Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome

Chun Kit Kwok,[1,4] Giovanni Marsico,[2,3,4] and Shankar Balasubramanian[2,3]

[1]Department of Chemistry, City University of Hong Kong, Hong Kong SAR, China

[2]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

[3]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, United Kingdom

Correspondence: sb10031@cam.ac.uk; ckkwok42@cityu.edu.hk

SUMMARY

RNA G-quadruplex (rG4) secondary structures are proposed to play key roles in fundamental biological processes that include the modulation of transcriptional, co-transcriptional, and posttranscriptional events. Recent methodological developments that include predictive algorithms and structure-based sequencing have enabled the detection and mapping of rG4 structures on a transcriptome-wide scale at high sensitivity and resolution. The data generated by these studies provide valuable insights into the potentially diverse roles of rG4s in biology and open up a number of mechanistic hypotheses. Herein we highlight these methodologies and discuss the associated findings in relation to rG4-related biological mechanisms.

Cold Spring Harbor Perspectives in Biology

Outline

1 Introduction

2 Computational approaches for searching RNA G-quadruplexes

3 Detecting RNA G-quadruplexes in the transcriptome in vitro

4 Detecting RNA G-quadruplexes in vivo

References

[4]These authors contributed equally to this work.

Editors: Thomas R. Cech, Joan A. Steitz, and John F. Atkins
Additional Perspectives on RNA Worlds available at www.cshperspectives.org

 doi: 10.1101/cshperspect.a032284
Cite this article as *Cold Spring Harb Perspect Biol* 2018;10:a032284

1 INTRODUCTION

In 1910 Bang reported that guanylic acid (GMP) could form a gel, suggestive of a self-assembly phenomenon (Bang 1910). Some 50 years later, Gellert and coworkers revealed that the fibers obtained from the dried GMP gel comprised what became known as the G-quartet motif (Gellert et al. 1962). A G-quartet involves four guanine bases interacting with each other via H-bonding, further stabilized by a central monovalent cation, such as K^+ or Na^+ (Fig. 1). The biological relevance of such non-Watson–Crick nucleic acid structures was largely ignored until in the late 1980s, when it was shown that G-rich sequences based on either the telomeric region of DNA or immunoglobulin switch region of DNA could form four-stranded structural motifs (Sen and Gilbert 1988), which became referred to as G-quadruplexes (G4s) (Fig. 1).

Cellular visualization of DNA G4s was reported in 2001 for the ciliate *Stylonychia lemnae*, by immunostaining the telomeric G4 (Schaffitzel et al. 2001). In 2013, a G4-specific single-chain antibody, blood group antigen H1 (BG4), was generated by phage display and used to visualize DNA G4s in the nuclei of human cells by immunofluorescence (Biffi et al. 2013). Shortly afterward, RNA G4s (rG4s) were visualized in the cytoplasm of human cells using BG4 and were also shown to be selectively stabilized by an rG4-specific ligand carboxypyridostatin (cPDS) (Biffi et al. 2014).

Sequence-based prediction approaches have enabled exploration of sequence and structural features characteristic of G4s in vitro at the level of the whole genome or transcriptome for a range of organisms. Early predictors (Huppert and Balasubramanian 2005; Todd et al. 2005) were informed by biophysical measurements conducted on G4 DNA oligonucleotides, and it has been largely assumed that rG4s conform to the same "rules." However, biophysical measurements have highlighted important differences between DNA and RNA G4s. The presence of the 2′-hydroxyl (2′-OH) group in the ribose can enable additional intramolecular interactions within the loops of rG4s and with water molecules, providing the rG4 with greater stability than the corresponding DNA G4 (Zhang et al. 2010). Furthermore, the 2′-OH causes a C3′-endo rather than a C2′-endo sugar pucker, generally favoring a parallel rG4 folding topology, whereas DNA G4s can fold into parallel, antiparallel, and mixed conformations (Zhang et al. 2010). These issues are described in detail elsewhere (Fay et al. 2017; Kwok and Merrick 2017). The higher propensity of RNA to fold into other single-strand secondary structures such as stem-loops and pseudoknots is a key feature for considering G4 formation in the context of transcripts.

Numerous studies suggest that rG4s may play significant roles in a myriad of biological processes (Bugaut and Balasubramanian 2012; Fay et al. 2017) and are linked to human diseases (Simone et al. 2015; Cammas and Millevoi 2017). rG4s in the 5′ untranslated region (UTR) of messenger RNA (mRNA) have been shown to impede translation (Kumari et al. 2007), and rG4s at 3′ UTR can suppress translation (Arora and Suess 2011; Crenshaw et al. 2015) and affect microRNA targeting (Stefanovic et al. 2015; Rouleau et al. 2017), alternative polyadenylation (Beaudoin and Perreault 2013), and RNA localization (Subramanian et al. 2011). rG4s in the coding DNA sequence (CDS) can suppress translation (Endoh et al. 2013) and stimulate ribosomal frameshifting (Endoh and Sugimoto 2013; Yu et al. 2014), whereas rG4 near splice junctions can influence alternative splicing (Marcel et al. 2011; Weldon et al. 2018), suggesting a co-transcriptional role. In addition, rG4 has also been shown to act like the canonical hairpin loop in the ρ-independent pathway and stimulate mitochondrial transcription termination (Wanrooij et al. 2010). It has also been shown that rG4 has stronger binding affinity to Polycomb repressive complex 2 (PRC2) than unstructured G-rich motif or duplex RNA (Wang et al. 2017). Recently, it has been reported that rG4s have potential to form in primary and precursor microRNAs (Mirihana Arachchilage et al. 2015; Kwok et al. 2016b; Rouleau et al. 2018), as well

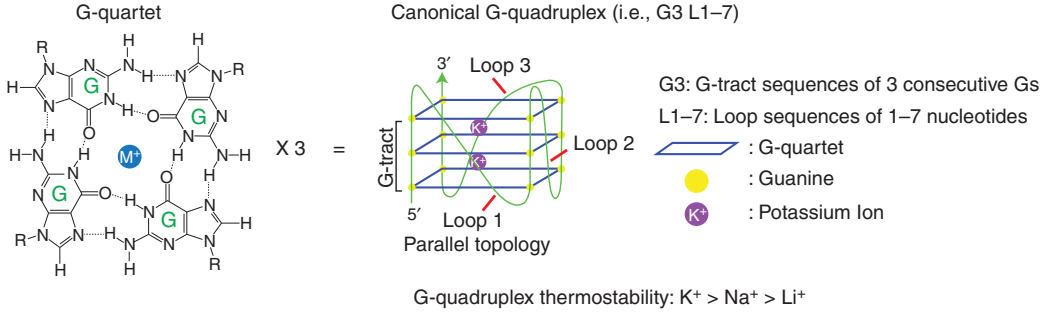

Figure 1. RNA G-quadruplex (rG4) structure. A G-quartet showing the hydrogen bonding and stabilizing cation. Three stacked G-quartet planes with connecting loop sequences form a canonical G-quadruplex (G4) (i.e., G3 L1–7).

as in long-noncoding RNA (Matsumura et al. 2017), suggesting roles of rG4 in the noncoding RNA transcriptome.

Herein, we describe recent methodological advances to detect rG4s in the transcriptome.

