RNA in 89 μ L of nuclease-free H₂O, was heated at 95 °C for 2 min, then immediately transferred to ice for 1 min. Then 10 μ L of ice-cold 10X RNA Folding Buffer (500 mM HEPES pH 7.5; 1 M KCl) was added, the solution was mixed, and then incubated at 37 °C for 15 min. Then, 1 μ L of 1M MgCl₂, pre-warmed at 37 °C, was added, the solution was mixed, and further incubated for 15 min. Probing was performed by adding 11 μ L of NAI (1M stock), incubating at 37 °C with moderate shaking for 10 min. For the control reaction, 11 μ L of neat DMSO was added. The reaction was quenched by adding 111 μ L of 1M DTT. RNA was recovered by purification on a Monarch RNA 10 μ g column (New England Biolabs).

For the library preparation, both DMSO- and NAI-treated samples, were fragmented by incubating at 94°C for 8 min, in a buffer containing a final concentration of 4mM MgCl₂. The fragmented RNA was then subjected to random-primed reverse transcription, as per standard SHAPE-MaP conditions. Briefly, fragmented RNA was mixed with 1 μL of 10 μM random decamers, and 1 μL of 10mM dNTPs, then incubated at 70°C for 5 min, and immediately transferred to ice for 1 min. Then, 4 μL of 5X RT Buffer (250mM Tris pH 8.0; 375mM KCl), 2 μL of DTT (0.1M), 1 μL of Superscript RNase inhibitor (Ambion), 1 μL of SuperScript II RT (ThermoFisher Scientific), and 1 μL 120mM MnCl₂ were added, the reactions were mixed, and incubated at 25°C for 10 min, followed by 2 hours at 42°C. After reaction cleanup, the buffer was replaced with the standard SuperScript II RT First Strand Buffer with MgCl₂, and the RNA-cDNA hybrids were used as input for the NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs). After having converted the RNA into dsDNA, the remainder of the library preparation was carried out using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs), as per manufacturer instructions.

All relevant data analysis steps were conducted using the RNA Framework (<https://github.com/dincarnato/RNAFramework>).⁶¹ SHAPE raw reactivities were calculated via the rf-norm module, by using the previously published method,⁶² followed by box-plot normalization (**GEO accession codes**: GSE191144).

Crystallization and structural determination

The *NRAS*-G4 variant, consisting of a G8U mutation (*NRAS*-G8U), was chemically synthesized by Dharmacon, and purified using denaturing polyacrylamide gel electrophoresis. The *NRAS*-G8U construct used for crystallization comprised of 22 nucleotides (5'-UGUGGGAUGGGCGGGUCUGGA-3'). Prior to crystallization, the *NRAS*-G8U was heated to 95°C for 1.5 min and allowed to cool overnight in 50 mM potassium chloride and 25 mM sodium cacodylate at pH 6.8. Crystals of *NRAS*-G8U were produced using the vapor diffusion method at 21°C by combining 1 μL each of the folded RNA and reservoir solution (20–25% PEG 3350, 80 mM sodium chloride and 100 mM BIS-TRIS pH 6.5). After two months, small tetragonal bipyramidal crystals (50 × 50 × 150 μm) were supplemented with 30% PEG 3350 and plunged into liquid nitrogen. X-ray diffraction data were collected at a wavelength of 1.104 Å at beamline 17-ID-C located at the Advanced Photon Source (Argonne National Laboratory), and reduced using the XDS suite,⁶⁵ AIMLESS⁶⁶ and POINTLESS.⁶⁷ The *NRAS*-G4 structure was solved by molecular replacement,⁶³ using an NMR solution structure of a parallel DNA G4⁶⁸ (PDB: 5NYS) as a search model (Top LLG = 74.4, Top TFZ = 8.7). The final model was produced from iterative cycles of rebuilding in COOT⁶⁹ with interspersed refinements using Phenix.refine.⁷⁰ The following refinement strategy was implemented: torsion-angle simulated annealing, XYZ refinement, and individual isotropic B-factor. Crystallographic and refinement statistics for *NRAS*-G8U can be found in Table S4. The *NRAS*-G8U structure was deposited in the PDB under accession code PDB: 7SXP.

