

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

Sarah K. Denny¹ and William J. Greenleaf^{1,2,3}

¹Stanford University Department of Genetics, Stanford, California 94305

²Stanford University Department of Applied Physics, Stanford, California 94025

³Chan Zuckerberg Biohub, San Francisco, California 94158

Correspondence: wjg@stanford.edu

SUMMARY

High-throughput sequencing methods have revolutionized our ability to catalog the diversity of RNAs and RNA–protein interactions that can exist in our cells. However, the relationship between RNA sequence, structure, and function is enormously complex, demonstrating the need for methods that can provide quantitative thermodynamic and kinetic measurements of macromolecular interaction with RNA, at a scale commensurate with the sequence diversity of RNA. Here, we discuss a class of methods that extend the core functionality of DNA sequencers to enable high-throughput measurements of RNA folding and RNA–protein interactions. Topics discussed include a description of the method and multiple applications to RNA-binding proteins, riboswitch design and engineering, and RNA tertiary structure energetics.

Outline

- | | |
|----------------|--------------|
| 1 Background | 4 Conclusion |
| 2 Techniques | References |
| 3 Applications | |

1 BACKGROUND

Our ability to catalog the immense diversity of RNAs that exist in the cell has accelerated at a spectacular rate with the advent of high-throughput sequencing methods, allowing the mapping of a universe of previously uncharacterized transcripts. Likewise, sequencing methods have enabled the mapping of the diverse proteins that interact with RNA in vivo, providing a window into a network of protein–RNA interactions critical to gene expression regulation (Butter et al. 2009; Castello et al. 2012). However, our understanding of the detailed relationship between the sequence of RNA and its structure and function is limited, both in terms of how complex RNAs fold into 3D structures and also how RNA sequence and structures determine protein–RNA interactions. A predictive, quantitative understanding for these functional determinants (both 3D structure and protein interaction) will be needed before we can truly understand, and engineer, RNA functions.

However, the sequence–structure–function relationship in RNA is complex—perhaps not as fundamentally combinatorial and interactive as protein, but much more challenging than double-stranded, or even single-stranded, DNA. RNA makes diverse intramolecular interactions—Watson-Crick base pairs, noncanonical base pairs, tertiary interactions, etc.—that define the 3D structures RNA takes on (Turner et al. 1988; Tinoco and Bustamante 1999), which in turn sets the landscape for its function. For example, critical RNAs in translation, splicing, and chromatin organization and compaction must form precise tertiary structures to achieve their biological function (Staley and Guthrie 1998; Kieft et al. 2001; Noller 2005). In higher eukaryotes, these biological functions are rarely achieved in the absence of RNA-binding proteins, which help to coordinate and regulate gene expression at multiple stages in the life cycle of an RNA (Moore 2005; Keene 2007), and further entangle the RNA folding process with the physical determinants of RNA–protein interactions. Furthermore, RNA sequences form conformationally dynamic structures, with biologically functional conformational transitions and multiple nonnative folded states (Winkler and Breaker 2005; Solomatina et al. 2010). Protein binding can affect the occupancy and rate of exchange between these structural states, while conversely the structure of the RNA helps to define protein binding affinity and specificity. Therefore, there exists a need for methods that can begin to provide generalizable, broad insights into the complex relationship between RNA sequence, dynamic conformations, and protein binding, creating a principled framework for understanding RNA’s diverse functions within the cell.

However, the challenge is vast. A productive approach to these problems requires measurements that can account

for these multiple structural states at an immense scale—commensurate with the combinatorial sequence complexity of nucleic acids. And although a variety of high-throughput methods have been developed to measure RNA structure and RNA–protein binding, few allow for reporting of these interactions in terms of the thermodynamic and kinetic constants that must govern physical behaviors in the cell. Transcriptome-wide RNA structure measurements have relied on proximity cross-linking (Ramani et al. 2015) or base accessibility (Ding et al. 2014), each of which provides constraints on the behavior of the ensemble of RNA structures, but these methods require averaging over many different structures and thus cannot quantify occupancy across different structural states and cannot be easily read out in terms of binding energies and kinetic constants. In vivo approaches to quantify RNA–protein interactions through cross-linking and immunoprecipitation (CLIP, eCLIP, iCLIP, etc.) (Hafner et al. 2010; Van Nostrand et al. 2016) have mapped enrichment of RNA-binding protein (RBP) binding at high throughput, but often suffer from poor sequence resolution, biased cross-linking efficiency, large variability in the abundance of different potential binding sites, and indirect targeting of binding interactions with an antibody. In vitro approaches can solve a number of these uncertainties, and medium-throughput approaches for quantitative measurement have been enabled by microfluidic methods (Martin et al. 2012). Higher-throughput measurements of relative RBP affinity have been achieved with sequencing-enrichment approaches like RNAcompete, RNA-bind-and-seq, RNA SELEX, HiTS-Kin, and HiTS-Eq (Ellington and Szostak 1990; Ray et al. 2009; Lambert et al. 2014; Jankowsky and Harris 2017; Lou et al. 2017). These methods have enabled measurements of relative RBP affinity across randomized libraries; however, they often require immense sequencing resources for readout and the enrichment readout is not as generalizable as direct measurements of biophysical parameters.

A recently developed class of techniques has enabled direct, fluorescence-based, quantitative measurement of binding affinity and kinetics to RNA, with the same or larger experimental capacity as sequencing-based approaches. Two highly related approaches in this class are RNA-MaP (quantitative analysis of RNA on a massively parallel array) (Buenrostro et al. 2014) and HiTS-RAP (high-throughput sequencing and RNA affinity profiling) (Tome et al. 2014); In this review, we will refer to this class of methods collectively as RNA-HiTS, for an RNA array on a high-throughput sequencer.

These methods make use of the core functionality of the high-throughput DNA sequencers (i.e., clonal amplification and massively parallel fluorescence quantification) to enable high-throughput, fluorescence-based detection of

nucleic acid substrates for diverse applications. RNA-HiTS was born from work on DNA-protein interactions that used a high-throughput sequencer to characterize protein binding to a vast swath of DNA sequences (Nutiu et al. 2011)—the first demonstration of the power of the sequencing instrument for quantitative determination of affinity landscapes of mutant nucleic acid sequences. Other applications of sequencing infrastructure include selection of specific oligonucleotide sequences within a pool of targets (Matzas et al. 2010), detection of mRNA sequence using fluorescently labeled transfer RNAs (tRNAs) (Uemura et al. 2010), and determination of the length of poly(A) tails among different classes of mRNAs (Subtelny et al. 2014). RNA-HiTS technology substituted diverse RNA targets for the original dsDNA targets in Nutiu et al. (2011), ultimately enabling equally comprehensive and quantitative characterization of binding to RNA (Buenrostro et al. 2014; Tome et al. 2014).