2 COMPUTATIONAL APPROACHES FOR SEARCHING RNA G-QUADRUPLEXES

2.1 Early Predictors

Biophysical data led to a motif for unimolecular G4s comprising stretches of guanines (G runs) separated by linker sequences (loops) of limited length. The algorithm Quadparser (Fig. 2A) (Huppert and Balasubramanian 2005) was used to search the motif ($G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) predicted to fold into stable G4s under near-physiological conditions (Hazel et al. 2004). A parallel study using the same consensus motif (Todd et al. 2005) analyzed the specific sequence and loop lengths within G4s to illuminate preferred structural and sequence features in the human genome. Whereas sequences that form G4s and escape this rule were known (Patel and Hosur 1999), a conservative choice of parameters initially provided more than 360,000 predicted G4s in the human genome. Lower stringency—for example, allowing two quartets (G_{2+} instead of G_{3+})—gave more than 8.5 million predicted G4s. A limitation of Quadparser was the fixed motif definition and the binary nature of the G4 prediction, whereas biophysics would suggest a variation in thermal stability between different G4s (Hazel et al. 2004). Other G4 predictors were described at about the same time. The quadruplex-forming G-rich sequence (QGRS) mapper (Fig. 2) (D'Antonio and Bagga

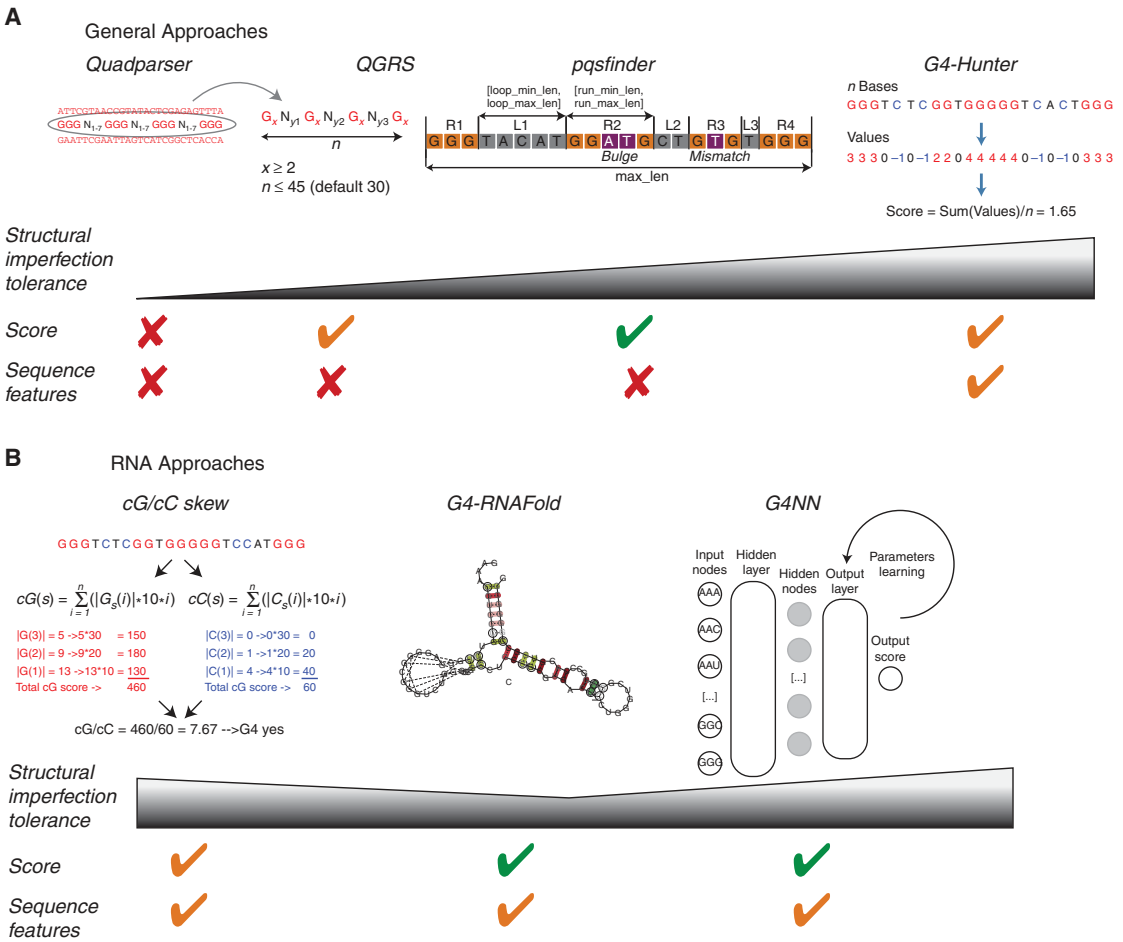

Figure 2. Computational prediction of G4s. (*A*) General appoaches: Approaches originally developed for DNA G4s. Key features are indicated below each method: Tolerance to structural "'imperfections'" (e.g., bulges, incomplete guanine run), with higher gray bar indicating more tolerance; if the method returns a score indicating the propensity to form a G4; if the method takes into account sequence features. Red cross, feature absent; orange tick, feature partially present; green tick, feature present. (*B*) RNA approaches: Approaches specifically developed for rG4s. Features and legend as for *A*.

2004) used an algorithm comprising a more flexible motif definition (G_xN_{y1} G_xN_{y2} G_xN_{y3} G_x, in which $x \geq 2$) and a ranking of predicted G4s according to the G score, defined as the likelihood to form a stable G4 based on favoring short loops, equal loop lengths, and a higher number of quartets. An alternative approach using a sliding window strategy was implemented in G-quadruplex potential (G4P) calculator (Eddy and Maizels 2006): The program computed the G4 DNA potential based on the density of G runs in a given sequence window and calculating the percentage that meets the searched criteria, returning a score independent of sequence length.

All such computational approaches have limitations. First, they did not use a large-scale experimental data set to validate the algorithm or optimize the critical parameters for scoring G4s. Rather, these decisions were based on extrapolating from a small number of biophysical studies with limited scope. Second, scoring functions are used interchangeably for DNA and RNA, ignoring critical differences between the respective G4s.

2.2 Recent Predictors

Most recent approaches allow greater motif flexibility, such as longer loops, bulges, or mismatches in the G4 motif (see Pqsfinder in Fig. 2A) (Hon et al. 2017) as supported by few experimentally validated structures (Mukundan and Phan 2013). Others penalized the presence of Cs proximal to or within the G4 motif (see G4-Hunter in Fig. 2A) (Bedrat et al. 2016), because Cs had been experimentally shown to base-pair with Gs to compete with G4 formation (Beaudoin et al. 2014). Quadron (Sahakyan et al. 2017) takes into account more than 200 sequence-based and structural features to classify via machine learning putative G4s into forming and nonforming ones. pqsfinder was trained using 392 in vitro experimentally validated sequences and validated using the larger G4-sequencing (G4-seq) data set generated by high-throughput sequencing (Chambers et al. 2015). G4-Hunter was also tested on these 392 sequences, plus experimental validation on the entire human mitochondrial genome, in which 165 potential G4s were tested in vitro. Quadron was instead trained on the entire G4-seq data set and assessed through a complex cross-validation scheme. Those recent approaches provided improved accuracy in predicting nonstandard G4 motifs for a more comprehensive identification of G4s in the genome.

2.3 RNA G-Quadruplex-Specific Predictors

The approaches described in the previous section are not RNA-specific and were not based on experimentally validated rG4-forming sequences. RNA-specific features include the 2′-OH, which confers higher stability, a preference for parallel conformation (Zhang et al. 2010), and the propensity for competing, alternative RNA secondary structures, and also long-range interactions. These key differences from DNA should be considered.

2.4 *cG/cC* Skew Approach

Cytosines proximal to guanines within the G4 motif can base-pair (C:G) and compete with the Hoogsteen base-pairing required for G-quartet formation (Beaudoin et al. 2014). An approach proposed to address this uses the *cG/cC* scoring scheme (Beaudoin et al. 2014), which penalizes the presence of Cs to account for their negative effect on G4 stability. This method calculates the ratio between two different factors, the *cG* and the *cC* score, each proportional to the number of G (or C) stretches, progressively weighted more for longer stretches, according to the formula

$$cG(s) = \sum_{i=1}^{n} (|G_s(i)| * 10 * i),$$

and similarly for the *cC* score (Fig. 2B). The experimental validation used two sets of more than 10 G4 sequences and led to an empirical threshold of 2–3 as *cG/cC* score for the formation of stable G4s. This scoring scheme overcomes the limitation of using rigid sequence motifs (i.e., two or three Gs and a loop of defined length). However, the parameterization is arbitrary: Both the scoring threshold and the multiplicative factors in the formulas are chosen based on heuristics that have not been rigorously justified. Another limitation is that only Gs and Cs are taken into account explicitly, whereas other nucleotides (A or T) or more complex sequence motifs (dinucleotides, k-mers) are not considered.