Cell viability assay

MCF-7 and SKE-MEL-2 cells were plated in 96-well culture plates at a density of 1 × 10⁴ cells/mL for MCF-7 cells, and 2.5 × 10⁴ cells/mL for SK-MEL-2 cells. After 24-h incubation, varying concentrations of compound **18** or DMSO (control) were added to the cells and further incubated for 48 hours. Following incubation, 3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was directly added into the wells and incubated again at 37°C with 5% CO₂ for 4 hours. The ability of cells to form formazan crystals by active mitochondrial respiration was determined using a Microplate reader at 540nm, after dissolving the crystals in an SDS-HCl solution. A blank measurement of the wells with media only was taken and subtracted accordingly. The cell viability percentages were calculated by normalizing against the untreated control cells. The IC₅₀ (inhibitory concentration to produce 50% cell death) values were determined by fitting the data in dose–response curves in Graphpad Prism. Data were presented as the mean ± SEM.

qPCR

MCF-7 and SKE-MEL-2 cells were seeded in 6-well culture plates at a density of 1 × 10⁶ cells/mL for MCF-7 cells, and 2.5 × 10⁶ cells/mL for SK-MEL-2 cells. After 24-hour incubation, different doses of compound **18** (1, 5, 10 and 25 μM) or DMSO (control) were added to the cells and further incubated for 48 hours. Total RNAs were isolated using E.Z.N.A.® MicroElute Total RNA Kit (Omega), and DNA was removed by on-membrane DNase I (Omega) digestion according to the manufacturer's protocol. Then 1 μg of RNA was used for cDNA synthesis. The cDNA was synthesized using qScript cDNA SuperMix (Quanta Biosciences) and oligo d(T) primer according to the manufacturer's protocol. The 20 μL reactions were incubated in a thermocycler (MiniAmp plus Applied Biosystem) for 5 min at 22°C, 30 min at 42°C, 5 min at 85°C, and then held at 4°C. 60 ng of the resulting cDNAs were subjected to qPCR using a Perfecta SYBR Green Super Mix (Quanta Biosciences) on an Eppendorf Mastercycler RealPlex2 in the presence of appropriate primers. The reactions were incubated at 95°C for 10 min, followed by 40 cycles of 95°C for 30 sec, 60°C for 30 sec and 72°C for 20 sec. The threshold cycle value (C_T) is the first cycle that shows a detectable increase in fluorescence due to the formation of PCR products and was used to determine the template amount in each sample. The relative fold change in expression was measured using the Livak method and were normalized to the relative level of mRNA in the control experiment.⁷¹ For example, to calculate the ΔΔ(C_T) between each gene of interest and the average of the control samples: Δ(C_T) = C_T(*NRAS*) – C_T(*GAPDH*); ΔΔ(C_T) = ΔC_T(treatment) – ΔC_T(control); fold change = 2^{-ΔΔ(C_T)}. The primers used are shown in the Table S5.

Western blotting

MCF-7 and SKE-MEL-2 cells were seeded in 6-well culture plates at a density of 1 × 10⁶ cells/mL for MCF-7 cells, and 2.5 × 10⁶ cells/mL for SK-MEL-2 cells. After 24-hour incubation, different doses of compound **18** (1, 5, 10 and 25 μM) or DMSO (control) were added

to the cells and further incubated for 48 hours. Proteins were extracted from the cells with RIPA buffer (RIPA, sodium orthovanadate, PMSF, protease inhibitor, and phosphatase inhibitors A and B), vortexed to homogenize, and sonicated with intervals of 1s on, 30 seconds off, for 1 minute. The amount of protein was quantified by a standard Bradford protocol. Then, 35 μg of protein was loaded into each well of 4–12% Bis-Tris Gels (Novex) and electrophoresed at 180 V for 60 min. Ponceau staining (Thermo Scientific) was performed to confirm equal loading and transfer. Blots were blocked with 1x Blocking buffer (thermo Scientific) for 1 hour and washed three times in 1X TBST for 10 min each. The NRAS and GAPDH proteins were detected by incubating with an NRAS (abcam ab-188369) antibody and GAPDH (ThermoFisher MA5-15738) antibody at 4 °C overnight. Horseradish peroxidase-conjugated rabbit anti-mouse IgG (abcam ab-6728) was used as the secondary antibody at 1:1000 dilutions for GAPDH and NRAS. Proteins were visualized by Western Blotting Luminol Reagent (Cell Signaling Technology 7003) in Image Quant LAS 4000 (GE healthcare).