2 TECHNIQUES

2.1 Illumina Platforms Generate Arrays of Sequence-Identified DNA Clusters on the Surface of a Chip

During the process of high-throughput sequencing on an Illumina GAII(x) DNA Sequencer, millions of individual DNA strands are first immobilized on a flow cell surface, then clonally amplified to generate a “cluster” of DNA containing approximately 1000 identical molecules (Fig. 1A). During the sequencing process, the nucleic acid sequence of each clonal cluster is determined by step-wise sequencing-by-synthesis of one strand of each DNA molecule on the surface of the flow cell. Fluorescently labeled nucleotides are incorporated such that each incorporation cycle allows determination of the identity of one base of the DNA sequence (Fig. 1A). Typical sequencing reactions proceed for at least 30 cycles per read (up to 300 cycles), and multiple reads can be performed starting from different priming sequences to

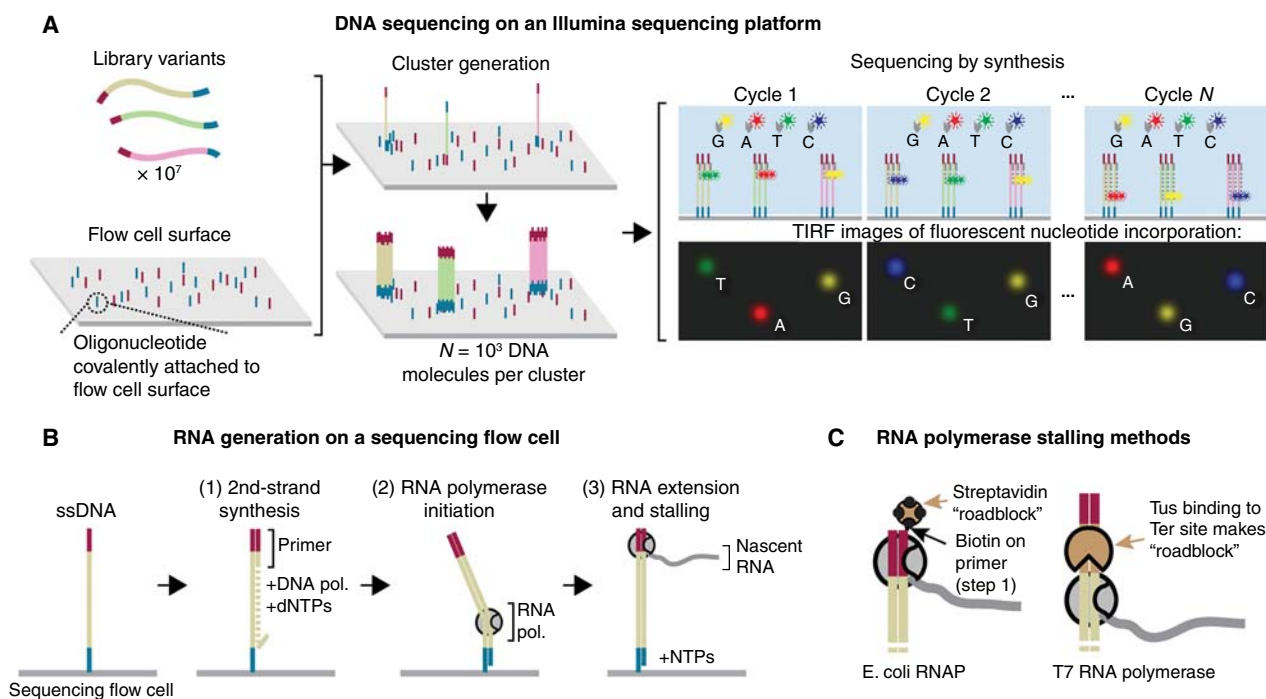


Figure 1. Harnessing the Illumina sequencing platform for systems-level biochemistry of RNA. (A) Schematic for DNA sequencing on an Illumina platform. DNA fragments with sequencing adapters on 5′ and 3′ ends anneal to complementary oligonucleotides covalently attached to the sequencing flow cell surface. Molecules are clonally amplified by polymerase chain reaction (PCR) to form “clusters” on the chip surface, which are then sequenced by sequential incorporation of reversibly terminated, fluorescent nucleotides. (B) RNA-HiTS extends the Illumina sequencing platform by transcribing clusters of DNA into RNA directly on the surface of the sequencing chip. (1) The second strand of DNA is regenerated with a DNA polymerase. (2) An RNA polymerase is initiated by allowing binding to an initiation sequence designed into the DNA fragment. (3) The RNA polymerase is allowed to extend until it stalls at a “roadblock,” resulting in stable display of the nascent RNA transcript (gray). (C) Two methods for stalling the RNA polymerase. (Left) A biotinylated primer is used for second-strand synthesis (see B); streptavidin is then bound to this biotin before transcription, making a “roadblock” that stalls the *Escherichia coli* RNA polymerase during the extension step. (Right) A Tus protein binds a Ter sequence element designed into each DNA fragment, which stalls T7 RNA polymerase.

detect barcodes for demultiplexing or to read from both ends of the molecule for paired-end sequencing.

The GAII hardware platform integrates fluidics handling and thermal control that enables molecular biology reactions to take place on the surface of the chip, as well as total internal reflection fluorescence (TIRF) imaging for detection of nucleotide incorporation across clusters in massive throughput. Initial implementations of RNA-HiTS methods hijacked these instruments directly (Buenrostro et al. 2014; Tome et al. 2014). However, the GAII platform is antiquated and unsupported, and newer, more economical Illumina instruments like the MiSeq allow much more rapid sequencing turnaround. However, these newer instruments are highly integrated and do not support the facile implementation of nonstandard molecular biology and imaging workflows. Recently, our laboratory has developed a modified, custom-built imaging station based on the original Illumina GAII that decouples the platform used for sequencing from the instrument used for custom biochemistry applications (She et al. 2017). This imaging station is built using components from the original Illumina GAII(x) but is designed to interface with Illumina MiSeq chips, with “home brew” image analysis software and hardware interfaces.

2.2 In Situ Transcription to Display Nascent Transcripts of RNA at High Throughput

The conversion of the native DNA array of a postsequenced chip to an RNA array requires in situ transcription of the DNA clusters on the surface of the flow cell. Broadly, this in situ transcription is performed in three steps: beginning with single-stranded DNA molecules covalently attached to the sequencing flow cell surface, (1) second-strand synthesis of the DNA molecules by primer annealing and extension by a DNA polymerase is carried out, (2) an RNA polymerase is initiated in a sequence-specific manner, and then (3) a nascent RNA transcript is extended until the RNA polymerase reaches a “roadblock” that stably halts the polymerase on the DNA template (Fig. 1B).

This “roadblock” has been implemented in two distinct ways. In the first, a biotinylated primer can be used to prime second-strand synthesis, and subsequently this biotin may be bound by streptavidin before RNA polymerase initiation and extension (Fig. 1C) (Buenrostro et al. 2014). The RNA polymerase (in this case *Escherichia coli* RNAP) stalls when it encounters this terminal biotin–streptavidin roadblock (Greenleaf et al. 2008), allowing stable display of the nascent RNA transcript. Furthermore, this workflow aims to allow only a single RNAP molecule to initiate at the engineered promoter, producing a single piece of RNA stably tethered

to its DNA template. In an alternate stalling scheme, the replication terminator protein Tus is bound to a 32-bp *Ter* sequence element engineered into the DNA library (Mohanty et al. 1996; Mulugu et al. 2001). T7 RNA polymerase then stalls when it encounters the bound Tus protein (Fig. 1C) (Tome et al. 2014).

