



**HAL**  
open science

# Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq

Sunny y Yang, David Monchaud, Judy M y Wong

► **To cite this version:**

Sunny y Yang, David Monchaud, Judy M y Wong. Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq. *Nature Protocols*, 2022, 17, pp.870-889. 10.1038/s41596-021-00671-6 . hal-03563720

**HAL Id: hal-03563720**

**<https://hal.science/hal-03563720v1>**

Submitted on 9 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq

Sunny Y. Yang<sup>1</sup>, David Monchaud<sup>2</sup>, & Judy M.Y. Wong<sup>1,\*</sup>

<sup>1</sup>*Faculty of Pharmaceutical Sciences, University of British Columbia, Pharmaceutical Sciences Building,  
2405 Wesbrook Mall, V6T 1Z3, Vancouver, Canada*

<sup>2</sup>*Institut de Chimie Moléculaire, ICMUB CNRS UMR 6302, UBFC Dijon, France*

\*correspondence: Dr. Judy Wong; Tel: 604-827-3314; Fax: 604-822-3035; Email: judy.wong@ubc.ca

**EDITORIAL SUMMARY** Guanine-rich RNAs folded into G-quadruplex structures (G4-RNAs) are captured from crosslinked RNA using a G4-specific chemical probe, BioTASQ, and mapped globally by sequencing.

**PROPOSED TWEET** #NewNProt for capturing RNA G-quadruplexes (G4-RNAs) using a G4-specific probe. @dmonchaud

**PROPOSED TEASER** Global mapping of G4 RNAs

**Related links**

**Key reference(s) using this protocol**

Yang, S. et al. Nat Commun 9, 4730 (2018): <https://doi.org/10.1038/s41467-018-07224-8>

Renard, I. et al. Nucleic Acids Research 47, 5502–5510 (2019): <https://doi.org/10.1093/nar/gkz215>

## 1 **ABSTRACT**

2 Guanine-rich RNAs can fold into four-stranded structures, termed G-quadruplexes (G4-RNAs), and  
3 participate in a wide range of biological processes. Here, we describe in detail the G4-RNA-specific  
4 precipitation (G4RP) protocol, which allows the transcriptomic profiling of G4-RNAs. The G4RP protocol  
5 consists of a chemical crosslinking step, followed by affinity capture with the G4-specific probe,  
6 BioTASQ. G4RP can be coupled with sequencing to capture a comprehensive global snapshot of folded  
7 G4-RNAs. This method can also be used to profile induced changes (i.e. through G4 ligand treatments)  
8 within the G4-RNA transcriptome. The entire protocol can be completed in 1-2 weeks and can be scaled  
9 up or down depending on the specific experimental goals. In addition to the protocol details, we also  
10 provide here a guide for optimization in different laboratory setups.

11

## 12 **INTRODUCTION**

13 G-quadruplexes (G4s) are non-canonical nucleic acid structures formed by Hoogsteen bonding of four  
14 guanines to form planar guanine quartet (G-quartet) units, which  $\pi$ -stack on each other to form the  
15 overall columnar structure<sup>1-2</sup>. G4s forming in DNAs (G4-DNAs)<sup>3</sup> and RNAs (G4-RNAs)<sup>4</sup> can affect the  
16 corresponding biological processes. G4-forming sequences are found within the genome and  
17 transcriptome of several prokaryotes, eukaryotes and viruses<sup>5</sup>. Importantly, G4-forming sequences are  
18 highly saturated at sites of disease-relevance such as oncogenes, telomere, and repeat expansion  
19 regions<sup>6</sup>. Biochemical<sup>7</sup>, bioinformatics<sup>8</sup>, optical imaging<sup>9</sup> and genomic techniques<sup>7,10</sup> have been used to  
20 show that G4 formation is widespread in both the human genome and the transcriptome. However,  
21 accurately quantifying and mapping by sequencing the G4-DNA<sup>11-14</sup> and G4-RNA sites<sup>15,16</sup> formed within  
22 living cells continues to be an ongoing challenge. It is important to note that G4s formation is often  
23 outcompeted by Watson-crick duplex formation, protein-binding, and G4-destabilizing helicase (G4-

24 helicase)<sup>17</sup> activities; thus, the G4 folding process requires windows of opportunities when the  
25 polynucleotides are single-stranded and in more openly accessible conformations. In support of this  
26 notion, it has been observed that open chromatin states are more favourable for the formation of G4-  
27 DNAs<sup>13</sup>. The high prevalence of potential G4-forming sequences in the human genome<sup>8</sup> suggests that  
28 formation of G4s likely permeates throughout most, if not all, important biological events in the flow of  
29 genetic information. G4-DNAs may play roles in various biological functions, including telomere  
30 maintenance, transcription regulation, epigenetic regulation, replication and recombination<sup>18</sup>. Despite  
31 the increasing volume of research in the G4 field, much of the previous focus has been on G4-DNA  
32 rather than G4-RNA. Given the established ability of RNA strands to fold into intricate structures, which  
33 govern their functions<sup>19</sup>, and the superior thermostability of G4-RNA<sup>4</sup>, it is anticipated that G4-RNAs  
34 constitute a significant proportion of all folded G4s within living cells. G4-RNA has also been implicated  
35 in a wide range of biological processes, including translation, RNA processing, splicing, and  
36 ribonucleoprotein (RNP) formation<sup>4</sup>. Here, we discuss some of the current techniques used to study G4s  
37 and highlight our recently developed affinity-capture technique, called the G4-RNA-specific precipitation  
38 (G4RP) with sequencing (G4RP-seq), for studying G4-RNAs on a transcriptome-wide level.

39

#### 40 **Current methods for studying genomic and transcriptomic G4s in cells**

41 While interest in G4s has been multiplying, there is still a limited number of techniques for studying G4s  
42 on a genome- and transcriptome-wide scale. For genome-wide studies, G4-seq<sup>15</sup> combines the  
43 polymerase-stop assay with next-generation sequencing under conditions that favour (using K<sup>+</sup> cations  
44 or G4-stabilizing ligand pyridostatin (PDS)<sup>20</sup>) or disfavour G4 formation (using Li<sup>+</sup> cations) in purified,  
45 protein-free DNA. As folded G4s can cause polymerase-stalling, quantification of these polymerization  
46 stop sites can infer G4 folding state and generate a genomic G4 map. While the resolution is high, a

47 major limitation of G4-seq is the in vitro experimental setting, which is not representative of the native  
48 chromatin state. Genome-wide G4 mapping by performing chromatin-immunoprecipitation sequencing  
49 (G4-ChIP-seq)<sup>13</sup> using the G4-specific antibody BG4<sup>11</sup>, a single-chain variable fragment antibody specific  
50 for G4, allows approximation of a more biologically relevant (i.e. “in vivo”) setting. In comparison, G4-  
51 seq observed over 750,000 potential G4 sites<sup>12</sup>, whereas G4-ChIP-seq reported only 10,000 enriched  
52 sites<sup>13</sup>. This large discrepancy between the two techniques emphasizes the importance of the local  
53 chromatin environment in the formation of G4-DNAs.

54 A recently developed sequencing technique, rG4-seq<sup>15</sup>, which is conceptually similar to G4-seq,  
55 has been implemented for studying transcriptome-wide G4-RNAs. This technique exploits the effect of  
56 G4-RNAs on the processivity of reverse-transcriptase (RT) in complementary DNA synthesis. By mapping  
57 and comparing RT polymerization stop sites under G4-favourable and unfavourable ionic conditions in  
58 vitro, thousands of potential G4-RNA sites were discovered. As the query of G4 formation using purified  
59 RNA will not recapitulate the contributions of other biological factors (e.g. RNA-binding proteins, other  
60 nucleic acid structures.), rG4-seq possesses much of the same limitations as G4-seq. To that end, the RT  
61 stop-profiling method has been adapted to study G4-RNAs in vivo with an additional step of treating  
62 cells with dimethyl sulfate (DMS) before purification and RT. DMS-modified guanine bases are unable to  
63 fold into G4s in vitro and thus can be used to infer folding state in vivo indirectly<sup>21</sup>. Surprisingly, DMS/RT-  
64 stop profiling showed that at equilibrium, G4-RNAs are nearly all unfolded in vivo, a significant  
65 departure from observations made using live-cell imaging<sup>9,22</sup> and in vitro quantification<sup>15</sup> methods. This  
66 apparent discrepancy is likely due to the nature of the DMS/RT-stop profiling assay, which may be more  
67 suitable for establishing the shift of G4 equilibrium rather than capturing a snapshot of the folded G4  
68 transcriptome. Drawing inspirations from G4-ChIP-seq, we developed the G4RP-seq protocol<sup>16</sup>, which  
69 allows a broad-level snapshot view of the folded G4 transcriptome. Using G4RP-seq, we showed that G4-  
70 RNAs fold in a sequence-dependent manner and that treatment with G4-stabilizing ligands enhances

71 these G4 folded states. Thus, the current model depicts G4-RNAs as short-lived structures which are  
72 continuously being folded and unfolded by various regulatory mechanisms inside living human cells.  
73 These findings present us with a new perspective in which G4-RNAs are portrayed more as events rather  
74 than simple physical structures.

## 75 **Development of the Biotinylated Template-Assembled Synthetic G-quartet (BioTASQ) and G4RP** 76 **Protocol**

77 A common recurring challenge in designing G4 ligands has been the lack of specificity due to unintended  
78 off-target binding. A class of biomimetic probes, called the Template-Assembled Synthetic G-quartet  
79 (TASQ)<sup>23</sup>, was developed to mitigate these off-target binding events. The chemical design of TASQ,  
80 inspired by the natural formation of G-quartets in G4s, consists of a central core with four arms, each  
81 attached to a guanine (Fig. 1a). These guanines self-assemble into a synthetic G-quartet upon  
82 association with a G4 target through end-quartet stacking. The use of the synthetic G-quartet itself as  
83 the recognition molecule provides high specificity and defined binding properties. A caveat of TASQ is its  
84 preference for targeting G4s with a parallel conformation due to the accessible end-quartets for TASQ  
85 binding without steric hindrance from the loops. This binding preference is especially useful for targeting  
86 G4-RNAs due to their parallel-conformational constraints. The extra utility of TASQ molecules is  
87 demonstrated by swapping the central core with different application-specific moieties<sup>23-26</sup>. The newest  
88 generation of TASQ molecules is a biotinylated version of TASQ (BioTASQ)<sup>16,27</sup>, allowing for affinity  
89 capture of G4 targets using streptavidin-biotin interactions. G4RP is derived from a standard RNA  
90 immunoprecipitation (RIP) protocol<sup>28</sup>; however, instead of antibody-protein recognition in RIP, G4RP  
91 uses a ligand-nucleic acid structure recognition with BioTASQ. Like RIP and other similar methods, the  
92 G4RP protocol can be coupled with high-throughput sequencing to profile transcriptome-wide G4  
93 distribution.

## 94 **Advantages and Limitations of G4RP**

95 Using G4 ligands as baits for G4 targets has previously been explored<sup>29</sup>; however, specificity for the  
96 intended G4 targets has been an ongoing challenge for most, if not all, G4 ligands<sup>30</sup>. BioTASQ was  
97 designed with these limitations in mind by providing higher specificity due to the strictly defined binding  
98 mechanism. The use of a synthetic G-quartet as the bait provides the best-predicted interactions given  
99 that G-quartet stacking interaction is obligatory and exclusive for all G4 formations and stabilization.

