Published in final edited form as:

Methods Enzymol. 2015 ; 558: 465–493. doi:10.1016/bs.mie.2015.02.007.

RNA Bind-n-Seq: Measuring the Binding Affinity Landscape of RNA Binding Proteins

Nicole J. Lambert^{1,*}, Alex D. Robertson^{1,2,3,*}, and Christopher B. Burge^{1,2,§}

¹Department of Biology, MIT, Cambridge MA 02142

²Program in Computational and Systems Biology, MIT, Cambridge MA 02142

Abstract

RNA binding proteins coordinate post-transcriptional control of gene expression, often through sequence-specific recognition of primary transcripts or mature messenger RNAs. Hundreds of RBPs are encoded in the human genome, most with undefined or incompletely defined biological roles. Understanding the function of these factors will require the identification of each RBP's distinct RNA binding specificity. RNA Bind-n-seq (RBNS) is a high-throughput, cost-effective in vitro method capable of resolving sequence and secondary structure preferences of RBPs. Dissociation constants can also be inferred from RBNS data when provided with additional experimental information. Here we describe the experimental procedures to perform RBNS, and discuss important parameters of the method and ways that the experiment can be tailored to the specific RBP under study. Additionally, we present the conceptual framework and execution of the freely available RBNS computational pipeline, and describe the outputs of the pipeline. Different approaches to quantify binding specificity, quality control metrics and estimation of binding constants are also covered.

Keywords

binding affinity; CLIP; dissociation constant; EMSA; high-throughput sequencing; RNA motif; RNA secondary structure; RNA splicing; surface plasmon resonance

Introduction

RNA binding proteins (RBPs) guide RNA splicing, translation control, mRNA localization and other forms of post-transcriptional gene regulation. Because of their key roles in proper gene expression and processing, genetic alterations to RBP proteins, the *cis*-elements they recognize or changes in their cellular concentrations have important implications for development and disease.

Most RBPs are thought to bind to specific RNA sequences and/or secondary structures. RBPs of the developmentally regulated RBFOX, MBNL and CELF splicing factor families

[§]Address correspondence to: cburge@mit.edu.

³Current address: Counsyl, South San Francisco, CA 94080

*These authors contributed equally

each have high affinity for specific short RNA motifs, while other RBPs recognize more degenerate motifs, e.g., the pervasive splicing regulator SRSF2 binds diverse sequences containing SSNG (S = G or C) with similar affinity (Daubner, Clery, Jayne, Stevenin, & Allain, 2012). Still other RBPs such as those of the adenosine deaminase acting on RNA (ADAR) family discriminate binding sites based on RNA secondary structure (Bass et al., 1997). In vivo binding data using UV crosslinking such as cross-linking immunoprecipitation (CLIP) identify directly bound RNA regions (Ule et al., 2003). In general it is difficult to resolve comprehensive binding portraits of RBPs using these in vivo approaches because binding sites are obscured by other regulatory and functional information pertinent to other RBPs and cellular machinery as well as inherent technical biases (Sugimoto et al., 2012). Binding signals reflecting synergy and competition of multiple factors can complicate the identification of a single RBP's binding potential. If the RBP has a known effect on mRNA stability, for example, then indirect information about RBP specificities can also be obtained by perturbing cellular RBP levels followed by global gene expression profiling.

Due to the difficulty in obtaining a comprehensive binding spectra from in vivo derived data, in vitro approaches have been widely used to quantify the diversity of motifs recognized by RBPs and to quantify the binding strength of RBPs to individual binding sites. Standard techniques to quantify absolute affinities of RBPs for specific RNAs include isothermal calorimetry (ITC), surface plasmon resonance (SPR) and electrophoretic mobility shift assay (EMSA), all of which have low throughput. Recent efforts to capture binding specificity landscapes using higher throughput approaches have been employed, generally at the expense of quantitative measurement.

Many systematic high-throughput methods to identify RBP specificity have been developed, each with different strengths and weaknesses. Systematic evolution of ligands by exponential enrichment (SELEX) uses iterative cycles of binding, RNA isolation and amplification followed by sequencing, allowing the identification of high affinity binding sites (Tuerk & Gold, 1990). Multi-round selection can be time-consuming and the nature of iterative binding/amplification cycles reduces the power to resolve moderate affinity binding sites. However, in some cases iterative binding steps may be necessary to detect very weak and/or degenerate binding signals. Recently described methods use an Illumina sequencer to directly measure fluorescent RBP binding to clusters of RNA molecules anchored to flowcell-attached DNA (Buenrostro et al., 2014; Tome et al., 2014). This methodology has a complicated experimental setup and is quite costly. However, affinity estimations may be more accurate due to the direct measurement of fluorescent RBP binding to RNA clusters. In a medium throughput approach that uses a microfluidic platform, RNA induced trapping of molecular interactions (RNA-MITOMI) directly measures protein-bound sequences, but requires specialized equipment and is not easily scalable (Martin et al., 2012). High-throughput identification of RBP binding motifs can also be accomplished with RNAcompete, which uses an in vitro RNA binding step followed by microarray hybridization of RBP bound sequences (Ray et al., 2009). RNAcompete has proven to be scalable and has been used to measure binding specificities of ~400 RBPs (Ray et al., 2013). Contextual and RNA secondary structural information important for RBP binding may be

more difficult to obtain due to the low incubation temperature typically used and the limited set of structural contexts.

RNA Bind-n-seq (RBNS) measures the sequence and structural preferences of RBPs (Lambert et al., 2014). Several binding reactions containing different quantities of a tagged RBP are incubated with random RNA, followed by high-throughput sequencing of RBP-bound RNA. It is a relatively cost-effective, high-throughput assay suitable for simultaneously resolving weak and strong binding motifs as well as contextual information. Additionally, dissociation constants (K_d values) can be estimated from RBNS data, when supplemented by additional experiments for calibration. The direct sequencing of bound RNAs, the lack of iterative binding steps and the relatively straightforward protocol distinguishes RBNS from other high-throughput methods. Here we will describe important parameters of an RBNS experiment and considerations to tailor RBNS to the experimenter's needs.

While RBNS can provide a wealth of information, careful analysis of the data is essential to quantify the specificity of the RBP. Furthermore, the subtle biases that must be accounted for differ greatly from genomic analyses. The analyses also differ from standard biochemical assays as the high throughput nature necessitates modeling of many simultaneous equilibria. In this chapter we walk through the computational analyses in detail and describe an open source implementation of these analyses. We discuss theory and practical considerations in running the computational analyses efficiently. Additionally, we describe analysis methods to calculate quality control metrics, motif enrichment values and the estimation of binding affinities for RBP bound motifs.