After transcription, DNA oligos may be annealed to common sequences on all transcripts to help stabilize the desired secondary structure of the RNA, to provide a readout of transcription efficiency (Fig. 2A), or to otherwise label the transcribed RNA.

2.3 Performing Kinetic and Thermodynamic Measurements and Quantifying Images

The TIRF imaging platform enables direct measurement of association of any fluorescently labeled binding partner to every RNA cluster on the surface of the chip. This direct observation of binding allows relatively straightforward measurement of kinetic and thermodynamic parameters. For example, the equilibrium dissociation constants (K_d) can be measured by applying different concentrations of the binding partner to the chip, waiting sufficient time for equilibration, then imaging fluorescence observed at each cluster (Fig. 2B,C). Quantifying the bound fluorescence across concentrations and fitting these data to a binding curve enables inference of the K_d and the standard free-energy change of forming the bound complex ($\Delta G^\circ = RT \log K_d$) (Fig. 2C). Dissociation rate constant (k_{off}) can likewise be directly measured by subsequently diluting the fluorescently bound binding partner from solution, then sequentially imaging of the loss of bound fluorescence over time. In principle, association rate constants can be measured in a similar manner, or can be calculated from $k_{\text{on}} = k_{\text{off}}/K_d$.

A necessary step to quantifying fluorescence of each cluster is to align the positions of clusters obtained from the sequencing data with each fluorescent image (Fig. 2D). An initial cross-correlation provides a rough alignment that accounts for overall offsets between the sequencing data and the images, and then subsequent iterations of cross-correlation of progressively finer grids of subtiles then account for any optical aberrations introduced in different imaging stations. Together, these values are fit to a continuous offset map that can be used to obtain precise alignment of sequence data and experimental images (She et al. 2017). Subsequently, subtiles are fit to a sum of 2D Gaussian functions to determine the integrated fluorescence of each cluster. This strategy also allows for a subtile-specific background fluorescence value to be determined to account for any differences in overall fluorescence in different regions of the chip.

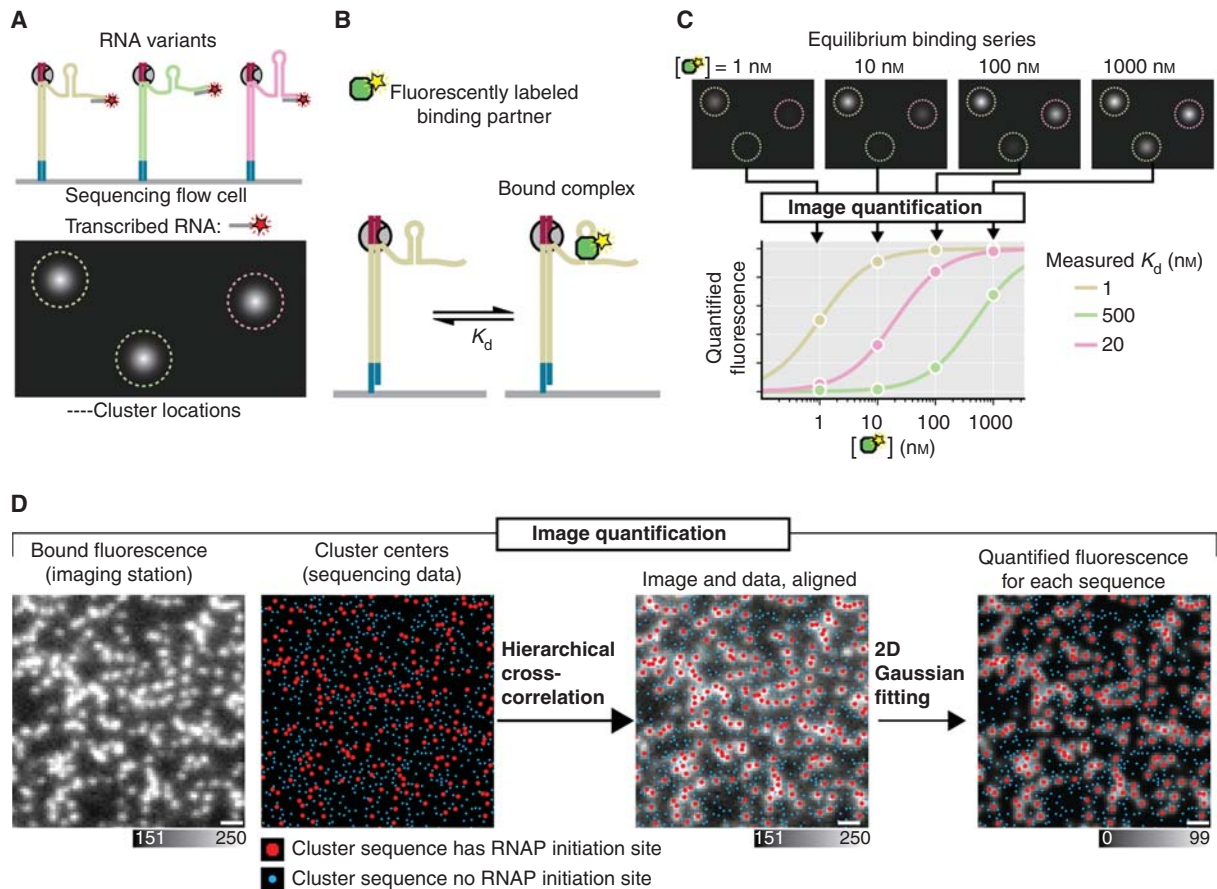


Figure 2. Performing kinetic and thermodynamic measurements and quantifying images. (A) After in situ transcription, transcribed RNAs may be labeled with fluorescent oligonucleotides complementary to a common sequence to all library variants (simulated data below). (B) Application of a fluorescently labeled binding partner to the flow cell enables detection of the thermodynamics and kinetics of binding to each transcribed library variant. (C) A typical equilibrium binding experiment proceeds by applying a specific concentration of the binding partner to the flow cell, waiting for equilibrium, detecting bound fluorescence at each cluster of RNA, and then repeating at multiple concentrations of binding partner. After image quantification, fluorescent measurements are fit to a binding isotherm to obtain the observed equilibrium dissociation constant, K_d . (D) Image quantification workflow. First, images are aligned with the sequencing data through hierarchical cross-correlation. Images are then fit to a sum of 2D Gaussians, with each Gaussian centered at the position of each cluster. Scale bar, 2.5 μm .

2.4 Library Considerations and Requirements

The first step in any RNA-HiTS experiment is to generate a diverse library of sequenceable DNA fragments that will serve as the basis of the RNA array. The generation of this DNA library has two distinct steps. First, one must design and synthesize the variable region that will be transcribed in situ. Next, this variant library must be converted into a sequenceable library. We will deal with each of these steps in turn.