100 Some of the key advantages and limitations of G4-RNA quantification assays stem from whether the  
101 approach is direct or indirect. Direct and indirect G4-RNA identification approaches provide different  
102 perspectives due to the dynamics of G4-RNA formation, and thus the appropriate method of choice  
103 depends on the specific scientific question. Table 1 outlines some of the key differences between the  
104 two types of methods. We consider G4RP a direct approach to detect G4s due to the probe's direct  
105 association with the G4 target. However, direct approaches do have the potential pitfall of probe-  
106 assisted stabilization of the G4 targets. Since BioTASQ has properties of a G4 ligand, it conceivably can  
107 stabilize G4 targets, resulting in quantification biases in capturing G4-RNAs that are not representative  
108 of their true biological states. Therefore, an *in vivo* crosslinking step is added prior to cell lysis and  
109 BioTASQ-assisted pulldown in G4RP, which is crucial in minimizing this bias by preventing G4-folding *in*  
110 *vitro*. In contrast, indirect approaches, including rG4-seq<sup>15</sup> and DMS/RT-stop profiling<sup>21</sup>, involve nucleic  
111 acid modifications prior to subsequent screening using reverse transcriptase. Indirect approaches  
112 heavily rely on the following two assumptions: first, modifying chemicals are assumed to access all  
113 cellular RNA and modify unprotected sites uniformly within the transcriptome of living cells, which is  
114 unlikely; second, *in vitro* folded G4s are presumed to be the sole cause of all RT stop sites, which may  
115 lead to overestimation of the actual number of G4 sites due to the inclusion of potential polymerase-  
116 stalling effects by other RNA structures<sup>19</sup>. Indirect approaches are likely biased against highly dynamic  
117 RNA structures as these readouts can only be estimated from highly stable structured RNAs under the

118 set of experimental conditions<sup>31</sup>. Depending on the experimental conditions, when the labelling  
119 chemicals are usually not the limiting reagent, chemical modification in live cells can often lead to  
120 overwhelmingly modified RNAs. Thus, G4RP is considered a more appropriate method for providing a  
121 less-biased view of transiently formed G4 structures. Finally, the currently reported indirect methods do  
122 not consider the formation of intermolecular G4-RNA since the principle of RT-stop profiling is based on  
123 the effect of the cis-acting intramolecular G4s. In this sense, G4RP provides measurements that are  
124 more representative of the biologically relevant interactions due to its ability to reporting on both intra-  
125 and inter-molecular folded G4-RNAs.

126 As indirect methods often provide high resolution data at near single base levels (e.g. chemical probing  
127 techniques, RT-stop profiling), a major limitation of G4RP is the lack of mapping resolution (>200 bases)  
128 due to the sonication step. The main challenge lies in striking a balance between over-sonication and  
129 under-sonication. Under-sonication prevents the proper release of cellular content, resulting in a  
130 reduction in resolution and lower pulldown efficiency. In contrast, over-sonication may increase  
131 mapping resolution but may also lead to disruption of the binding structure (G4 end stack) necessary for  
132 BioTASQ binding. Optimization to reach the specific goals of the experiment is highly recommended.  
133 While sequence analysis of the BioTASQ-enriched RNA fraction in G4RP shows robust enrichment of G-  
134 rich sequences, it should be noted that a minor portion of indirect interactions could be crosslinked and  
135 captured.

136 While G4RP does offer several advantages over the indirect polymerase-dependent methods, the choice  
137 of G4-RNA quantification method depends on the biological questions of interest. It is important to  
138 emphasize that no individual method is without limitations and caveats, and that necessary future steps  
139 should be invested in combining complementary methods in order to mitigate some of their individual  
140 limitations, thus providing a more unbiased view of the G4-RNA transcriptome.



<b>Table 1. Indirect vs direct methods in studying transcriptomic G4-RNAs</b>		
	<b>Indirect method (rG4-seq, DMS/RT-profiling)</b>	<b>Direct method (G4RP-seq)</b>
<b>Readout method</b>	Reverse-transcription stalling + sequencing	Probe-binding/extraction + sequencing
<b>Methods to capture biological state</b>	Modification of nucleic acids in live cells	Chemical crosslinking
<b>Resolution</b>	High	Low
<b>Targets</b>	-Intramolecular G4s	-Intramolecular G4s -Intermolecular G4s
<b>Enrichment method</b>	Poly (A) enrichment	Ribo-depletion
<b>Biases/ Limitations</b>	-in vitro biases -DMS modification could lead to underestimation of dynamic structures -Overestimation of G4 structures by counting stops induced by other structures -Not able to read 5' G4 if there is a strong RT stop caused by 3' G4 signal	-Probe-assisted stabilization could lead to overestimation of G4 structures -a small portion of crosslinked indirect interactions could be captured

## 142 **Future applications**

143 Theoretically, G4RP could be combined with chemical modification and in vitro RT-stop profiling to  
 144 generate a high-resolution map of transcriptome-wide G4-RNA sites; however, sample processing  
 145 compatibilities and assay optimization between these methods have not yet been tested. RT-stop  
 146 profiling could be performed in parallel with G4RP on the same set of samples and subsequently  
 147 analyzed and compared. Alternatively, RT-stop and G4RP could be performed sequentially by profiling  
 148 BioTASQ-captured RNA to improve resolution. It is also yet to be determined how chemical modification  
 149 (e.g. through DMS), if incorporated, could affect G4RP. While G4RP-seq could be used to infer in vivo  
 150 folded state, rG4-seq (or equivalent RT-profiling methods) could be used to determine at high resolution

151 the positions in which intramolecular G4s are more likely to form in vitro. Further optimization of the  
152 sonication step (i.e. increasing sonication duration) and/or adding ribonucleases could potentially  
153 improve resolutions of G4RP-seq by decreasing fragment sizes. However, these optimization strategies  
154 are limited to improving the resolution of intramolecular G4s. Intermolecular G4s remain a challenge to  
155 map with the currently known techniques. The use of different crosslinking agents (e.g. psoralen) with  
156 different effective distances could also be informative to elucidating a complete picture of the G4-RNA  
157 interactions. Integration of G4RP-seq with other methods for identifying RNA-RNA interactions, such as  
158 LIGR-seq<sup>32</sup>, MARIO<sup>33</sup>, SPLASH<sup>34</sup>, and PARIS<sup>35</sup>, may aid in identifying long-range interactions through G4s.  
159 G4RP may also be potentially useful, when coupled with proteomics, in identifying protein partners that  
160 may interact with G4-RNAs. While the G4RP protocol has initially been optimized for human cells, it can  
161 conceivably be adapted and optimized for other model organisms, including other eukaryotes,  
162 prokaryotes and even viruses<sup>5,36</sup>. Given the unique G4 profiles of different organisms and species, G4RP-  
163 seq may be valuable for studying the evolution of G4-interactome and potential inter-organism  
164 interactions (e.g. host-pathogen)<sup>37</sup>.

## 165 **Experimental Design**

166 **Selection of cell models and treatment endpoints.** G4RP is expected to be compatible with most  
167 mammalian cell lines. The samples should be chosen with appropriate controls to ensure an unbiased  
168 comparison (e.g. vector controls for overexpression experiments). Comparisons between cell lines  
169 should include ones with similar genomic backgrounds. Typically, for drug treatments (e.g. small-  
170 molecule compounds such as G4 ligands: BRACO-19 and RHPS4<sup>16,27</sup>), a dose-response profile for an  
171 observable endpoint should be generated to assess the appropriate doses for subsequent experiments.  
172 This would ensure that the intracellular concentration of the ligand is sufficient to cause a biologically  
173 observable change in an endpoint (e.g. cytotoxicity) without leading to excessive cell death and

174 subsequently the loss of sample. We suggest using subtoxic or low toxic dose (LD5-25; dose producing 5-  
175 25% of cell deaths). For classical G4 ligands, such as BRACO-19 and RHPS4, we have observed optimal  
176 concentrations ranging from 0.5 to 20  $\mu$ M depending on the cell line. For treatments, the vehicle (i.e.  
177 solvent for compound) should be used as the control. Another factor of consideration is the duration of  
178 the treatment. Typical treatment time points include 6h, 12h, 24h, 48h and 72h. A different endpoint  
179 measure, other than cytotoxicity, may be chosen depending on the experimental goals.

180 **Chemical crosslinking.** Crosslinking could also impact sonication efficiency; however, we advise against  
181 deviating from our recommended crosslinking duration of 5 to 10 min. Most formaldehyde  
182 formulations, including those with methanol additive, are effective for crosslinking. Methanol increases  
183 cell permeabilization, which may increase crosslinking efficiency. Since formaldehyde is sensitive to light  
184 and air, we suggest storing it in aliquots in a dark and sealed environment. Fresh formaldehyde without  
185 methanol, sold in ampules, can also be considered to decrease potential variability.

186 **Sonication and fragmentation.** Fragmentation of the RNA material is required for the G4RP. Typically,  
187 the range of fragments should fall between 200-800 nt with centering around 400-500 nt. Gel  
188 electrophoresis coupled with RNA-staining dye is a simple method for test. Other techniques requiring  
189 more advanced equipment, such as a Bioanalyzer, can also be used. We suggest using mechanical  
190 sonication as it is more consistent and sonication devices are commonly available in most institutions.  
191 However, different sonicators (e.g. Covaris ultrasonicator, Diagenode Bioruptor, Branson Sonifier, or  
192 other equivalent equipment) will require additional optimization. We recommend testing a range of  
193 sonication durations, with the optimal total sonication duration likely to be between 2 to 4 minutes.  
194 Different sonicators may have a recommended range in terms of the input number of cells. Please check  
195 with the manufacturer-supplied manual or recommended protocols as a guide for the initial  
196 optimization. Sonicators used for G4RP should also be equipped with a temperature control system to

197 prevent heating and premature crosslink reversal. Mechanical sonication serves two purposes: lysing the  
198 cell and fragmenting the nucleic acids. Use of any other non-mechanical shearing alternatives (e.g.  
199 detergent-based lysis and nuclease-mediated fragmentation) will require a different set of optimization  
200 and additional considerations of potential biases.

201 **Bead binding.** Our protocol is optimized using magnetic streptavidin-coupled beads, specifically the  
202 Promega Streptavidin MagneSphere. Other brands or types of streptavidin-coupled beads (agarose) are  
203 likely to be compatible with BioTASQ as well; however, additional optimization and validation may be  
204 necessary. BioTASQ concentration should greatly exceed the binding capacity to ensure the capture is  
205 driven by BioTASQ, not by non-specific binding to the beads. Typically, each G4RP capture uses 10-20  
206 nmol of BioTASQ with 50-100 µg of beads. Increasing the amount of beads or switching to high-capacity  
207 beads can increase the capture yield. When altering the amounts of beads used, negative controls must  
208 be tested in parallel.

