Expanding the DNA-encoded library toolbox: identifying small molecules targeting RNA

Qiuxia Chen [ID][†]**, You Li** [ID][†]**, Chunrong Lin**[†]**, Liu Chen, Hao Luo, Shuai Xia, Chuan Liu, Xuemin Cheng, Chengzhong Liu, Jin Li and Dengfeng Dou** [ID][*]

HitGen Inc., Shuangliu District, Chengdu, China

ABSTRACT

DNA-encoded library (DEL) technology is a powerful tool for small molecule identification in drug discovery, yet the reported DEL selection strategies were applied primarily on protein targets in either purified form or in cellular context. To expand the application of this technology, we employed DEL selection on an RNA target HIV-1 TAR (trans-acting responsive region), but found that the majority of signals were resulted from false positive DNA–RNA binding. We thus developed an optimized selection strategy utilizing RNA patches and competitive elution to minimize unwanted DNA binding, followed by k-mer analysis and motif search to differentiate false positive signal. This optimized strategy resulted in a very clean background in a DEL selection against *Escherichia coli* FMN Riboswitch, and the enriched compounds were determined with double digit nanomolar binding affinity, as well as similar potency in functional FMN competition assay. These results demonstrated the feasibility of small molecule identification against RNA targets using DEL selection. The developed experimental and computational strategy provided a promising opportunity for RNA ligand screening and expanded the application of DEL selection to a much wider context in drug discovery.

INTRODUCTION

DNA-encoded library (DEL) consists of a mixture of small molecules conjugated to unique DNA tags, and the structure information is encoded within DNA sequences to allow screening of billion compounds simultaneously in one single tube. First proposed on 1992 by Brenner and Lerner ([1]), DEL has progressed along the years with the development of combinatorial chemistry and next generation sequencing. With its high throughput, low cost and fast turn over, DEL selection possesses great value for hit identification in drug discovery and has been adopted by major pharmaceutical companies. Drug candidates originated from DEL have entered different stages of clinical trials ([2]).

To date, DEL selection is mostly an affinity-based selection. Typical affinity selection process includes incubation of DEL molecules with the target, washing off the weak or non-specific binders, revealing the enriched structures by subsequent sequencing, and data analysis. Various screening methodologies allowed the expansion of screening objects from recombinant proteins to living cells. For DEL selections performed on purified proteins with affinity tags, solid support including magnetic beads or resin is utilized to immobilize the protein and separate the DEL molecules bound from the bulk DEL solution ([3],[4]). Cell based selection by overexpression of target protein ([5]), labeling the protein with a complementary DNA ([6]), or increasing ligand valence ([7]) enables its performance with cell membrane proteins. Cell penetrating peptide ([8]) and injection ([9]) have been utilized to facilitate the interaction of DEL molecules with cellular proteins. These practices have achieved significant scientific milestones, despite the fact that they were only successful on protein targets while the necessity of screening other target classes starts to draw attention.

The majority of approved drug targets are proteins. However, ∼70% of human genome encodes non-coding RNAs, among which microRNAs and long non-coding RNAs are also validated therapeutic drug targets ([10]). Extending drug targets to RNAs could expand the landscape of drug discovery industry significantly. In addition, some proteins are considered undruggable as they are intrinsically disordered or lack typical small molecule binding pockets, while targeting the structured RNA motifs of their corresponding mRNAs provides a promising strategy to modulate protein expression levels involved in diseases ([11]). Also, targeting viral or bacterial RNAs offers opportunities to modulate their biological processes and facilitate the development of antiviral or antibacterial drugs.

[*]To whom correspondence should be addressed. Tel: +86 28 85197385 8700; Email: df.dou@hitgen.com
[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.
Present address: Lead Generation Unit, HitGen Inc., No. 8, Huigu 1st East Road, Chengdu 610200, China.

As targeting RNA is emerging as an attractive strategy in drug discovery practices, RNA targeted ligands have been developed and entered different stages in drug discovery. Risdiplam targeting SMN2 pre-mRNA splicing for the treatment of spinal muscular atrophy (SMA) (12) has been approved by FDA. Multiple screening methodologies including high-throughput screening (13), structure based design (14), fragment based screening (15), integrated statistical and computational approaches (16–18) have been utilized for RNA targeting small molecule identification. However, DEL selection on RNA targets has not been detailed in peer-reviewed papers. In fact, most reported DEL selection strategies are applied on protein targets, except in one case of selection on DNA target G-Quartets, which consists of specific guanine rich repeats with a unique four-stranded structure and is thermodynamically stable (19). Another paper (20) mentioned the utilization of Vipergen yoctoReactor DEL, which is very different from the DEL in our case as well as the ones investigated in literature, for discovering a ligand of Aptamer 21. More importantly, the selection details and performance were not revealed.