Experimental Method

In the RBNS assay, several individual binding reactions are performed in parallel, each reaction performed with a different RBP concentration (Fig. 1A). This is akin to an EMSA, where we anticipate RNA motifs to be bound in a RBP concentration-dependent manner. Unlike an EMSA assay, relative binding affinities can be determined for many oligonucleotides (*k*mers) simultaneously. Additionally, two negative controls are typically performed: sequencing of the input library to assess the randomness of the input pool, and a no protein binding experiment to detect potential apparatus-selected biases. RBNS can be used to measure an RBP's affinity among a pool of random sequences, a designed set of array-synthesized oligonucleotides or natural transcripts, as desired. The following methodology will detail the experimental and computational methods of RBNS assuming use of a randomized RNA pool.

Prepare RBNS Reagents

***in vitro* transcribe RNA pool**—Input RNA is *in vitro* transcribed using the RBNS T7 template, a synthetic DNA oligo containing a random 40mer flanked by Illumina primers and a T7 promoter sequence (primer sequences are listed in Table 1). If available, hand mixed rather than machine dispensed random nucleotides is recommended for the synthesis of the RBNS T7 template for the equal representation of all nucleotides. To mimic a double-stranded T7 promoter, the T7 promoter oligo is annealed to an equimolar quantity of RBNS

T7 template, by heating the two oligos at 65° C for 5 minutes and then allowing the mixture to cool at room temperature for 2 minutes. This template is used as input to a T7 in vitro transcription reaction, using commercially available kits according to manufacturer's instructions. Formamide is added to the in vitro transcription reaction (to a 25% final volume), then run on a 6% TBE-urea polyacrylamide gel. Dyes can be added for easier visualization during sample loading if the dyes are not expected to run near the expected product size.

Length of library random region—The length of the input RNA is an important aspect of the experimental design. Increasing the length of the RNA pool has the effect of diluting the signal for simple linear motifs (discussed below), but increases power to detect the effect of RNA secondary structure on binding, since longer sequences have greater structure potential. For example, suppose that an RBP binds to a specific 5mer within a 40mer oligo. When analyzing enrichment of short motifs in the bound versus input pool, recovery of this 40mer will increment the count of the specific 5mer (the “signal”) by one, but will also increment the counts of the 35 (= 40 – 5) other 5mers present in the 40mer (the “noise”), resulting in “dilution” of the signal by a factor of 1:36. However, if the oligo were of size 10, then the analogous dilution factor is only 1:6. We have compared RBNS results using different oligo sizes for RBFOX2. We observed that for the top motif, UGCAUG, the relative enrichment in the bound pool relative to the input pool – termed the “R value” (Fig. 1B) – was indeed several-fold higher when using the shorter library (Fig. 2A). However, importantly, the rank order of enriched motifs and the estimated binding affinities were very similar between the two oligo sizes. Therefore, when the signal above background is sufficient, using longer oligo sizes results in little loss of linear motif information, and 40mers obviously contain a much greater diversity of secondary structures than 10mers, so there is potential to learn more.

Gel purification of in vitro transcribed RNA—The gel is removed from the glass plates and placed on a fluor coated TLC plate that is wrapped in saran wrap. The band is visualized with a handheld UV light source and excised with a clean razor blade. In preparation for band isolation, use a clean 18 or 22 gauge needle to puncture a hole in the bottom of a 0.5 ml eppendorf tube and then place into a larger 1.5 ml eppendorf tube (gel-shredder). After isolating the band, place it into the gel-shredder, and spin the gel slices at room temperature at max speed for 10 minutes. Shredded gel pieces should be pelleted at the bottom of the 1.5 ml tube. Remove and discard the smaller tube. Resuspend shredded gel pieces in 400 ul of gel extraction buffer. Extract the RNA by rotating at 4 degrees overnight. The following day, pipet gel and buffer solution into a microcentrifuge filtration tube and remove the gel pieces by spinning the mixture at 14,000 × g for 2 minutes. Place the flow-through into a clean 1.5 mL tube and add 400 ul of isopropanol. Place the mixture on dry ice for 20 minutes and spin for 25 minutes at 4° C at max speed on a tabletop centrifuge. Remove ethanol, and wash the pellet with cold 70% ethanol.

Purified RBP QC—A purified RBP with an affinity tag is a requirement for the assay. A tandem affinity glutathione S-transferase/streptavidin-binding peptide (GST-SBP) fusion tag is used in our system, though other tags can be used for this assay. We express the RBP in a

bacterial system and purify it via the GST tag. Protein purification will not be covered due to the complex and variable nature of the process (see (“Guide to Protein Purification, Second Edition,” 2009)). The SBP tag is then used for protein capture in RBNS selection steps. Prior to performing RBNS, the RBP’s affinity for RNA containing a known or putative binding site is determined by EMSA. This is an important step for confirming protein function and also to determine suitable protein concentrations to use for RBNS. For cases where no putative binding site is known for an RBP, we generally assume functionality if the purified protein is full-length and soluble.

RBNS assay

Prepare Buffers

<i>Binding Buffer</i> *	25mM tris pH 7.5, 150mM KCl, 3mM MgCl2, 0.01% tween, 1 mg/mL BSA, 1 mM DTT
<i>Wash Buffer</i>	25mM tris pH 7.5, 150 mM KCl, 60 ug/mL BSA, 0.5 mM EDTA, 0.01% tween
<i>Elution Buffer</i>	10mM tris pH 7.0, 1mM EDTA, 1%SDS
<i>Gel Extraction Buffer</i>	10mM tris pH 7.0, 300mM NaCl, 2mM EDTA
<i>TE</i>	10mM tris pH 7.0, 1mM EDTA

* 30 ug/mL poly I/C can optionally be added to the binding buffer as a non-specific competitor

Protein equilibration—Identical binding reactions are set up in parallel, each containing a different RBP concentration. Typical concentrations used in our lab are 0 nM, 5 nM, 20 nM, 80 nM, 320 nM and 1300 nM. These concentrations capture the range of binding affinities of most studied RBPs. If the RBP is expected to have a very high or low affinity for its binding substrates the concentrations should be varied accordingly. As the protein concentration increases, R values will initially increase, reach a maximum at a moderate protein concentration and then steadily decrease until reaching background levels, for reasons that we have outlined elsewhere (Lambert et al., 2014). Such a unimodal R value curve is observed in both simulated and actual RBNS data. It is important to select a wide range of RBP concentrations to capture the concentrations with maximum information, as each protein shows a different binding profile. Two negative control conditions are typically performed, a no protein RBNS condition and sequencing of the input RNA pool. Sequencing the input pool controls for biases in nucleotide composition in the input RNA pool and the no protein control can detect potential biases in binding to the apparatus. The protein concentrations are assembled in 250 ul of Binding Buffer and placed on a rotator for 30 minutes at a selected temperature.