2.5 G-Quadruplex RNAfold

A thermodynamically based approach was introduced within the RNAfold tool of the Vienna package (Fig. 2B) (Lorenz et al. 2013). This approach aims to explicitly address the issue of competing alternative secondary structures by calculating the energy function due to the rG4 and estimating the overall minimal energy (i.e., higher stability) structure for a given RNA sequence. This approach essentially models the ΔG as a logarithmic function of the total length of linkers between G runs and incorporates this term into a simplified energy function. The investigators found that the majority of putative quadruplex-forming sequences in the human genome are likely to fold into non-G4 secondary structures instead. This is not in agreement with recent experimental reports showing rG4 formation in transcripts in vitro (Guo and Bartel 2016; Kwok et al.

2016a). This is perhaps attributable to the folding algorithm having limited knowledge of the energy function for longer or asymmetric loops and to the limited training set. Moreover, kinetically trapped structures may be missed by a thermodynamic approach, especially if the Cs are downstream from the Gs.

2.6 G4RNA Screener

A recent data mining approach, G4RNA screener (see G4NN in Fig. 2B) (Garant et al. 2017), is based on training an artificial neural network on a compendium of RNA sequences (149 G4 and 179 non-G4), investigated by the literature for G4 folding, plus 200 sequences randomly taken from the transcriptome. The approach was tested on nearly 4000 in vitro detected rG4s (Kwok et al. 2016a), considered as a positive set, and compared for classification performance with the *cG/cC* and the G4-Hunter algorithms, yielding comparable or better outcomes. This approach paves the way to more comprehensive approaches in which complex sequence features other than just C stretches can be factored in the prediction. However, the neural network predictor is a black box, which does not readily provide insights into the predictive features determining G4 formation.

3 DETECTING RNA G-QUADRUPLEXES IN THE TRANSCRIPTOME IN VITRO

Broadly, three categories of transcript-specific methods have been developed for the detection of rG4s (Kwok and Merrick 2017): (i) biophysics on short synthetic oligos, including circular dichroism (CD) spectroscopy (Paramasivan et al. 2007), ultraviolet (UV) or fluorescence resonance energy transfer (FRET) thermal melting analysis (Mergny et al. 1998; Mergny et al. 2001), structural analysis by nuclear magnetic resonance (NMR) spectroscopy (Webba da Silva 2007), and fluorescence assay (Kwok et al. 2013c; Renaud de la Faverie et al. 2014); (ii) chemical approaches on longer in vitro transcribed transcripts, including in-line probing (Beaudoin et al. 2013), selective 2′-OH acylation with lithium ion-based primer extension (SHALiPE) (Kwok et al. 2016b), dimethyl sulfate with lithium ion-based primer extension (DMSLiPE) (Kwok et al. 2016b), footprinting of long 7-deazaguanine-substituted RNAs (FOLDeR) (Weldon et al. 2017); and (iii) approaches on transcripts extracted from cells, including reverse transcriptase stalling (RTS) and RTS-ligation mediated polymerase chain reaction (PCR) (Kwok and Balasubramanian 2015).

These approaches collectively provide insights into the structure and folding of rG4s for candidate sequences; however, they are low-throughput and preclude a global, transcriptome-wide analysis of rG4s.

The pausing by a DNA/RNA polymerase has been exploited to detect nucleic acid secondary structures, including G4s (Woodford et al. 1994). To clarify whether the structure is a G4, one can compare polymerase pausing under conditions that differentially stabilize folded G4 structures, for example physiological K^+ conditions (that help stabilize folded G4s) versus Li^+ or Na^+ conditions of the same ionic strength (which lead to less stable G4s). This principle has been adapted for high-throughput sequencing (G4-seq) to generate a map of DNA G4s in the human genome (Chambers et al. 2015). Sequencing rG4s was enabled by detecting RTS at rG4 structures under K^+ conditions, which is not observable under Li^+ conditions. There is also the option of extending the principle to strong RTS at rG4s on the inclusion of an rG4 stabilizing small molecule (Kwok and Balasubramanian 2015). This general approach was developed into an in vitro transcriptome-wide rG4 profiling method called rG4-sequencing (rG4-seq) and applied to purified poly(A)-enriched HeLa RNAs (Kwok et al. 2016a). Here, complementary DNA (cDNA) fragments are prepared into libraries for next-generation sequencing (Fig. 3A). At a similar time an in vitro method called reverse transcriptase (RT) stop profiling was reported (Fig. 3B), which used similar concepts to obtain an in vitro rG4 map in purified poly(A)-enriched RNAs from mouse embryonic stem cells, as well as human HEK293T and HeLa cells (Guo and Bartel 2016). This map was then used as reference in the rest of their study to assess rG4 formation in vivo via chemical mapping (Guo and Bartel 2016), which will be discussed later.

Here, we discuss the experimental workflows and bioinformatics pipelines for the rG4-seq and RT stop profiling (see also Fig. 3), and then we summarize biological implications of the in vitro studies. Although the two methods are conceptually similar and share some steps, they differ in aspects of the experimental design, as discussed in the next section.

3.1 Comparison of Experimental Procedures in rG4-seq and RT Stop Profiling

A key difference is that for RT stop profiling, 60–80-nt RNA fragments were selected, whereas for rG4-seq, an average size of 250-nt RNA was obtained. Given canonical rG4s (G3L1–7) are ~25-nt-long, the length of the flanking sequences can affect the rG4 folding propensity (Beaudoin et al. 2014). As evidence suggests that RNA can have long-range intramolecular base-pairing (Sugimoto et al. 2015), as well as rG4 motifs that cover a long range (Jodoin et al. 2014), the use of a longer flanking sequence may help reveal

