G-quadruplex probing in native RNA context using nanopore current signatures

Anya Korsakova[1,2,*,†], Tiphaine Thébault[2,*], Kah Wai Lim[2], Anh Tuân Phan[2,3,†]

[1] Calico Life Sciences LLC, South San Francisco, CA 94080, USA

[2] School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore

[3] NTU Institute of Structural Biology, Nanyang Technological University, 636921, Singapore

* These authors contributed equally

† Correspondence: anya@calicolabs.com, phantuan@ntu.edu.sg

Abstract

G-quadruplex RNA (rG4) structures play crucial roles in cellular regulation, but their detection in long RNA remains challenging. Current methods rely on indirect measurements through chemical modifications of RNA or bulk spectroscopic techniques. Here, we demonstrate direct rG4 detection in native RNA contexts using Oxford Nanopore sequencing. By analyzing nanopore current stalling events, we identify rG4 signatures in synthetic RNA constructs ranging from 131 to 748 nucleotides, in longer native RNA molecules, and in total RNA. We distinguish between canonical, bulged, and irregular rG4 topologies, offering a novel approach for structural RNA analysis without chemical modifications. This method paves the way for high-throughput, direct detection of RNA structures in their native context.

G-quadruplexes (G4) are secondary structures found both in DNA[1] and RNA[2]. G4s are comprised of stacked tetrads of guanines connected with Hoogsteen hydrogen bonds within tetrads and stabilized by a cation in the central channel. There are more than 100,000 potential G-quadruplex-forming sequences in human RNA[3], and at least 3000 mRNAs harbor RNA G4 (rG4) *in vitro*[4]. rG4 formation was found to play regulatory roles in cells[5], is associated with

telomere homeostasis[6], mRNA translation[7] and pre-mRNA splicing[8]. The 5'-UTR G4 formation in human mRNA was shown to inhibit oncogene transcription[3].

Current methods for rG4 detection, including chemical probing[9], spectroscopic techniques, and predictive algorithms, have limitations. They often rely on short oligonucleotides, lack native sequence context, or provide only indirect measurements. Chemical ligands can be applied to long RNAs, but they favor forced G4 formation and are selective. Studying short G4 forming oligonucleotides, on the other hand, lacks a larger sequence context, while predictive algorithms for rG4 are based on limited experimental data. While recent advances in DNA G4 detection using chromatin immunoprecipitation coupled with sequencing[10] and RNA structure inference through SHAPE[11,12] have improved our understanding, a method for direct detection of rG4 structures in long, native RNA remains elusive.

Oxford Nanopore Technologies' direct RNA sequencing[13], offers new possibilities for RNA structure analysis. The platform's ability to sequence long RNA molecules and detect nucleotide modifications[14,15] suggests potential for direct structural inference. Previous studies have demonstrated G-quadruplex trapping in custom-engineered nanopores[16] and observed G4 unfolding dynamics[17–19], but these approaches have not been applied to native RNA contexts.

Here, we present a novel method for direct detection of rG4 structures in long RNA molecules using Oxford Nanopore sequencing. By analyzing nanopore current stalling events, we identify rG4 signatures in synthetic constructs and total RNA, distinguishing between different rG4 topologies. This approach offers a unique opportunity to study rG4 structures in their native context without chemical modifications, potentially revolutionizing our understanding of RNA structural dynamics.

a

Sample	Median read length	Median read identity	Median PHRED quality
RNA9	168	0.813	7.34
RNA3	144	0.818	6.85
RNA3+PhenDC3	193	0.831	4.07
RNA6	75	0.706	6.36
RNA5	126	0.810	7.05