Variable DNA sequences that serve as the core of a RNA variant library can be created in three ways: (1) a full or partial randomization of nucleic acid synthesis, (2) an array-based programmed library synthesis, and (3) the se-

lection of a diverse set of natural sequences. A completely randomized nucleic acid sequence library is often not feasible—experiments can generally probe on the order of 10^6 different sequences at most (assuming an average of 10 replicate clusters per sequence variant and 10^7 total clusters per chip), which corresponds to a completely random sequence of only 10 bases. Structured RNA motifs like hairpin stems will often exceed this limit and have considerable sequence constraints to maintain base-pairing, making targeted sequence variations as opposed to random N -mers often more informative and interpretable. One way to focus investigational throughput around interesting functional variations is to produce a library of sequence variants “centered” on a known consensus se-

quence motif with error-prone polymerase chain reaction (PCR) (Cirino et al. 2003; Tome et al. 2014), or with doped-in synthesis (Buenrostro et al. 2014), each of which can introduce single, double, and some fraction of triple and higher-order mutations to the known motif. However, these classes of “dirty” synthesis are constrained by the statistics of random incorporation: In general, the consensus sequence will be highly represented, whereas higher-order mutations will have a much lower representation depending on sequence length and the error rate.

More recently, high-throughput oligo synthesis platforms have given researchers access to programmed parallel synthesis of up to 10^6 unique oligonucleotide sequences. These commercially available technologies allow researchers to design a library of sequence variants that will each be synthesized independently with approximately equal representation. This synthesis method allows the designer to ask multiple questions within each library—that is, by introducing variation in multiple known consensus sequences to assess evolutionary paths between the sequences or by changing the sequences flanking the consensus sequence to assess the context dependence of mutational effects. Currently, commercially available oligonucleotide pools are generally limited to sequences <200 nt in length, and whereas synthesis error rates can introduce additional unwanted variation, efforts for improving fidelity are ongoing.

A final method takes advantage of the diversity of sequence variants inherent to biological genomes. For example, the *Saccharomyces cerevisiae* genome can be enzymatically digested to form the variable transcribed region in an RNA-HiTS experiment (She et al. 2017). This approach enabled assessment of binding to targets embedded in their physiologically relevant, local sequence context that could affect local structure formation.

The second step in library generation requires the construction of a DNA fragment amenable to sequencing and transcription, which is achieved by adding common sequences to the entire library of variants at the 5′ and 3′ end. Each library variant must have an RNAP initiation site at the 5′ end of the variable transcribed region, and if using Ter/Tus stalling, the 32-bp Ter element at the 3′ end of the transcribed region. Both stalling methods require a sufficient number of base pairs between the stall site of the RNAP polymerase and the 3′ end of the RNA target, as the RNAP polymerase transcripts have a footprint that leaves approximately 25 bases of RNA within the exit channel (Greenleaf et al. 2008), preventing access by the fluorescently labeled binding partner.

The DNA sequencing error rate can present an issue for an RNA-HiTS experiment, especially for a library of highly related sequences, as sequencing errors can lead to clusters being assigned to the wrong molecular variant, producing

spurious results. A per-base error rate of ~0.1% or more can be expected for Illumina sequencing. For a variable region of 100 nt, such a rate might produce misassignment of a substantial fraction (1/10th) of library variants. One strategy to remove the impact of sequencing error is to introduce a unique molecular identifier (UMI) to each molecule within the library, comprising a random sequence of 16 nt. On incorporation of this UMI to each member of the library, the library is diluted to $\sim 8 \times 10^5$ molecules and then reamplified, creating a “bottlenecked” population of PCR amplicons, with each UMI represented many times in the library. Sequencing this bottlenecked library allows for confident assessment of the specific sequence variant associated with each UMI, and downstream analysis requires only the sequence of the UMI to identify the molecular variant.

In general, the number of sequenced clusters should exceed the number of variants by at least 10-fold, in order to make multiple distributed measurements for each molecular variant in the library to enable a principled understanding of measurement noise.

3 APPLICATIONS

3.1 Sequence and Structural Determinants of Protein–RNA Interactions

3.1.1 MS2 Coat Protein to Mutated RNA Targets

An overarching goal of RNA-HiTS technology has been to make a generalizable, predictive model of protein–RNA interaction affinity across any possible RNA sequence variant. The first protein targeted for this analysis was the MS2 coat protein (Buenrostro et al. 2014), a protein that was initially discovered to play a bifunctional role in viral coat assembly and translational repression in the bacteriophage MS2. The high affinity and specificity of this interaction has made it useful in biotechnology applications including affinity purification, live-cell RNA imaging, and synthetic biology (Bardwell and Wickens 1990; Bertrand et al. 1998). The MS2 protein binds an RNA motif consisting of a stem loop structure with a single bulged residue (Fig. 3A). Starting with this consensus sequence, RNA-HiTS enabled the quantitative measurement of MS2 interaction with all single, double, and a subset of triple and higher-order mutants to comprehensively assess the contribution of each residue to the free-energy landscape of the binding process. The effects of each single mutant revealed high mutational sensitivity to a subset of single stranded residues important for docking into the single-stranded protein. Double-mutant effects revealed epistasis between specific base-paired residues, showing the importance of maintaining base-pairing throughout the stem for MS2 protein recognition and binding.