209 **Negative control and normalization.** There are two types of controls used in G4RP. For the negative  
210 control, G4RP is performed in the same manner with biotin instead of BioTASQ. The yield of purified  
211 RNA from the negative G4RP control sample is typically much lower and serves as a control for non-  
212 specific bindings to the beads. The negative control is necessary for early optimization but is optional  
213 after the protocol has been optimized and only negligible background is detected. The input control, a  
214 small aliquot of the material collected prior to BioTASQ binding, is critical and required for all  
215 experiments. In contrast with genomic DNA, cellular RNA content can vary widely, and this input control  
216 serves as an internal control for each G4RP experiment during the normalization step of the analyses.

217 **PCR Primer selection.** We have included a list of recommended targets for checking the efficiency of  
218 G4RP (Table 2). However, if there are other targets of interest, additional sets of primers may be used.  
219 These primers should undergo standard validation for specificity to the target region and PCR efficiency.

220 Typically, at least three biological replicates should be used to quantify each locus. For sequencing, we  
221 recommend having multiple replicates (at least 4-6). Based on the results of qPCR screening, the best 2-  
222 3 replicates can be used for library preparation. Any newly discovered region of interest from  
223 sequencing analyses should also be validated using qPCR and appropriate primers.

224 **Library preparation and sequencing depth.** For library preparation of the G4RP samples, we highly  
225 recommend the use of ribodepletion over poly-A enrichment. Due to the fragmentation step in G4RP-  
226 seq, poly-A selection will likely result in loss of signal and a significant 3' bias. Always use the minimal  
227 number of PCR cycles for the library generation. In the case of low amount of input, consider pooling  
228 multiple parallel-processed samples to increase concentration. Sequencing depth will depend on the  
229 experimental question. We recommend 15-20 million uniquely mapped reads; however, this could be  
230 scaled up to 50 million reads to detect rarer events or changes. Typically, sequencing 2-3 replicates for  
231 each BioTASQ-enriched sample is sufficient for cost-effective data analysis. More replicates would  
232 improve the robustness of the statistical analysis but are unlikely to change the overall conclusions of  
233 the experiment. Sequencing of the negative control is optional and is unlikely to result in any useful data  
234 due to the low levels of the transcriptome of interest and significant rRNA background. Each condition  
235 must be accompanied by the sequencing of at least one input control. Ideally, for the most robust  
236 normalization, each BioTASQ-enriched sample can be accompanied by its respective input control.

237

## 238 **MATERIALS**

### 239 **Biological materials**

240 Cell lines of interest. We have used G4RP successfully in variety of transformed or cancer-derived human  
241 cell lines of different types, including MCF7 ([https://scicrunch.org/resolver/CVCL\\_0031](https://scicrunch.org/resolver/CVCL_0031)), MDA MB-231

242 ([https://scicrunch.org/resolver/CVCL\\_0062](https://scicrunch.org/resolver/CVCL_0062)), HELA ([https://scicrunch.org/resolver/CVCL\\_0030](https://scicrunch.org/resolver/CVCL_0030)), HEK293  
243 ([https://scicrunch.org/resolver/CVCL\\_0045](https://scicrunch.org/resolver/CVCL_0045)), and HT29 ([https://scicrunch.org/resolver/CVCL\\_0320](https://scicrunch.org/resolver/CVCL_0320)).

244 With proper optimization of later steps, we expect G4RP to be compatible with most organisms. Here,  
245 our protocol focuses on cultured cell lines. The use of tissue may also be compatible but will require  
246 additional optimization of tissue homogenization, crosslinking and sonication steps.

247 CAUTION: The cell lines used in your research should be regularly checked to ensure they are authentic  
248 and are not infected with mycoplasma.

## 249 **Reagents**

250 CRITICAL Most standard molecular and cell biology grade reagents can be obtained from various  
251 vendors and do not need to be from the ones specified here.

### 252 Cell culturing

- 253 • Dulbecco's Modified Eagle Medium (DMEM) (Gibco, cat. no. LS11965092)
- 254 • Fetal bovine serum (FBS) (GE Lifescience, cat. no. SH3039602)
- 255 • Penicillin-streptomycin (P/S) (Gibco, cat. no. 151140148)

256

### 257 Standard buffer components in the G4RP protocol

- 258 • Nuclease-free water (Thermos, cat. no. AM9922 or vendors of choice or made in-house using  
259 DEPC treatment)
- 260 • Diethyl pyrocarbonate (DEPC) (Sigma, cat. no. D5758)
- 261 • Formaldehyde (37% wt/vol) (Sigma, cat. no. 252549)

262 CRITICAL: Keep away from light and air; aliquot before use.

263 CAUTION: Formaldehyde is toxic through ingestion, inhalation and skin contact. Always wear  
264 appropriate personal protective equipment and operate only inside a fume hood. All waste  
265 material should be collected separately and disposed according to institutional procedures.

266 • Glycine (Fisher, cat. no. G45-212)

267 • 10X PBS (Fisher, cat no. BP39920)

268 • HEPES (Sigma, cat. no. H3375)

269 • NaCl (Fisher, cat. no. S271-10)

270 • EDTA (Sigma, cat. no. E5134)

271 • EGTA (Sigma, cat. no. E3889)

272 • KCl (Fisher, cat. no. P330)

273 • DTT (Fisher, cat. no. R0861)

274 • IPEGAL (NP40) (Sigma, cat. no. I3021)

275 • Tris (Fisher, cat. no. BP152)

276 • Chloroform (Sigma, cat. no. C2432)

277 CAUTION: Chloroform is toxic if inhaled or swallowed. Always wear appropriate personal  
278 protective equipment and operate only inside a fume hood. All waste material should be  
279 collected separately and disposed according to institutional procedures.

280 • Isopropyl alcohol (Sigma, cat. no. I9516)

281 • Ethyl alcohol

282 • RiboLock RNA inhibitor (ThermoFisher, cat. no. EO0381)

283 • Protease-inhibitor cocktail (PIC) (optional; usually supplied as 100X; NEB #5871S)

284 • TRIZOL (Invitrogen, cat. no. 15596026)

285 CAUTION: TRIZOL is a corrosive material and is toxic through ingestion, inhalation and skin  
286 contact. Always wear appropriate personal protective equipment and operate only inside a  
287 fume hood. All waste material should be collected separately and disposed according to  
288 institutional procedures.

289 G4RP-specific reagents

- 290 • Biotin (Sigma, cat. no. 2031)
- 291 • BioTASQ (available soon through Merck/MilliporeSigma)
- 292 • Promega MagneSphere Paramagnetic (Streptavidin beads) (Promega, cat. no. Z5481)

### 293 Reverse Transcription and PCR

- 294 • Superscript III (Invitrogen, cat. no. 18080093)
- 295 • Random hexamer pd(N)6 (Invitrogen, cat no. N8080127)
- 296 • dNTP mix (NEB, cat no. N0447L)
- 297 • SYBR green PCR-mix (BioRad, cat. no. 1725120)

### 298 Agarose Gel reagents

- 299 • UltraPure Agarose (ThermoFisher, cat. no. 16500100)
- 300 • Bleach (Clorox)
- 301 • SYBR Green II RNA Gel Stain (ThermoFisher, S7564)

### 302 Sequencing kits

- 303 • RNEasy Plus RNA purification kit (Qiagen, cat. no. #74134)
- 304 CRITICAL: Illumina Ribo-Zero has been discontinued and incorporated into the TruSeq Stranded  
305 Total RNA kit. Use either the old Illumina Ribo-Zero + TruSeq RNA library kit or the new Illumina  
306 Stranded Total RNA library prep kit.
- 307
- 308 • Old version: Illumina Ribo-Zero + TruSeq RNA library prep kit (Illumina, cat. no. RS-122-2001)
- 309 • New version: TruSeq Stranded Total RNA library prep (Illumina, cat. no. 20020596)

### 310 **Equipment**

#### 311 Lab equipment

- 312 • Standard cell culture incubator (e.g. HERACell, Thermos)
- 313 • Standard Benchtop centrifuge (e.g. Eppendorf)
- 314 • Temperature controlled centrifuge (or a standard unit in a refrigerated environment)
- 315 • Standard laboratory rocker



- 316 • Standard benchtop aspirator
- 317 • Cell sonicator (e.g. Covaris Ultrasonicator, Diagenode Bioruptor, Branson Sonifier)
- 318 • Tube rotator
- 319 • Nanodrop spectrophotometer (Thermos)
- 320 • Thermal cycler (Bio-Rad)
- 321 • Real-time PCR detection system (Bio-Rad)
- 322 • Gel imager (Alpha Innotech)
- 323 • Illumina HiSeq, NextSeq or NovaSeq sequencer

324 Other consumables

- 325 • Cell and tissue culture dishes (vendor of choice)
- 326 • Cell culture treated flasks (vendor of choice)
- 327 • Cell scrapers (Corning, cat. no. 3011)
- 328 • Wide-mouth tips (VWR, cat. no. 89049)
- 329 • Covaris microtube AFA Fiber Pre-Slit Snap-Cap (Covaris, cat. no. 520045)
- 330 • PCR 8-tube strip and caps (Bio-Rad, TLS0851)
- 331 • Hard-shell 96-well PCR plate (Bio-Rad, HSP9601)

332

333 Bioinformatics Tools

- 334 • Trimmomatic<sup>38</sup> - <https://github.com/timflutre/trimmomatic>
- 335 • HISAT2<sup>39</sup> ([https://scicrunch.org/resolver/SCR\\_015530](https://scicrunch.org/resolver/SCR_015530)) - <http://daehwankimlab.github.io/hisat2/>
- 336 • Integrative Genomics Viewer<sup>40</sup> ([https://scicrunch.org/resolver/SCR\\_011793](https://scicrunch.org/resolver/SCR_011793)) -
- 337 <http://software.broadinstitute.org/software/igv/>
- 338 • HTSeq-count<sup>41</sup> ([https://scicrunch.org/resolver/SCR\\_005514](https://scicrunch.org/resolver/SCR_005514)) -
- 339 <https://htseq.readthedocs.io/en/master/>

340 • DESeq2<sup>42</sup> ([https://scicrunch.org/resolver/SCR\\_015687](https://scicrunch.org/resolver/SCR_015687)) -

341 <https://doi.org/doi:10.18129/B9.bioc.DESeq2>

## 342 **Reagent Setup**

343 **CRITICAL:** All buffers should be sterile and carefully prepared using RNase-free water.

### 344 **DEPC H<sub>2</sub>O**

345 0.1% vol/vol DEPC-treated water can be made by adding 1 mL DEPC to 1000 mL of ddH<sub>2</sub>O. Mix  
346 and incubate at room temperature (20-22°C) for 1h before autoclaving.

### 347 **DEPC-PBS**

348 Dilute 10X PBS (RNase-free) to 1X in DEPC H<sub>2</sub>O or nuclease-free water from a vendor. Store at  
349 room temperature.

350 **1X Fixing buffer** (50 mM HEPES KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH  
351 8.0 in DEPC H<sub>2</sub>O)

352 Prepare a 5X stock solution and store at room temperature for at least 1 year.

### 353 **1M Glycine**

354 Prepare in DEPC H<sub>2</sub>O and filter sterilize. Store at 4°C for up to 2 months.

355 **G4RP Buffer** (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% vol/vol NP40 in  
356 DEPC H<sub>2</sub>O or nuclease-free water)

357 Prepare the stock buffer and store at 4°C for at least 1 year. A working stock should be made for  
358 every new experiment by transferring the required amount for the experiment to a conical tube  
359 and adding fresh RNase inhibitor. Use the manufacturer-recommended amount of RNAase  
360 (typically ~100 U/mL of buffer). Optionally, fresh 1X PIC may be added, especially if the  
361 downstream application involves protein.