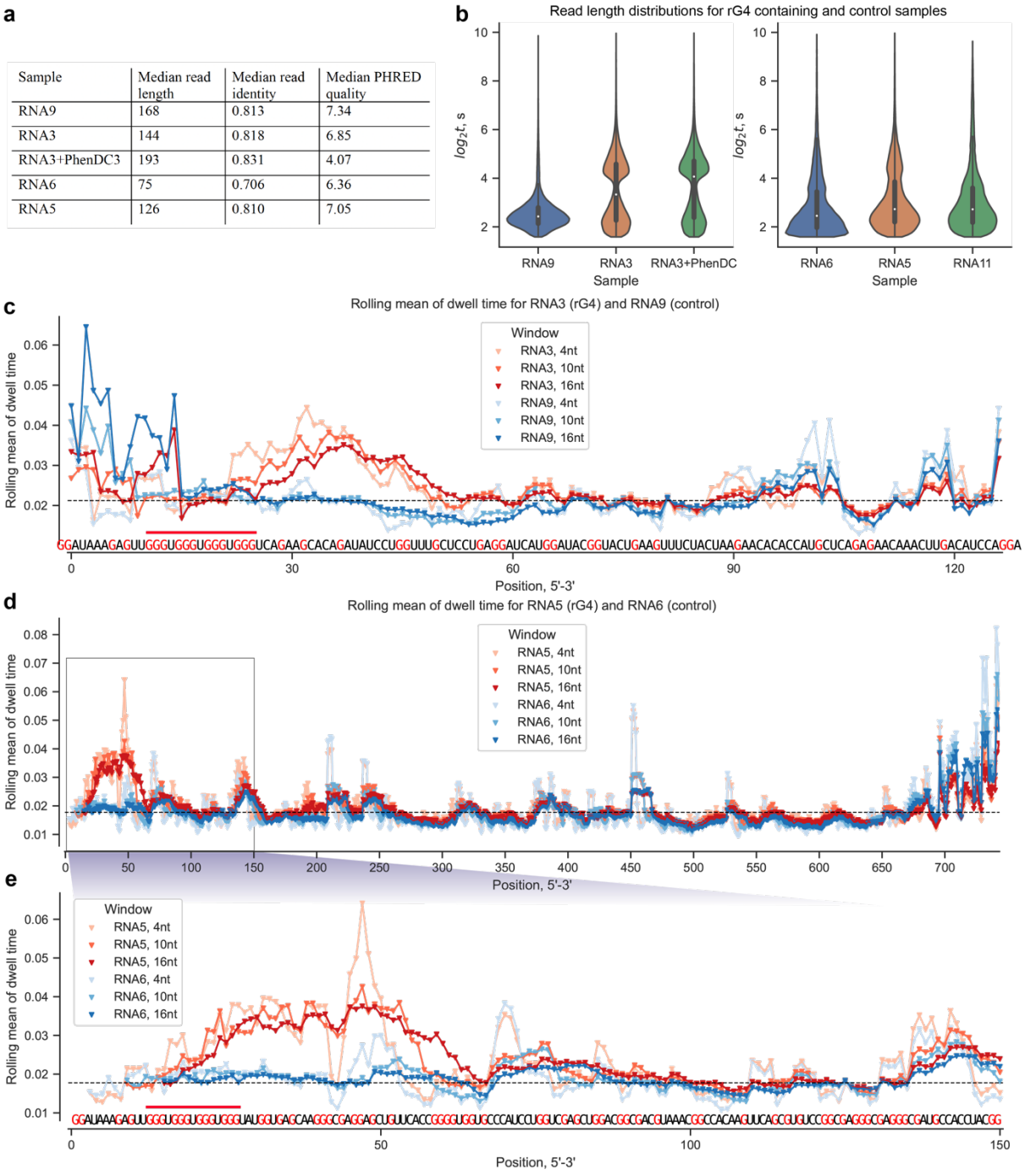


Fig. 1: Read statistics and moving average plots of kmer dwell time for the RNA samples with rG4 and mutated counterparts. a) Experimental statistics for the samples. b) Violin plots of read time, normalized by read length in nucleotides, $log_2 t$ (s), for RNA9, RNA3, RNA3+PhenDC, RNA6, RNA5, RNA11. Notice the secondary peaks, prominent for RNA3 experiments, and noticeable for RNA5 sample; these peaks may be attributed to rG4 unfolding inside the nanopore vestibule. Plots of moving averages of the kmer dwell time with window sizes of 4nt, 10nt, 16nt for: c) RNA3 containing a rG4 motif (sequence shown) and RNA9 containing a mutated rG4 motif (sequence not shown); d) Zoomed out point plot for RNA5-RNA6 moving averages; e) RNA5 containing a rG4 motif (sequence shown) and RNA5 containing a mutated rG4 motif (sequence not shown). The averages in both cases diverge

before approaching the G4 from the 3' end. The rG4 motifs are highlighted in red overline. Mean signal over all positions is shown in dashed black line.

We first investigated rG4 detection in synthetic RNA constructs of varying lengths (131-748 nucleotides) containing known rG4-forming sequences or their mutated counterparts (Fig. 1a, Supplementary Fig. S1). Nanopore sequencing of these constructs revealed distinct current stalling events associated with rG4 structures (Supplementary Fig. S2. The rG4-containing RNA3 and RNA5 exhibited secondary peaks in molecule escape time distributions, which were absent or less pronounced in their mutated counterparts RNA9 and RNA6 (Fig. 1b). Analysis of kmer dwell times showed a sharp increase at the 3' end of the rG4 motifs in rG4-containing samples, contrasting with the mutated controls (Fig. 1c-e). The current stalling starts 6 nucleotides before the nanopore detects the first guanines of the G4 motif. This is consistent with previous observations of G4 completely unfolding before reaching the *trans*-depot of the nanopore assembly[20]. The addition of the rG4-stabilizing ligand PhenDC3 further accentuated this effect (Supplementary Fig. S3).

To validate our method in more biologically relevant contexts, we examined two longer RNA constructs: a derivative of the BCL-x gene containing multiple G4-forming motifs [21,22], and a semi-synthetic transcript (NRQ3) with the NRAS rG4 sequence (NRQ) 5'-GGGAGGGGCGGGUCUGGG-3' that was shown to modulate translation[23]. Both constructs displayed prominent dwell time peaks at the 3' end of known G4 motifs (Supplementary Fig. S4), demonstrating the method's applicability to complex, native RNA sequences.

Extending our analysis to total human mRNA from A431 cells, we bioinformatically identified rG4 motifs in aligned transcripts and grouped them into canonical, bulged, and irregular types from the total of 10,384 aligned nanopore reads (see Materials and Methods). Canonical ("classical") G4 motifs are characterized by the uninterrupted G-runs and loops of up to 12 nucleotides, bulged G4 motifs contain one or more G-run interruptions, and irregular G4 motifs have short G-runs of 1-2 nucleotides interrupted by 1-2 other nucleotides. Among all transcripts with nonzero coverage, there was a total of 1,456 unique rG4-containing transcripts with 4,748 rG4 instances: 1,225 canonical, 2,083 bulged and 1,440 irregular G4s. 1,844 nanopore reads aligned to these transcripts and were analyzed. Despite the complexity of the sample and competing RNA secondary structures, we observed distinct nanopore current stalling signatures for each rG4 type (Fig. 2). Canonical rG4s showed a bimodal distribution with a sharp peak preceding the rG4 3' tail, while bulged and irregular rG4s displayed single peaks with varying positions and intensities. These distinct signatures likely reflect differences

in rG4 stability and topology, providing a novel means of classifying rG4 structures in native RNA contexts.