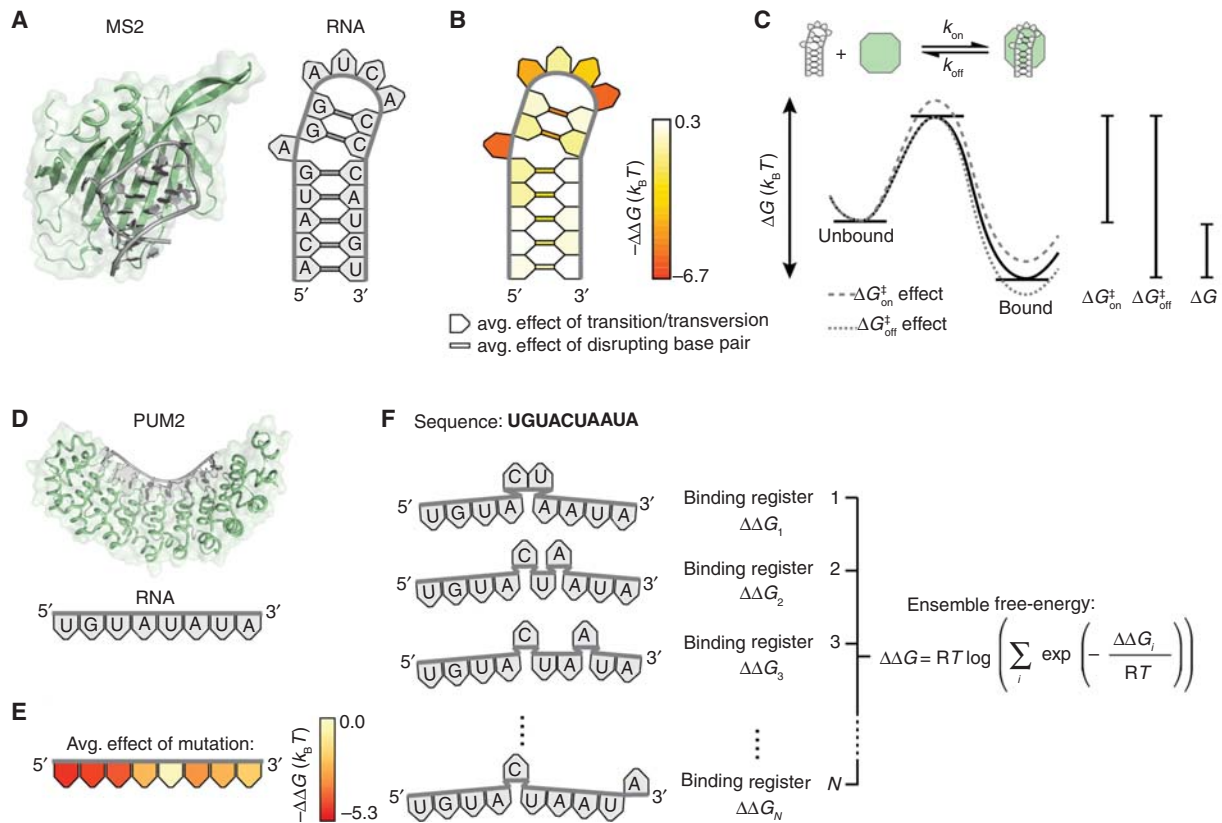


Figure 3. Sequence and structural determinants of protein–RNA interactions. (A) (Left) Crystal structure of the MS2 coat protein–RNA interaction (PDB: 2BU1) (Grahn et al. 2001). (Right) Primary sequence and secondary structure of the consensus MS2 RNA binding site. (B) Linear regression analysis of single and double mutants of the consensus MS2 RNA binding site attributed effects to either primary sequence (base colors) or to secondary structure (base pair colors) (Buenrostro et al. 2014). (C) Free-energy diagram showing the unbound, transition, and bound states. Each ΔG^\ddagger term is linked to the kinetics of forming the complex, with $\Delta G = \Delta G_{\text{off}}^\ddagger - \Delta G_{\text{on}}^\ddagger$. Mutations can affect the ΔG by affecting $\Delta G_{\text{off}}^\ddagger$ (dotted line), $\Delta G_{\text{on}}^\ddagger$ (dashed line), or both. (D) Crystal structure of the RNA-binding domain of human PUM2 protein in complex with RNA (PDB: 3Q0Q) (Lu and Hall 2011), with the primary sequence of the consensus site indicated. (E) The average effect of each single mutation to the consensus site (data from I Jarmoskaite, SK Denny, P Vaidyanathan, et al., in prep.). (F) Multiple binding registers can give rise to a stable interaction between PUM2 and RNA for a given RNA sequence. Residues pointing down contribute to protein binding, whereas residues pointing up are “flipped out.” The $\Delta\Delta G$ of each binding register is calculated with an energy term for each residue contacting the protein, with a penalty for having flipped out residues, as well as other coupling terms. The final effect for a given sequence is the ensemble free-energy of each possible binding register.

These observations allow direct quantification of primary sequence and secondary structure determinants of protein recognition and binding using a simple regression model that estimates the effect of a transition or a transversion of each residue, and the effect of converting a base pair to either a noncanonical base pair or to an otherwise disrupted base pair (Fig. 3B). This full model was predictive of the higher-order mutation effects.

The free-energy change in forming a bound complex can be decomposed into two contributions: the free-energy change between the starting state and the transition state ($\Delta G_{\text{on}}^\ddagger$), and the free-energy change between the transition state and the final state ($\Delta G_{\text{off}}^\ddagger$), such that $\Delta G = \Delta G_{\text{off}}^\ddagger - \Delta G_{\text{on}}^\ddagger$

(Fig. 3C). Each of these ΔG^\ddagger terms is linked to the kinetics of forming the complex, with the first step affecting the association rate constant, and the second step affecting the dissociation rate constant: $\Delta G_{\text{on}}^\ddagger = RT \log k_{\text{on}}$ and $\Delta G_{\text{off}}^\ddagger = RT \log k_{\text{off}}$. Thus, measuring the kinetic rate constants in addition to equilibrium binding allows quantification of mutational effects on each of these steps of the binding process (Fig. 3C). In the MS2 coat protein experiment, dissociation rate constants (k_{off}) were measured for each mutant, revealing that mutation effects predominantly resulted from changes in the association rate constant, k_{on} , suggesting that MS2 must wait for the formation of a binding competent RNA secondary structure before association.

Mutational effects can be used to constrain the functional landscape in which evolution must operate to generate or maintain high affinity binding. Many double mutants within the MS2 hairpin had very small effects and thus represent equivalent “solutions” to the evolutionary “problem” of creating a stable RNA–protein interaction. However, the mutational paths that connect each pair of double mutants were observed to have very different “roughness” in this functional binding space, depending on the order in which the mutations accumulated. These types of data provide quantitative insight into complexities and path dependences of natural and engineered methods for evolving high-affinity binding partners.

3.1.2 RNA Aptamers for GFP and NELF

RNA-HiTS allows characterization of synthetic RNA–protein interactions. For example, Tome and colleagues characterized the specificity of two RNA aptamers that were each originally identified through *in vitro* selection: aptamers to the green fluorescent protein (GFP) and to the *Drosophila* negative elongation factor (NELF-E) (Shui et al. 2012; Pagano et al. 2014; Tome et al. 2014). Variants of each aptamer were generated with error-prone PCR, resulting in single, double, and some higher order mutants of each sequence. Examining the effects on binding affinity of each aptamer to their respective proteins identified regions with substantial sensitivity to single point mutants. The summed effects of these single point mutations were predictive of the effects of higher-order mutations in the NELF aptamer, supporting an additive model whereby each residue contributes additively to the affinity of this interaction. However, this linearity was not observed in the GFP aptamer, which had sensitivity to changes in base-paired regions, leading to significant epistasis between residues. These observations show the ability of RNA-HiTS to quantify and define the potentially diverse binding mechanisms that may come out of an *in vitro* selection. Mutational analysis of the GFP aptamer also revealed two point mutants with several-fold higher affinity than the original aptamer, showing RNA-HiTS’s capacity to optimize affinity over the original selected aptamer. We envision that, in the future, the RNA-HiTS technology may be used as part of the *in vitro* selection process (i.e., to characterize a broader diversity of putative RNA targets at earlier rounds of selection).

3.1.3 Puf Binding to Designed Library of RNA

Puf family proteins are versatile posttranscriptional regulators found throughout eukaryotes (Quenault et al. 2011). The specificity of Puf–RNA interactions is thought to be driven by an ~8-nt consensus sequence found within

the 3′ UTR of genes (Wickens et al. 2002; Van Nostrand et al. 2016). Binding preference for this consensus motif is conferred by a highly modular protein structure composed of eight amino acid repeats that each interact with a single RNA nucleotide (Fig. 3D) (Wang et al. 2001). Thus, this system provides an ideal testbed for developing an additive model aimed at predicting binding free-energy from primary nucleic acid sequence.

However, the high-precision measurements afforded by RNA-HiTS analysis revealed a substantially more complex picture of the binding process, even for this simple interaction, than could be captured by a simple linear model (I Jarmoskaite, SK Denny, P Vaidyanathan P, et al., in prep.). These complexities stemmed from three primary sources: (1) RNA secondary structure formation can affect the accessibility of the binding site; (2) alternate binding registers with “flipped-out” residues can form stable complexes; and finally (3) subtle, yet pervasive, energetic coupling can occur between residues within the binding site. Systematic identification and quantification of these distinct contributions resulted in a two-step model that first enumerates the possible binding configurations, then predicts the binding free-energy of each configuration using a linearly additive model with coupling terms (Fig. 3E,F). The final observed affinity for any sequence is thus derived from the ensemble of free energies for all binding configurations (Fig. 3F), modulated by secondary structure formation as predicted by algorithms such as Vienna fold (Gruber et al. 2008). This model predicted Puf occupancy that was linearly related to the observed occupancy of these sites as determined in an *in vivo* cross-linking experiment, highlighting both the usage of this fully predictive model and supporting the importance of thermodynamics in *in vivo* settings (Van Nostrand et al. 2016).