### 362 **1 mM Biotin**

363 Make 1 mM stock using pure ddH<sub>2</sub>O. Can be stored at 4°C for at least 3 months.

364 **1 mM BioTASQ**

365 Make 1 mM stock using pure ddH<sub>2</sub>O. Can be stored at 4°C for at least 6 months.

366 **Streptavidin MagneSphere Paramagnetic particles (Promega)**

367 Always mix well before transferring beads with a pipette using wide-mouth pipette tips. The  
368 magnetic beads are stored in a stabilizing buffer (with PBS/ 0.02% wt/vol sodium azide/ 1mg/mL  
369 BSA). You may gently mix well with tapping and aliquot to small volumes (e.g. 30 µL) for later  
370 use. For a new experiment, prepare the beads by washing in 1 mL of DEPC-PBS 3 times with  
371 gentle tapping; use a magnetic rack, aspirate the buffer and resuspend in fresh DEPC-PBS to  
372 make a working stock (for one supplied 0.6 mL tube of beads at 1 mg/mL, resuspend in 100 µL of  
373 DEPC-PBS to make 6 mg/mL). While prepared beads may be stored short term (<1 week) at 4°C,  
374 it is highly recommended that fresh beads are prepared for each experiment for optimal results.

375 **1X TAE (40 mM Tris, 20 mM acetic acid, 1mM EDTA, pH 8)**

376 Make 50X stock and store at room temperature.

377 **PROCEDURE**

378 **CRITICAL** The G4RP protocol itself can be completed in two days. Prior to G4RP, cell culture and live-cell  
379 chemical treatment steps can take a few days to a week to set up. It is recommended that the weekly  
380 schedule is planned to avoid any interruption in the experimental pipeline. In case there are time  
381 constraints or other factors that cause anticipated interruptions, we have included a recommended  
382 pause point in which the experiments can be temporarily halted and resumed later. Please refer to Fig.  
383 1b for an overview of the protocol.

384 **Seeding cells and treatment with G4 ligands TIMING: 2 days to 1 week**

385 1) Seed and treat cells as necessary in DMEM culture medium supplemented with FBS and P/S (or with  
386 the culture medium suitable for your cell line).

387 Generally, for most cell lines, 250,000-1,000,000 cells (minimum 150,000 cells) are seeded per 10 cm  
388 dish for a 3-4 days experiment (including seeding, cell growth and 2-3 day treatments). Seeding  
389 density can be scaled for a shorter or longer experimental period. The exact timeframes need to be  
390 optimized for different cells and treatments. Typically, the cells are harvested while in the log-phase  
391 growth with confluence at around 60-90%.

392 CRITICAL STEP: If G4 ligand treatment is used, make sure a suitable dose is chosen. Cell lines can  
393 respond differently to different G4 ligands. A dose-finding experiment should be performed, and a  
394 low-toxic (LD15 to LD25) dose should be chosen for subsequent experiments.

395 **Crosslinking TIMING: 1-2 h**

396 CAUTION: Formaldehyde is highly toxic. Always handle high concentrations (>4% formaldehyde) of  
397 formaldehyde inside the chemical fume hood. Diluted (1-4%) formaldehyde may be used outside of  
398 hood briefly (<10 min). After the crosslinking step, all formaldehyde waste should be placed in a  
399 separate waste container in the fume hood to be disposed of appropriately according to institutional  
400 procedures.

401 2) Make fresh 1% vol/vol formaldehyde in 1X Fixing Buffer before a new experiment.

402 3) Aspirate media and wash once with 10 mL of DEPC-PBS

403 4) Aspirate PBS and add 10 mL of 1% formaldehyde/1X Fixing Buffer to a 10 cm dish for 5 min on a  
404 rocker at room temperature.

405 5) Quench crosslinking by adding glycine to a final concentration of 125 mM and incubate 5 min on a  
406 rocker at room temp.

407 CRITICAL STEP: The duration of the formaldehyde crosslinking step is important. Try to keep it as  
408 consistent as possible between samples. Stagger the addition of formaldehyde and quenching if  
409 processing several samples at the same time. You may want to optimize the duration of crosslinking.

410 It should generally be between 5 to 10 minutes. Typically, crosslinking over 10 minutes is not  
411 recommended as over-crosslinking may limit the accessibility of epitopes.

412 CRITICAL STEP: After the crosslinking step, keep the sample at 4°C as much as possible.

413 Centrifugation should be done at 4°C as well. All buffers from this point on need to be RNase-free.

414 6) Remove the fixing solution, add 0.8 mL of DEPC-PBS (pre-chilled) and scrape cells into an Eppendorf  
415 tube. Keep on ice.

416 7) Spin tubes at 1500 rpm (200 x g) for 3 min at 4°C.

417 8) Aspirate supernatant (avoid disturbing the cell pellet) and wash again with 1 mL of DEPC-PBS.

418 Aspirate the supernatant and discard.

419 PAUSE POINT: You may freeze down the cell pellets at this stage and store them at -20°C to be

420 processed later. The pellets can be stored for at least 2 months. However, for optimal results, we

421 recommend that you do not pause at this point as processing fresh samples generally gives a better and

422 more consistent yield for the final RNA targets from G4RP.

#### 423 **Sonication (TIMING: 1-2.5 h)**

424 9) Resuspend the pellet in 400 µL G4RP buffer (supplemented with 100U/mL RNase inhibitor; it is  
425 optional to add 1XPIC). If cell pellets were frozen, thaw on ice for 30 min before resuspending in  
426 G4RP buffer.

427 ? TROUBLESHOOTING

428 10) Sonicate on a Covaris m220 focused ultrasonicator using 10% duty factor “Chromatin Shearing”

429 preset for 2 min at 7°C temperature setpoint. The Covaris AFA Fiber microtube-130 is used for the

430 sonication. Mix well before adding the resuspended cells to the tube. Make sure there are no

431 bubbles when adding the sample to the AFA microtubes prior to sonication. Each sonication tube

432 should contain exactly 130  $\mu$ L corresponding to approximately 2-3 million cells. Repeat until all the  
433 resuspended cells are sonicated.

434 CRITICAL STEP: Optimization with specific sonicators (Covaris, Diagenode, and other brands) is  
435 required. The sonication protocol should be similar to that recommended for ChIP according to  
436 sonicator-specific published protocols. We recommend that different durations of sonication be  
437 tested. Generally, the sonication duration should not exceed that recommended for ChIP, and a  
438 good starting point for optimizing the duration is to use half of what is recommended for ChIP. Make  
439 sure the temperature is kept around 2-7°C. Sonication can be checked by proceeding to Steps 22-36  
440 and running a 1% bleach gel<sup>43</sup> as outlined in Mini-procedure Box (Fig. 2a).

441 Mini-procedure Box: RNA bleach gel

- 442 1. Make a gel mix with a final concentration of 1X TAE buffer and 1.5% Agarose in ddH<sub>2</sub>O.
- 443 2. Microwave until boiling then swirl to mix
- 444 3. Add bleach to final concentration of 1%, microwave to boil again. Swirl to mix.
- 445 4. Pour gel mix into gel cast. Let gel cool.
- 446 5. Load RNA samples with 1X loading dye. Run at 100V for 45 min.
- 447 6. Post-stain using SYBR Green dye diluted 1:5000 in 1X TAE buffer for 30 min away from light.
- 448 Image gel on a gel imager.

449 ? TROUBLESHOOTING

- 450 11) Spin sonicated lysate at 13200 rpm (16100 x g) for 10 min at 4°C. Move the supernatant to new  
451 tubes. The supernatant (sonicate), containing the cellular materials that are successfully released  
452 from sheared cells, will be used for the subsequent steps.

453 PAUSE POINT: It is possible to freeze down the sonicate for later processing. Consider using this  
454 pause point rather than the previous one if the issue with the clumpy frozen cell pellets arises.  
455 However, for optimal results, it is recommended to continue directly to the next step.

456 **Binding and washes TIMING: O/N + 4 h**

- 457 12) Each sonicate of ~390  $\mu\text{L}$  can be aliquoted into two tubes (each with 180  $\mu\text{L}$ ): one for the biotin  
458 control and the other for BioTASQ. If the volume is insufficient (<360  $\mu\text{L}$ ), the volume can be  
459 adjusted by adding the small amount of the remaining G4RP buffer, prepared previously, to a final  
460 volume of 390  $\mu\text{L}$ . Save 8.75  $\mu\text{L}$  of the remaining sonicate as a 5% input control. The input control  
461 can be stored temporarily at 4°C then processed later with the other samples.
- 462 13) Add 20  $\mu\text{L}$  of 1 mM stock biotin (20 nmol) to a final concentration of 100  $\mu\text{M}$  to the control tube.  
463 Add 20  $\mu\text{L}$  of 1 mM stock BioTASQ (20 nmol) to a final concentration of 100  $\mu\text{M}$  to the G4RP tube.
- 464 14) Incubate overnight at 4°C on a rotator.
- 465 15) Next day, use wide-mouth tips to add 10  $\mu\text{L}$  of MagneSphere magnetic beads (60  $\mu\text{g}$ ) to each tube.  
466 CRITICAL STEP: The amount of magnetic bead added can be optimized. The recommended amount  
467 is between 20-100  $\mu\text{g}$ , depending on the brand and make of the magnetic beads. An optimization  
468 experiment using different amounts of beads is highly recommended. Generally, a higher amount  
469 increases yield but also increases non-specific background binding. Always mix well before  
470 transferring the beads to another tube.
- 471 16) Incubate for 2 h at 4°C on a rotator.
- 472 17) Use the magnet to pull beads to the side by leaving tubes on the magnet for 2 min. Shift the tubes  
473 slightly to pull beads closer together. Aspirate carefully without removing the beads.
- 474 18) Wash beads 3 times using 1 mL of G4RP buffer. Put on rotator at room temp for 5 min for each  
475 wash.
- 476 19) Wash 2 times in 1 mL of DEPC-PBS for 2 min each.

477 20) Resuspend beads in 100  $\mu$ L of DEPC-PBS (supplemented with RNase inhibitor). Take out the 5%  
478 input saved from Step 12 and add 100  $\mu$ L DEPC-PBS (with RNase inhibitor) to it as well.

479 21) Incubate at 70°C for 1 h to reverse crosslinks.

480 **RNA purification TIMING: 2 h**

481 CAUTION: TRIZOL and chloroform should be handled in the chemical fume hood. Waste should be  
482 disposed of appropriately according to institutional procedures.

483 22) Add 1 mL of TRIZOL to each tube directly to the beads and mix gently by flicking the tubes 10-15  
484 times.

485 23) Incubate the sample in TRIZOL for 5-10 min at room temp to dissociate any ribonucleoproteins  
486 (RNPs) and other structures.

487 24) Add 0.2 mL of chloroform. Make sure the cap is sealed and vigorously shake for 15 s. Leave at room  
488 temp for 3 min.

489 25) Centrifuge at 12000 x g for 15 min at 4°C. The mixture will separate into the top aqueous phase, the  
490 middle interphase and the lower pink-coloured organic phase. The aqueous phase contains the RNA.