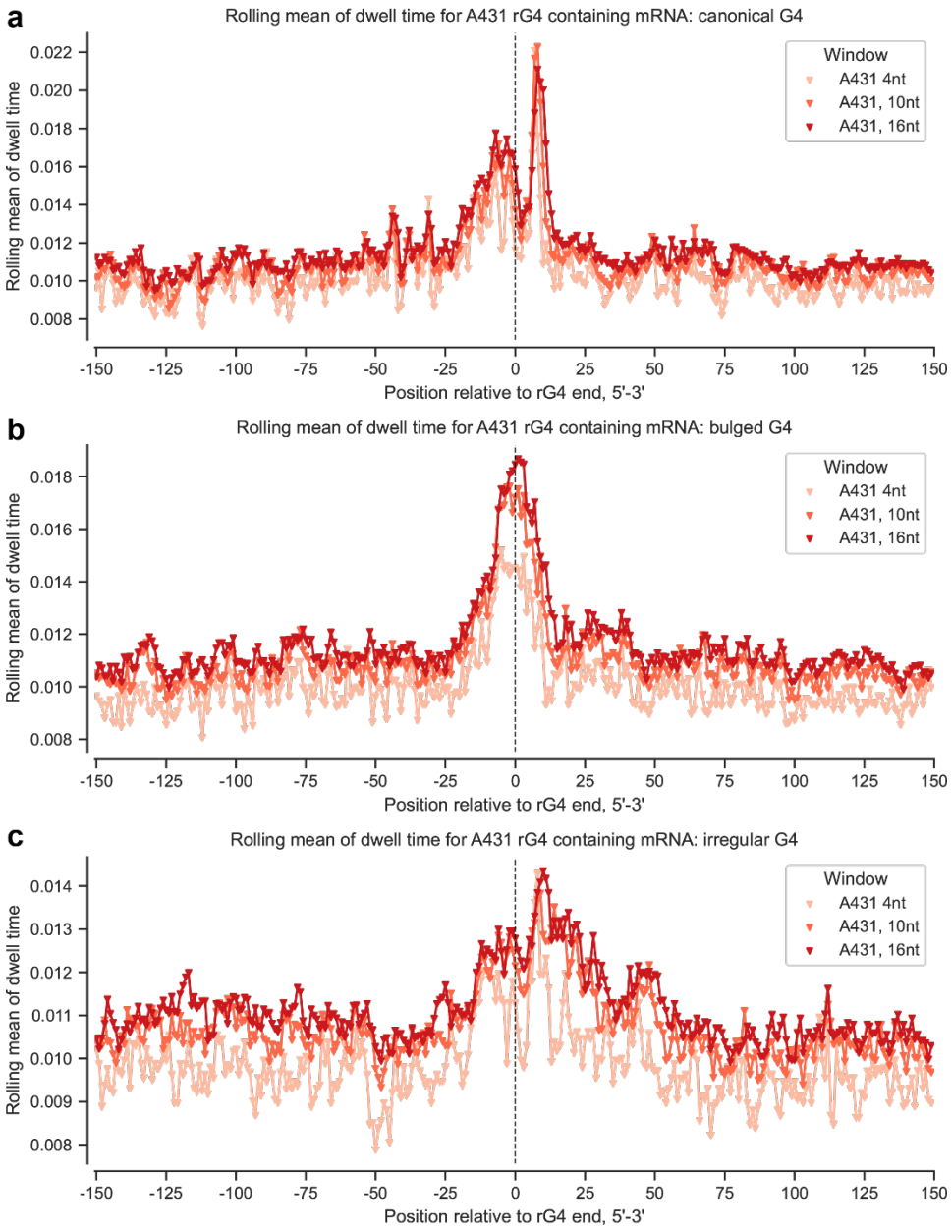


Fig. 2: Distributions of dwell time around rG4 3' ends in various A431 mRNAs show a G4-associated stalling peak. Moving averages of the kmer dwell times with window sizes of 4nt, 10nt, 16nt for kmers from mRNA reads aligned to transcripts with a) canonical G4 motifs, b) bulged G4 motifs, c) irregular G4 motifs. Aligned reads from various positions in different transcripts were grouped by the G4 type and stitched at the G4 motif 3' end. G4 motifs and underlying sequences may be different.

Our findings demonstrate the potential of nanopore sequencing for direct detection and classification of rG4 structures in long RNA molecules. This method overcomes limitations of existing techniques by providing structural information in native sequence contexts without chemical modifications. The ability to distinguish between different rG4 topologies in total RNA opens new avenues for studying the roles of these structures in gene regulation and cellular processes. Future work should focus on increasing the resolution and sensitivity of this method, potentially through modifications to the nanopore protein or improved signal processing algorithms. Additionally, combining this approach with targeted RNA enrichment techniques could enable the study of low-abundance transcripts and their rG4 structures.

In conclusion, we present a novel, direct method for detecting and classifying rG4 structures in native RNA using nanopore sequencing. This approach has the potential to revolutionize our understanding of RNA structural dynamics and their biological implications.

Conflicts of interest

There are no conflicts to declare.

Materials and methods

G4-containing RNA sample synthesis

A derivative of pETDuet-1 plasmid was used to extract the target DNA sequences with synthetically incorporated G4. The sequences were chosen from *RHAU140* gene and *CFP* gene. G4 (5'-TGGGTGGGTGGGTGGGT-3') and mutated G4 (5'-TGAGTGAGTGAGTGAGT-3') motifs along with a T7 promoter (5'-TAATACGACTCACTATAG-3') sequence were inserted in the 5' primer for PCR-mediated extension. Full sequences of the primers available in the Supplementary Data. PCR reactions were performed in a total volume of 100 µl with 50-150 ng of the plasmid template, 10 µl dNTP (2 mM), 0.3 µl of each primer (100 mM), 10 µl of 10X Pfu Buffer and 2 µl Pfu DNA polymerase with 35 cycles of amplification. 2% agarose gel with was used for quality control, stained by thiazole orange fluorescent dye and imaged under UV light with AlphaImager (Alpha Innotech). Subsequent DNA purification was performed using QIAquick PCR Purification Kit (Qiagen) following the standard protocol. Elution volume was decreased to 20 µl of DEPC water instead of 30-50 µl to get higher product concentration.