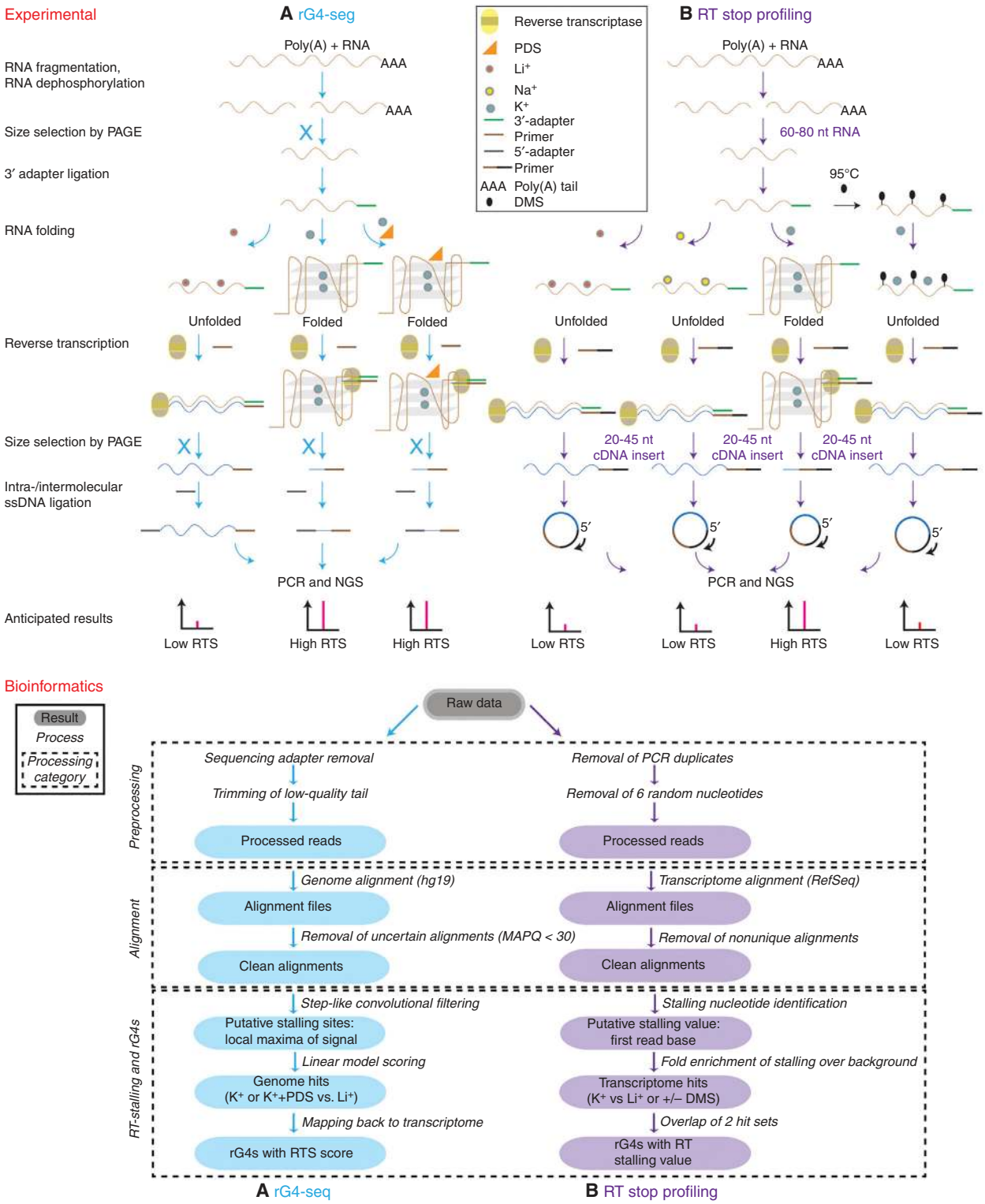

Figure 3. Reverse transcriptase stalling (RTS) sequencing approach to map rG4s. (*Top*) The experimental flowchart with relevant processes and results is illustrated for (*A*) rG4-sequencing (rG4-seq) (*left*, cyan) and for (*B*) reverse transcriptase (RT) stop profiling (*right*, purple). Key steps are shown. (*Bottom*) The step-by-step bioinformatics pipeline with relevant processes and results is illustrated for (*A*) rG4-seq (*left*, cyan) and for (*B*) RT stop profiling (*right*, purple). Starting from the raw data, the three major processing categories are shown (dashed boxes): Reads preprocessing; reads alignment; RTS score calculation, and identification of rG4s as hit regions.

Cold Spring Harbor Perspectives in Biology

such interactions and better reflect the natural sequence context.

The second difference is that for rG4-seq, an rG4 stabilizing ligand, pyridostatin (PDS), was used, which can increase the RTS (Kwok and Balasubramanian 2015). For RT stop profiling, dimethyl sulfate (DMS) was used under denaturing conditions (95°C, 0 mM K^+). Under such conditions the N7 position of Gs, which would otherwise be hydrogen-bonded in an intact G4, is methylated to m7G, which is typically not inhibitory to the RT (Wells et al. 2000). After methylation of Gs, it is assumed that rG4s cannot be refolded.

The third major difference is the single-stranded DNA (ssDNA) ligation strategy used. For RT stop profiling, CircLigase-mediated intramolecular ssDNA ligation was performed. For rG4-seq, a T4 DNA ligase-mediated intermolecular ssDNA ligation was performed. The latter approach showed less bias in nucleotide preference and is more cost efficient than the CircLigase approach (Kwok et al. 2013a; Ritchey et al. 2017).

3.2 Comparison of Bioinformatics Procedures in rG4-seq and RT Stop Profiling

In rG4-seq, reads are aligned to the genome and significant stalling sites are later assigned to the most abundant isoform, an assignment that can introduce uncertainty. In RT stop profiling, transcriptome alignment is performed and ambiguous reads mapping to multiple isoforms are removed, potentially introducing coverage biases. For these reasons, both methods could suffer from false positives/negatives.

rG4-seq (Fig. 3A) computes RT stops with a two-step procedure: Coverage signal is processed by convolutional filters to identify candidate stalling positions, which are then statistically assessed using a linear model that contrasts the positive condition (i.e., K^+ or K^+ + PDS) against the Li^+ negative control. This approach is able to reduce false positives and identify bona fide G4 motifs but has the drawback that a small number of structures displaying high stability also under Li^+ conditions (estimated 72 putative rG4s) could be false negatives.

RT stop profiling (Fig. 3B) examines the nucleotide immediately adjacent to the stalling by calculating a fold enrichment value between the number of reads stalled and the background read density. RT-stalling sites specific to K^+ (the rG4-stabilizing condition) were identified by comparing to conditions using less rG4-stabilizing cations (Na^+ and Li^+). Independently, extracted RNA was treated with DMS under in vitro denaturing conditions (i.e., 95°C, 0 mM K^+), followed by RT with K^+, providing another diagnostic feature for rG4 presence, because Gs will not easily methylate within a rG4. The combination of these two approaches identified bona fide rG4 structures in vitro; however, the lack of replicates and the choice of an arbitrary threshold could potentially hamper confident detection.

3.3 Biological Validation and Findings

To support the approaches in the two methods, both laboratories have reported enrichment in Gs in the sequence immediately upstream of the RTS site, which would be expected for rG4-based stalling. In addition, for rG4-seq, in vitro SHALiPE was performed to verify several individual candidates, and for RT stop profiling, CD spectroscopic experiments were used to confirm selected rG4 candidates. Although the two methods use different experimental and bioinformatics steps, both appear to be robust.

Together, both methods identified thousands of rG4 structures in the human transcriptome, in vitro, for the first time. The rG4s in human mRNAs were enriched in UTRs, suggestive of role(s) in translation consistent with a number of publications on specific transcripts (Kumari et al. 2007; Arora and Suess 2011; Crenshaw et al. 2015). Also, rG4s were enriched near microRNA (miRNA) target sites and polyadenylation sites, suggesting potential roles in miRNA-mediated regulation and alternative polyadenylation, which is in agreement with recent findings on individual transcripts (Beaudoin and Perreault 2013; Stefanovic et al. 2015; Rouleau et al. 2017). Notably, rG4-seq revealed a cluster of G4 sequences that are conserved among eukaryotes and are overrepresented in genes that are annotated with RNA processing and RNA stability, which warrant further investigation. For RT stop profiling, 7852 nonoverlapping rG4 regions were detected in at least one of the two human cell lines (HEK293T and HeLa), and 4935 were identified in both. Factors such as sequencing coverage of transcripts, differences in natural RNA abundance, sequence difference in HEK293T and HeLa transcriptomes, in addition to technical variability could explain differences in rG4 detected across cell lines. The overlap between the HeLa sites and those identified by rG4-seq in the same cell line was ~45% when compared with K^+ in rG4-seq and >65% when compared with the larger rG4 data set generated on further stabilization in the K^+ + PDS conditions. This constitutes a reasonable overlap of rG4s measured in vitro considering the studies were performed in different laboratories using different experimental approaches and computational analyses.

The in vitro experimental rG4 maps provide a resource to further investigate rG4 structure and function in vivo. The main limitations in these data sets arise from rG4s that cannot yet be detected, such as those that are insufficiently stable or extremely stable, those that form from long-

range rG4 interactions, and those in low abundance RNAs. It will also be interesting to go beyond the transcriptomes of mouse and human and map rG4s in other organisms to expand the repertoire of rG4s identified in vitro, which may enhance our ability to understand and predict the relationship between sequence and rG4 folding.