Temperature considerations—Protein equilibration, RNA binding and RBP pull-down are performed at a single chosen temperature. To systematically explore the influence of temperature on the binding profile of RBFOX2, RBNS was performed at room temperature, 30° and 37° C (Fig. 2B). Temperature had modest effects on the enrichment of RBFOX’s canonical and secondary motif, though the influence of temperature on binding may be somewhat RBP-specific. A slight increase in R values was observed at higher temperatures, likely because RNA secondary structure is reduced, increasing the accessibility of binding sites.

Prepare beads—After protein equilibration reactions are assembled, streptavidin magnetic beads (Invitrogen) are washed and prepared for pulldown of the SBP-tagged RBP. For each binding reaction, 75 μ l of streptavidin bead slurry is used. Sufficient bead slurry for all binding reactions is placed in a 1.5 ml Eppendorf tube. The tube is placed on a standing magnet until all visible magnetic beads are pulled out of solution and the supernatant is removed and discarded. The beads are washed 3 times with 1 ml of wash buffer, discarding the wash buffer after each wash step. The beads are then resuspended in Binding Buffer (the same volume initially removed from vial) along with Supersin (0.5 μ l/reaction). The beads are then placed on a rotator until needed. As many beads are not guaranteed RNase-free, the equilibration of the beads with RNase inhibitors inactivates RNases that maybe present. The inclusion of BSA will in theory coat non-specific/”sticky” components of the binding apparatus.

RNA binding—Following protein equilibration the RNA pool is added to each of the equilibrated protein mixtures to a final concentration of 1 μ M, along with 2 μ l of Supersin. Each mixture is then placed back on the rotator at the temperature of choice and incubated for 1 hour.

The final concentration of RNA used in each binding reaction can be varied based on experimenter’s wishes and reagent availability. To test the influence of input RNA concentration on the measurement of binding affinities, RBFOX2 RBNS experiments were performed with either 10 nM, 100 nM or 1000 nM total RNA concentrations (Fig. 2C). The consensus motif was clearly detected in all tested conditions. However, higher input RNA concentrations yielded elevated R values for all concentrations tested. For instance, at a low RBFOX concentration (2 nM) there was a 3-fold increase in R value for a top motif when the input RNA concentration was increased from 10 to 1000 nM (Fig. 2C). A possible explanation for this effect is that the top RNA motifs may be readily saturated when total RNA concentration is low, limiting the dynamic range of enrichment values.

RBP pull-down—After protein/RNA binding, the tagged protein is affinity selected with the washed magnetic beads. In preparation for RBP selection, the beads are aliquoted into separate labeled tubes (75 μ l/tube). The buffer is removed from the beads and the RNA/protein mixture is directly added to the beads. The bead/RNA/RBP solution is placed back onto the rotator for 1 hour.

RBP elution—After the RBP mixture is incubated with streptavidin beads, the beads are pulled out of solution, washed and the bound material eluted. First, the beads are placed on the magnetic strip and the supernatant is carefully removed and discarded. Then the beads are resuspended in 1 ml of wash buffer, placed back on the standing magnet and the wash buffer discarded. The number and stringency of the wash steps have a considerable impact on the results. Highest R values for RBFOX were seen when the wash buffer contained low salts (data not shown). We typically do a single wash step in an attempt to capture moderate affinity binding sites and maintain steady-state-like conditions.

After the beads are washed, the RBP and bound RNA are eluted. Immediately after washing the beads, 100 microliters of elution buffer is added to the beads. The beads are then heated

at 70° C for 10 minutes and the supernatant collected and placed in a clean tube. Bound RNA is then purified away from the RBP by phenol/chloroform isolation and ethanol precipitation (see above). The resulting pellet is then resuspended in 10 µl of TE.

Quantify concentration of RBP bound RNA (for affinity estimations)—If RBNS data will be used to estimate binding affinities, the amount of recovered RNA from each RBNS concentration must be quantified. Assuming, all tagged RBP was pulled out of solution and that all RBP-bound RNA was subsequently purified, the concentration of RBPs bound to RNA can be estimated. Quantitation of bound RNA can be performed by running 20% of the purified RBP-bound RNA on a Bioanalyzer (Agilent), according to manufacturer's instructions. If the peak corresponding to the expected library size is quantified, the concentration of RBP bound RNA can be estimated by controlling for the amount loaded onto the Bioanalyzer and the volume of the initial binding reaction. This parameter can also be measured via quantitative RT-PCR.

Sequencing library preparation—Library preparation steps include reverse-transcription and PCR. Half of the isolated RNA from each RBNS condition is input into a 20 µl Superscript III reverse transcription reaction (Invitrogen), with 2 pmol of RT primer. This reaction is incubated at 55° C for 1 hour. An additional RT reaction is conducted with 0.5 pmol of the input RNA pool as the RT template.

After reverse transcription, the resulting cDNA is amplified via PCR. Between 1–5 µl of cDNA is added into a 25 µl PCR reaction. The input quantity must be empirically determined because it is influenced by the amount of RNA bound by the protein and the amount of co-eluting non-specific RNA bound to the apparatus. The required amount of cDNA input into the PCR typically decreases as the RBP concentration used increases. As a starting point, use 1 µl of cDNA into each PCR reaction and perform 8–10 PCR cycles. A different barcoded PCR primer is used to amplify each RBNS condition for downstream sample multiplexing (Table 1).