NRQ insertion through site-directed mutagenesis

To create a long molecule (~1000nt) containing rG4 in the middle, the NRAS G4 (NRQ) was extracted from the pSKC11-UTRQ plasmid and re-inserted to the pETDuet-1 plasmid. In the original plasmid, the NRQ sequence: 5'-GGGAGGGGCGGGTCTGGG-3' was located too close to the T7 promoter and therefore would be too close to the molecule tail end, therefore, the NRQ fragment was amplified with T7 promoter disruption via PCR. Digestion enzyme recognition sites were added with primers, primer sequences used: 5'-GTCGAATTCACTATAGGGACG-3' (EcoRI), 5'-GCTCATATGCAGTTGCTCTCC-3' (NdeI). Colony PCR was done with Pfu polymerase (Promega) with annealing step at 55 °C and 35 amplification cycles. EcoRI and NdeI are compatible for a double digestion reaction. The digestion reaction for the plasmid and double stranded amplicon inserts to create sticky ends was performed at 37 °C with NEBuffer 2.1 to minimize tail end activity with the following recipe: 1 µg pD52 plasmid or 500ng amplicon insert, 5 µl NEBuffer 2.1 (10X), 1 µl EcoRI, 1 µl NdeI, in 50 µl nuclease-free water. Enzymatic purifications were performed with QIAquick PCR purification kit after each reaction. To join the sticky ends and insert the G4 sequence, a ligation reaction was performed: 150 ng plasmid (digested), 200 ng insert, 1 µl 10X ligation buffer, T4 DNA ligase (NEB) 0.5 µl, 1hr at RT. Cell transformation was then performed as follows: 50 µl of competing cells (E.Coli + Ca^{2+}), were kept on ice for 10 min after -80 °C freezer, 2-4 µl DNA was added, incubated for 15 min on ice, then 60s at 42 °C, put back on ice for 3 min, 300 µl of SOC was added and incubated for 15 at 37 °C with 700 rpm mixing. The product was amplified from bacterial plasmids with PCR: Pfu polymerase (Promega), annealing temperature 51 °C, 35 amplification cycles, with primers 5'-TAATACGACTCACTATAG-3', 5'-GCTAGTTATTGCTCAGCGG-3'. The result was checked on 2% agarose gel, ran at 120mV in 1X TAE, subsequently stained with thiazole orange dye and visualized under UV light (AlphaInnotech). DNA sequencing for confirmation of reaction success was carried out with 1st Base Sequencing, Singapore.

In vitro transcription

In vitro transcription (IVT) reactions for the synthetic DNA constructs were performed using T7 RNA polymerase (Thermo Fisher) according to the standard protocol, with 500 ng instead of 1 µg in a total reaction volume 50 µl. RNA were isolated using

phenol:chloroform:isoamyl alcohol (25:24:1) reagent (Sigma-Aldrich) with one round of chloroform extraction and overnight isopropanol precipitation. 8% TBE-Urea-PAGE gel (37.5:1 BioRad) with 8M Urea was run for RNA quality control. RNA was denatured at 90 °C and loaded with RNA loading dye (47.5% formamide, 0.01% SDS, 0.01% bromophenol blue, 0.005% Xylene Cyanol, 0.5mM EDTA); gels were stained with SYBR Gold fluorescent dye and imaged under UV light with AlphaImager.

RNA poly-A tailing

RNA polyA-tailing was performed with E.coli poly(A) polymerase (New England Biolabs) following the recipe: 1 µg of RNA template, 2 µl 10X buffer, 2 µl (10 mM) ATP, 3 µl E.Coli polyA polymerase; the enzyme quantity was changed from 1 µl from 3 µl, and the reaction was incubated for 15 min instead of 30 min, as recommended by Dr. Thidathip Wongsurawat in personal communication. EDTA was added to a final concentration of 10 mM to stop the reaction. The goal was set to obtain 30-50nt-long polyA tails. After the polyA-tailing reaction RNA samples were purified with Agencourt RNAClean XP (Beckman Coulter) magnetic beads following the supplied protocol.

DNA and RNA quantification

Nanodrop 2000 (Thermo Fisher Scientific) was used for DNA and RNA quantification. RNA sample purity was assessed using 260nm/280nm absorbance ratio to be between 2.0 and 2.2.

Oxford Nanopore SQK-RNA001/SQK-RNA002 library preparation and direct RNA sequencing

During the library preparation, the reverse transcription step was skipped, according to the Oxford Nanopore protocol. Both R9.4 and R9.5 chemistries were used. Sample quantity of the synthesized libraries ranged between 100-400 ng. For RNA3+PhenDC sequencing, molar template:ligand ratio used is 1:2. SQK-RNA001 sequencing kit was used for all experiments. All enzymatic purification steps were conducted using Agencourt RNAClean XP (Beckman Coulter) magnetic beads with 80% ethanol. Sequencing was performed at ambient

conditions for at least 4 hours for each sample, and a minimum of 150,000 reads were obtained for each experiment. All reads were base-called with *guppy* v.5.0.11 with no quality filtering, and mapped with *minimap2* with parameters *'-ax* splice *-uf -k14'*. Tables with nanopore current events corresponding to k-mers were constructed with nanopolish *eventalign*. The event tables were then re-aggregated with nanopolishComp *eventalign collapse* tool.

Analysis of total RNA nanopore current stalling and G4 definitions used

Total RNA nanopore current stalling analyses was performed through alignment of the A431 total RNA to GENCODE transcripts release 41 for the GRCh38 human genome. Then, a regular expression search for G4 motifs was performed in the transcript sequences with non-zero coverage. The following regular expressions were used: $[G_{3+}L_{1-12}]_{3+}G_{3+}$ – classical G4 pattern with extended loop length; $[GN_{0-1}GN_{0-1}GL_{1-3}]_{3+}GN_{0-1}GN_{0-1}G$ – "bulged" G4 pattern with possible singular "bulges" of nucleotides, where at least one continuous run of three guanines was required; $[G_{1-2}N_{1-2}]_{7+}G_{1-2}$, $\sum G_i \geq 12$ – "irregular" G4 pattern including less canonical sequences. After tracing back the nanopore current events with nanopolishComp *eventalign collapse* utility, nanopore current events for the rG4 containing reads were re-aligned again based on the location of the rG4.