4 DETECTING RNA G-QUADRUPLEXES IN VIVO

The folding propensity of rG4s in cells has been recently explored by two methods (Guo and Bartel 2016). The first uses methylation by DMS, followed by RT stop profiling under K^+ conditions, and then next-generation sequencing. The second uses selective 2′-OH acylation analyzed by primer extension (SHAPE) using 2-methylnicotinic acid imidazolide (NAI), followed by RT stop profiling under Na^+ conditions, and then next-generation sequencing. Here, we discuss the experimental details of both methods (see also Fig. 4A,B), the bioinformatics pipelines applied to the data, and finally the biological findings.

4.1 Experimental Pipeline and Assessment of DMS-seq + RT Stop Profiling

To probe rG4 formation, mouse embryonic stem cell (mESC) cells were treated with ∼800 mM (8%) DMS for 5 min at 37°C to methylate the N7 position of Gs (see Table 1 for comparison with related studies). For any rG4s folded in cells, Gs involved in rG4 G-quartet formation are protected from methylation, thus when RT profiling is performed on the extracted RNA under K^+ conditions bona fide rG4 structures will form, causing RTS. Conversely, if the rG4 motif is unfolded in cells, the associated Gs are methylated, and subsequently the extracted RNA does not form rG4 or cause RTS (Guo and Bartel 2016).

DMS can also methylate N1 of A and N3 of C causing rG4-independent RTS at those sites (Wells et al. 2000) and ∼70% of global RT stops were at A or C for in vitro and in vivo DMS-treated samples. An individual rG4 exemplar that was ectopically expressed (G3A2) also showed similar A and C modification induced by DMS between in vitro and in vivo conditions along the gel. Given the RT stop signal at an rG4 site in vitro, at 0 mM K^+, was diminished (although not quite baseline) compared with in vitro, at 150 mM K^+, the in vitro DMS assay was deemed to be within the experimental dynamic range (i.e., not overmethylated). As RT stop at A and C was comparable in vitro and in vivo the in vivo DMS probing was also considered to be within the dynamic range of the assay (i.e., not overmethylated by DMS). It should be noted that the RT stop at G, primarily caused by rG4s stalling, constitutes 20% of all RT stops for in vitro DMS reaction (at both 0 mM K^+ and 150 mM K^+), whereas in vivo the RT stop at G is <10% of total, suggesting that the Gs are methylated to a greater extent in vivo, leading to less rG4-mediated RT stalling. As DMS methylation of N7 of G has much faster kinetics than methylation of N1 of A or N3 of C (Lawley and Brookes 1963). RNA structure probing should ideally be performed under single-hit kinetics conditions (see Table 1). Thus, the dynamic range of the DMS assay and the condition of single-hit kinetics should ideally be carefully balanced for each cell line and species under study. On considering reaction conditions used for in vivo RNA structurome experiments, the DMS concentration used by Guo and Bartel was relatively high leaving open the possibility of some overmethylation in that study (Table 1).

4.2 Experimental Pipeline and Assessment NAI-seq + RT Stop Profiling

Another in-cell structure-probing method involves the use of SHAPE reagents that acylate the RNA 2′-OH group with reaction kinetics influenced by its local flexibility (Wilkinson et al. 2006). Typically, unpaired and unconstrained nucleotides are kinetically more susceptible to 2′ acylation (Weeks 2010). As the SHAPE reagent reacts with all four nucleotides, it provides structural information on RNA at single-nucleotide resolution. One of the SHAPE reagents, NAI, was developed for in vivo probing of RNA structure (Kwok et al. 2013b; Spitale et al. 2013). Recently, we developed SHALiPE that used NAI, followed by lithium ion–mediated reverse transcription to probe rG4 in vitro (Kwok et al. 2016b). Interestingly, unlike Watson–Crick base pairs that typically lower the SHAPE reactivity, the in vitro formation of rG4 increased the SHAPE reactivity at the 3′ G position of the first three G-tracts (Kwok et al. 2016b).

In the study of Guo and Bartel (Guo and Bartel 2016), mESC cells were treated with 80 mM NAI for 15 min at 37°C (single-hit conditions, see Table 1), followed by RT stop profiling under Na^+ conditions, and then compared with in vitro NAI conditions (at both 0 mM K^+ and 150 mM K^+). Given Na^+ can also stabilize rG4, reverse transcription under Li^+ conditions would have been more optimal. Based on observations under rG4-forming conditions in vitro (150 mM K^+; Guo and Bartel 2016; Kwok et al. 2016b), if rG4 folding in cells is similar to in vitro, under the NAI reaction conditions (150 mM K^+), it would be expected that the 3′ G of each G-tract in rG4 and the loop residues of rG4s would be susceptible to modification, whereas the other Gs involved in rG4 formation would be modified to a lower degree. Guo and Bartel provided an explanation for the ectopically expressed rG4 examples by analysis of the crystal structure of the rG4 formed by telomeric repeat-containing