In this paper, we intend to demonstrate the feasibility of DEL selections against RNA targets. Whether our DEL and conventional selection strategy are compatible with RNA targets is the first question that needs to be addressed. As all the DEL molecules are attached to DNA tags, would the DNA tags interfere with RNA-small molecule interaction? To answer this question, we chose human immunodeficiency virus type-1 (HIV-1) trans-acting responsive region (TAR) RNA, which is a validated anti-viral RNA target, to test the feasibility of DEL selection. Significant false-positive signals due to DNA–RNA interaction were observed which extensively complicated data analysis. To solve this technical issue, we developed an algorithm to statistically differentiate DNA–RNA binding signals from small molecule-RNA binding. In order to minimize the DNA–RNA interaction during selection, we designed and utilized RNA patches in combination with competitive elution during the DEL selection. The above strategy allowed us to successfully identify the signal of positive control compound during DEL selection, with a significant reduction in false-positive signals.

With this novel methodology described above, we employed the optimized selection process on an anti-bacterial RNA target, *Escherichia coli* (*E.coli*) flavin mononucleotide (FMN) Riboswitch, to further validate our selection strategy. Consequently, compounds synthesized from this selection were identified as double digit nanomolar hits by functional assay and isothermal titration calorimetry (ITC). We further compared the performance of DNA patches and found it was slightly less effective than RNA patches, but still significantly better than selection without patches. Taken together, we have developed a selection strategy as well as an analytical algorithm to successfully apply DEL selection on RNA targets. This approach is proven robust with positive binders identified in FMN Riboswitch selection, and should be feasible to employ on other biological significant RNA drug targets.

MATERIALS AND METHODS

Materials and RNA preparation

3′Biotin tagged TAR RNA (5′-GGCAGAUCUGAGCCU GGGAGCUCUCUGCC-3′ biotin), 5-FAM and 5′ propargylglycine tat peptide (5′ pra-AAARKKRRQRRRAAA, Pra-tat, propargylglycine is introduced for DNA conjugation) (21) were synthesized by GenScript.

E.coli FMN Riboswitch with and without poly A were prepared by in vitro transcription. *E.coli* FMN Riboswitch sequence with poly A (5′TAATACGACT CACTATAGGGCTTATTCTCAGGGCGGGGCGAA ATTCCCCACCGGCGGTAAATCAACTCAGTTGA AAGCCCGCGAGCGCTTTGGGTGCGAACTCA AAGGACAGCAGATCCGGTGTAATTCCGGGG CCGACGGTTAGAGTCCGGATGGGAGAGAGT AACGAAAAAAAAAAAAAAAAAAAAAAAAAA-3′) were introduced to pcDNA3.4 at clone site SacI and XbaI. DNA templates were prepared by PCR from the plasmid using forward primer (TAATACGACTCAC-TATAGGG) incorporating the T7 promoter and three different reverse primers (R1 CGTTACTCTCTCC-CATCCG, R2 TTTTTTTTTTTTTTTTTTTTTTTTTTT or R3 TTTTTTTTTTTTTTTTTTTTTTTTTTTCGT-TACTCTCTCCCA). In vitro transcription was carried out by using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific K0441) and purified by VAHTS RNA Clean Beads (Vazyme N412-02) following the protocol from the manufacturer. FMN Riboswitch was prepared in 50 mM Tris, 100 mM KCl pH 7.4 buffer and annealed each time before use by heating at 95 °C for 5 min and put at RT (room temperature) for at least 15 min.

RNA immobilization test

Fixed amount of RNA was incubated with various volumes of beads in selection buffer without ssDNA (sheared salmon sperm DNA) at RT for 30 min, then the flow through (FL) was collected, the beads were rinsed and transferred in a new tube with selection buffer. The FL and beads were then loaded on Urea gels (Invitrogen EC6875BOX), analyzed by electrophoresis in TBE buffer and further visualized by staining with SYBR Gold nucleic acid stain (ThermoFisher S11494).

Fluorescent ligand binding on immobilized RNA

The binding of fluorescent tool compound on immobilized RNA was performed by incubation of fluorescent tool compounds (FAM labeled tat peptide or FMN) with immobilized TAR or FMN Riboswitch for 1 h at RT. The flow through (FL) was collected, then the immobilized RNA was washed 5 times by corresponding selection buffer for 1 min each time, beads were either heated at 100 °C for 10 min or incubated with 50 μM Pra-tat or 100 μM Ribocil-C for competitive elution. The fluorescence of each portion was detected by microplate reader at Ex 485 nm/Em 520 nm for FAM labeled tat peptide and Ex 455 nm/Em 525 nm for FMN.