After PCR, the sequencing library is size-selected to exclude primer dimers and PCR artifacts. Loading dye is added to each reaction and then loaded onto a 8% TBE gel (alternating with empty lanes) and run at 200 V for about 45 minutes. The gels are stained with SYBR gold for ~5 minutes. The sharp band at the expected size (at ~180 bp if using the 40mer pool) is excised with a clean razor and gel purified. Follow the same procedure outlined above except elute the DNA from the shredded gel pieces at 65° C for 1 hour. The pellet is resuspended in 10 µl of TE, samples are deep sequenced (Illumina single-end 40 nt, multiplexed to give ~20 million reads per concentration). PCR samples corresponding to higher RBP concentration conditions may need to be repeated due to over-cycling. Over-cycling is apparent from the appearance of PCR products with higher than expected molecular weights and, in extreme cases, the concomitant depletion of the correctly sized PCR product. This can be remedied by decreasing the number of PCR cycles and/or by decreasing the amount of cDNA used in the PCR reaction. Conditions with no or low RBP concentration may have very faint or even no PCR bands under initially tested conditions. In this case, the input cDNA quantity and/or the number of PCR cycles should be increased.

Variations on input RNA pool composition—The RNA pool can be tailored to answer different questions about an RBP's binding affinity.

- i. *Randomized RNA* can be used for an unbiased motif inquiry. This approach tests the greatest diversity of RNA sequences, yielding unbiased analysis of RNA affinity.
- ii. *Pre-designed sequences* can be used to resolve binding affinities to CLIP clusters, orthologous sequences, mutated potential binding sites, etc. Thousands of DNA oligos with a T7 promoter can be synthesized on a custom microarray and the transcribed product used as input for RBNS.
- iii. *Fragmented mRNA* could be size selected and used as a RBNS input, limiting potential binding measurements to expressed RNA in a given cell type and to constrain binding to naturally occurring sequences.

The input RNA source should be tailored to the experimenter's interest, as the results will yield different types of information. An RBP's absolute biophysical preferences will be better assayed using randomized RNA, where all possible k mers ~ 10 nts are represented many times. The use of pre-designed oligos or mRNA – which will typically have much lower diversity than random 40mers – can yield an RBP's affinity for an extended oligo or RNA fragment rather than simply a motif. However, depending on the complexity of the input pool, binding motifs may be apparent, but more difficult to comprehensively survey.

Computational Analysis—The computational analysis of RBNS data requires a specific set of software tools. We have organized the necessary computational tools for a basic analysis of an RBNS experiment into an open-sourced pipeline (https://github.com/alexrson/rbns_pipeline) available for general use. This section details the computational steps necessary for analyzing RBNS data, the usage of our computational pipeline implementing these steps, the sequence data format for input, and interpretation of the pipeline's output. For an RBNS experiment done on an RBP, the pipeline performs several analyses. The flow of the analyses of the data is illustrated in Figure 3. Starting from a raw multiplexed fastQ file the pipeline can perform the following calculations, some of which are optional.

1. Demultiplex the reads by barcode
2. Count the k mers for each barcode for each user indicated length of k
3. Perform QC checks of the background and no input experiments
4. Calculate the enrichment (RBNS R) of each k mer over background
5. Calculate the fraction of reads in which each k mer is present
6. Run the streaming k mer assignment algorithm (SKA) over the data
7. Analyze RNA structural preferences of the RBP
8. Calculating binding affinities (K_d values) for k mers which show moderate to strong binding
9. Create various plots for the above analyses

As certain steps in the pipeline may prove especially lengthy for large data sets, the pipeline may be configured for parallelization across a compute cluster.

The pipeline can be downloaded and installed on a UNIX or Linux computer by following the instructions at https://github.com/alexrsn/rbns_pipeline.

If you wish to perform structural calculations, you will need to install the Vienna RNAfold package, following the instructions on the providers' website: <http://www.tbi.univie.ac.at/RNA/index.html#download>.

Once the package and its dependencies have been installed, the RBNS pipeline can be run from source on a dataset. The optional flags for the pipeline are listed in Table 2. Running the main script with the `-h` flag will display the help text for the RBNS pipeline along with the current options and settings (\$ indicates the shell prompt):

```
$ ./rbns_main.py -h
```

Pipeline Inputs—The pipeline requires two input files: the raw multiplexed sequence data for the experiment and a settings file detailing the settings for the experiment. In general, multiplexing all the concentrations needed for an RBNS experiment in a single Illumina HiSeq lane or NextSeq run provides adequate depth for analysis of k mers up to about 10 bases in size, so the pipeline assumes that all the data for a factor are present in a single lane. For example, we typically run libraries for two proteins, with 5 non-zero protein concentrations each, plus a zero protein control (total of 11 libraries), in a single HiSeq 2000 lane, which typically yields over 200 million short reads. Given 11 multiplexed libraries one can expect about 20 million reads per library, which is adequate for most analyses. Probing binding to k mers of greater length requires exponentially greater read depth to obtain sufficient sampling since the number of distinct k mers grows as 4^k . For practical reasons we choose to use short reads, typically of length 20 to 40 nt.

The settings are specified in a settings file that includes such information as the concentrations of the RBP in each experiment, the barcodes associated with each concentration, the desired values of k to be analyzed, and information to show in the outputted graphs. Several example settings files are included in the RBNS pipeline package, and the settings are summarized in Table 2.

k mer counting methods—The central question of a RBNS experiment is usually: what RNA sequences does the protein bind to? Answering this question quantitatively requires a couple of algorithms, implemented in the pipeline.

Naïve method—The naïve method of quantifying k mer enrichment is intended to give a simple, straightforward estimate of the level of enrichment for all k mers in the library over the input RNA pool. The naïve method simply counts the number of occurrences of each of the 4^k k mers over all the reads in each RBNS condition. These counts are then normalized by dividing by the total count of all k mers to give the k mer frequency (fraction of total counts represented by the k mer). If the input pool of oligonucleotides were perfectly

uniformly distributed across k mers, then these fractions would provide useful measures as is. However, since some biases are always present in the randomly synthesized input pool, these biases are corrected for by performing the same counting procedure on the input pool and then calculating the ratio of each k mer's frequency in the selected pool to its frequency in the input pool. This gives the simple enrichment ratio, or RBNS R value, which is a good quantitative measure of a protein's relative affinity for different k mers at a given protein concentration.

By default, the RBNS pipeline performs the naïve count and RBNS R calculations for all k mers of specific sizes and outputs these to tables. It also optionally creates graphs of R values for k mers of interest as a function of protein concentration, among other analyses.

Calculating B values—One problem with R values is that there is an upper limit on the maximum possible R value that depends on the size of the k mer relative to the oligonucleotide size used in the experiment (so R values underestimate affinity differences). This is because all of the k mers present in the bound oligonucleotide are counted equally when measuring R values, though the protein might have bound to only one. RBNS R values of “offset” motifs (motifs that overlap a bound motif extensively, e.g., at $k-1$ positions) will also be enriched, even if they are not sufficient to confer binding. By making a few simple assumptions, a value called B can be calculated from the R value that provides a more accurate scale for relative binding affinity.