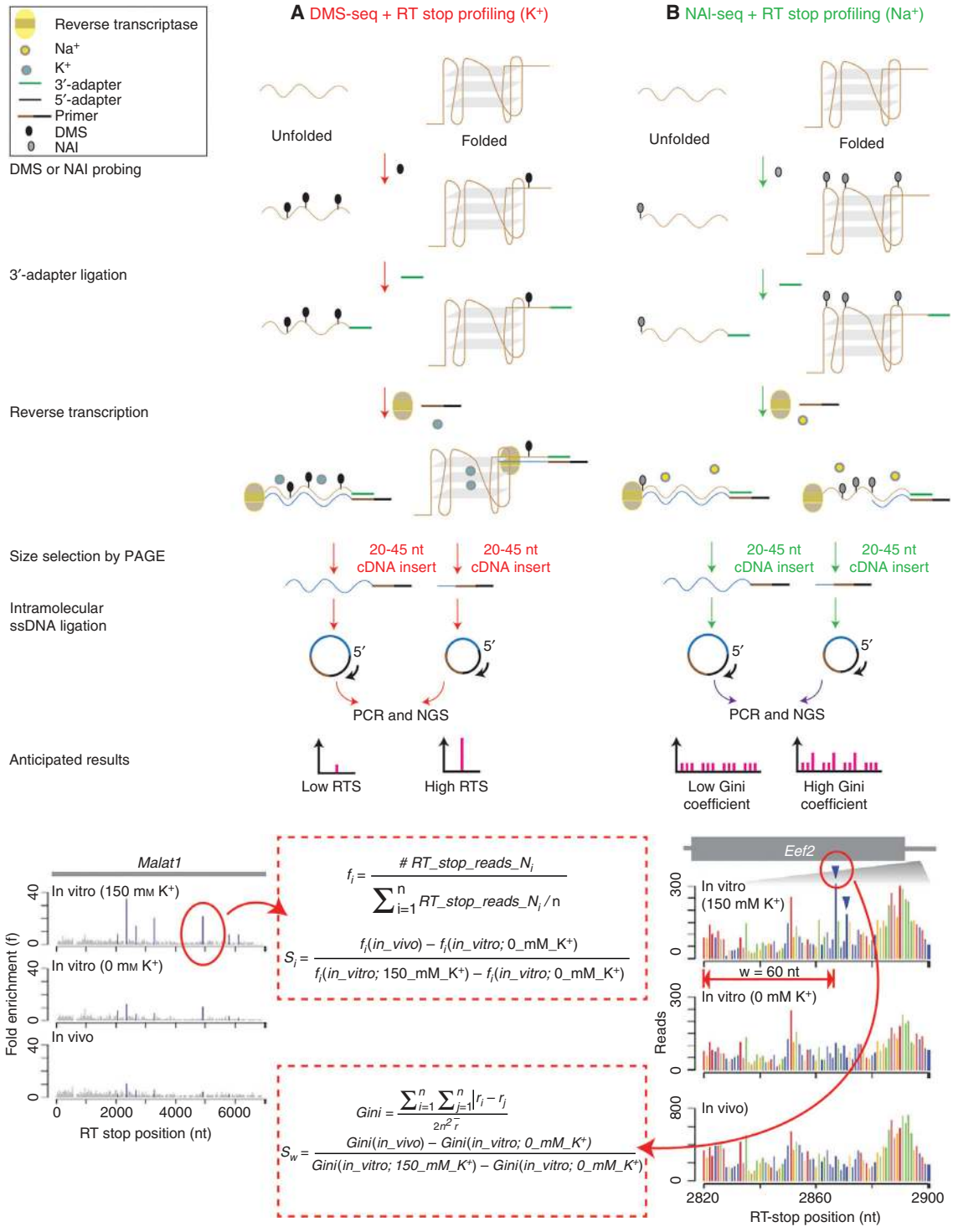

Figure 4. Chemical probing sequencing approach to map rG4s. (*Top*) Experimental flowchart of processes and results for (*A*) dimethyl sulfate sequencing (DMS-seq) + RT stop profiling (K^+) (*left*, red) and for (*B*) 2-methylnicotinic acid imidazolide sequencing (NAI-seq) + RT stop profiling (Na^+) (*right*, green). (*Bottom left*) Graphical representation of the scoring procedure for DMS-seq followed by RT stop profiling, with the formulas to calculate the fold enrichment score (f) at each nucleotide (i) and the in vivo folding scores (S), as indicated by the red arrow and red dashed box. (*Bottom right*) Graphical representation of the scoring procedure for NAI-seq followed by RT stop profiling, with the formulas to calculate the Gini index for a window (w) of size 60 upstream of stalling sites over putative rG4s previously detected by RT stop in vitro, as indicated by the red arrow and red dashed box below the DMS-seq one.

Table 1. Chemical probes and reaction conditions used in in vivo transcriptome-wide RNA structure probing studies

In vivo probe	Reaction conditions (Conc., time, temp.)	System	Single-hit kinetics?[a]	Reference
DMS	75 mM, 15 min, 22°C	*Arabidopsis thaliana*	Yes	Ding et al. 2014
DMS	200–400 mM, 2–4 min, 30°C; 260 mM, 40 min, 10°C	*Saccharomyces cerevisiae*	No	Rouskin et al. 2014
DMS	200–260 mM, 4 min, 37°C	Human K562, fibroblast	No	
DMS	100 mM, 2 min, 30°C	*S. cerevisiae*	Yes	Talkish et al. 2014
NAI-N3	100 mM, 15 min, 37°C	mESC	Yes	Spitale et al. 2015
DMS	500 mM, 4 min, 30°C	*S. cerevisiae*	No	Zubradt et al. 2017
DMS	200 mM, 4–5 min, 37°C	HEK293T	No	
DMS	5 M, 5 min, 26°C	*Drosophila melanogaster*	No	
DMS	800 mM, 5 min, 37°C	mESC, HEK293T, *Escherichia coli*	No	Guo and Bartel 2016
DMS	800 mM, 5 min, 30°C	*S. cerevisiae*	No	
NAI	80 mM, 15 min, 37°C	mESC	Yes	
NAI	80 mM, 15 min, 30°C	*S. cerevisiae*	Yes	
DMS	75 mM, 10 min, 22°C	*Oryza sativa*	Yes	Ritchey et al. 2017
DMS	340 mM, 5 min, 37°C	HEK293T	No	Wu and Bartel 2017
NAI	100 mM, 10 min, 22°C	*S. cerevisiae*	Yes	Selega et al. 2017
DMS	25 mM, 15 min, 28°C	*O. sativa*	Yes	Deng et al. 2018

DMS, dimethyl sulfate; NAI, 2-methylnicotinic acid imidazolide; Conc., concentration; temp., temperature; mESC, mouse embryonic stem cell.

[a]One modification every ~300 nt RNA region (McGinnis et al. 2009) or 75%–90% unmodified RNA (Wan et al. 2013; Ding et al. 2015). With chemical-based RNA structure probing experiments, single-hit kinetics is preferred as the modification of the first nucleotide/site that can induce conformational changes that can cause modification of non-native nucleotides/sites and, therefore, lead to experimental artifacts and inaccurate data interpretation. 100% DMS is ~10 M.

RNA (TERRA) (Collie et al. 2010), and found that the 2′-OH group of the 3′ G of the G-tracts was exposed (Guo and Bartel 2016). Conversely, if the rG4 is fully unfolded in cells, the Gs associated with the sequence motif will be modified more evenly by NAI.

4.3 Bioinformatic Pipeline of DMS-seq + RT Stop Profiling and NAI-seq + RT Stop Profiling

The computational analysis of the DMS-seq profiling is very similar to RT stop profiling (Fig. 4A). For the NAI-seq analysis, Gini coefficients were calculated for rG4-containing regions (Fig. 4B). Essentially, the Gini index measures inequality within a distribution: A value of 0 expresses perfect equality (e.g., nucleotides have the same stalling frequency), whereas a value of 1 expresses maximal inequality (i.e., nucleotide-specific stalling). However, outliers due to experimental noise could affect Gini estimates, and transcripts with vastly different reactivity profiles might have the same coefficient, making it difficult to use it as a comparative feature (Choudhary et al. 2017).

General-purpose methods and metrics for analyzing RT stop profiling data exist (Aviran and Pachter 2014; Choudhary et al. 2017; Li et al. 2017). These approaches use libraries from unmodified transcripts to distinguish RT drop-off noise from chemical incorporation and estimate of reactivity at single-nucleotide level. It may be worth revisiting rG4 mapping by chemical mapping using these methods and analyses.

4.4 Biological Validation and Findings

The most noteworthy outcome from the DMS-seq + RT stop profiling method in mouse, human, and yeast was that the RT stop profile for most rG4 motif regions was similar to what was observed at 0 mM K^+ in vitro (assumed, rG4-unfolded state), as opposed to what was observed at 150 mM K^+ in vitro (assumed, rG4-folded state), suggesting either that most rG4 regions are unfolded in these eukaryotes (Guo and Bartel 2016), or that the method samples the unfolded state during the time course of the reaction. In experiments intended to perturb rG4 folding, the rG4-stabilizing ligand PDS caused a small, but detectable increase in global rG4 folding, and the deletion of a known rG4 helicase, DEAH box protein 36 (DHX36), caused little detectable change in global rG4 folding. The absence of a "positive control" in the context of these experiments, despite several rational attempts to perturb the system or introduce rG4s exogenously, does leave open the possibility that rG4 structures in the transcript population may exist but were undetectable under the experimental conditions used, DMS and SHAPE structural probing assays (Guo and Bartel 2016) measure whole-cell, ensemble RNA structural conformations within the reaction time frame. Information on the structural conformation of individual RNAs and the dynamic structural interconversion, subpopulation, and heterogeneity within and across cells may be lost.

Previous computational analyses predicted a role of DNA G4 motifs in *Escherichia coli* gene regulation (Rawal

et al. 2006) and a conserved DNA G4-hairpin-duplex-switch with a potential regulatory role, in the same species (Kaplan et al. 2016). In the study by Guo and Bartel (Guo and Bartel 2016) a bioinformatics search over several bacteria transcriptomes such as *E. coli*, *Pseudomonas putida,* and *Synechococcus* indicated that rG4 regions are generally depleted in bacteria, although this may reflect that bacteria do not comprise much noncoding RNA (UTRs, introns, lncRNA), whereas rG4s are typically enriched in mammals. As endogenous rG4 is rare, G4-forming RNAs comprising G3A2 (GGGAAGGGAAGGGAAGGG) or G3U (UUUGGGU GGGUGGGUGGG) were appended to the 3′ UTR of mCherry CDS and the PCR product was inserted to the expression vector plasmid (pCR2.1 backbone), and ectopically introduced to *E. coli*. The investigators showed that, in contrast to eukaryotes, ectopically introduced rG4s can be folded in *E. coli*.