DNA conjugate synthesis and binding test on immobilized RNA

The DNA conjugate was synthesized by coupling the pratat peptide to the short DNA linker by click chemistry, after the conjugation was confirmed by mass spec, the ligand was purified by HPLC and further ligated to the longer DNA tags to make the full-length DNA conjugate, a control DNA tag without coupling to any compound was also made. The concentration of the conjugate was quantified by qPCR and purity was determined by Agilent Bioanalyzer. The procedure of conjugate binding with the immobilized RNA test was similar to fluorescent ligand binding test, with the exception of quantifying the binding of on-DNA conjugate by qPCR, all qPCR experiments were conducted at 95°C for 10 min, then 35 cycles of 95°C 10 s, 55°C 10 s, 72°C 10 s using SYBR Green master mix (Thermo A25778).

Blocking of DEL DNA tags by RNA or DNA patches

The RNA patches for TAR RNA were designed by fragmenting the sequence of TAR RNA as following, the RNA patch 1 is the first 12 nucleotides (nt) of TAR RNA, the subsequent patches are with 3 nt shift from the previous patch. To block the DEL DNA tags, DEL was dissolved in selection buffer and incubated with the first RNA patch at RT for 30 mins, then the subsequent RNA patches were added one by one after 30 min incubation with the DEL tags.

For *E. coli* FMN Riboswitch, as the sequence is much longer, the RNA patches mainly covered the loop or bulge sequences. DNA patches with the same sequences were also made. DEL DNA blocking was prepared by incubation of the RNA or DNA patches with DEL at RT for 30 min as described above.

The full list of patch sequences designed for TAR RNA and FMN Riboswitch can be found in Supplementary Table S1, the patches were synthesized by GenScript or HitGen.

DEL selection

HitGen DELs were used in the selection and automated selection was carried out using a KingFisher Duo Prime Purification System (ThermoFisher) in 96-well plate. 10 μM biotin-TAR RNA was incubated with 10.38 billion member DEL sample in selection buffer (50 mM Tris, 80 mM KCl, 0.3 mg/ml ssDNA, 0.01% Tween 20, pH 7.5) at RT for 1 h followed 30 min incubation with Pierce™ Streptavidin Magnetic Beads (Thermo Scientific, cat# 88817). Immobilized RNA was washed 1 min at RT in 500 μl wash buffer (50 mM Tris, 80 mM KCl, 0.3 mg/ml ssDNA, 0.01% Tween 20, 400 μM biotin, pH 7.5) for 5 times. Retained DEL members were recovered by heat elution in elution buffer (50 mM Tris, 160 mM KCl, pH 7.5) at 95 °C for 10 min or competitive elution by incubation with 50 μM Pra-tat in buffer containing 50 mM Tris, 80 mM KCl, pH 7.5. After the first round of selection, second round was repeated with the eluted portion of previous round used as the input to the successive round with fresh RNA. For competitive elution groups, the eluted samples were purified by QIAquick kit (Qiagen 28306) to remove Pra-tat peptide. After each round,

the output was quantified by qPCR. After two rounds, the selection was done and the output was amplified by PCR, then sequencing was performed for PCR amplified samples on an Illumina NovaSeq 6000 DNA sequencer.

DEL selection for *E. coli* FMN Riboswitch was carried out in a similar procedure with differences illustrated below. 1 nmole *E. coli* FMN Riboswitch with poly A was immobilized by Oligo d(T)25 MagBeads (NEB S1419S) in selection buffer (50 mM Tris, 100 mM KCl, 2 mM MgCl₂, 0.3 mg/ml ssDNA, 0.05% Tween 20, pH 7.4), immobilized RNA was incubated with 46.3 billion member DEL in selection buffer for 1 h at RT, followed by five washes in selection buffer. Retained DEL members were recovered by heat elution in elution buffer (50 mM Tris, 100 mM KCl, pH 7.4) at 95 °C for 10 min or competitive elution by incubation with 100 μM Ribocil-C in elution buffer. Ribocil-C was removed by QIAquick kit (Qiagen 28306) following the protocols provided by the manufacturer.

The DEL size was larger for FMN riboswitch screening as the HitGen DEL size was growing. Therefore, the DELs used for TAR RNA screening is only a subset to that of the FMN-Riboswitch screening experiment. All selections were summarized in supplementary Table S2.