RNA seq

MCF-7 and SKE-MEL-2 cells were seeded in 6-well culture plates at a density of 1×10^6 cells/mL for MCF-7 cells, and 2.5×10^6 cells/mL for SK-MEL-2 cells. After 24-hour incubation, 1 μM compound **18** or 0.5% DMSO (control) were added to the cells and were further incubated for 48 hours. Total RNAs were isolated using E.Z.N.A.® MicroElute Total RNA Kit (Omega), and DNA was removed by on-membrane DNase I (Omega) digestion according to the manufacturer's protocol. Isolated RNA integrity and quality were determined with a Bioanalyzer (Agilent 2100) by using the Agilent RNA 6000 Nano kit. Then, samples were submitted to the Center for Cancer Research Sequencing Facility (CCR-SF). 500 ng of total RNA was used as the input for mRNA capture with oligo-dT coated magnetic beads. The library was prepared using an Illumina TruSeq Stranded mRNA Library Kit according to manufacturer's protocol and final purified products were quantitated by qPCR before cluster generation and sequencing. Samples were sequenced using a 2x 76 cycles run on a NextSeq500 sequencer. The samples have 21 to 56 million pass filter reads with more than 94% of bases above the quality score of Q30.

Upon generation of FASTQ files by CCR-SF, the sequencing analysis was performed using the Center for Cancer Research Collaborative Bioinformatics Resource (CCBR) RNA-Seq pipeline (<https://github.com/CCBR/Pipelinier>). The CCBP Pipelinier was then deployed on the National Institutes of Health (NIH) HPC Biowulf cluster (<https://hpc.nih.gov>). In brief, the CCBP Pipelinier trims the sequencing adapters using the Cutadapt tool before alignment with the reference genome (Human - hg38) using STAR's two-pass alignment system. The mapping statistics were calculated using Picard software. The average mapping rate of all samples was 95% and unique alignment was above 86%. There were between 3.49–7.70% unmapped reads, and the samples had 0.20% ribosomal bases. Percent coding bases were between 56–61%, percent UTR bases were 32–36%, and mRNA bases were between 90–94% for all the samples. Library complexity was measured in terms of unique fragments in the mapped reads using Picard's MarkDuplicate utility. The samples had 83–87% non-duplicate reads. Gene expression quantification analysis was performed using the RSEM tool, yielding raw and normalized counts. Upon generation of gene counts by RSEM, a CPM (counts per million) filter (CPM < 0.5) was used to remove lowly expressed genes. Then, the DESeq2 package was used to perform comparative analysis and derive fold change in expression and false discovery rate (FDR) adjusted p-values needed for generation of volcano plots. In addition to the output from the CCBP Pipelinier, we applied an additional Transcripts Per Kilobase Million (TPM) filter to only keep genes that have TPM > 3 in more than two samples when comparing an experimental set to DMSO controls. Only genes passing this TPM filter were shown in the volcano plots and used for gene ontology (GO) and pathway analysis.

Enrichment analysis on pathways and GO of significant differentially expressed genes were performed using Enrichr (maayanlab.-cloud/Enrichr). Significantly upregulated genes were identified as genes with fold change > 1.5 compared to DMSO with FDR < 0.05. Significantly downregulated genes were chosen using fold change < -1.5 compared to DMSO with FDR < 0.05. Significant differentially expressed genes were then fed into Enrichr and "KEGG 2021 Human" was chosen under the "Pathways" tab for the pathway enrichment analysis and "GO Biological Process 2021" was chosen under the "Ontologies" tab for GO analysis. Genes with highest ranked Enrichr "Combined score" were explored for further analysis.

GEO accession codes: GSE191144.