4.5 Concluding Remarks

These recent approaches to detect and map rG4 structures have provided maps of rG4s that can form from in vitro refolded cellular transcripts. The global picture emerging from the initial in-cell mapping experiments suggest that rG4s are generally unfolded within a cellular context in mammalian cells, with a G4-stabilizing ligand and also knockdown of the G4-specific helicase not causing a detectable shift in global rG4 formation. This initial picture actually closely mirrors what has been observed for DNA G4s in the human genome for which of the totality of G4 structures mapped in vitro by sequencing (Chambers et al. 2015) only a very small minority (∼1%) were detected in chromatin by chromatin immunoprecipitation (ChIP) using a G4-antibody (Hansel-Hertsch et al. 2016). Careful control of the global formation of rG4 structures, by protein binding or catalytic function, would be consistent with the various regulatory functions proposed for rG4s, that would logically require dynamic control between folded and unfolded states. There exists the possibility of heterogeneity between cells and also within transcript copies in a given cell, with regard to rG4 folding. While chemical mapping provides a useful means to assess the global picture, an approach with improved temporal resolution may be required to observe dynamic structures, and enrichment approaches might be necessary to select/detect structures that exist as subpopulations. Recent disclosures (Murat et al. 2017; Sauer et al. 2017) have reported measureable cellular effects of knocking down expression of rG4-targeting helicase DHX36 on translation and usage of alternative upstream open reading frames. A greater acknowledgment of rG4s as dynamic structures coupled with an improved understanding of functionally important protein–rG4 interactions is now needed for advancement of this area.

ACKNOWLEDGMENTS

We thank Professor Phil Bevilacqua and Professor Sharon Aviran for critically reading this manuscript and for their helpful comments. The Balasubramanian laboratory is supported by European Research Council Advanced Grant No. 339778, a Wellcome Trust Senior Investigator Award, and core funding from Cancer Research UK. The Kwok laboratory is supported by City University of Hong Kong Project No. 9610363, 7200520, Croucher Foundation Project No. 9500030, and Hong Kong RGC Project No. CityU 21302317, N_CityU110/17.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

Arora A, Suess B. 2011. An RNA G-quadruplex in the 3′ UTR of the proto-oncogene PIM1 represses translation. *RNA Biol* **8:** 802–805.
Aviran S, Pachter L. 2014. Rational experiment design for sequencing-based RNA structure mapping. *RNA* **20:** 1864–1877.
Bang I. 1910. Untersuchungen über die Guanylsäure. *Biochem Zeit* **26:** 293–311.
Beaudoin JD, Perreault JP. 2013. Exploring mRNA 3′-UTR G-quadruplexes: Evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res* **41:** 5898–5911.
Beaudoin JD, Jodoin R, Perreault JP. 2013. In-line probing of RNA G-quadruplexes. *Methods* **64:** 79–87.
Beaudoin JD, Jodoin R, Perreault JP. 2014. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res* **42:** 1209–1223.
Bedrat A, Lacroix L, Mergny JL. 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* **44:** 1746–1759.
Biffi G, Tannahill D, McCafferty J, Balasubramanian S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5:** 182–186.
Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. 2014. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem* **6:** 75–80.
Bugaut A, Balasubramanian S. 2012. 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Res* **40:** 4727–4741.
Cammas A, Millevoi S. 2017. RNA G-quadruplexes: Emerging mechanisms in disease. *Nucleic Acids Res* **45:** 1584–1595.
Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. 2015. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* **33:** 877–881.
Choudhary K, Deng F, Aviran S. 2017. Comparative and integrative analysis of RNA structural profiling data: Current practices and emerging questions. *Quant Biol* **5:** 3–24.
Collie GW, Haider SM, Neidle S, Parkinson GN. 2010. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res* **38:** 5569–5580.
Crenshaw E, Leung BP, Kwok CK, Sharoni M, Olson K, Sebastian NP, Ansaloni S, Schweitzer-Stenner R, Akins MR, Bevilacqua PC, et al. 2015. Amyloid precursor protein translation is regulated by a 3′UTR guanine quadruplex. *PLoS One* **10:** e0143160.

D'Antonio L, Bagga P. 2004. Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences. In *Proceedings of the IEEE Computational Systems Bioinformatics Conference*, pp. 590–591. IEEE, Stanford, CA.

Deng H, Cheema J, Zhang H, Woolfenden H, Norris M, Liu Z, Liu Q, Yang X, Yang M, Deng X, et al. 2018. Rice in vivo RNA structurome reveals RNA secondary structure conservation and divergence in plants. *Mol Plant* **11:** 602–622.

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505:** 696–700.

Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. 2015. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat Protoc* **10:** 1050–1066.

Eddy J, Maizels N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* **34:** 3887–3896.

Endoh T, Sugimoto N. 2013. Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame. *Anal Chem* **85:** 11435–11439.

Endoh T, Kawasaki Y, Sugimoto N. 2013. Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew Chem Int Ed* **52:** 5522–5526.

Fay MM, Lyons SM, Ivanov P. 2017. RNA G-quadruplexes in biology: Principles and molecular mechanisms. *J Mol Biol* **429:** 2127–2147.

Garant JM, Perreault JP, Scott MS. 2017. Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* **33:** 3532–3537.

Gellert M, Lipsett MN, Davies DR. 1962. Helix formation by guanylic acid. *Proc Natl Acad Sci* **48:** 2013–2018.

Guo JU, Bartel DP. 2016. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353:** aaf5371.

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

Hazel P, Huppert J, Balasubramanian S, Neidle S. 2004. Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc* **126:** 16405–16415.

Hon J, Martinek T, Zendulka J, Lexa M. 2017. Pqsfinder: An exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* **33:** 3373–3379.

Huppert JL, Balasubramanian S. 2005. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33:** 2908–2916.

Jodoin R, Bauer L, Garant JM, Mahdi Laaref A, Phaneuf F, Perreault JP. 2014. The folding of 5′-UTR human G-quadruplexes possessing a long central loop. *RNA* **20:** 1129–1141.

Kaplan OI, Berber B, Hekim N, Doluca O. 2016. G-quadruplex prediction in *E. coli* genome reveals a conserved putative G-quadruplex-Hairpin-Duplex switch. *Nucleic Acids Res* **44:** 9083–9095.

Kumari S, Bugaut A, Huppert JL, Balasubramanian S. 2007. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* **3:** 218–221.

Kwok CK, Balasubramanian S. 2015. Targeted detection of G-quadruplexes in cellular RNAs. *Angew Chem Int Ed* **54:** 6751–6754.

Kwok CK, Merrick CJ. 2017. G-quadruplexes: Prediction, characterization, and biological application. *Trends Biotechnol* **35:** 997–1013.

Kwok CK, Ding Y, Sherlock ME, Assmann SM, Bevilacqua PC. 2013a. A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. *Anal Biochem* **435:** 181–186.

Kwok CK, Ding Y, Tang Y, Assmann SM, Bevilacqua PC. 2013b. Determination of in vivo RNA structure in low-abundance transcripts. *Nat Commun* **4:** 2971.

Kwok CK, Sherlock ME, Bevilacqua PC. 2013c. Effect of loop sequence and loop length on the intrinsic fluorescence of G-quadruplexes. *Biochemistry* **52:** 3019–3021.

Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. 2016a. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* **13:** 841–844.

Kwok CK, Sahakyan AB, Balasubramanian S. 2016b. Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor microRNA. *Angew Chem Int Ed* **55:** 8958–8961.

Lawley P, Brookes P. 1963. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem J* **89:** 127–138.

Li B, Tambe A, Aviran S, Pachter L. 2017. PROBer provides a general toolkit for analyzing sequencing-based toeprinting assays. *Cell Syst* **4:** 568–574.e7.