Data analysis and feature identification

Short reads in FastQ format for each sample were derived from demultiplexed Illumina sequencing data. An in-house DNA-Encoded library data analysis toolkit, namely DELTA toolkit, was used for downstream sequence analysis and feature identification. Detailed protocols for data preprocessing, sequence decoding, molecule counting using DELTA toolkit are provided in supplementary documents (Supplementary method 1, Figures S1 and S2).

Features in a DEL selection context are defined as groups of molecules sharing at least one common building block. It is crucial to accurately quantify these structurally related clusters and their enrichments. To minimize the experimental variations, we use the 'minus $\log_{10}$-transformed Poisson probability product score with normalization' scoring function for feature identification. Its abbreviation, o0ooooooooo (namely PolyO), is a symbolic form of a line feature. PolyO determines the background noise level dynamically based on the sequencing depth and the DEL size. For a given feature, it calculates the fold-change of its preliminary PolyO score over the background noise, in order to get the final PolyO score. Therefore, it is a normalized score adjusted for samples with different DEL input and sequencing depth. The mathematic equations to calculate PolyO scores from compound copy counts are provided in Supplementary method 1. Features with a PolyO score ≥4 are defined as significantly enriched features.

Identification of features enriched by DNA–RNA interaction

The DNA tags of HitGen DELs consist of both double strand region and single strand region. Targets interacting with different regions result in different signal readout (Figure 1). For instance, the unique molecule identifier (UMIs) region is a single strand, randomly synthesized sequence region that serves as an indicator of PCR duplicates in the
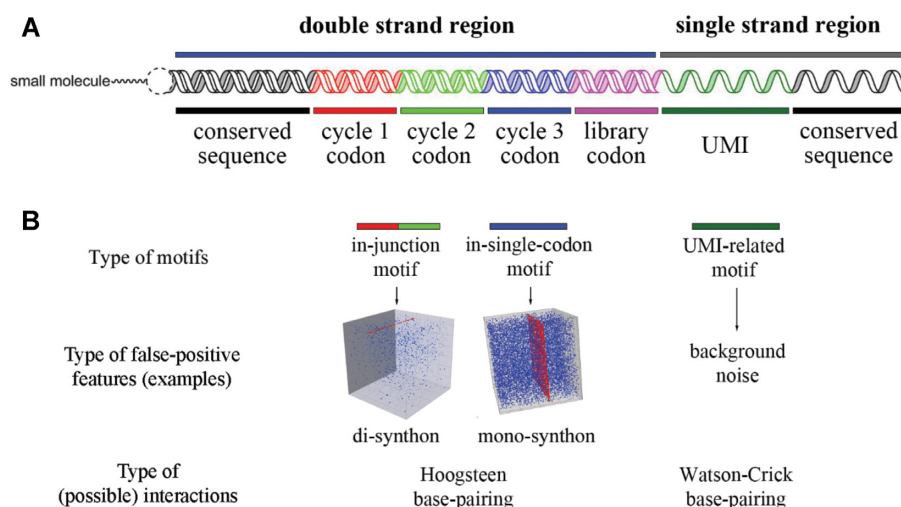
Figure 1. HitGen DEL DNA tag build and example signals derived from interactions between targets and different regions in DNA tags. (**A**) An example of a 3-cycle DEL DNA tag. This DNA tag is consisted of conserved sequences on both ends with coding sequences in the middle. The cycles 1–3 codon (colored in red, green and blue, respective) are double strand region that help to track the building blocks used when making this DEL, while the library codon (double strand region, colored in purple) encodes DEL chemistries and/or scaffold structures. The Unique Molecule Identifier (UMI) region is a single strand region with a randomly synthesized nucleotide of length 12. (**B**) In-junction motifs found in codon junctions (e.g. cycles 1 and 2) lead to the enrichment of false-positive di-synthon features, while in-single-codon motifs often result in enriched false positive mono-synthon features. Motifs in these coding regions are likely to be enriched by interactions between RNA targets with the Hoogsteen edge of double strand DNA. Comparatively, targets interacting with the UMI region are likely to be happened on the Watson–Crick edge of the single strand DNA, and will lead to an increased level of background noise.

data analysis step (22). Since UMIs do not encode structures, interactions between the UMIs and the RNA target do not contribute to feature enrichment. A computational method focusing on the k-mer profile of the UMI region is developed to quantify the level of DNA–RNA interaction. Specifically, UMI sequences of all decoded reads were extracted and fragmented using a sliding window of size k (preferably six to eight). The occurrences of k-mers were calculated and further normalized into ratios by dividing the counts by the total number of k-mers. K-mer profiles from any two samples were compared using scatter plots to identify certain k-mers that are preferentially enriched in one sample but not in another.