3.2 Protein–RNA Interactions across Transcriptome Targets

The investigation of the relationship between protein binding in the context of *in vivo*–derived RNAs is another exciting application of RNA-HiTS. The immense capacity of these instruments opens the possibility of placing an entire eukaryotic genome on the chip, and probing binding to every theoretically possible transcript. This approach, termed the transcribed genome array (TGA), was used to examine binding preferences of the RNA binding domain of Vts1 from *S. cerevisiae* (She et al. 2017), a member of the Smaug family of proteins with conserved RNA binding domains implicated in mediating RNA decay throughout eukaryotes (Aviv et al. 2003; Oberstrass et al. 2006; Rendl et al. 2008, 2012; Riordan et al. 2011). The entire *S. cerevisiae* genome was fragmented into ~100-nt fragments result-

ing in coverage of $>30\times$ per nucleotide. This library of overlapping fragments formed the substrate for in vitro transcription in the RNA-HiTS experiment.

Using this unbiased, transcriptome-wide approach, approximately 300 sites within the genome were expected to be significantly occupied within the cell, based on the apparent K_d s observed and the estimated physiological protein concentration (130 nM); two-thirds of these sites fell into the transcribed strand of an open reading frame. These gene targets were significantly overlapping with those from in vivo cross-linking and immunoprecipitation (RIP-seq) experiments, suggesting that in vitro identified targets recapitulate in vivo binding events in aggregate (Aviv et al. 2006; Hogan et al. 2008). The functional relevance of these binding events was assessed by comparing the expression between *vtS1A* and wild-type cells by RNA-seq. These data showed that genes associated with TGA-identified binding sites had significantly increased average expression, although RIP-identified targets displayed no significant average expression change. Targets common to both analyses had the greatest expression increase, suggesting that TGA may help uncover the true positives within RIP-identified targets. In addition, TGA identified many new targets, virtually all of which were lowly expressed in vivo, highlighting a crucial difference between TGA and in vivo immunoprecipitation experiments: Cross-linking and pulldown experiments are often limited to probing only targets within genes with high expression in the conditions that the cells had been grown. Instead, TGA allows investigation of affinity to all possible transcripts simultaneously.

3.3 Design and Engineering of RNA Devices Using Massively Parallel Rational Design

The flexibility of RNA-HiTS enables investigation of RNA beyond its capacity to bind proteins. One exciting area of application is in the field of biological engineering wherein RNA is used as a programmable tool for control of biological systems. The motivation for this molecular engineering stemmed from the discovery of natural RNA riboswitches, which are conceptually composed of two pieces: the structured aptamer that binds a small-molecule “effector,” and a second region that undergoes a conformational change in response to the effector, generally resulting in a change of gene expression through various mechanisms (Tucker and Breaker 2005). The separation between these two functions (i.e., the aptamer domain and the resulting functional consequence) supports that RNAs might enable modular programming of specific computations within the cell (Kim and Smolke 2017). However, the design of these devices is currently limited by imperfect predictive under-



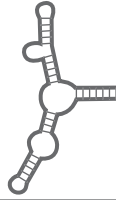

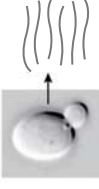
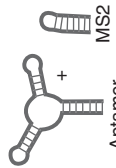








standing of these aptamers and their coupling to conformational changes.

To make progress on this question, the RNA-HiTS platform was used to measure the response of tens of thousands of rationally designed RNA riboswitches. These designs came from scientific “enthusiasts” playing the online game EteRNA. This crowd-based rational approach enables massively parallel, yet hypothesis-driven science, as originally shown for RNA secondary structure folding (Lee et al. 2014). EteRNA players were tasked with designing RNA sequences that would satisfy the “challenge” of forming the prespecified aptamer domain and changing another domain’s conformation in response to the effector molecule (see eternagame.org/web/). The player-generated designs were tested using RNA-HiTS to detect the response of each riboswitch, resulting in a score for each riboswitch; players could then use these scores in subsequent rounds to optimize their designs. In combination with RNA-HiTS, this crowd-sourced approach produced thermodynamically optimal sensors, with binding of the aptamer domain fully coupled to changes in the conformation of the readout domain (JOL Andresen et al., in prep.).

3.4 Understanding Sequence Contributions to RNA Tertiary Structure

A last compelling application of RNA-HiTS is in developing a predictive model for the RNA folding process. Although the first step of hierarchical RNA folding (i.e., secondary structure formation) may be predicted fairly accurately for the lowest free-energy state with readily available software (Turner et al. 1988; Gruber et al. 2008), models to predict RNA tertiary structure from this secondary structure are substantially more limited in scope. Current approaches have focused on obtaining structural characterization of the “modules” of RNA tertiary structure—that is, recurrent RNA elements like helices and junctions that have similar crystal structures in different RNAs—and then using these structures like Legos to infer the conformation of larger RNAs (Jaeger et al. 2001; Petrov et al. 2013; Miao and Westhof 2017). However, the inherent flexibility of RNA limits the applicability of these approaches to achieve thermodynamic understanding of the stability of tertiary structures (Herschlag et al. 2015). Furthermore, this approach is limited by the small number of elements for which high-quality crystallographic information exists.

To approach these challenges, RNA-HiTS was used to measure the sequence dependence of tertiary structure formation at high throughput (Denny et al. 2018; Yesselman et al. 2018). RNA elements like helices and two-way junctions were introduced as “guest” motifs into a “host” system—together, they formed a minimal RNA tertiary as-

Substrate	MS2	PUM2	GFPapt	NELFapt	Vts1	Riboswitches	tectoRNAs
	 <p>Single, double, and higher-order mutants of consensus sequence.</p>	 <p>Single, double mutants of consensus sequence, in multiple flanking sequence and structure contexts.</p>	 <p>Single and double mutants of in vitro selected sequence</p>	 <p>Single and double mutants of in vitro selected sequence</p>	 <p>Genomic fragments of <i>S. cerevisiae</i></p>	 <p>Aptamer binding site MS2</p> <p>Rationally designed riboswitches from online game EteRNA</p>	 <p>Variants of RNA helices and junctions inserted into tectoRNA scaffold</p>
Binding partner	 <p>MS2 coat protein</p>	 <p>PUM2 RNA-binding domain</p>	 <p>EGFP protein</p>	 <p>NELF-E protein</p>	 <p>Vts1 RNA-binding domain</p>	 <p>MS2 coat protein +/- effector</p>	 <p>tectoRNA-binding partner</p>
Observations	K_D, k_{off}	K_D, k_{off}	K_D	K_D	$K_{d, obs}$	K_D	K_D, k_{off}
Interpretations	Primary sequence and secondary structure determinants of binding, in a linear regression.	Primary sequence effects, alternate binding registers, secondary structure effects, and other epistatic contributions.	Primary sequence effects from single mutants; these effects were not predictive of double-mutant effects.	Primary sequence effects from single mutants; the sum of each single-mutant effect predicted double-mutant effects.	Binding sites of Vts1 in the yeast genome; comparison with functional gene targets.	Riboswitches that fully couple binding of aptamer domain to change in conformation of MS2 binding site.	Classes of conformational behavior of RNA junctions determined through thermodynamic fingerprints.