491 26) Transfer the aqueous phase to a new tube by tilting the tube at 45° and slow pipetting. Avoid  
492 disrupting the interphase/organic phase.

493 27) Add 0.5 mL of fresh isopropanol to the tube with the aqueous phase. Incubate at room temp for 10  
494 min.

495 28) Centrifuge for 10 min at 12000 x g at 4°C.

496 CRITICAL STEP: The RNA pellet will be very small and unlikely to be seen by naked eye. Do not touch  
497 the bottom of the tube with the pipette. Pipette out slowly by drawing against the side of the tube.

498 29) Carefully and slowly discard the supernatant but leave 10-15  $\mu$ L of liquid at the bottom.

499 30) Add 1 mL of fresh 75% (vol/vol) ethanol.



500 PAUSE POINT: You may store pellets in ethanol at 4°C for up to 3 days or -20°C for months.

501 31) Vortex briefly and centrifuge at 7500 x g for 5 min at 4°C.

502 CRITICAL STEP: It is unlikely that you will see a pellet here, as the amount of RNA is very low. Avoid  
503 touching the pipette tip to the bottom or creating any motion that could disturb the pellet. The  
504 addition of 25 µg of linear acrylamide during isopropanol precipitation step may also be considered  
505 to assist with recovery and visualization.

506 32) Slowly remove the ethanol leaving approximately 10 µL of liquid at the bottom of the tube.

507 33) Keep the lids of the tubes open and air dry for 10 min at room temp.

508 34) Add 30 µL of RNase-free water and pipette up and down.

509 35) Incubate at 55°C for 15 min.

510 36) Check RNA concentration and quality using Nanodrop.

511 ? TROUBLESHOOTING

512 PAUSE POINT: RNA samples can be stored long term at -70°C. Freezing and thawing should be  
513 minimized to avoid RNA degradation. We recommend to immediately process the reverse  
514 transcription-qPCR to check for G4RP efficiency. The remaining stored samples may be used to  
515 process for library preparation after successful validation through RT-qPCR.

516 **Reverse transcription and qPCR quantification TIMING: 4-5 h**

517 CRITICAL: After adding different components, always mix by gently flicking the tube, followed by a  
518 quick pulse with a centrifuge. Do not vortex.

519 37) Thaw purified RNA (including all G4RP, biotin control, and input control samples) from Step 36 on  
520 ice.

521 38) A master mix can be made with pdN6 and dNTP. Add 2 µL to each tube containing RNA and mix:

<b>Component (1<sup>st</sup> part)</b>	<b>Amount (μL)</b>	<b>Final concentration in RT reaction</b>
Purified RNA sample from Step 36	12.5	varies
50 μM random hexamer (pdN6)	1	2.5 μM
10 mM dNTP mix	1	500 μM

522

523 39) Incubate at 65°C for 5 min.

524 40) Place on ice for 5 min.

525 41) Make a master mix of the following components and gently mix by flicking. Add 20 μL to each tube:

<b>Component (2<sup>nd</sup> part)</b>	<b>Amount (μL)</b>	<b>Final concentration in RT reaction</b>
5X first strand buffer	4	1X
0.1 M DTT	1	5 mM
RNAse inhibitor (40U/μL)	0.25	10U
Superscript III (200U/μL)	0.25	50U
Total	20	

526

527 42) Gently mix by flicking the RT reaction tube and pulse spin by brief centrifugation.

528 43) Incubate at room temp for 5 min.

529 44) Then, incubate at 55°C for 45 min. These incubations (steps 43-45) can be performed on a  
530 thermocycler.

531 45) Inactivate reaction by incubating at 70°C for 15 min and cool down to room temp. Proceed to real-  
532 time qPCR with the appropriate primer set and SYBR mix of your choice. Table 2 includes primer sets  
533 for several genes that have been tested in MCF7, MDA MB-231 and HEK293 cell lines. However,  
534 depending on your experimental objectives and models, you may require a different set of controls.

535 PAUSE POINT: The cDNA samples can be stored at -20°C long term.

536 46) Make a master mix with the appropriate primer set. Typically, we suggest the use of two to three  
537 technical replicates. Use qPCR plates or strip tubes compatible with the real-time qPCR machine. An  
538 example of a single qPCR reaction is as follows:

539

Component	Amount (μL)	Final concentration in RT reaction
forward primer (10 μM)	0.5	500 nM
reverse primer (10 μM)	0.5	500 nM
2X SYBR mix	5	1X
PCR-grade H <sub>2</sub> O	3	-
G4RP RT sample (add last)	1	varies
Total	10	

540

541 47) Run the real-time qPCR using the default amplification protocol for your instrument. It should  
542 typically take around 20-30 cycles to reach exponential phase (~2-3 hours).

543 48) Normalization is performed by calculating the percentage of G4RP or biotin control sample against  
544 its respective input control using the following formula:

$$G4RP \text{ qPCR signal} = 5 * (2^{\text{Average}(Cq_{input}) - Cq_{G4RP \text{ or biotin}}})$$

545 The G4RP qPCR signal is expressed as % of total amount of input going through G4RP that is  
546 effectively captured by BioTASQ. For each technical replicate, find the difference between the average  
547 of the Cq input and the individual technical replicate. Then, multiply by two to power of the previously  
548 calculated value (difference in the Cq values) by 5 (due to using 5% of the total input for normalization  
549 purposes).

550 ? TROUBLESHOOTING

Gene	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
<i>NRAS</i>	ATGACTGAGTACAACTGGTGGT	CATGTATTGGTCTCTCATGGCAC
<i>VEGFA</i>	CCTTGCCTTGCTGCTCTACC	AGATGTCCACCAGGGTCTCG
<i>MALAT1</i>	AAAGCAAGGTCTCCCCACAAG	GGTCTGTGCTAGATCAAAAGGCA

551

## 552 **Ribo-depletion and Library Preparation (TIMING: 6-8 h)**

553 49) Thaw the remaining TRIZOL-purified RNA samples (biotin-control, G4RP, and input; from step 36)

554 that have been validated successfully with RT-qPCR (steps 37-48) on ice.

555 50) It is recommended to have an additional column purification step before RNA-seq library  
556 preparation to remove any potential residual contamination (e.g. phenol, guanidium salt, etc.) and  
557 DNA contaminant. We recommend using the Qiagen RNeasy Plus Mini Kit and following the  
558 manufacturer's protocol. Elute at the low volume (<25  $\mu$ L) in RNase-free water. Quantify RNA for  
559 concentration using Qubit fluorometer.

560 51) For library preparation, use the Illumina TruSeq RNA Library prep kit (with Ribo-Zero) following  
561 manufacturer's protocol. We recommend a minimum RNA input of >50 ng for the library  
562 preparation process. To minimize PCR bias, higher concentration (>200 ng) is preferred if sufficient  
563 input material is available. Briefly, the steps are as follows:

- 564 • Deplete rRNA using Ribo-Zero oligos and beads.
- 565 • Thermally fragment RNA briefly at 4 min.
- 566 • Synthesize the first-strand of cDNA followed by the second cDNA strand.
- 567 • Adenylate 3' ends and ligate adapters.
- 568 • Enrich DNA fragments by PCR amplification. Use the minimum number of PCR cycles if  
569 possible (5-7 cycles).
- 570 • Quality-check and quantify library using the BioAnalyzer or equivalent methods.

571 CRITICAL STEP: If possible, use a higher amount (200-1000 ng) of input to improve library quality and  
572 reduce PCR cycles. G4 sequences may be sensitive to PCR by inhibiting polymerase processivity and  
573 could be underrepresented after a higher number of cycles. If the input is too low for individual  
574 samples, it is recommended to pool multiple replicate samples together.

575 CRITICAL STEP: The fragmentation step can be modified depending on the fragment size. Due to the  
576 sonication step earlier, the average length of RNA fragments will be reduced. To avoid over-

577 fragmentation, the duration of the thermal fragmentation should be reduced. We recommend 2-4  
578 min of thermal fragmentation.

579 52) Pool sequencing libraries at equivalent proportion and sequence on a HiSeq, NextSeq or NovaSeq  
580 platform using pair-end and read length of 75bp or higher. We recommend the sequencing depth of  
581 each sample to be 15-20 million reads.

## 582 **Sequencing Bioinformatics analyses TIMING: varies**

583 CRITICAL An overview of the G4RP-seq analysis steps is depicted in Fig. 2b. For investigators who are less  
584 experienced with bioinformatics or lack a local workstation computer (minimum requirement: >8 core  
585 processor, >32 GB memory, >3 TB hard drive) performing data analysis through a remote Galaxy server<sup>44</sup>  
586 (usegalaxy.org), which contains all the required bioinformatics packages, is highly recommended.

587 53) Quality-check raw FASTQ file using FASTQC and trim reads using Trimmomatic (38) (or equivalent  
588 read-trimming tool) with default settings.

589 54) Align reads to the target genome (e.g. human genome hg19 or hg38 assembly) using HISAT2 (39)  
590 using default settings. The output file should be formatted as the compressed binary version of  
591 sequence alignment files (BAM format). Integrative Genome Viewer (IGV) (40) can then be used to  
592 visualize these alignments. To load in IGV, index files will need to be generated from igvtools for  
593 each BAM file.

594 55) Prepare and download a gene annotation file (in GTF format) from UCSC. This can be accessed at  
595 <http://genome.ucsc.edu/cgi-bin/hgTables> . Make sure to choose the right genome and assembly  
596 version. Choose “Genes and Gene Predictions” for the group and “NCBI RefSeq” in the track. To use  
597 gene symbols as an attribute of interest, choose “refFlat” for the table. The parameters may be  
598 changed depending on what attributes you want to be listed. Select “genome” for region and GTF as  
599 output format. The resulting GTF file can be either downloaded directly or sent as output to Galaxy.

600 56) Count features and annotate the aligned reads using HTSeq-count (41) with the prepared GTF file.  
601 Use “Union” mode and set values as follows: feature type=exon, ID attribute=gene\_id. Depending  
602 on the library prep kit used, you can set the strandedness mode either “stranded” or “non-  
603 stranded.” The “stranded” setting would keep data on the directionality of the transcript.  
604 57) Use the raw count output from HTSeq-count as input for DESeq2 (42) for normalization and  
605 differential expression analysis.  
606 58) Run DESeq2 for the BioTASQ sample vs input control for each treatment condition to generate  
607 relative BioTASQ enrichment scores (ES), which are log fold change readout between the test  
608 conditions for each transcript. The gene transcripts can be ranked according to their enrichment  
609 score. It is important to note that all subsequent comparisons are relative measures.  
610 59) Compare the enrichment scores between conditions to quantify ligand-induced changes in G4 RNA  
611 structures. The enrichment score change ( $\Delta$ ES) is calculated as follows: ES (treated) - ES (untreated).  
612 60) Using Excel, filter transcripts by counts (as a measure of RNA abundance) and fold change. The filter  
613 threshold can differ depending on the quality and depth of sequencing.