DNA binding motifs in the coding region are classified into in-single-codon motifs and in-junction motifs (Figure 1). Mono-synthon features exclusively enriched in the target samples were used to identify in-single-codon motifs. Particularly, DNA codons encoding the enriched mono-synthons were extracted and searched within all DELs. We define if the same codon encoding different building blocks are found to be enriched in at least two DELs, then this codon is likely to be an in-single-codon motif. To search for in-junction motifs, for each feature, all corresponding reads are piled up and a matrix containing all position A%, T%, C% and G% are calculated. The matrix is then compared with the baseline per base ATCG content. A cut-off of 1000-fold bit difference (calculated using Shannon entropy (23)) is set to identify the longest in-junction motifs (Supplementary method 1). Predicted in-junction motifs and in-single-codon motifs were aligned back to the consensus sequence of enriched features. Features are labeled as DNA-binding features if the consensus sequence is closely related with any of the predicted motifs.

Surface plasmon resonance (SPR)

SPR measurement was carried out on BIAcore T200 (Cytiva) at 25 °C in buffer containing 50 mM Tris, 80 mM KCl, 0.01% Tween 20, pH 7.5, 1% DMSO. The biotin-TAR RNA was immobilized on a series S streptavidin (SA) chip and the binding affinity of analyte was measured by passing serially increased concentrations of FAM-tat peptide at the flow rate of 30 μl/min. The final response was calculated by subtracting the response from reference channel and the response at zero concentration, binding affinity was fitted by steady state fitting model from the built-in software.

Fluorescence based FMN binding and competition assay

The assay was adopted from literature (24). Briefly, for FMN binding assay, serial dilution of RNAs was prepared and incubated with FMN at final concentration of 150 nM in assay buffer (50 mM Tris, 100 mM KCl, 2 mM $MgCl_2$, pH 7.4). After 1 h incubation at RT, fluorescence was detected by microplate reader at Ex 455 nm/ Em 525 nm. For testing the effect of patches, same concentration of RNA or DNA patches was incubated with FMN Riboswitch in selection buffer. For FMN competition assay, serial dilution of compounds was prepared in assay buffer and incubated with 30 nM *E. coli* FMN Riboswitch at RT for 15 min in 384-well plate. Then FMN was added to the samples to make final concentration at 25 nM and the mixture was incubated at RT for 1 h. Fluorescence was detected and IC_{50} was fitted by GraphPad.

Isothermal titration calorimetry (ITC)

ITC experiments were performed on MicroCal PEAQ-ITC (Malvern Panalytical) at 25°C. Compounds were diluted in

assay buffer (50 mM Tris, 100 mM KCl, 2 mM MgCl$_2$, pH 7.4) at 50 μM or 30 μM, FMN Riboswitch RNA was prepared at 5 μM. The initial injection of 0.4 μl compounds was titrated and excluded in data analysis, then 18 injections of 2 μl compounds were added to the cell containing RNA with duration of 4 s at intervals of 150 s while stirring at 750 rpm. The thermogram and thermodynamic parameters were fitted by the built-in software.

RESULTS

DNA–RNA binding introduced significant false positive signals in DEL selection for TAR

Motivated by the extensive published investigation on HIV-1 TAR (25,26), we first focused on this target for DEL selection. TAR RNA plays an important part in the replication of HIV-1 by binding with the transactivator protein tat, small molecules disrupting tat-TAR interaction would inhibit HIV-1 replication and exhibit anti-viral activity. Before selection, the activity of biotin tagged TAR RNA was validated by SPR by testing the binding with FAM-tat peptide (Figure 2A). The immobilized activity under DEL selection condition was further demonstrated by binding with the FAM labeled tat peptide after successful immobilization of the RNA with streptavidin beads (Figure 2B). As shown in Figure 2C, the FAM-tat peptide flow through (FL) fluorescence intensity of immobilized TAR was significantly lower than that of the blank beads, indicating more tat peptide bound with the immobilized TAR RNA, much higher fluorescence intensity for the group with RNA was observed after we heated the beads to denature the RNA and release the molecules bound, suggesting the immobilized RNA bound well with the FAM-tat peptide. We further evaluated the feasibility of selection by testing the enrichment of the tat DNA conjugate by heat elution and competitive elution with Pra-tat. Good enrichment of the conjugate was observed on immobilized TAR RNA for heat elution and competitive elution, this further boosted our confidence for applying TAR with DEL selection (Figure 2D).

Encouraged by the pre-selection result, we performed DEL selection with our standard protocol by incubation of the 10.38 billion member DEL with TAR RNA, then immobilized to magnetic beads and washed extensively to remove weak binders. Bound molecules were released by heat denaturation and used as input for subsequent round of selection. Sequencing and data analysis were consecutively performed after the selection endpoint was reached.