Take the following simple model of an RBP binding to a heterogeneous RNA population. A RBP binds with high affinity to only one k mer and with a low and constant affinity for all other k mers. Suppose that in the binding equilibrium for a given concentration of RBP the protein is B times more likely to bind its preferred k mer than to any other. However, there are 4^{k-1} times as many low-binding k mers, assuming a uniform pool of oligonucleotides. Also, for every high affinity k mer pulled down, $\sim \lambda - k$ other k mers will also be recovered, where λ is the length of the oligo. From these assumptions one can calculate the value of R in terms of B as:

$$R \approx 1 + (4^k B / ((B + 4^k) (\lambda - k + 1)))$$

One can then solve for B in terms of R. The full derivation is provided in supplementary information of (Lambert et al., 2014).

$$B \approx (4^k - 1) (R - 1) (\lambda - k + 1) / (4^k + \lambda - k - R(\lambda - k + 1))$$

As R approaches its theoretical upper limit, the denominator becomes small and B becomes large. For example, RBFOX2 had an R value of 22 for the 6mer UGCAUG in an experiment using a random 40mer pool with 365 nM concentration of RBFOX2. Applying the equation above yields a B value of ~ 900 , implying ~ 900 -fold higher binding of RBFOX2 to UGCAUG than to nonspecific 6mers under the tested conditions.

Streaming *k*mer Assignment—There are several reasons why R values are not completely satisfactory from a theoretical perspective. First, they are highly dependent on the protein concentration: typically they increase with protein concentration, reach a maximum, and then decline. Second, they are related to but not directly proportional to the actual binding affinities of motifs. Third, a *k*mer that merely overlaps with a strong binding *k*mer will often have an elevated R value even if it does not bind to the protein of interest by itself. For these reasons and others we seek to assign binding directly to different *k*mers using a computational approach called streaming *k*mer assignment or SKA (Fig. 4). A fuller description of how this is accounted for in the equilibrium modeling is in the next section.

The gist of SKA is that every read in a selected library is fractionally assigned to the *k*mers present in its sequence based on the continually updated knowledge of the weight of each *k*mer in the library. This algorithm is based on the assumption that each read in the library is present because a protein molecule bound to one of the *k*mers located within the bound oligonucleotide. Initially, we do not know which *k*mer was specifically bound by the RBP so we equally credit each of the *k*mers present in the read. However, as we process more and more *k*mers we achieve better and better knowledge of which *k*mers are strongly binding and assign a larger and larger fraction of each read's weight to the strongest *k*mers present within the read's sequence. The set of weights for all *k*mers converges to give the bound library fraction, the fraction of all protein-RNA binding sites that are located at each *k*mer sequence. This is a fairly important quantity as it gives insight into the equilibrium makeup of the protein and the set of all *k*mer sequences. What follows is a much more in-depth description of the SKA algorithm and its implementation details in the RBNS pipeline.

There are three steps in SKA:

1. Initialization
2. First ("rough") pass through reads, and
3. Subsequent ("refining") passes through the reads.

Each step generates increasingly accurate estimates of the proportion of protein that is bound to each *k*mer sequence. We call our current estimate of these proportions the *k*mers' weights. As the algorithm progresses, the weights converge. Initially, prior to looking at the data, all weights are initialized to a constant value. In our implementation we choose to initialize with a starting value of 1, equivalent to saying that the *k*mer is responsible for binding of one read.

The second step, the rough pass, steps through each of the bound reads and fractionally assigns its weight to the *k*mers present in its sequence, in proportion to their current weights. So for each read we identify all the *k*mers present in its sequence. We then sum up the current weights of each of the *k*mers. The weights of these *k*mers are then increased by their current weight normalized by the sum of the weights of all the *k*mers present in the read. Thus, for example, if there are 10 *k*mers present in single copy in the first read, then each of these *k*mers' weights will be incremented equally by 1/10 to a value of 1.1. For the second read, however, if one of the *k*mers present in the read's sequence was also present in the first then it will be assigned a larger fraction of the second read's weight: 1.1/10.1,

whereas each of the other k mers will be incremented by only $1/10$. Through these steps over millions of reads, one gets a good idea of the binding fraction for each k mer. However, the fractional assignments are biased by the initialization of equal weight to all k mers, especially for the early reads. For this reason we do a subsequent pass or passes through the reads.

In subsequent passes the weights are more finely honed. In each subsequent pass after the first the reads are fractionally assigned to k mers using the weights from the end of the previous pass. These fractional assignments are in turn used to fractionally assign reads in the next pass and so on until convergence. After convergence, the counts are normalized by the total number of reads for that library to give the fraction of bound RNA attributable to each k mer. When analyzing RBFOX2 at a concentration of 370 nM binding to 6mers we observe that 8.2% of RNA is attributable to binding to UGCAUG followed by other GCAUG- and GCACG-containing 6mers. Note that this does not directly correspond to the expected fraction bound *in vivo*, since the pool of RNA differs in composition (e.g., GCACG is rarer than GCAUG, in part because it contains a hypermutable CpG dinucleotide).

The RBNS pipeline optionally runs SKA on all protein-selected libraries for specified values of k . We chose to analyze binding of 5mers, 6mers and 7mers of RBFOX2 using SKA. Running the algorithm is fairly computationally intensive and is ideally performed on a computer cluster, but the burden may be reduced by reducing the number of passes SKA makes through the reads. In general, SKA converges in fewer passes when the library is strongly biased (because the protein has high specificity at the given concentration) or when there are a large total number of reads in the library (> 10 million). For RBFOX2 data with ~ 20 million reads per library we found that the algorithm fully converged after only two passes. Larger numbers of passes may be required for proteins with lower sequence specificity or for datasets with lower coverage.

Presence Fractions—The SKA algorithm provides an estimate of the proportion of reads have each k mer within their sequence. This should be run on the input read sequences. Thus, while SKA gives values proportional to the *bound* concentrations of each k mer, the presence fraction indicates what proportion of reads contain each k mer, which is then in turn directly proportional to the *total* concentration of the k mer at equilibrium.

The presence fractions can be easily determined by normalizing the naïve counts for each k mer by the total number of reads in the input library. Note that the calculation of R values accounts for biases present in the input library, but values calculated by SKA do not control for these biases.