#### 614 **TIMING**

615 Step 1, Cell culture phase (including seeding and treatment): 2 days to 1 week (depending on cell  
616 culturing experimental setup)

617 Steps 2-8, Crosslink: 1-2 hours (depending on the number of samples)

618 Steps 9-11, Sonication: 1-2.5 hours (depending on the number of samples)

619 Steps 12-21, Binding and washing: overnight + 4 hours

620 Step 22-36, RNA purification: 2 hours

621 Step 37-48, Reverse transcription and qPCR quantification: 4-5 hours

622 Step 49-51, Ribodepletion and library preparation: 6-8 hours

623 Step 52-60, Sequencing and bioinformatics – varies (depending on sequencing queues and

624 computational power)

625

626

### 627 TROUBLESHOOTING

628 Please refer to Table 3 for advice on troubleshooting.

<b>Step</b>	<b>Problem</b>	<b>Possible Reason</b>	<b>Solution</b>
9	Cells are too clumpy and difficult to resuspend	Freezing after formaldehyde may cause clumping	Do not freeze down pellet and directly proceed to the sonication step.
10	No bands corresponding to purified RNA from sonication check fraction visible on gel	Crosslink reversal is insufficient	Increase duration of the crosslinking reversal step to 2h.
	Either faint bands or bands with apparently high molecular weight observed in purified RNA of sonication check fraction	Cells not sonicated properly.  Cellular components not released into the supernatant	Reoptimize the sonication step. Increase sonication duration.  Add 0.1% (wt/vol) SDS to resuspended cells before sonication
36	Low A260/280 ratio for RNA	Contamination from TRIZOL components (e.g. Phenol), protein or DNA.	Wash twice, instead of once, with ethanol during the TRIZOL extraction. This will normally not impact qPCR quantification. Perform an additional column purification on RNA before sequencing library preparation.

	Low RNA yield of all extracted samples including input control	Cells not sonicated properly; cellular components not released into the supernatant.	Reoptimize the sonication step. Increase sonication duration. Adding 0.1% SDS to resuspended cells before sonication may also help.
		RNase contamination	Remake new buffers and ensure no RNase contamination is present. Add in fresh RNase inhibitors in every new experiment.
		Problem with TRIZOL RNA purification	Check reagents used for RNA purification (i.e. TRIZOL, chloroform, isopropanol, ethanol). Change to fresh reagents.
		Insufficient reversal of crosslinking leading to loss of RNA to RNPs during TRIZOL step	Increase duration of the crosslinking reversal step to 2h.
48	High signal in biotin control	Concentration of streptavidin beads too high	Add lower amounts of beads Optimize pull-down by testing different amounts of beads at the lower range (20-60 µg)
		Input too high due to too many cells being used which reduces signal-to-background ratio	Decrease number of cells for sonication or dilute cells in higher volume of shearing buffer prior to sonication
48	Lower enrichment of BioTASQ sample as detected by RT-qPCR (G4RP qPCR signal is <0.5%)	Incorrect concentration of BioTASQ and beads are used	Confirm that the stock concentrations are correct. Remake stock concentration if uncertain.
		Insufficient amounts of beads	Increase the amounts of beads; optimize pull-down using different amount of beads at the higher range (50-250 µg)
		Over crosslinking leading to loss of accessibility to the G4 epitopes	Reduce crosslinking duration; it should never exceed 10 min. Some formaldehyde contains additives (e.g. methanol) may increase cell membrane permeability and crosslinking efficiency, consider using fresh formaldehyde without additives. Make fresh glycine to quench effectively. Slightly increase glycine concentration.



		RNAse contamination	Remake new buffers and ensure no RNase contamination is present. Add in fresh RNAse inhibitors in every new experiment.
Low ligand-induced change detected		Under crosslinking leading to bias towards BioTASQ stabilization	Increase crosslinking duration in the range between 5 min to 10 min. Keep samples on ice or 4°C after crosslinking.
		Treatment with an insufficient dose of G4 ligand	Cell lines can exhibit varying degrees of sensitivity to G4 ligands. Perform a dose-response curve to profile the cytotoxicity of a specific G4 ligand in a specific cell line and chose a subtoxic dose. Reconsider the treatment duration. Typical timepoints include 24h and 48h. Depending on the stability of the ligand, new stocks may need to be produced or purchased.
Low qPCR signal (high Ct values; >35 cycles) for all samples		Problems with reverse transcription and/or qPCR	Check reagents for RT-qPCR. Run a positive control with a different confirmed sample.

629

630

631

632 **ANTICIPATED RESULTS**

633 **Sonication check**

634 Due to the different sonication techniques used by different laboratory groups, it is advised that a  
635 sonication check be performed using different sonication conditions. For example, for the Covaris m220  
636 Ultrasonicator, we divided a crosslinked sample and performed parallel comparisons between different  
637 sonication conditions (sonication duration: 0 min, 2 min and 4 min) under the same settings. Sonicated  
638 samples can be heated for 1h to reverse crosslinks, and the RNA can be extracted using TRIZOL or RNA  
639 column purifications. The purified RNA can then be observed on an RNA bleach gel stained with SYBR

640 green nucleic acid gel stain<sup>43</sup>. As depicted in Fig. 2a, absence of sonication would result in an inefficient  
641 release of RNA from crosslinked material. Note that crosslink reversal is inefficient in unsonicated  
642 material and thus would result in a loss of extracted RNA, as shown by the lack of RNA band on the gel  
643 for the 0 min condition. Generally, a sonication duration of 2-4 min is sufficient, yielding a range of  
644 bands between 200-800 nt. Longer sonication duration leads to a shorter average length of RNA  
645 fragments, as seen by the shifting of the median band length between the 2 and 4 minutes of sonication  
646 conditions. It is important to note that, while longer sonication may help increase resolution, it may also  
647 disrupt some interactions. Thus, a balanced approach is required to prevent under- or over-sonication.

#### 648 **RNA yield from G4RP**

649 After TRIZOL-purification, the purified RNA can be quantified using Nanodrop or other comparable  
650 methods. A lower A260/280 ratio (<1.8) is usually indicative of contamination from TRIZOL extraction.  
651 However, A260/280 value may not be reliable if the sample concentration is too low. In that case, a  
652 more sensitive method, such as Qubit, may be employed. Typically, the concentration of the 5% input  
653 sample is around 100 ng/μL (with a 30 μL yield = 3 μg). The concentration of the BioTASQ-enriched  
654 sample varies around 100-200 ng/μL (30 μL yield of 3-6 μg). The concentration of the biotin control  
655 sample is much lower at ~50 ng/μL (30 μL yield of 1.5 μg), as the biotin control consists mostly of non-  
656 specific rRNA binding to the beads. Regardless of purity or concentration, we highly recommend column  
657 purification before further processing (i.e. ribodepletion and library preparation) for RNA-seq.

#### 658 **G4RP-qPCR quantification**

659 RT-qPCR quantification of the G4RP samples typically yields high signal-to-background ratios (>20 fold  
660 enrichment) when comparing % input yield from BioTASQ versus the biotin control sample (Fig. 2b).  
661 Ligand-induced change ranges from 2-fold to as high as 12-fold, depending on the RNA target and G4  
662 ligand used.

## 663 **G4RP-seq analysis**

664 After processing the raw sequencing files through a series of bioinformatics tools (Fig. 3a), the  
665 differential expression values (output from DESeq2) derived from comparing the BioTASQ sample with  
666 its corresponding input control, are used to determine the BioTASQ enrichment score. A quick indicator  
667 of a successful G4RP-seq is to rank the transcripts by enrichment score and check for an increase in GC  
668 content. At baseline levels (derived from samples not treated with a G4 ligand), the enrichment score is  
669 expected to be positively correlated with the GC% of the gene transcript (Fig.3b). Signals of aligned  
670 reads can be visualized using IGV (Fig. 3c). The biotin control is expected to show negligible levels with  
671 reads primarily aligned to rRNA regions. By comparing the BioTASQ signal with that of the input control,  
672 the relative level of BioTASQ interaction with the target can be elucidated (Fig. 3c). An example of a high  
673 G4-ligand-inducible target is the long non-coding RNA, MALAT1. The raw signals show much higher  
674 alignments in the G4-ligand-treated samples compared to that of the untreated sample.

675

## 676 **FIGURE LEGENDS**

677

678 **Figure 1. Overview of G4RP. a,** Chemical structure of BioTASQ in its open (in absence of a G4 target) and  
679 closed conformations (in the presence of a G4 target). **b,** Schematics of the G4RP workflow. 1) Cells are  
680 cultured in 10 cm culture dishes. If required, cells may be treated with G4 ligands at this stage to  
681 increase global G4 levels. 2) Cells are crosslinked using formaldehyde. 3) Cells are harvested by scraping  
682 and sonicated using an ultrasonicator. 4) BioTASQ is added to the sonicate. 5) Streptavidin beads are  
683 added. 6) Beads are washed multiple times. 7) Beads are heated to reverse the crosslinks. 8) TRIZOL is  
684 added to the sample, and the RNA fraction is purified. 9) Reverse-transcription is performed on the RNA  
685 sample before quantification using real-time quantitative PCR.

686 **Figure 2. Expected results from G4RP. a**, An example of an bleach RNA gel (1.5% agarose w/v, 1%  
687 bleach, v/v, 1X TAE; post-staining with SYBR Green Gel Stain) for crosslinked samples that have  
688 undergone different durations of sonication (0, 2, 4 min). **b**, (Top) G4RP-qPCR signal of BioTASQ vs biotin  
689 control for RNA transcript targets, VEGFA and NRAS. Multiple t-test was performed. (Bottom) G4 ligand  
690 (BRACO-19<sup>45</sup> and RHPS4<sup>46</sup>)-induced changes in the G4RP-qPCR signals for targets, VEGFA, NRAS and  
691 MALAT1. One-way ANOVA with Dunnett's multiple correction was performed for each gene to compare  
692 each condition against its respective non-treated control. Each point represents a biological replicate.  
693 Three biological replicates were used to quantify the means. Error bars represent SEM. These data are  
694 derived from a previous published work<sup>16</sup>.

695 **Figure 3. Expected results from G4RP-seq. a**, Schematic of the G4RP-seq bioinformatics analyses. **b**,  
696 Linear regression plot comparing the BioTASQ Enrichment Score with GC content of respective gene  
697 transcripts at baseline (derived from untreated sample). Each dot represents a gene transcript (n=5083).  
698 Analysis shows a significant non-zero relationship with a positive slope ( $p < 0.0001$ ) and a moderate  
699 positive correlation ( $R = 0.47$ ). **c**, Signal of aligned sequencing reads in the MALAT1 region for (Left) biotin  
700 control, BioTASQ in untreated sample, BioTASQ in BRACO-19-treated sample and BioTASQ in RHPS4-  
701 treated sample and (Right) the corresponding input controls. The G4RP-seq datasets used are available  
702 under the accession number GSE112898.

### 703 DATA AVAILABILITY

704 All relevant data are available upon request. The G4RP-qPCR dataset (Fig. 2) is publicly available on  
705 FigShare (<https://doi.org/10.6084/m9.figshare.16684870>). The raw sequencing data used in the  
706 examples shown in Fig. 3 in this protocol are publicly available and accessible at GSE112898.

707

708 **ACKNOWLEDGEMENTS**

709 This work is supported by the Research Reinvestment Funds from the University of British Columbia and  
710 Agence Nationale de la Recherche (ANR-17-CE17-0010-01). Pauline Lejault is warmly acknowledged for  
711 her critical analysis of the G4RP protocol and of the manuscript.