Nanopore current blockage detection utility

Raw nanopore current stalling detection is based on the z-score of the current moving average. An averaging window (400 time units) is set and the current mean is continuously measured: μ_i. If the difference in the moving average of the new data point and a point lagging by 100 units of time is below a fraction of the current (z-score): $\left|\mu_i^{400} - \mu_{i-100}^{400}\right| < z * \mu_i^{400}$, the first occurrence of the stalling is recorded. The z-score was heuristically determined at 1.5%. A stalling event is ultimately recorded if its length exceeds 5000 time units ("samples"), which is equivalent to 5000/r = 1.66 s, where r is the sampling rate: r = 3012 sample/s. Python scripts are available at https://github.com/anyakors/porebump.

References

1. Murat, P. & Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA. *Current Opinion in Genetics and Development* vol. 25 22–29 Preprint at https://doi.org/10.1016/j.gde.2013.10.012 (2014).

2. Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews: RNA* vol. 3 495–507 Preprint at https://doi.org/10.1002/wrna.1113 (2012).

3. Song, J., Perreault, J.-P., Topisirovic, I. & Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **4**, e1244031 (2016).

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

5. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* vol. 19 414–422 Preprint at https://doi.org/10.1016/j.tcb.2009.05.002 (2009).

6. Gomez, D. *et al.* Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Research* **32**, 371–379 (2004).

7. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Research* vol. 40 4727–4741 Preprint at https://doi.org/10.1093/nar/gks068 (2012).

8. Didiot, M. C. *et al.* The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Research* **36**, 4902–4912 (2008).

9. Chang, T. *et al.* Detection of G-Quadruplex Structures Formed by G-Rich Sequences from Rice Genome and Transcriptome Using Combined Probes. *Analytical Chemistry* **89**, 8162–8169 (2017).

10. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature Protocols* (2018) doi:10.1038/nprot.2017.150.

11. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610–1616 (2006).

12. Aw, J. G. A. *et al. Determination of Isoform-Specific RNA Structure with Nanopore Long Reads*. *Nature Biotechnology* vol. 39 (2021).

13. Garalde, D. R. *et al.* Highly parallel direct RN A sequencing on an array of nanopores. *Nature Methods* **15**, 201–206 (2018).

14. Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *bioRxiv* (2019) doi:10.1101/843136.

15. Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nature Biotechnology* **39**, 1394–1402 (2021).

16. Shim, J. W. & Gu, L. Q. Encapsulating a single G-quadruplex aptamer in a protein nanocavity. *Journal of Physical Chemistry B* **112**, 8354–8360 (2008).

17. Shim, J. W., Tan, Q. & Gu, L. Q. Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity. *Nucleic Acids Research* **37**, 972–982 (2009).

18. An, N., Fleming, A. M. & Burrows, C. J. Interactions of the human telomere sequence with the nanocavity of the α-hemolysin ion channel reveal structure-dependent electrical signatures for hybrid folds. *Journal of the American Chemical Society* **135**, 8562–8570 (2013).

19. An, N., Fleming, A. M., Middleton, E. G. & Burrows, C. J. Single-molecule investigation of G-quadruplex folds of the human telomere sequence in a protein nanocavity. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 14325–14331 (2014).

20. Shim, J. & Gu, L. Q. Single-molecule investigation of G-quadruplex using a nanopore sensor. *Methods* vol. 57 40–46 Preprint at https://doi.org/10.1016/j.ymeth.2012.03.026 (2012).

21. Weldon, C. *et al.* Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nature Chemical Biology* **13**, 18–20 (2017).

22. Weldon, C. *et al.* Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Research* **46**, 886–896 (2018).

23. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology* **3**, 218–221 (2007).

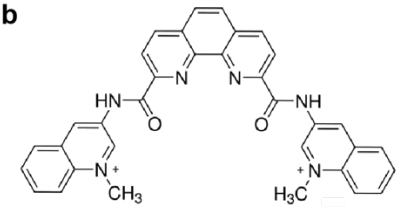
24. Aksimentiev, A., Heng, J. B., Timp, G. & Schulten, K. Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophysical Journal* **87**, 2086–2097 (2004).

Supplementary Figures

a

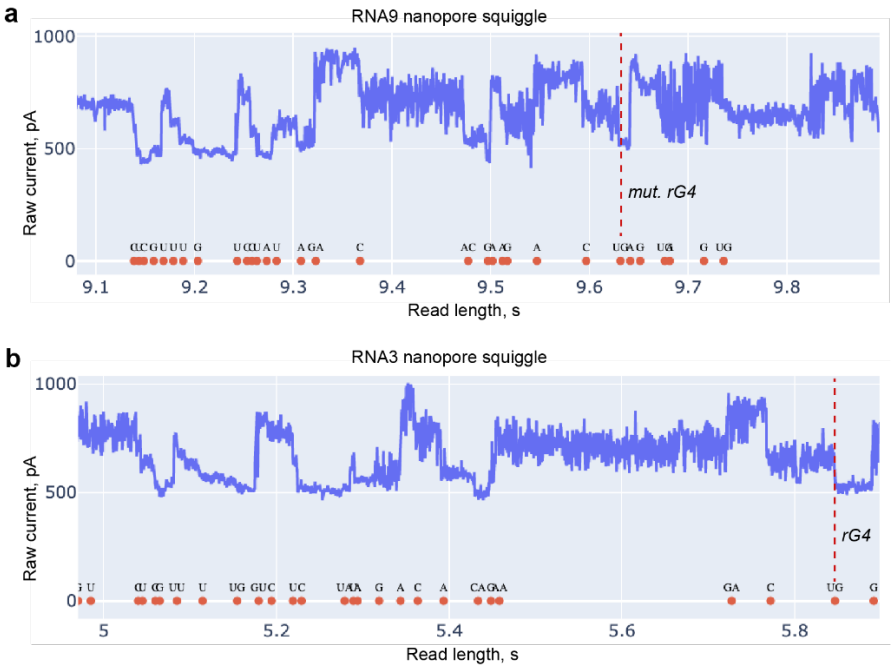
RNA3	131nt	synthetic PQS insert	*RHAU*
RNA9	131nt	mutated PQS insert	*RHAU*
RNA5	748nt	synthetic PQS insert	*CFP*
RNA6	748nt	mutated PQS insert	*CFP*
RNA11	723nt	random, G-rich	*CFP*