RNA-binding equilibrium—Throughout this chapter it is worthwhile to bear in mind the following model of how the selection step of RBNS works, specifically with regard to the binding equilibrium.

In the experiment, an RBP is present and RNA oligos of length λ . The protein is assumed to bind RNA sequences of length k ($k < \lambda$) with a certain equilibrium constant specific to each k mer. Binding and dissociation between the protein and the 4^k possible k mers are assumed

to be in simultaneous equilibrium, giving a system of 4^k dissociation equilibrium equations of the form:

$$K_{d(kmer)} = [RBP][free\ kmer] / [RBP:kmer]$$

During RBP capture, we magnetically select sequences containing a bound $kmer$ (corresponding to the denominators of the binding equilibria) and perform sequencing. The challenge then is to estimate from the full length read sequences of bound oligos what are the dissociation constants for the bound $kmers$.

Relative Binding Affinity Determination—This section describes the calculations necessary to determine relative binding affinities for bound $kmers$. Relative affinities can be estimated accurately over several orders of magnitude, but weakly bound motifs have higher signal to noise and cannot be estimated accurately. In practice, we do not consider motifs that bind with affinity more than $1000\times$ lower than the strongest binding motif.

We define the relative binding constant of a $kmer$ as the ratio of the $kmer$'s dissociation constant to the \ dissociation constant of the highest affinity $kmer$. This can be thought of as how many fold weaker the binding is if all else (accessibility, copy number, sequence context) is equal. The relative binding constant is expressible in terms of quantities that can be calculated from RBNS data:

$$K_{d,relative} = [L_{kmer,free}] / [L_{best,free}] \times [RL_{best}] / [RL_{kmer}]$$

In order to calculate dissociation constants for individual $kmers$, one must first calculate the concentrations of bound $kmer$ and unbound $kmer$ for the strongest binding $kmer$ as well as other strong binders. This is done as follows. In the binding reaction, the total concentration of a $kmer$ is the concentration of oligos that contain that $kmer$. We do not double count multiple occurrences in a single oligo, because we assume binding to one site is sufficient to ensure binding. Therefore the total concentration of $kmer$ in solution is the concentration of the oligo times the fraction of oligos that contain each $kmer$. This fraction is simply the $kmer$ presence fraction, which is calculated by the pipeline.

One also needs to know the total concentration of bound RNA oligo in the binding equilibrium step, which can be measured by taking a fluorescent trace after the selection step as described in the experimental section. The concentration of bound $kmer$ is then simply the concentration of bound oligo multiplied by the SKA fraction for that $kmer$.

The concentration of free RBP, which appears in the definition formula for absolute dissociation constants, could be determined by taking the difference between the concentration of bound oligo and the known concentration of protein. However, relative uncertainties in these values are increased in the subtraction operation, and it is usually not known what fraction of protein molecules are properly folded. Therefore, we recommend calculating *relative* dissociation constants, defined as the ratio of the K_d of a $kmer$ to the K_d of the highest affinity $kmer$. In this case the free RBP concentration term cancels out, and the

effects of other systematic biases such as varying proportions of unfolded protein or salt concentrations are also reduced. This allows the relative K_d values to be compared to values from traditional low throughput measurements such as gel shift (EMSA) and SPR. $K_{d,relative}$ values can be calculated from any selected library, in principle, but we find the values most accurate in the library where the enrichment of the strongest binding k mer is maximized. This usually occurs at an intermediate concentration of RBP where the concentration of bound RNA is close to the K_d . Since this information is not known in advance, this is one of the advantages of performing RBNS with multiple concentrations in parallel.

An example calculation for RBFOX is shown in Table 3. The upper portion shows the input information about protein concentration and total RNA:protein complex concentration which is needed to make the intermediate calculations for each k mer. The middle portion of the table shows the calculation using the SKA data for UGCAUGU, the strongest binding 7mer and UGCACGU, an additional 7mer for which we generated SPR data. We show experiments here covering a wide range of RBP concentrations, but in practice a smaller number of experiments covering the middle of the RBP concentration range is sufficient. While the total concentration of RNA oligos containing UGCAUGU is assumed to be constant across experiments, the proportion of complexed UGCAUGU oligos increases to a maximum at 365nM RBFOX. While most of the experiments we have shown have low signal and therefore underestimate $K_{d,relative}$, towards the middle of the concentration range (shaded row) is selected because the signal for this library is maximized. At this concentration the $K_{d,relative}$ value of UGCACGU, 4.13, can be compared with the value calculated by SPR of 6.43. Overall, there is a close correlation between $K_{d,relative}$ measured by RBNS and SPR ($r = 0.94$, $P < 0.001$) (Lambert et al., 2014).

RBNS Quality Control—There are several modes by which RBNS experiments can fail. The following quality control checks should be performed:

1. Some k mers should be significantly enriched above background (assuming there is reason to believe that the protein binds short RNA motifs)
2. RBNS R values of k mers calculated for adjacent (similar) RBP concentrations should be well correlated, particular for k mers with R values that are above background
3. If RBNS R values of top k mers are plotted as a function of protein concentration across a broad range of concentrations, a unimodal curve should be evident

Conclusions

RBNS is a straightforward, fairly cost-effective experiment to quantitatively assess comprehensive sequence and structural specificities of RBPs. The experimental and computational methods outlined here enable identification of bound motifs and estimation of binding constants for an RBP of interest. The iterative binding and amplification used in SELEX approaches identify high affinity motifs, but have little power to measure moderate affinity motifs and contextual binding preferences. We have previously shown that RBNS and CLIP data can provide complementary information. CLIP-seq surveys in vivo binding

sites, but is subject to crosslinking biases, has relatively high background and rarely if ever achieves saturation, limiting the identification of binding sites. RBNS more directly assesses binding specificity, and can be used to distinguish likely true positive and false positive CLIP motifs.