712 **AUTHOR CONTRIBUTIONS**

713 S.Y.Y., D.M. and J.M.Y.W designed the experiments. S.Y.Y. and J.M.Y.W. wrote the manuscript with  
714 inputs from D.M.

715 **COMPETING INTERESTS**

716 The authors declare no competing interests.

717 **REFERENCES**

- 718 1. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology  
719 and structure. *Nucleic Acids Research* **34**, 5402–5415 (2006).
- 720 2. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-  
721 Quadruplexes. *Trends in Chemistry* **2**, 123–136 (2020).
- 722 3. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids*  
723 *Research* **43**, 8627–8637 (2015).
- 724 4. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular  
725 Mechanisms. *Journal of Molecular Biology* **429**, 2127–2147 (2017).
- 726 5. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple  
727 species. *Nucleic Acids Research* **47**, 3862–3874 (2019).
- 728 6. Maizels, N. G4-associated human diseases. *EMBO reports* **16**, 910–922 (2015).

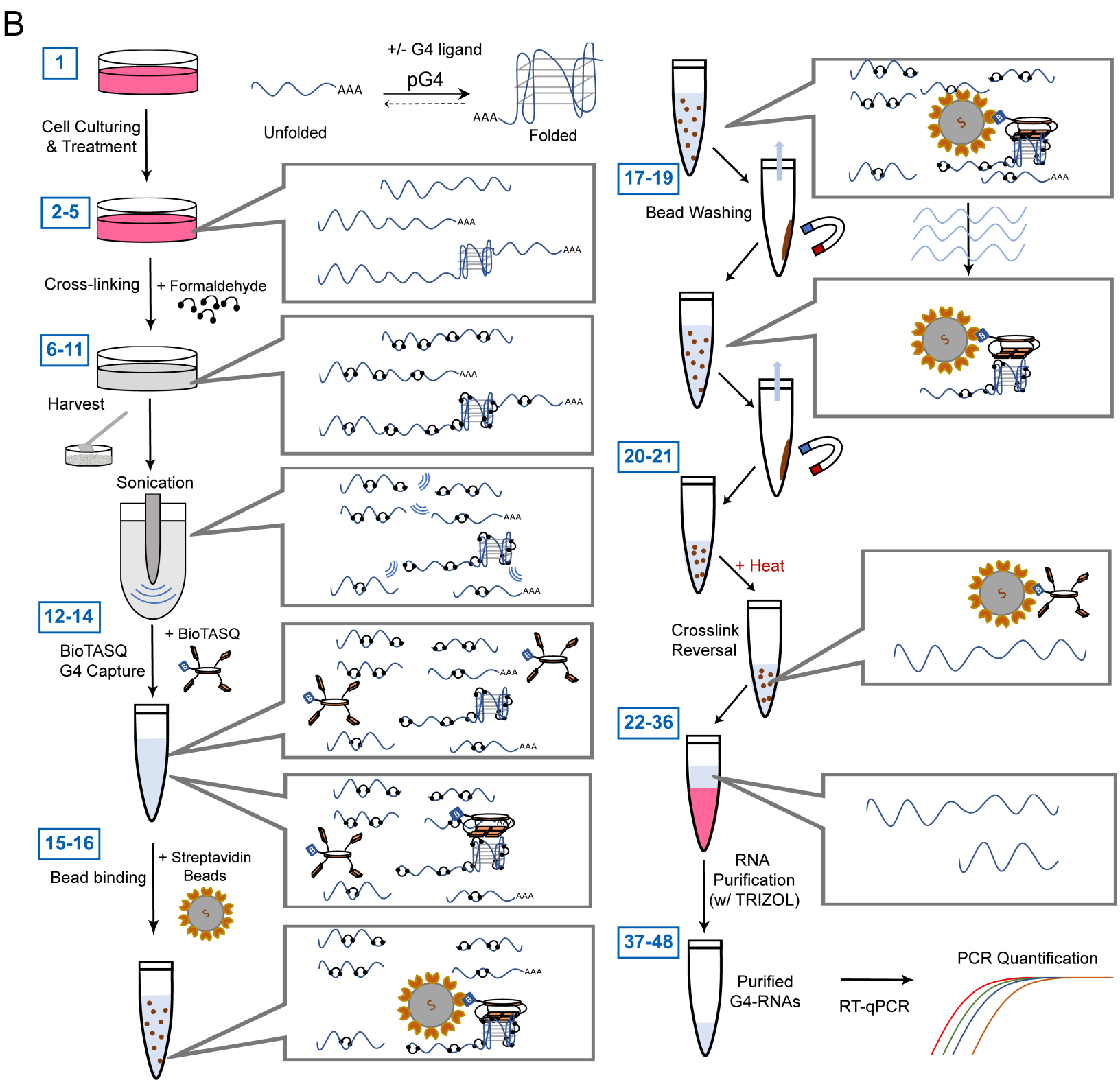
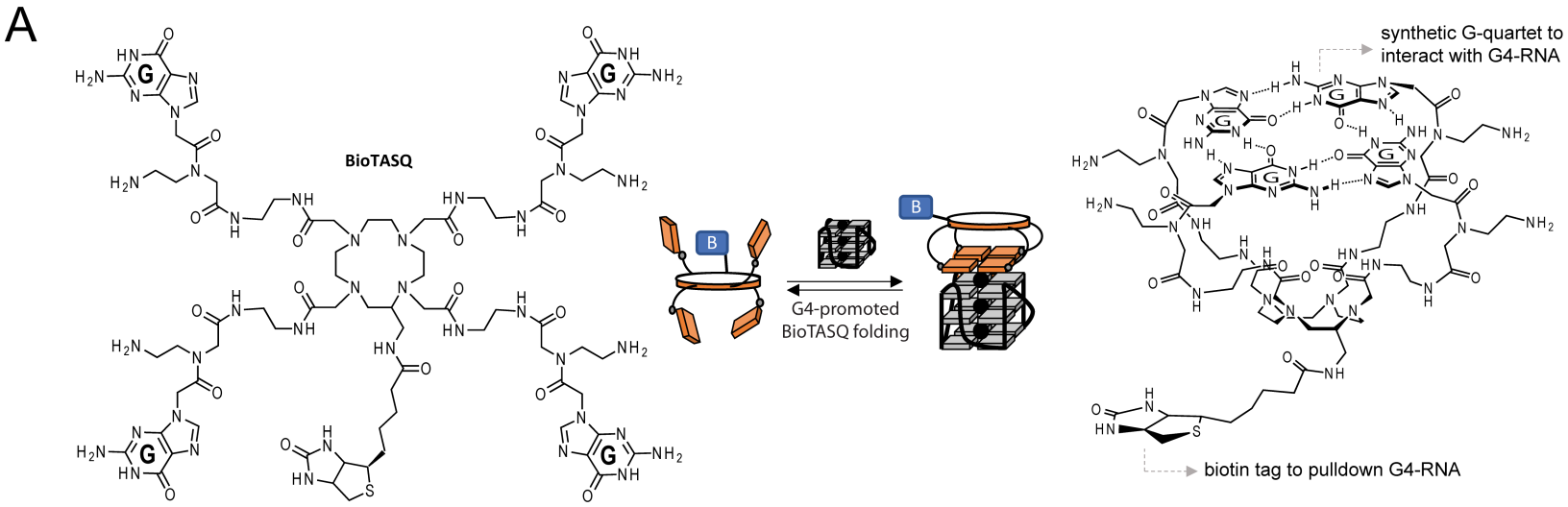
- 729 7. Raguseo, F., Chowdhury, S., Minard, A. & Di Antonio, M. Chemical-biology approaches to probe DNA  
730 and RNA G-quadruplex structures in the genome. *Chemical Communications* **56**, 1317–1324 (2020).
- 731 8. Puig Lombardi, E. & Londoño-Vallejo, A. A guide to computational methods for G-quadruplex  
732 prediction. *Nucleic Acids Research* **48**, 1–15 (2019).
- 733 9. Monchaud, D. Quadruplex detection in human cells. *Annual Reports in Medicinal Chemistry* 133–160  
734 (2020) doi:10.1016/bs.armc.2020.04.007.
- 735 10. Kwok, C. K. & Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological  
736 Application. *Trends in Biotechnology* **35**, 997–1013 (2017).
- 737 11. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-  
738 quadruplex structures in human cells. *Nature Chemistry* **5**, 182–186 (2013).
- 739 12. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human  
740 genome. *Nature Biotechnology* **33**, 877–881 (2015).
- 741 13. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide  
742 mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-  
743 throughput sequencing. *Nature Protocols* **13**, 551–564 (2018).
- 744 14. Zheng, K. *et al.* Detection of genomic G-quadruplexes in living cells using a small artificial  
745 protein. *Nucleic Acids Research* (2020) doi:10.1093/nar/gkaa841.
- 746 15. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals  
747 widespread formation of G-quadruplex structures in the human transcriptome. *Nature Methods* **13**,  
748 841–844 (2016).
- 749 16. Yang, S. Y. *et al.* Transcriptome-wide identification of transient RNA G-quadruplexes in human  
750 cells. *Nature Communications* **9**, (2018).
- 751 17. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R. M. & Mergny, J.-L. G-quadruplexes and  
752 helicases. *Nucleic Acids Research* **44**, 1989–2006 (2016).

- 753 18. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions  
754 of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology* **21**, 459–474 (2020).
- 755 19. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the  
756 cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20**, 474–489 (2019).
- 757 20. Rodriguez, R. *et al.* A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-  
758 Damage Response at Telomeres. *Journal of the American Chemical Society* **130**, 15758–15759  
759 (2008).
- 760 21. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted  
761 in bacteria. *Science* **353**, aaf5371–aaf5371 (2016).
- 762 22. Umar, M. I., Ji, D., Chan, C.-Y. & Kwok, C. K. G-Quadruplex-Based Fluorescent Turn-On Ligands and  
763 Aptamers: From Development to Applications. *Molecules* **24**, 2416 (2019).
- 764 23. Stefan, L. & Monchaud, D. Applications of guanine quartets in nanotechnology and chemical  
765 biology. *Nature Reviews Chemistry* **3**, 650–668 (2019).
- 766 24. Haudecoeur, R., Stefan, L., Denat, F. & Monchaud, D. A Model of Smart G-Quadruplex  
767 Ligand. *Journal of the American Chemical Society* **135**, 550–553 (2013).
- 768 25. Laguerre, A. *et al.* A Twice-As-Smart Synthetic G-Quartet: PyroTASQ Is Both a Smart Quadruplex  
769 Ligand and a Smart Fluorescent Probe. *Journal of the American Chemical Society* **136**, 12406–12414  
770 (2014).
- 771 26. Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *Journal of the American Chemical*  
772 *Society* **137**, 8521–8525 (2015).
- 773 27. Renard, I. *et al.* Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in  
774 vitro and in vivo through the G4RP protocol. *Nucleic Acids Research* **47**, 5502–5510 (2019).
- 775 28. Gilbert, C. & Svejstrup, J. Q. RNA Immunoprecipitation for Determining RNA-Protein Associations In  
776 Vivo. *Current Protocols in Molecular Biology* **75**, 27.4.1-27.4.11 (2006).