Proteomic analysis

MCF-7 and SKE-MEL-2 cells were seeded in 6-well culture plates at a density of 1×10^6 cells/mL for MCF-7 cells, and 2.5×10^6 cells/mL for SK-MEL-2 cells. After 24-hour incubation, 1 μM compound **18** or 0.5% DMSO (control) were added to the cells and were further incubated for 48 hours. Cell pellets were lysed and digested using the EasyPep Kit (Thermo A40006) as described below. Each sample pellet was resuspended in 200 μL of EasyPep lysis buffer and protein concentration was determined by the BCA method. For each sample 20 μg was treated with 50 μL each of reducing solution and alkylating solution provided with the EasyPep kit, incubated at 25 °C for 1 hr, then treated with 20 μL of 100ng/ μL Trypsin/LysC (Thermo A40007). Samples were then incubated at 37 °C with shaking for 24hrs after which point 40 μL of 4.2 $\mu\text{g}/\mu\text{L}$ TMTpro 18-plex (Thermo A52045) reagent was added to each sample and incubated for 1hr at 25 °C with shaking. Excess TMTpro was quenched with 50 μL of 5% hydroxylamine, 20% formic acid for 10min and samples were then combined. Samples were cleaned using EasyPep Mini columns as described in the manual. Eluted peptides were dried in a speed-vac.

To analyze the digested peptides by LC/MS, peptides from each cell line were resuspended in 0.1% FA and analyzed in duplicate using a Dionex U3000 RSLC in front of a Orbitrap Eclipse (Thermo) equipped with an EasySpray ion source and FAIMSTM interface. Solvent A consisted of 0.1% FA in water and Solvent B consisted of 0.1% FA in 80% ACN. The loading pump consisted of Solvent A and was operated at 7 $\mu\text{L}/\text{min}$ for the first 6 minutes of the run, then dropped to 2 $\mu\text{L}/\text{min}$ when the valve was switched to bring the

trap column (Acclaim™ PepMap™ 100 C18 HPLC Column, 3 μ m, 75 μ m I.D., 2cm, PN 164535) in-line with the analytical column (EasySpray C18 HPLC Column, 2 μ m, 75 μ m I.D., 25cm, PN ES902). The gradient pump was operated at a flow rate of 300nL/min. Each of the two injections per cell line consisted of the same LC gradient conditions and global MS parameters, with only the FAIMS compensation voltages (CVs) changed for each method. Each run used a linear LC gradient of 5-7% B for 1 min, 7-30% B for 134 min, 30-50% B for 35 min, 50-95% B for 4 min, holding at 95% B for 7 min, then re-equilibration of analytical column at 300nL/min at 5% B for 17 min. All MS injections employed the TopSpeed method with 4 FAIMS compensation voltages (CVs) and a 0.75 second cycle time for each CV (3 second cycle time total) that consisted of the following: spray voltage was 2200V and ion transfer temperature was 300°C. MS1 scans were acquired in the Orbitrap with a resolution of 120,000, AGC of 4e5 ions, max injection time of 50 ms, and mass range of 350-1600 m/z; MS2 scans were acquired in the Orbitrap using the TurboTMT method with a resolution of 15,000, AGC of 1.25e5, max injection time of 22ms, HCD energy of 30%, isolation width of 1.6Da, intensity threshold of 2.5e4 and charges 2-5 for MS2 selection. Advanced Peak Determination, Monoisotopic Precursor selection (MIPS), and EASY-IC for internal calibration were enabled and dynamic exclusion was set to a count of 1 for 15sec in all methods. The only difference in the methods was the CVs used. Method 1 used CVs of -45, -55, -65, -75 and Method 2 used CVs of -50, -60, -70, -80.

Then, a database search and post-processing analysis was performed. The two MS files for each cell line were searched together with Proteome Discoverer 2.4 using the Sequest node. Data was searched against the Uniprot Human database from Feb 2020 using a full tryptic digest, 2 max missed cleavages, minimum peptide length of 6 amino acids and maximum peptide length of 40 amino acids, an MS1 mass tolerance of 10 ppm, MS2 mass tolerance of 0.02 Da, fixed modifications for TMTpro (+304.207) on lysine and the peptide N-terminus, variable oxidation on methionine (+15.995 Da) and fixed carbamidomethyl (+57.021) on cysteine. Percolator was used for FDR analysis and TMTpro reporter ions were quantified using the Reporter Ion Quantifier node and normalized on total peptide amount. For the final group comparisons only peptides that were observed in 6 or more total samples were included.

Data repositories

MassIVE database (massive.ucsd.edu) and accession <https://doi.org/10.25345/C5VX0673N>.