b

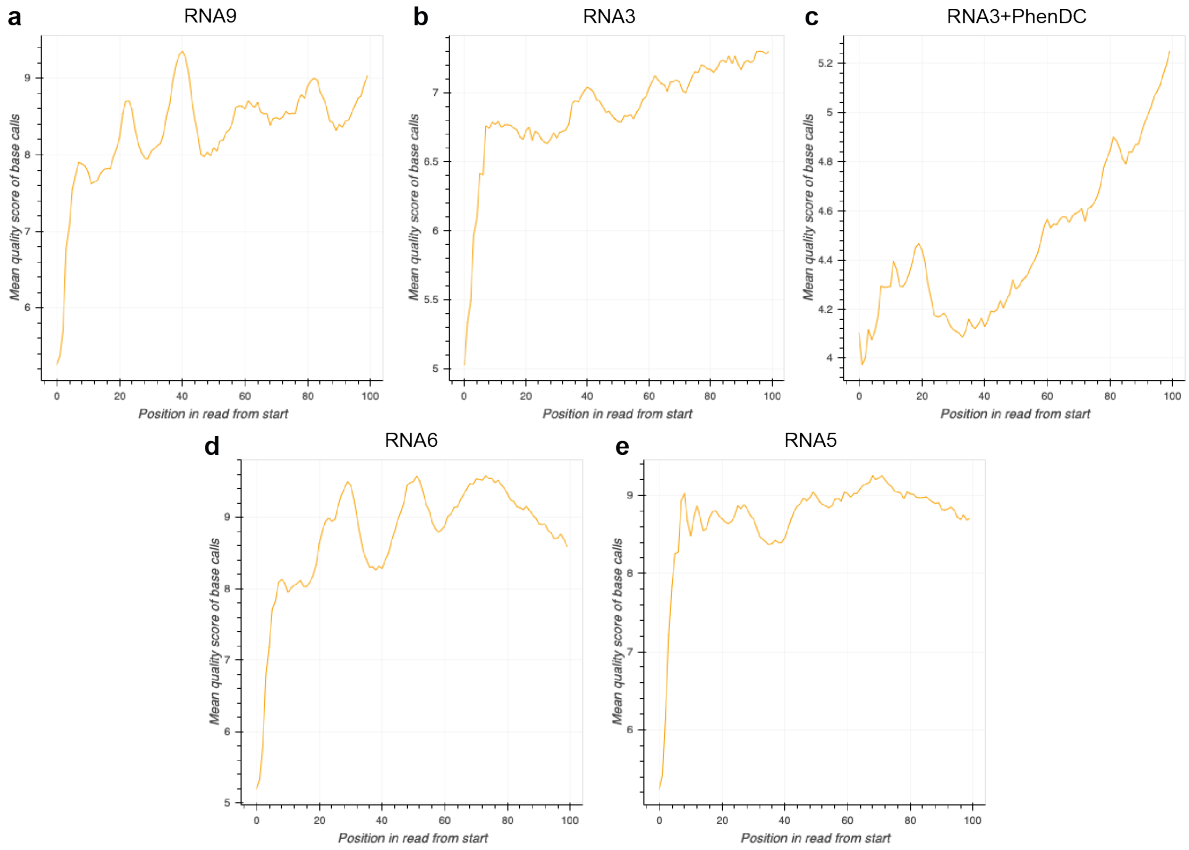


Phen-DC3

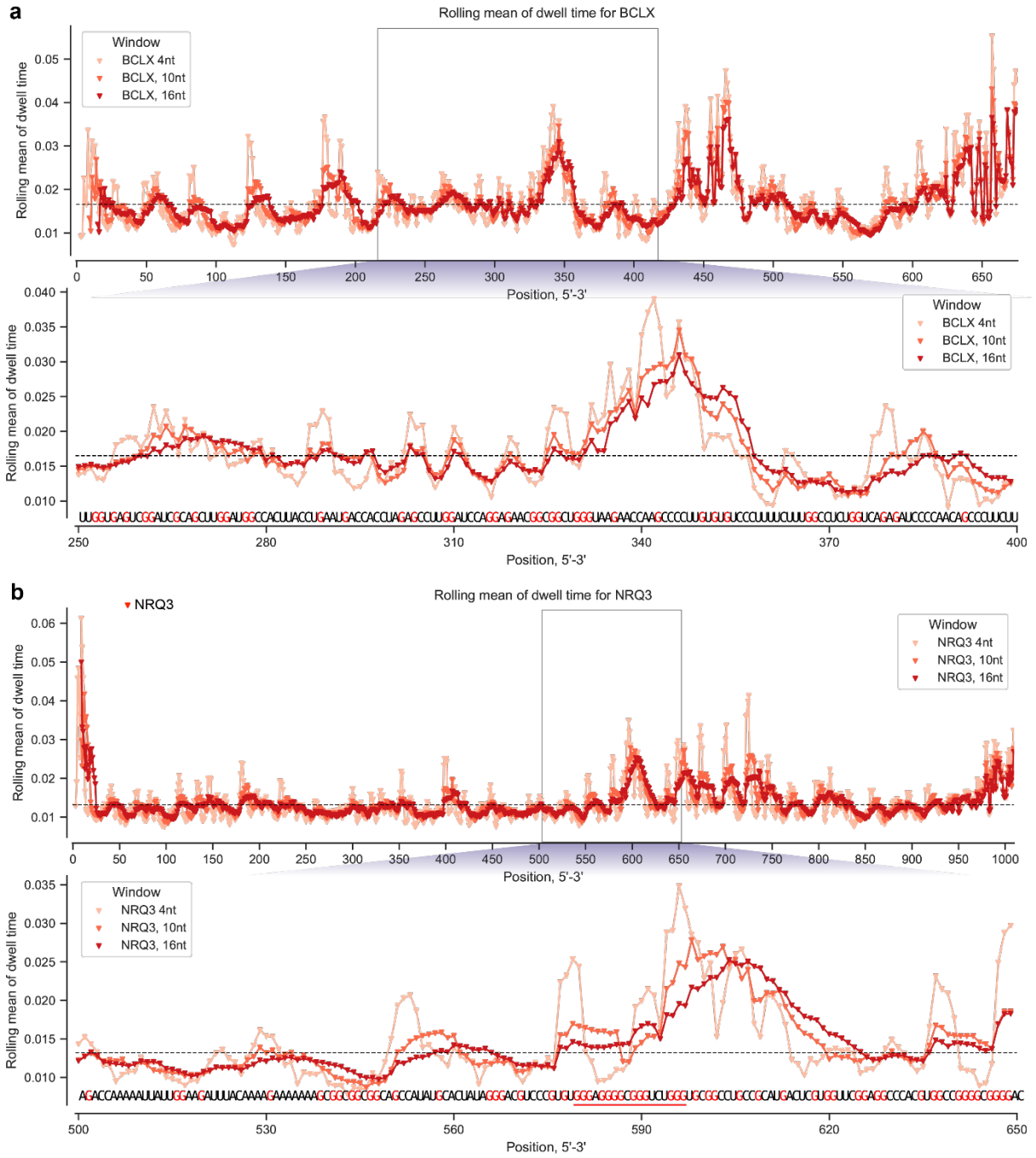
Supplementary Fig. S1: RNA samples used for the rG4 detection experiments, and the respective controls. a) Five sequences were selected from a pD52 plasmid containing *RHAU* and *CFP* genes. Two sequences are short (131nt), with one sequence containing site-directed insert of the rG4 motif $r(GGGU)_4$ (RNA3) and the mutated rG4 motif $r(GAGU)_4$ (RNA9). Similarly, RNA5 and RNA6 are selected from *CFP* gene, but have a length of 748 nt each. RNA11 is a randomly selected sequence with G-rich domains. b) PhenDC3 is a G4 stabilizing ligand of the bisquinolinium family.