Many variations can be introduced to tailor RBNS to the RBP or scientific question of interest. For example, potential synergistic or competitive binding can be assayed by performing RBNS in the presence of two recombinant RBPs, one tagged with the epitope used for pulldown in RBNS and the other untagged (N. J. L. and C. B. B., unpublished data). RBNS performed in this manner has the potential to determine contexts where one RBP can displace another or enhance its binding. As discussed above, the input RNA in RBNS can also be varied. Instead of randomized RNA, fragmented mRNA or in vitro transcribed RNA from custom array-synthesized oligonucleotides can be used in the binding reaction. This seemingly minor variation substantially alters the interpretation of the results; while the reduced sequence complexity relative to random sequences may decrease the precision of motif identification, it can also enable assessment of affinity relative to specific segments of the transcriptome, increasing the dynamic range of R values observed and potentially enabling assessment of subtle features such as cooperative binding (N. J. L. and C. B. B., unpublished data). The computational pipeline for analysis is under active development. Current efforts are aimed at automatic inference of the optimal *k*mer size, of the number of distinct motif classes bound by the RBP, and of the influence of RNA secondary structure on binding. Updated versions of the pipeline will be posted at the github site referenced above.

References

- Bass BL, Nishikura K, Keller W, Seeburg PH, Emeson RB, O'Connell MA, ... Herbert A. A standardized nomenclature for adenosine deaminases that act on RNA. *RNA*. 1997; 3(9):947–949. [PubMed: 9292492]
- Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol*. 2014; 32(6):562–568. DOI: 10.1038/nbt.2880 [PubMed: 24727714]
- Daubner GM, Clery A, Jayne S, Stevenin J, Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J*. 2012; 31(1):162–174. DOI: 10.1038/emboj.2011.367 [PubMed: 22002536]
- Guide to Protein Purification. *Guide to Protein Purification* (1). 2009; 463:1–851.
- Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell*. 2014; 54(5):887–900. DOI: 10.1016/j.molcel.2014.04.016 [PubMed: 24837674]
- Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods*. 2012; 9(12):1192–1194. DOI: 10.1038/nmeth.2225 [PubMed: 23142872]
- Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, ... Hughes TR. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol*. 2009; 27(7):667–670. DOI: 10.1038/nbt.1550 [PubMed: 19561594]
- Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, ... Hughes TR. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*. 2013; 499(7457):172–177. DOI: 10.1038/nature12311 [PubMed: 23846655]

- Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M, Ule J. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 2012; 13(8):R67.doi: 10.1186/gb-2012-13-8-r67 [PubMed: 22863408]
- Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, Lis JT. Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat Methods.* 2014; 11(6): 683–688. DOI: 10.1038/nmeth.2970 [PubMed: 24809628]
- Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science.* 1990; 249(4968):505–510. [PubMed: 2200121]
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science.* 2003; 302(5648):1212–1215. DOI: 10.1126/science.1090095 [PubMed: 14615540]

Figure 1. Overview of RBNS method and R value calculation

A) RBNS method overview: several individual binding reactions are performed in parallel (typically 5–10), each using a different RBP concentration. RNAs containing high affinity or low affinity sites are shown.

B) Calculating enrichment (“R”) values from raw RBNS data.

Figure 2. Effects of input oligonucleotide length, temperature, and input RNA concentration on RBNS R values

A) R value of primary 5mer motif, GCATG, and secondary motif, GCACG, for RBFOX2 using input oligonucleotide pools of length 10 nt and 40 (both at 1 μ M total RNA concentration).

B) R value of top 6mer motif, TGCATG, for RBFOX2 at different protein concentrations and temperatures (RT = room temperature).

C) R value of top 6mer motif, TGCATG, at different protein concentrations (indicated by color) and different total input RNA concentrations (indicated below).

Figure 3. Detailed flow chart of the RBNS computational pipeline. Pipeline function names are followed by parentheses, with the paths to the associated files indicated in parentheses.

A. Initialize weights (w) of all 5mers

$$w_{AAAAA} = 1, w_{AAAAC} = 1, w_{AAAAG} = 1, \dots$$

B. Process Read #1

	initial weight	Read #1 relative weight	updated weight
AUGCAUGANNNNNN			
AUGCA.....	1	1/10	1.1
.UGCAU.....	1	1/10	1.1
..GCAUG.....	1	1/10	1.1
...CAUGA.....	1	1/10	1.1
....AUGAN.....	1	1/10	1.1

+ 5 more

C. Process Read #2

	initial weight	Read #2 relative weight	updated weight
NNAGCAUGGGNNNN			
NNAGC.....	1	1/10.1	1.099
.NAGCA.....	1	1/10.1	1.099
..AGCAU.....	1	1/10.1	1.099
...GCAUG.....	1	1.1/10.1	1.21
....CAUGG.....	1	1/10.1	1.099
.....AUGGG.....	1	1/10.1	1.099

+ 4 more

Figure 4. Schematic of streaming kmer assignment (SKA) algorithm

Illustrates how passes through the sequence reads are used to update the weights of kmers in the SKA algorithm.

Table 1

Primer sequences used in RBNS

Primer Name	Sequence
T7 promoter oligo	TAATACGACTCACTATAGGG
RBNS T7 template	CCTTGACACCCGAGAATTCCA-(N)40-GATCGTCGGACTGTAGAACTCCCTATAGTGAGTCGTATTA
RT primer	GCCTTGGCACCCGAGAATTCCA
RNA PCR (RP1)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGACGATC
indexed PCR primer	CAAGCAGAAGACGGCATACGAGAT-barcode-GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