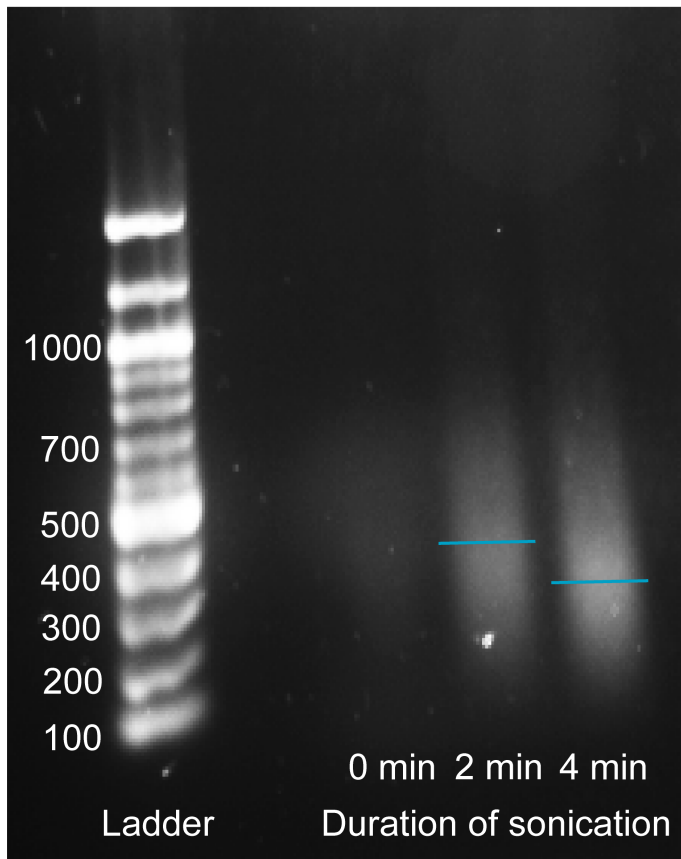
- 777 29. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex  
778 isolation from human cells. *Nature Chemistry* **2**, 1095–1098 (2010).
- 779 30. Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *Journal of Medicinal Chemistry* **59**,  
780 5987–6011 (2016).
- 781 31. Kubota, M., Tran, C. & Spitale, R. C. Progress and challenges for chemical probing of RNA structure  
782 inside living cells. *Nature Chemical Biology* **11**, 933–941 (2015).
- 783 32. Sharma, E., Sterne-Weiler, T., O’Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA  
784 Interactions. *Molecular Cell* **62**, 618–626 (2016).
- 785 33. Nguyen, T. C. *et al.* Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nature*  
786 *Communications* **7**, (2016).
- 787 34. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-  
788 Order Organization and Regulation. *Molecular Cell* **62**, 603–617 (2016).
- 789 35. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**,  
790 1267–1279 (2016).
- 791 36. Ruggiero, E., & Richter, S. N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral  
792 therapy. *Nucleic acids research*, **46**, 3270-3283 (2018).
- 793 37. Saranathan, N., & Vivekanandan, P. G-quadruplexes: more than just a kink in microbial  
794 genomes. *Trends in microbiology*, **27**, 148-163 (2019).
- 795 38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence  
796 data. *Bioinformatics* **30**, 2114–2120 (2014).
- 797 39. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory  
798 requirements. *Nature Methods* **12**, 357–360 (2015).
- 799 40. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-  
800 performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192  
801 (2012).



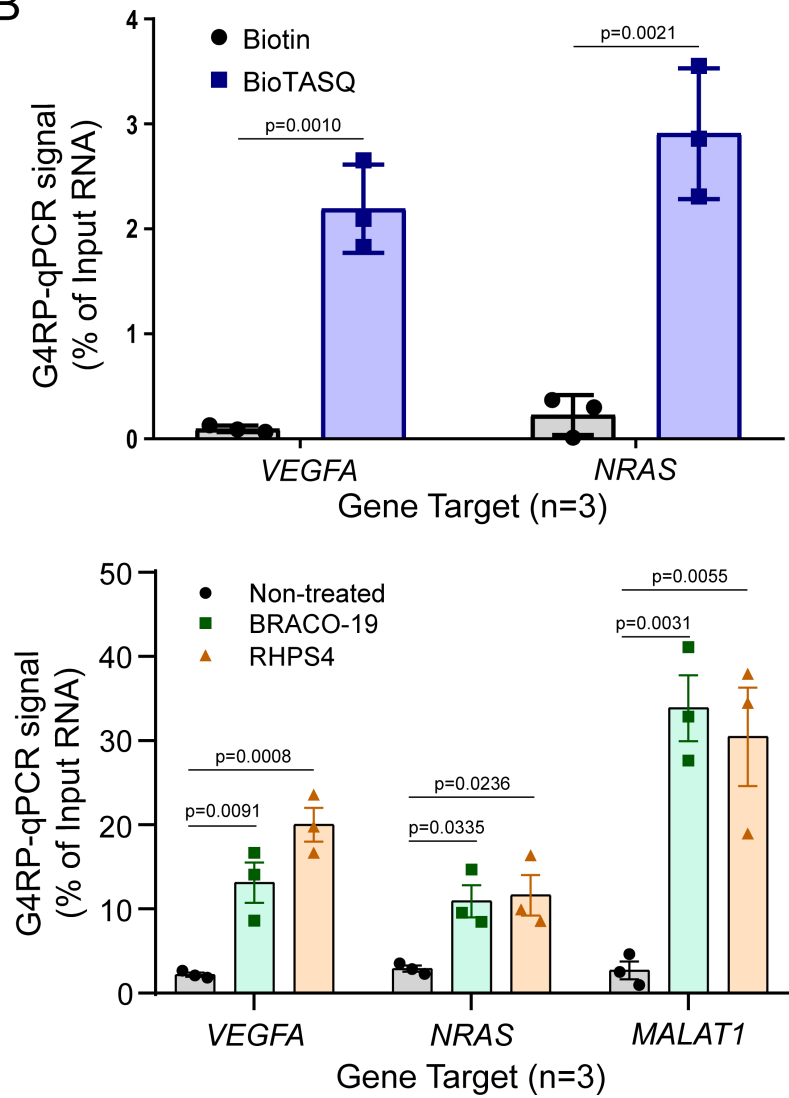
- 802 41. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput  
803 sequencing data. *Bioinformatics* **31**, 166–169 (2014).
- 804 42. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq  
805 data with DESeq2. *Genome Biology* **15**, (2014).
- 806 43. Aranda, P. S., LaJoie, D. M. & Jorcyk, C. L. Bleach gel: A simple agarose gel for analyzing RNA  
807 quality. *ELECTROPHORESIS* **33**, 366–369 (2012).
- 808 44. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy Team, T. Galaxy: a comprehensive approach for  
809 supporting accessible, reproducible, and transparent computational research in the life  
810 sciences. *Genome Biology* **11**, R86 (2010).
- 811 45. Read, M. *et al.* Structure-based design of selective and potent G quadruplex-mediated telomerase  
812 inhibitors. *Proceedings of the National Academy of Sciences* **98**, 4844–4849 (2001).
- 813 46. Heald, R. A. *et al.* Antitumor Polycyclic Acridines. 8.1Synthesis and Telomerase-Inhibitory Activity of  
814 Methylated Pentacyclic Acridinium Salts. *Journal of Medicinal Chemistry* **45**, 590–597 (2002).
- 815



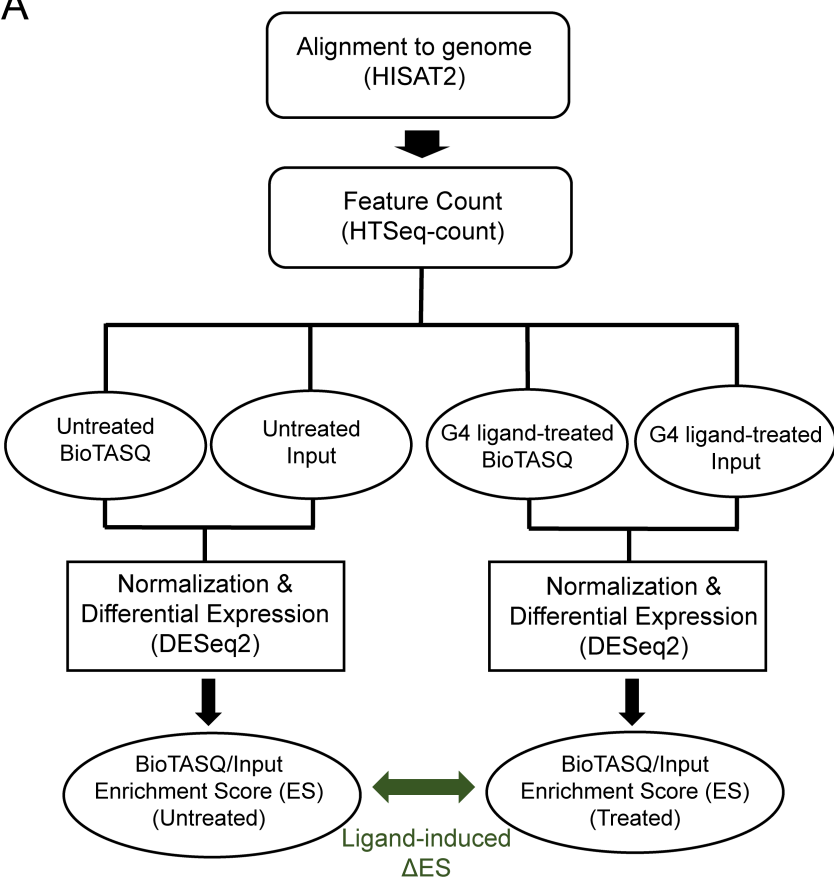
A



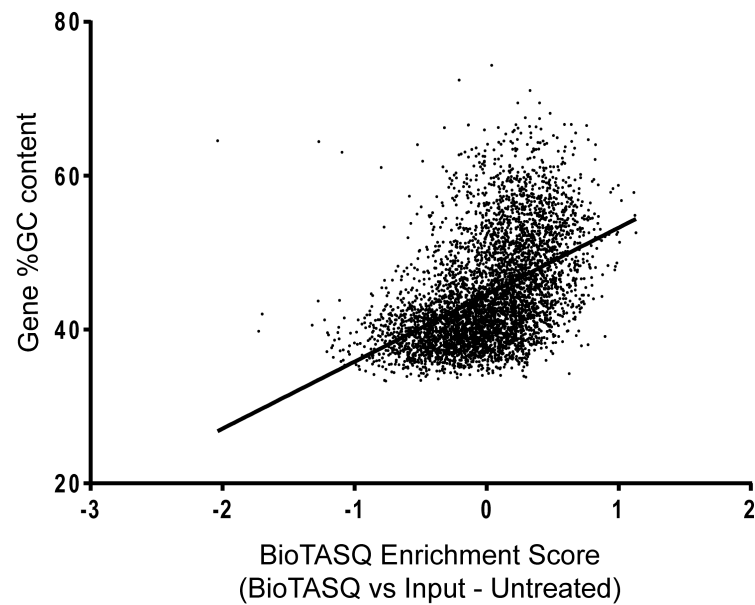
B



A



B



C