Lorenz R, Bernhart SH, Qin J, Honer zu Siederdissen C, Tanzer A, Amman F, Hofacker IL, Stadler PF. 2013. 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. *IEEE/ACM Trans Comput Biol Bioinform* **10:** 832–844.

Marcel V, Tran PL, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou MP, Hall J, Mergny JL, Hainaut P, Van Dyck E. 2011. G-quadruplex structures in TP53 intron 3: Role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32:** 271–278.

Matsumura K, Kawasaki Y, Miyamoto M, Kamoshida Y, Nakamura J, Negishi L, Suda S, Akiyama T. 2017. The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene* **36:** 1191–1199.

McGinnis JL, Duncan CDS, Weeks KM, Daniel H. 2009. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. in *Methods in Enzymology*, pp. 67–89. Academic, New York.

Mergny JL, Phan AT, Lacroix L. 1998. Following G-quartet formation by UV-spectroscopy. *FEBS Lett* **435:** 74–78.

Mergny JL, Lacroix L, Teulade-Fichou MP, Hounsou C, Guittat L, Hoarau M, Arimondo PB, Vigneron JP, Lehn JM, Riou JF, et al. 2001. Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc Natl Acad Sci* **98:** 3062–3067.

Mirihana Arachchilage G, Dassanayake AC, Basu S. 2015. A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem Biol* **22:** 262–272.

Mukundan VT, Phan AT. 2013. Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences. *J Am Chem Soc* **135:** 5017–5028.

Murat P, Marsico G, Herdy B, Ghanbarian A, Portella G, Balasubramanian S. 2017. RNA G-quadruplexes mark repressive upstream open reading frames in human mRNAs. *bioRxiv* doi: 10.1101/223073.

Paramasivan S, Rujan I, Bolton PH. 2007. Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods* **43:** 324–331.

Patel PK, Hosur RV. 1999. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res* **27:** 2457–2464.

Rawal P, Kummarasetti VBR, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S. 2006. Genome-wide prediction of G4 DNA as regulatory motifs: Role in *Escherichia coli* global regulation. *Genome Research* **16:** 644–655.

Renaud de la Faverie A, Guedin A, Bedrat A, Yatsunyk LA, Mergny JL. 2014. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res* **42:** e65.

Ritchey LE, Su Z, Tang Y, Tack DC, Assmann SM, Bevilacqua PC. 2017. Structure-seq2: Sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res* **45:** e135.

Rouleau S, Glouzon JS, Brumwell A, Bisaillon M, Perreault JP. 2017. 3′ UTR G-quadruplexes regulate miRNA binding. *RNA* **23:** 1172–1179.

Rouleau SG, Garant JM, Bolduc F, Bisaillon M, Perreault JP. 2018. G-quadruplexes influence pri-microRNA processing. *RNA Biol* **15:** 198–206.

Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505:** 701–705.

Sahakyan AB, Chambers VS, Marsico G, Santner T, Di Antonio M, Balasubramanian S. 2017. Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci Rep* **7:** 14535.

Sauer M, Juranek SA, Kazemier HG, Benhalevy D, Wang X, Hafner M, Paeschke K. 2017. DHX36 binding at G-rich sites in mRNA untranslated regions promotes translation. *bioRxiv* doi: 10.1101/215541.

Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc Natl Acad Sci* **98:** 8572–8577.

Selega A, Sirocchi C, Iosub I, Granneman S, Sanguinetti G. 2017. Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. *Nat Methods* **14:** 83–89.

Sen D, Gilbert W. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334:** 364–366.

Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. 2015. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett* **589:** 1653–1668.

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. *Nat Chem Biol* **9:** 18–20.

Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519:** 486–490.

Stefanovic S, Bassell GJ, Mihailescu MR. 2015. G quadruplex RNA structures in PSD-95 mRNA: Potential regulators of miR-125a seed binding site accessibility. *RNA* **21:** 48–60.

Subramanian M, Rage F, Tabet R, Flatter E, Mandel JL, Moine H. 2011. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep* **12:** 697–704.

Sugimoto Y, Vigilante A, Darbo E, Zirra A, Militti C, D'Ambrogio A, Luscombe NM, Ule J. 2015. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **519:** 491–494.

Talkish J, May G, Lin Y, Woolford JL Jr., McManus CJ. 2014. Mod-seq: High-throughput sequencing for chemical probing of RNA structure. *RNA* **20:** 713–720.

Todd AK, Johnston M, Neidle S. 2005. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **33:** 2901–2907.

Wan Y, Qu K, Ouyang Z, Chang HY. 2013. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat Protoc* **8:** 849–869.

Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C. 2017. Targeting of Polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol Cell* **65:** 1056–1067.

Wanrooij PH, Uhler JP, Simonsson T, Falkenberg M, Gustafsson CM. 2010. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc Natl Acad Sci* **107:** 16072–16077.

Webba da Silva M. 2007. NMR methods for studying quadruplex nucleic acids. *Methods* **43:** 264–277.

Weeks KM. 2010. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20:** 295–304.

Weldon C, Behm-Ansmant I, Hurley LH, Burley GA, Branlant C, Eperon IC, Dominguez C. 2017. Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat Chem Biol* **13:** 18–20.

Weldon C, Dacanay JG, Gokhale V, Boddupally PVL, Behm-Ansmant I, Burley GA, Branlant C, Hurley LH, Dominguez C, Eperon IC. 2018. Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res* **46:** 886–896.

Wells SE, Hughes JMX, Igel AH, Ares M. 2000. Use of dimethyl sulfate to probe RNA structure in vivo. *Method Enzymol* **318:** 479–493.

Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1:** 1610–1616.

Woodford KJ, Howell RM, Usdin K. 1994. A novel K^+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J Biol Chem* **269:** 27029–27035.

Wu X, Bartel DP. 2017. Widespread influence of 3′-end structures on mammalian mRNA processing and stability. *Cell* **169:** 905–917.

Yu CH, Teulade-Fichou MP, Olsthoorn RC. 2014. Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res* **42:** 1887–1892.

Zhang D-H, Fujimoto T, Saxena S, Yu H-Q, Miyoshi D, Sugimoto N. 2010. Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry* **49:** 4554–4563.

Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14:** 75–82.

Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome

Chun Kit Kwok, Giovanni Marsico and Shankar Balasubramanian

Cold Spring Harb Perspect Biol 2018; doi: 10.1101/cshperspect.a032284

Subject Collection RNA Worlds

Alternate RNA Structures
Marie Teng-Pei Wu and Victoria D'Souza

Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression
Patrick McDonel and Mitchell Guttman

Principles and Practices of Hybridization Capture Experiments to Study Long Noncoding RNAs That Act on Chromatin
Matthew D. Simon and Martin Machyna

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers
Sarah K. Denny and William J. Greenleaf

Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation
Melanie Weisser and Nenad Ban

Nascent RNA and the Coordination of Splicing with Transcription
Karla M. Neugebauer

Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes
Alexander Leitner, Georg Dorn and Frédéric H.-T. Allain

Structural Biology of Telomerase
Yaqiang Wang, Lukas Susac and Juli Feigon

Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes
Berthold Kastner, Cindy L. Will, Holger Stark, et al.

What Are 3′ UTRs Doing?
Christine Mayr

Single-Molecule Analysis of Reverse Transcriptase Enzymes
Linnea I. Jansson and Michael D. Stone

CRISPR Tools for Systematic Studies of RNA Regulation
Jesse Engreitz, Omar Abudayyeh, Jonathan Gootenberg, et al.

Relating Structure and Dynamics in RNA Biology
Kevin P. Larsen, Junhong Choi, Arjun Prabhakar, et al.

Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics
Alexander I. Taylor, Gillian Houlihan and Philipp Holliger

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/

Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA
Bastian Linder and Samie R. Jaffrey

Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast
Clemens Plaschka, Andrew J. Newman and Kiyoshi Nagai