Supplementary Fig. S2: Nanopore current fragments for RNA9 and RNA3 demonstrate the effect of G4 stalling. In the absence of any structured tail, the readout of RNA9 continues at the mutated rG4 site. Nanopore blockage is defined as a stretch of constant nanopore current with low dispersion. In the event of nanopore blockage, the mean μ and standard deviation σ of the current are not changing by more than 2-3%. Short blockages, comparable with several nucleotides' passage time, occur often, and are associated with stochastic molecule movements[1], while long stalling events may, in our hypothesis, be attributed to unfolding rG4 and other secondary structures. For RNA3, a stalling of ≈0.25s is incurred. The blockage occurs 6 nucleotides before the actual start of the motif, likely showing the accommodation of the folded G4 and the event of unfolding before the nucleotides reach the current measurement electrode.

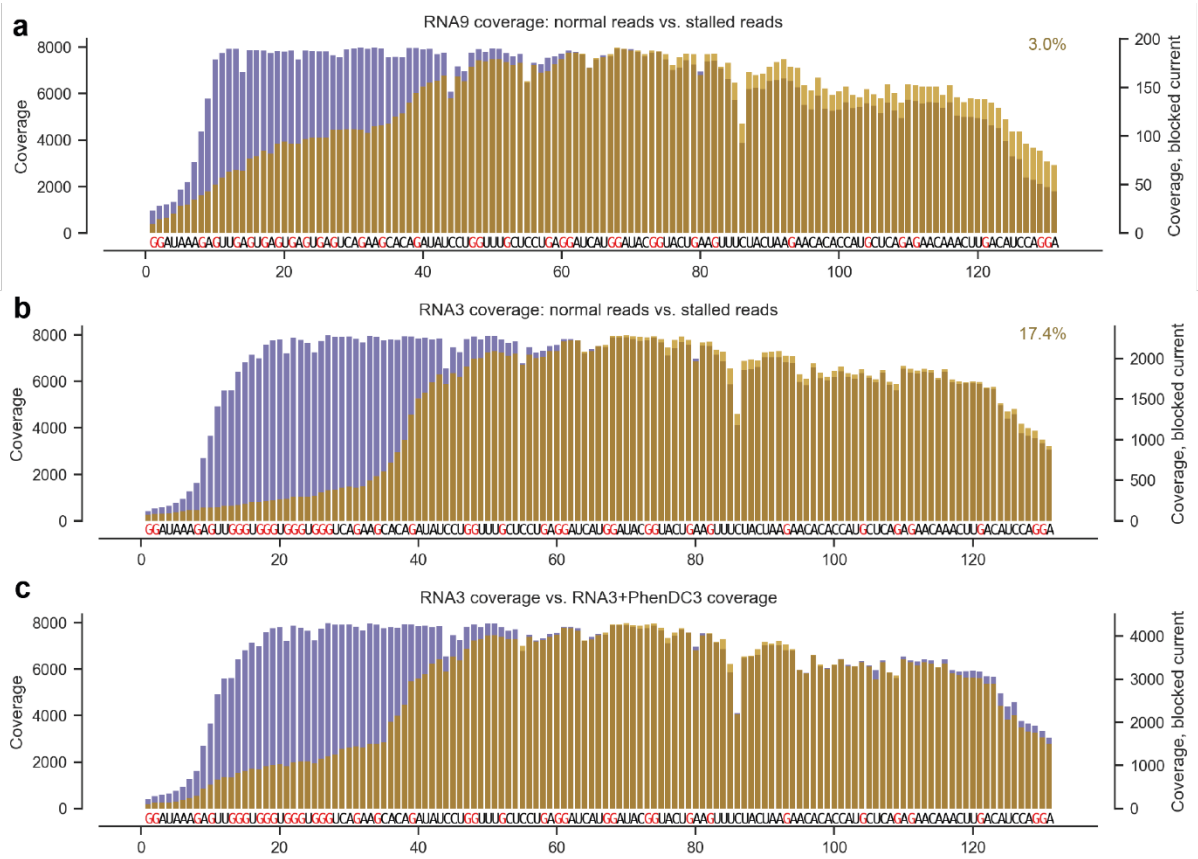


Supplementary Figure S3: Mean quality scores of base calls per sample position for RNA9, RNA3, RNA3+PhenDC3, RNA6 and RNA5. Both RNA3 and RNA5 are containing a G4 motif, but only RNA3+PhenDC3 has a substantial score impairment at the 5' end (coordinates closer to 0).