Originally designed for DNA binding motif identification in protein target DEL selection, the UMI profiling method was applied to TAR RNA screening experiment. All 8-mers identified in UMI regions were also aligned against the TAR RNA. Each 8-mer, represented by a dot in the scatter plot, is color-coded by its alignment score with TAR, using a scoring matrix of [match: +2, mismatch: –3, gap open: –5, gap extension: –1] (27). Apparently, significant levels of DNA binding events were identified from the UMI profiling result. More importantly, highly enriched 8-mers in the TAR selection sample tend to have a better alignment with TAR, suggesting that most DNA–RNA interactions follow

base pairing rules with certain level of mismatch tolerance (bottom-left scatter plot in Figure 3A).

Although the UMI profiling method could be used as an indicator of DNA–RNA binding, this UMI-RNA interaction actually contributes little to feature enrichments. To test if features specifically enriched in TAR RNA samples were derived from DNA–RNA interaction, we applied our in-single-codon motif and in-junction motif search algorithms (supplementary method 1, figure S3) and identified several highly enriched motifs in coding regions. A representative example of in-junction motifs was shown in Figure 2E and 2F. The DEL0934-26-0-0 plane (here '0' acts as a wildcard for any valid codon indices) was specifically enriched in the TAR RNA sample but not in the blank control. Though DNA codons used in each DEL are usually shuffled so that A%, T%, C%, and G% at each position is close to 25%, R2 codons of most enriched compounds on the DEL0934-26-0-0 plane started with AG. We concluded that these compounds were enriched because their DNA tags, rather than the small molecules, were interacting with the TAR RNA from base 21 to 29 through the 'GGCAGA-GAG' motif. The secondary structure of TAR is predicted and illustrated using RNAfold and forna webservers (28), respectively.

Using TAR RNA selections as examples, the informatics tools were proven to be effective for identifying DNA-binding events and the underline DNA-binding motifs. However, there is still an urgent need for reducing the level of DNA-binding experimentally, especially for a more effective use of the sequencing throughput.

Reduction of DNA binding by inclusion of RNA patches and competitive elution

We took up the challenge to optimize our selection strategy to reduce the false positive signals. As most of the false positive signals were introduced by the interaction of RNA sequences with our DEL DNA tags, we reasoned that pre-incubation of the RNA fragments from the RNA target with DELs can reduce the non-specific binding. To verify this hypothesis, we synthesized RNA patches and performed another selection with and without RNA patch blocking of DEL. To further increase the specificity of selection output, we employed competitive elution along for comparison with heat elution.