Table 2

Parameter settings used in RBNS pipeline

Setting name	Description
fastq	The raw sequence data in fastQ format must be specified with a path to the file. It may optionally be compressed in gzip format. The multiplexed-sequencing barcodes are expected to be in the read identifiers as is usual with Illumina output.
mismatches allowed in barcode	The number of mismatches allowed in barcode (maximal Hamming distance from the known barcode) should be specified. This should be 1 in most cases, but possibly 0 if the barcodes are especially short (<5nt) or 2 if they are especially long (>10nt).
barcodes	The barcodes relevant to an experiment must be specified. Since multiple may be run within a single HiSeq lane, only the barcodes for the protein of interest and the input RNA should be specified here. The order of the barcodes specifies the order of certain output plots. The barcode of the library must be specified specifically as well and included in this list.
read_len	The read length should be specified as the length of the randomized region of RNA used for the Bind-n-Seq experiment.
3' trim	If the actual read in the input fastQ is not the same length as the tested region, which end to trim should be specified with trim 3p and trim 5p. This may be the case, for example, when the size of the randomized oligo is smaller than the total read length and the sequencer reports the sequence of the oligo and primers concatenated.
Experiment name	An experiment name is specified to indicate label plots and the output directory of the whole run.
protein name	The name of the protein should be specified in the desired format for labeling of plots.
Concentrations	The concentrations of protein in the Bind-n-Seq experiment should be specified in nanomolar units in the same order as the list of barcodes.
Inhibitor concentration	The inhibitor (e.g. poly-IC) concentration (if any), input RNA concentration, number of washes and temperature can also be set in this way. This setting only affects plot labels.
Barcodes for sorting	In order to sort kmers based on affinity one must select the concentration(s) of protein at which to do the sorting. This can be specified. Since it is unknown in advance where the greatest disparity between affinities is, one can refine after doing an initial run of the pipeline.
RNA concentration	The concentration of RNA oligo in the binding reaction in nanomolar units.
Relevant experimental variables	If the exact technical parameters of the experiment are still being worked out, this setting allows the user to specify which experimental variables (such as temperature or number of washes) should be labeled on the generated plots. Generally, these will be the ones that vary between library
washes	Number of washes done for each library. We generally recommend one wash. This setting only affects plot labels.
temperature	The temperature of each binding reaction in the same order as the barcodes are specified.
library sequence barcode	The barcode sequence of the input library.
k for concordance check	The RBNS pipeline does various concordance checks between libraries; the kmer length for doing these checks should be specified.
number of kmers for enrichment humps	The number of top kmers to plot on various default plots can be specified.
naïve count	Doing the naïve count is optional and should be specified.
stream count	Doing the streamed counts aka SKA is optional and should be specified.
presence count	Doing the presence counts is optional and should be specified.
ks for naïve/stream/presence counts	Which values of k to do for each counting method: naïve, SKA and presence. For each counting method the counting will take longer the more values of k that are requested. It is not recommended to do values of k > 10 as the sampling is likely too low.
Force naïve/stream/presence recount	Force the pipeline to recount rather than take the results already generated by a previous pipeline run. This should be done if, for example, the source code is changed. This can be specified individually for each counting method.
max reads to split	For testing purposes, one can do the whole pipeline on a subset of the reads. This can be done by specifying the number of reads one wishes to subsample.

Setting name	Description
known motif	If there is a known primary motif for the protein of interest it should be specified to display it differently in the output plots. If there are other motifs of interest, they may be specified as well.
Motifs of interest	Additional motifs to include in output plots.
Free energy limits	To study the effects of RNA structure on protein binding, the RBNS pipeline optionally includes a simple analysis, binning of reads by free energy of folding. The edges of these free energy bins can be set here. Kmer enrichment is analyzed separately in each of the free energy bins to give insight into the protein's structural preferences.
experiments to compare	The pipeline can also make automated comparisons to other RBNS experiments, such as a homolog of the protein. This can be set by specifying the path to the settings file for the other experiment(s).
Results Directory	The directory containing the output of the pipeline must be set.
Work Directory	A work directory may also be set to store intermediate files during the process of running the pipeline. This directory is written to and read from frequently. Usually, this is the same as the results directory but can be changed to a local mount to tune performance.
Error Directory	If the pipeline runs into an error the directory that it is logged in can be specified here

Table 3

Using SKA to calculate relative binding affinity

Example $K_{d,relative}$ calculations for UGCACGU. In each table the shaded row indicates the RBNS experiment with maximal signal as measured by the maximum R value.

The upper portion of the table shows the experimental data that goes into calculating $K_{d,relative}$ for each RBNS experiment for our example of RBFOX. This includes the concentration of RBP, total concentration of RNA oligo and the estimated total concentration of RNA:protein complex from Bioanalyzer trace.

The middle portion of the table shows calculations for how to use the output of SKA to calculate the ratio of bound to unbound oligo for the top 7mer, UGCAUGU. The streaming fraction is taken from the output of SKA. The *total* concentration of UGCAUGU containing oligo, $[L_{UGCAUGU}]$ is estimated from the input library experiment; it is the product of the presence fraction of the kmer and the total concentration of oligo, $[L_{total}]$. The concentration of *bound* UGCAUGU, $[RL_{UGCAUGU}]$ containing oligo is estimated from the streaming fraction; it is the product of the streaming fraction and $[RL_{total}]$ from A. The final column shows the ratio of free to bound oligo.

The lower portion is similar to the middle portion but for the secondary kmer UGCACGU for which we have measured the K_d via SPR. The final column shows the $K_{d,relative}$ for UGCAUG. It is the ratio of free UGCACGU to bound UGCACGU divided by the ratio of free UGCAUGU to bound UGCAUGU (see formula in text).

Bind N Seq Library	[RBFOX] (nM)	$[L_{total}]$ (nM)	$[RL_{total}]$ (nM)
1	1.5	1000	0.37
2	4.5	1000	0.40
3	14	1000	1.10
4	40.5	1000	0.73
5	121	1000	5.44
6	365	1000	5.92
7	1100	1000	12.92
8	3300	1000	16.97
9	9800	1000	28.92

UGCAUGU			
Bind N Seq Library	Streaming Fraction	L_{UGCAUG} (nM)	RL_{UGCAUG} (nM)
1	6.23E-06	1.4	2.27E-06
2	6.37E-06	1.4	2.56E-06

UGCAUGU				
Bind N Seq Library	Streaming Fraction	L _{UGCAUG} (nM)	RL _{UGCAUG} (nM)	L _{UGCAUGUfree} /RL _{UGCAUGU}
3	8.45E-06	1.4	9.25E-06	1.53E+05
4	5.42E-04	1.4	3.96E-04	3.57E+03
5	8.24E-03	1.4	4.48E-02	3.06E+01
6	1.28E-02	1.4	7.57E-02	1.77E+01
7	3.55E-03	1.4	4.59E-02	2.99E+01
8	1.55E-03	1.4	2.63E-02	5.28E+01
9	1.01E-03	1.4	2.93E-02	4.74E+01

UGCACGU				
Bind N Seq Library	Streaming Fraction	L _{UGCACGU} (nM)	RL _{UGCACGU} (nM)	L _{UGCACGUfree} /RL _{UGCACGU}
1	6.9E-06	1.6	2.5E-06	6.21E+05
2	7.0E-06	1.6	2.8E-06	5.53E+05
3	8.5E-06	1.6	9.3E-06	1.67E+05
4	1.2E-05	1.6	8.7E-06	1.79E+05
5	7.82E-04	1.6	4.25E-03	3.65E+02
6	3.5E-03	1.6	2.1E-02	7.32E+01
7	2.1E-03	1.6	2.7E-02	5.62E+01
8	1.1E-03	1.6	2.0E-02	7.88E+01
9	9.8E-04	1.6	2.8E-02	5.37E+01

K_{D,relative}