Supplementary Fig. S4: Plots of moving average of the dwell time of moving kmers for each position for *BCL-x* and NRQ3 constructs. a) Moving averages of the kmer dwell time with window sizes of 4nt, 10nt, 16nt for *BCL-x* construct of length 681nt. The construct contains alternative 5' splice sites of exon 2 and multiple G4 forming motifs; top: full length, bottom: zoomed in locus with a dwell time peak, coinciding with the 3' end of the G4 experimentally detected earlier[2]. b) Moving averages of the kmer dwell time with window sizes of 4nt, 10nt, 16nt for the NRQ3 construct of length 1023nt. The construct contains an rG4 motif from the *NRAS* gene; top: full length, bottom: zoomed in locus with a dwell time peak before the NRQ 3' end. Further sequence contains G-rich regions (see full sequence in Supplementary Data) possibly forming G4s as well, explaining following peaks on the 3'

end. The rG4 motifs are highlighted in red overline. Mean signal over all positions is shown in dashed black line.



Supplementary Fig. S5: RNA9, RNA3 and RNA3+PhenDC coverage after alignment and stalling detection. a) RNA9 read coverage plot after alignment of 20,000 reads (purple, left y-axis) and alignment of the reads with stalling (orange, right y-axis). b) RNA3 read coverage plot after alignment of 20,000 reads (purple, left y-axis) and alignment of the reads with stalling (orange, right y-axis). c) RNA3 read coverage plot after alignment of 20,000 reads (purple, left y-axis) and alignment of original RNA+PhenDC reads. The coverage pattern of RNA3+PhenDC3 resembles the RNA3 coverage with stalling detection. Sequences of the RNA constructs are shown, guanines are highlighted in red. Percent of reads incorporating blockage is shown on top for a) and b), while the overlay of alignments in c) is from two differently designed experiments, therefore percent blockage does not apply. In total, 17.4% of RNA3 reads contained blockage events longer than 1.66s, while only 3.0% of mutated RNA9 had blocked reads.

Supplementary Data

Table A1: RNA sequences used in the direct rG4 sensing experiments.

RNA3

GGAUAAAGAGUUGGGUGGGGUGGGGUGGGUCAGAAGCACAGAUAUCCUGGUUUGCUCCUGAGGAU
CAUGGAUACGGUACUGAAGUUUCUACUAAGAACACACCAUGCUCAGAGAACAAACUUGACAUC
CAGGA

RNA9

GGAUAAAGAGUUGAGUGAGUGAGUGAGUCAGAAGCACAGAUAUCCUGGUUUGCUCCUGAGGAU
CAUGGAUACGGUACUGAAGUUUCUACUAAGAACACACCAUGCUCAGAGAACAAACUUGACAUC
CAGGA

RNA5

GGAUAAAGAGUUGGGUGGGGUGGGGUGGGUAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGG
UGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCG
AGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCU
GCCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUGGGGCGUGCAGUGCUUCAGCCGCUAC
CCCGACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGC
GCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCG
ACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGG
GGCACAAGCUGGAGUACAACUACAUCAGCCACAACGUCUAUAUCACCGCCGACAAGCAGAAGA
ACGGCAUCAAGGCCAACUUCAAGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGA
CCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUG
AGCACCCAGUCCAAGCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGU
UCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAA

RNA6

GGAUAAAGAGUUGAGUGAGUGAGUGAGUAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGG
UGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCG
AGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCU
GCCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUGGGGCGUGCAGUGCUUCAGCCGCUAC
CCCGACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGC
GCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCG
ACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGG
GGCACAAGCUGGAGUACAACUACAUCAGCCACAACGUCUAUAUCACCGCCGACAAGCAGAAGA
ACGGCAUCAAGGCCAACUUCAAGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGA
CCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUG
AGCACCCAGUCCAAGCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGU
UCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAA

RNA11

GGAAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGAC
GGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGC
AAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGA
CCACCCUGACCUGGGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUU
CUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGACGACGGC
AACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUG
AAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUGGAGUACAACUACAUC
AGCCACAACGUCUAUAUCACCGCCGACAAGCAGAAGAACGGCAUCAAGGCCAACUUCAAGAUCC

GCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGG
CGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCAAGCUGAGCAAAGAC
CCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCG
GCAUGGACGAGCUGUACAAGUAA

Table A2: primer sequences used in the direct rG4 sensing experiments.

Fw-9	TCACTATAGGATAAAGAGTTGAGTGAGTGAGTGAGTCAGAAGCACAGATATCCTGGTTTG
Fw-3	TCACTATAGGATAAAGAGTTGGGTGGGTGGGTGGGTCAGAAGCACAGATATCCTGGTTTG
Fw-11	GCGGCCTCTAATACGACTCACTATAGGAATGGTGAGCAAGGGCGAGGAGCTG
Rv-56	TTACTTGTACAGCTCGTCCATGCCGAGAGTG
Rv-39	TCCTGGATGTCAAGTTTGTTCTC
Fw-6	GCGGCCTCTAATACGACTCACTATAGGATAAAGAGTTGAGTGAGTGAGTGAGTATG
Fw-5	GCGGCCTCTAATACGACTCACTATAGGATAAAGAGTTGGGTGGGTGGGTGGGTATG

References

1. Aksimentiev, A., Heng, J. B., Timp, G. & Schulten, K. Microscopic kinetics of DNA translocation through synthetic nanopores. *Biophys J* 87, 2086–2097 (2004).
2. Weldon, C. *et al.* Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat Chem Biol* 13, 18–20 (2017).