We have analyzed the UMI profiles for all four samples ($\pm$RNA patches and $\pm$Competitive elution) by comparing the 8-mer occurrences with the matched controls (Figure 3A). A weak correlation (Pearson correlation $r = 0.303$) was observed for the sample without using RNA patches and competitive elution, suggesting high level of DNA-binding events. The correlation was slightly improved by 0.051 (Pearson correlation $r = 0.354$) with the use of competitive elution alone. The correlation was drastically increased by the use of RNA patches, reaching a correlation of 0.634 and 0.559, for samples with and without competitive elution, respectively. This indicates that customized RNA patches and competitive elution are capable of reducing DNA binding events dramatically. Notably, the correlation scores across these samples are negatively correlated with the distance shifted from the $y = x$ line (dash lines in
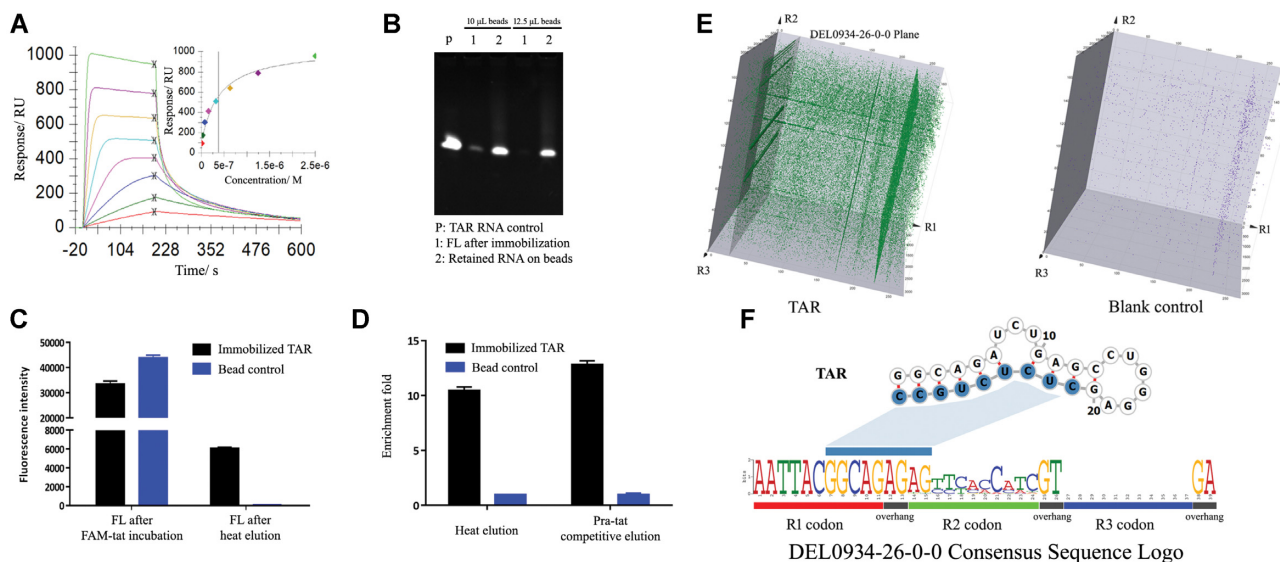
Figure 2. TAR RNA validation before DEL selection and schematic of DNA–RNA binding. (**A**) SPR demonstrated TAR RNA was active and showed binding with FAM-tat peptide with K_D of 0.38 μM. (**B**) Gel electrophoresis showed biotin-TAR RNA could be successfully immobilized by streptavidin beads, 125 pmol TAR RNA was tested with 10 and 12.5 μl beads in the assay, most RNA retained on beads with little or no RNA observed in the flow through (FL). (**C**) FAM-tat peptide binding assay revealed that immobilized TAR RNA was active under DEL selection condition with significant higher amount of FAM-tat peptide binding to RNA when compared with blank beads in heat elution, the flow through (FL) after FAM-tat incubation was lower, also indicated more FAM-tat bound with immobilized RNA than with blank beads. (**D**) tat DNA conjugate showed significant enrichment on immobilized TAR RNA by qPCR, indicating the immobilized activity of TAR. Both heat elution and Pra-tat competitive elution showed comparable enrichment, indicating the binding was mediated by tat-TAR, not by DNA–TAR interaction. Values of (C,D) were expressed as mean ± SD from duplicates. (**E**) A DEL containing representative DNA-binding features enriched in TAR but not in the blank control was illustrated in the cubic view. Several di-synthon features on the DEL0934-26-0-0 plane are highlighted in the figure. (**F**) An in-junction motif, GGCAGAGAG, is aligned to the last 9 bases of TAR. Therefore, all di-synthon features encoded by R1 = 26 codon and R2 codon started with AG were all enriched.

the scatter plot in Figure 3) to the cluster of k-mer species with low alignment score, indicating RNA–DNA interactions in such experiment consumed the majority of the sequencing throughput, which lead to a reduced sequencing depth for small molecule–target interactions and other non-DNA binding interactions.

To check the impact of using RNA patches and competitive elution on features enriched by small molecules interacting with the target, we selected a list of known DNA binding features (extracted from enriched features with their DNA sequences perfectly matching to one of the predicted motifs) as false positives (due to DNA–RNA interaction), and utilized features enriched in the blank controls as true positives (small molecules bind to the magnetic beads). The full list of predicted motifs and their alignments with TAR are provided in Table 1. We compared the enrichment of true positives and false positives using UpSet charts (29) for samples with/without RNA patches (denoted as P + or P–), and for samples with/without competitive elution (labelled as C + or C−) (Figure 3B). Results have shown that DNA binding features were most enriched in P−/C− sample, with 10,559 DNA binding features enriched. Using either RNA patches or competitive elution, the vast majority of DNA binding features were removed, leaving only 26 and 12 enriched features, respectively. This number was further reduced to 1, when both of the RNA patches and competitive elution were applied. In contrast, P+/C + sample reported the highest number (621) of beads binder features, comparing with 569 for P+/C- sample. The number of beads binding features enriched in P−/C + and

P−/C− are equivalent, with a count of 320 and 324, respectively. This matches with our expectation that when DNA-binding events were removed, sequencing throughput originally occupied by DNA-binding features are assigned to small molecule derived features. Notably, P+/C− treatment works better than P−/C+ and P−/C− treatments, which is in accordance with the correlation study. RNA patches and competitive elution synergize well with each other in terms of reducing DNA-binding while also retaining true positive features. Therefore, we concluded that P+/C+ can effectively remove DNA-binding features without affecting the enrichment of small molecule derived features. While no active compound was identified from the DEL selection, the tat DNA conjugate spiked-in the DEL with this selection was successfully pulled out, with the enrichment fold comparable in the groups with and without RNA patch blocking (supplementary Figure S5), indicating the introduction of RNA patches in selection did not affect the enrichment of real binders in DEL selection.