RT-PCR

All the cell lines (MCF-7, SKE-MEL-2, SiHa, H5518T, DU142, T47D, Bt 549, HEK-293T, HEK-293, ADR, H226, UOK 262, LS174T and HT1080) were grown in T75 flask with proper medium according to the ATCC protocol. Once the cells reached 80% confluence, total RNAs were isolated using TRIzol reagent (ThermoFisher scientific), and DNA was removed by TURBO DNase (ThermoFisher scientific) digestion according to the manufacturer's protocol. Then first strand cDNA was synthesized by Superscript first-strand synthesis kit (Invitrogen) with oligodT primers according to the manufactured protocol. Briefly, 1 μ g of RNA and oligodT were mixed together and heated at 65°C for 5min to denature the RNA. Then samples were placed in ice for 10 min. In a separate tube, 2X reaction mix was prepared by adding the 2 μ L of 10X RT buffer (containing the LiCl), 4 μ L of 25mM MgCl₂, 2 μ L of 0.1M DTT and 1 μ L of RNaseOUT. The 9 μ L of reaction mix was added into the RNA/primer mix samples and incubated at 42°C for 2min. Then 1 μ L of SuperScript II RT was added into the samples and incubated at 42 °C for 50 min, followed by incubated at 70 °C for 15 min to heat inactivate the enzyme. Then 2 μ L of cDNA was used for PCR amplification with Platinum SuperFi II DNA polymerase (Invitrogen). The reaction was incubated at 98°C for 30 sec, followed by 25 cycles of 98°C for 10 sec, 63°C for 35 sec and 72°C for 30 sec. Then samples were loaded into the 2% agarose gel and run for 1hour. The specific set of primers that were used to amplify the 5' UTR and CDS of the NRAS mRNA and CDS of NRAS-G4-FL reporter construct mRNA have shown in the [Table S5](#).

RT-qPCR

All cell lines (MCF-7, SKE-MEL-2, SiHa, H5518T, DU142, T47D, Bt 549, HEK-293T, HEK-293, ADR, H226, UOK 262, LS174T and HT1080) were grown according to the ATCC protocol. Total RNAs were isolated using TRIzol reagent (ThermoFisher scientific), and DNA was removed by TURBO DNase (ThermoFisher scientific) digestion according to the manufacturer's protocol. First strand cDNA was synthesized by Superscript First-Strand Synthesis Kit (Invitrogen) with oligodT primers according to the manufacturer's protocol as described above (RT-PCR section). Then 60 ng of cDNA was subjected to qPCR using a Perfecta SYBR Green Super Mix (Quanta Biosciences) on an Eppendorf Mastercycler RealPlex2 in the presence of appropriate primers. Threshold cycle (C_T) values were used to determine the template amount in each sample. The relative occurrence of G4-containing NRAS mRNA was calculated by using a previously established method.⁷² The primers used are shown in the [Table S5](#).

Molecular docking

ICM Molsoft version 3.9-2d was used to identify the potential binding modes of compound 18 on NRAS-G4. Docking was performed using the NRAS-G4 crystal structure (PDB: 7SXP) and small molecule **18**, with full ligand conformational flexibility and a thoroughness value of 500. The top 100 scored conformations were collected, and the final docked models were selected based on the proximity of ligand to the bases identified in our structure probing experiment.

5'-RACE

To map TSSs, 2 μ g of polyA RNA from HEK293 cells were fragmented to an average size of 200 nt by incubating at 94°C for 5 min in the presence of 4mM MgCl₂. Endogenous 5'P sites and 2'-3' cyclic P sites were resolved by treatment with 1U rSAP (New England Biolabs) in CutSmart buffer at 37°C for 30 min. Capped RNA fragments were then decapped using 5U Cap-Clip Acid

Pyrophosphatase (Cellsript) at 37°C for 1 hour. Treatment with Cap-Clip leaves a 5'P that can be exploited for direct adaptor ligation. Hence, only capped fragments will be ligated to both a 5' and a 3' adapter, and as such, amplified. RNA fragments were then used as input for the NEBNext® Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data were presented as mean \pm SEM for three independent experiments. Statistical analyses and graphical presentations were performed using GraphPad Prism 7.0 software. Statistical analyses were performed by unpaired t-tests. The significance level was defined as $p < 0.05$. All details are provided in the figure legends and/or in the [STAR Methods](#).