Has endogenous RNA binding? ■ Yes ■ No

Figure 4. Summary of RNA-HITS applications.

sembly (tectoRNA), whose formation quantitatively depends on the conformational preferences of the integrated element (Jaeger and Leontis 2000; Geary et al. 2008). For two-way junction elements, “thermodynamic fingerprints” were constructed that represented a multidimensional readout of the junction’s conformational behavior. Unbiased clustering of junction thermodynamic fingerprints revealed classes of junctions with similar conformational preferences. Differences in conformational behavior were largely driven by differences in secondary structure (as expected from Bailor et al. 2010), but also by previously unrecognized primary sequence attributes like mismatch identity. Conformational classes were also useful for inferring structural behavior. Existing crystallographic conformations were associated with specific thermodynamic fingerprints, ultimately enabling the identification of structural ensembles describing the flexible and dynamic behavior of junctions with similar thermodynamic fingerprints. These “bootstrapped” ensembles, when combined with ensembles for RNA helices and a statistical mechanics model for tectoRNA assembly formation, were predictive of measured thermodynamic fingerprints and outperformed the use of static structures for predicting these behaviors.

4 CONCLUSION

Initial applications of the RNA-HiTS platform focused on dissecting the sequence and structural determinants of RNA–protein interactions. Collectively, these studies allowed accurate quantification of the subtle contributions to binding affinity of primary sequence effects, alternate binding registers, secondary structure effects, and other epistatic relationships between residues not directly involved in base-pairing (Fig. 4). The quantitative contributions of each of these elements were identified in the case of the GFP aptamer and Vts1; and these contributions were further modeled for any arbitrary sequence in the cases of MS2, NELF-E, and PUM2. In the future, diverse other protein–RNA interactions may be interrogated in such a system, allowing a more complete mapping of the diverse physical mechanisms that determine RNA–protein interactions.

RNA-HiTS data can complement *in vivo* cross-linking and immunoprecipitation experiments, as seen in the cases of Vts1 and Pum2, by identifying likely false-positive targets, as well as binding sites in lowly expressed transcripts that are not able to be probed with cross-linking methods. The manner by which the *in vitro* occupancy predicted by RNA-HiTS measurements is affected by additional effects *in vivo*—such as RNA localization and sequestration, cooperativity with other RNA binding proteins, as well as helicases that might periodically “reset” the binding state—remain to be systematically addressed. However, a system-

atic understanding of these other effects is fundamentally impossible without a “baseline” of expected occupancy of sites derived from thermodynamic parameters measured in the absence of these confounders.

More broadly, we anticipate other areas of applications of “systems biochemistry” approaches like RNA-HiTS. Multiple groups have used sequencing arrays to investigate DNA–protein interactions (Nutiu et al. 2011; Boyle et al. 2017), and we anticipate that extending this approach with multicolor imaging capabilities will begin to unravel the nature of transcription factor binding cooperativity, and nucleic acid binding cooperative more generally. Beyond transcription factors, proteins that can be engineered to bind diverse nucleic acid targets (e.g., Cas9, Cpf1, TALEN, Cas13, and AGO proteins) are of special interest. The potential binding targets of each of these programmable proteins span an immense combinatorial space, making high-throughput and quantitative investigations a necessary step in characterizing their behavior for applied and therapeutic purposes.

With methods such as RNA-HiTS, the biochemical and biophysical realms have been brought into the high-throughput era. Now, we as RNA biologists must direct the immense bandwidth afforded by these approaches to “carve nature at its joints.” Thus, a significant challenge remains in designing and engineering RNA libraries to provide maximal insight into the physical underpinnings of complex RNA behaviors.

REFERENCES

- Aviv T, Lin Z, Lau S, Rendl LM, Sichi F, Smibert CA. 2003. The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nat Struct Biol* **10**: 614–621.
- Aviv T, Lin Z, Ben-Ari G, Smibert CA, Sichi F. 2006. Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. *Nat Struct Mol Biol* **13**: 168–176.
- Bailor MH, Sun X, Al-Hashimi HM. 2010. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**: 202–206.
- Bardwell VJ, Wickens M. 1990. Purification of RNA and RNA–protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* **18**: 6587–6594.
- Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **2**: 437–445.
- Boyle EA, Andreasson JOL, Chircus LM, Sternberg SH, Wu MJ, Guegler CK, Doudna JA, Greenleaf WJ. 2017. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci* **114**: 5461–5466.
- Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ. 2014. Quantitative analysis of RNA–protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* **32**: 562–568.
- Butter F, Scheibe M, Mörl M, Mann M. 2009. Unbiased RNA–protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci* **106**: 10626–10631.

- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.
- Cirino PC, Mayer KM, Umeno D. 2003. Generating mutant libraries using error-prone PCR. *Methods Mol Biol* **231**: 3–9.
- Denny SK, Bisaria N, Yesselman JD, Das R, Herschlag D, Greenleaf WJ. 2018. High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell* **174**: 377–390.e20.
- 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.
- Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**: 818–822.
- Geary C, Baudrey S, Jaeger L. 2008. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res* **36**: 1138–1152.
- Grahn E, Moss T, Helgstrand C, Fridborg K, Sundaram M, Tars K, Lago H, Stonehouse NJ, Davis DR, Stockley PG, et al. 2001. Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *RNA* **7**: 1616–1627.
- Greenleaf WJ, Frieda KL, Foster DAN, Woodside MT, Block SM. 2008. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319**: 630–633.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Res* **36**: W70–W74.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129–141.
- Herschlag D, Allred BE, Gowrishankar S. 2015. From static to dynamic: The need for structural ensembles and a predictive model of RNA folding and function. *Curr Opin Struct Biol* **30**: 125–133.
- Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. 2008. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6**: e255.
- Jaeger L, Leontis NB. 2000. Tecto-RNA: One-dimensional self-assembly through tertiary interactions. *Angew Chem Int Ed* **39**: 2521–2524.
- Jaeger L, Westhof E, Leontis NB. 2001. TectoRNA: Modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res* **29**: 455–463.
- Jankowsky E, Harris ME. 2017. Mapping specificity landscapes of RNA–protein interactions by high throughput sequencing. *Methods* **118–119**: 111–118.
- Keene JD. 2007. RNA regulons: Coordination of post-transcriptional events. *Nat Rev Genet* **8**: 533–543.
- Kieft JS, Zhou K, Jubin R, Doudna JA. 2001. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**: 194–206.
- Kim CM, Smolke CD. 2017. Biomedical applications of RNA-based devices. *Curr Opin Biomed Eng* **4**: 106–115.
- Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. 2014. RNA Bind-n-Seq: Quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54**: 887–900.
- Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H, Limpaecher A, Yoon S, Treuille A, Das R, et al. 2014. RNA design rules from a massive open laboratory. *Proc Natl Acad Sci* **111**: 2122–2127.
- Lou T-F, Weidmann CA, Killingsworth J, Tanaka Hall TM, Goldstrohm AC, Campbell ZT. 2017. Integrated analysis of RNA-binding protein complexes using in vitro selection and high-throughput sequencing and sequence specificity landscapes (SEQRS). *Methods* **118–119**: 171–181.
- Lu G, Hall TMT. 2011. Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* **19**: 361–367.
- Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. 2012. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods* **9**: 1192–1194.
- Matzas M, Stähler PF, Kefer N, Siebelt N, Boisguérin V, Leonard JT, Keller A, Stähler CF, Häberle P, Gharizadeh B, et al. 2010. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat Biotechnol* **28**: 1291–1294.
- Miao Z, Westhof E. 2017. RNA structure: Advances and assessment of 3D structure prediction. *Annu Rev Biophys* **46**: 483–503.
- Mohanty BK, Sahoo T, Bastia D. 1996. The relationship between sequence-specific termination of DNA replication and transcription. *EMBO J* **15**: 2530–2539.
- Moore MJ. 2005. From birth to death: The complex lives of eukaryotic mRNAs. *Science* **309**: 1514–1518.
- Mulugu S, Potnis A, Shamsuzzaman, Taylor J, Alexander K, Bastia D. 2001. Mechanism of termination of DNA replication of *Escherichia coli* involves helicase–contrahelicase interaction. *Proc Natl Acad Sci* **98**: 9569–9574.
- Noller HF. 2005. RNA structure: Reading the ribosome. *Science* **309**: 1508–1514.
- Nutiu R, Friedman RC, Luo S, Khrebtkova I, Silva D, Li R, Zhang L, Schroth GP, Burge CB. 2011. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat Biotechnol* **29**: 659–664.
- Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH-T. 2006. Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat Struct Mol Biol* **13**: 160–167.
- Pagano JM, Kwak H, Waters CT, Sprouse RO, White BS, Ozer A, Szeto K, Shalloway D, Craighead HG, Lis JT. 2014. Defining NELF-E RNA binding in HIV-1 and promoter-proximal pause regions. *PLoS Genet* **10**: e1004090.
- Petrov AI, Zirbel CL, Leontis NB. 2013. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **19**: 1327–1340.
- Quenault T, Lithgow T, Traven A. 2011. PUF proteins: Repression, activation and mRNA localization. *Trends Cell Biol* **21**: 104–112.
- Ramani V, Qiu R, Shendure J. 2015. High-throughput determination of RNA structure by proximity ligation. *Nat Biotechnol* **33**: 980–984.
- Ray D, Kazan H, Chan ET, Peña Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR. 2009. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* **27**: 667–670.
- Rendl LM, Bieman MA, Smibert CA. 2008. *S. cerevisiae* Vts1p induces deadenylation-dependent transcript degradation and interacts with the Ccr4p-Pop2p-Not deadenylase complex. *RNA* **14**: 1328–1336.
- Rendl LM, Bieman MA, Vari HK, Smibert CA. 2012. The eIF4E-binding protein Eap1p functions in Vts1p-mediated transcript decay. *PLoS ONE* **7**: e47121.
- Riordan DP, Herschlag D, Brown PO. 2011. Identification of RNA recognition elements in the *Saccharomyces cerevisiae* transcriptome. *Nucleic Acids Res* **39**: 1501–1509.
- She R, Chakravarty AK, Layton CJ, Chircus LM, Andreasson JOL, Damaraju N, McMahon PL, Buenrostro JD, Jarosz DF, Greenleaf WJ. 2017. Comprehensive and quantitative mapping of RNA–protein interactions across a transcribed eukaryotic genome. *Proc Natl Acad Sci* **114**: 3619–3624.
- Shui B, Ozer A, Zipfel W, Sahu N, Singh A, Lis JT, Shi H, Kotlikoff MI. 2012. RNA aptamers that functionally interact with green fluorescent protein and its derivatives. *Nucleic Acids Res* **40**: e39.
- Solomatin SV, Greenfeld M, Chu S, Herschlag D. 2010. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**: 681–684.
- Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**: 315–326.
- Subtelný AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**: 66–71.
- Tinoco I, Bustamante C. 1999. How RNA folds. *J Mol Biol* **293**: 271–281.

- Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, Lis JT. 2014. Comprehensive analysis of RNA–protein interactions by high-throughput sequencing–RNA affinity profiling. *Nat Methods* **11**: 683–688.
- Tucker BJ, Breaker RR. 2005. Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* **15**: 342–348.
- Turner DH, Sugimoto N, Freier SM. 1988. RNA structure prediction. *Ann Rev Biophys Biophys Chem* **17**: 167–192.
- Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD. 2010. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**: 1012–1017.
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**: 508–514.
- Wang X, Zamore PD, Hall TM. 2001. Crystal structure of a Pumilio homology domain. *Mol Cell* **7**: 855–865.
- Wickens M, Bernstein DS, Kimble J, Parker R. 2002. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**: 150–157.
- Winkler WC, Breaker RR. 2005. Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**: 487–517.
- Yesselman JD, Denny SK, Bisaria N, Herschlag D, Greenleaf WJ, Das R. 2018. RNA tertiary structure energetics predicted by an ensemble model of the RNA double helix. *bioRxiv* doi: 10.1101/341107.



Cold Spring Harbor Perspectives in Biology

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

Sarah K. Denny and William J. Greenleaf

Cold Spring Harb Perspect Biol published online October 15, 2018

Subject Collection [RNA Worlds](#)

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

Sarah K. Denny and William J. Greenleaf

The Future of Cross-Linking and Immunoprecipitation (CLIP)

Jernej Ule, Hun-Way Hwang and Robert B. Darnell

The Story of RNA Folding, as Told in Epochs

Daniel Herschlag, Steve Bonilla and Namita Bisaria

What Are 3' UTRs Doing?

Christine Mayr

RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements

Matthijs M. Jore, Stan J.J. Brouns and John van der Oost

Evolution of Protein Synthesis from an RNA World

Harry F. Noller

Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations

Susan Gottesman and Gisela Storz

Setting the Stage: The History, Chemistry, and Geobiology behind RNA

Steven A. Benner, Hyo-Joong Kim and Zunyi Yang

Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome

Chun Kit Kwok, Giovanni Marsico and Shankar Balasubramanian

Protocells and RNA Self-Replication

Gerald F. Joyce and Jack W. Szostak

Technique Development for Probing RNA Structure In Vivo and Genome-Wide

Philip C. Bevilacqua and Sarah M. Assmann

The RNA Worlds in Context

Thomas R. Cech

The Origins of the RNA World

Michael P Robertson and Gerald F Joyce

Aptamers and the RNA World, Past and Present

Larry Gold, Nebojsa Janjic, Thale Jarvis, et al.

Riboswitches and the RNA World

Ronald R. Breaker

Ancestral Roles of Small RNAs: An Ago-Centric Perspective

Leemor Joshua-Tor and Gregory J. Hannon

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



All Modifications and
Oligo Types Synthesized

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?

Your wish is our command.



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



**All Modifications and
Oligo Types Synthesized**

Oligo Modifications?
Your wish is our command.

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