DEL selection of FMN riboswitch by optimized DEL selection strategy

To demonstrate the performance of the optimized DEL selection strategy, we further performed DEL selection on another RNA target, *E. coli* FMN Riboswitch. Riboswitches are bacterial-specific regulatory elements that located in the untranslated regions (UTRs) of mRNAs consisting of an aptamer ligand binding domain and an expression platform. Binding or dissociation of the cognate ligand FMN
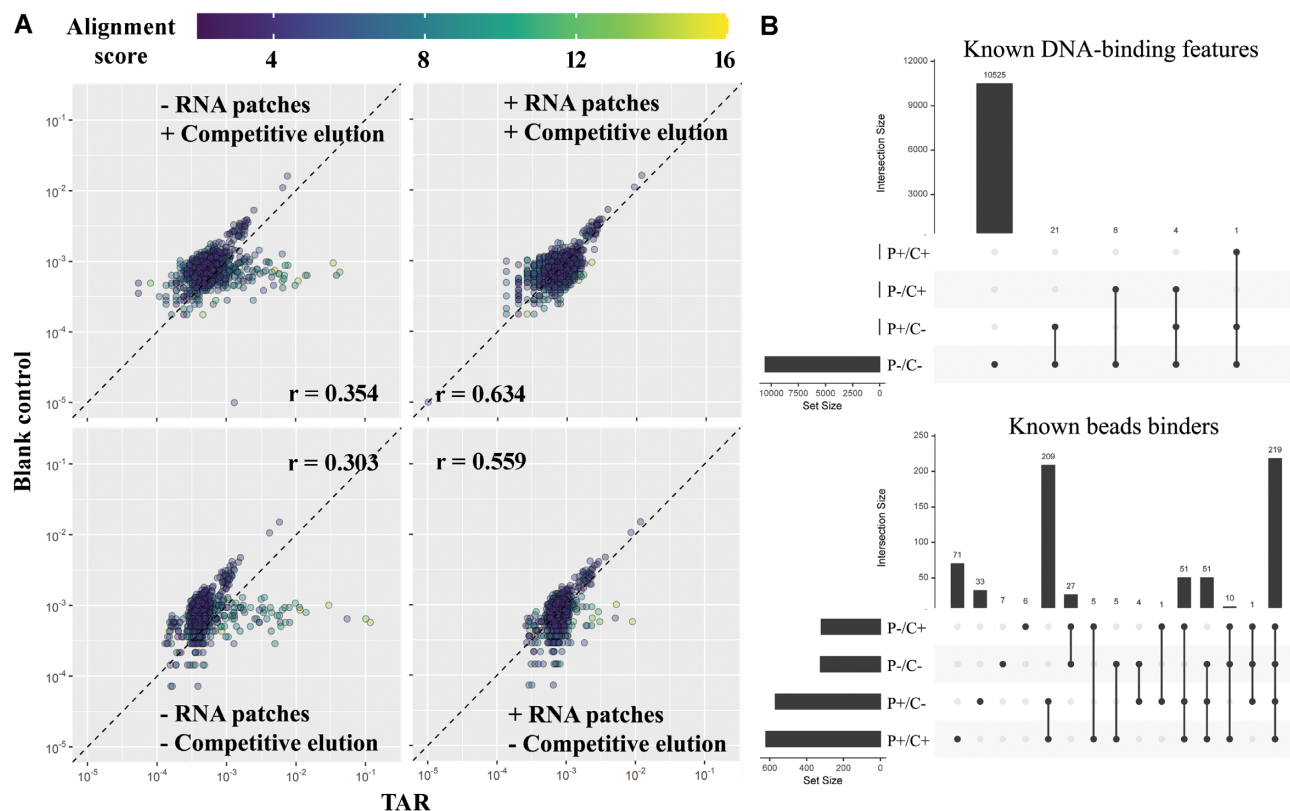
Figure 3. Levels of DNA binding measured by UMI k-mer profiling and UpSet charts for the TAR selection experiment. (**A**) The scatter plots of the UMI *k*-mer profiles for the target sample (x-axis) versus the corresponding blank control (y-axis). Each dot in the scatter plot represents a unique *k*-mer, with its normalized ratio plotted on the x-axis and y-axis for TAR and the blank control, respectively. Four TAR samples were illustrated in the chart. Top-left: no RNA patches but with Ribocil-C competitive elution; Top-right: with RNA patches and with Ribocil-C competitive elution; Bottom-left: no RNA patches and with heat elution; Bottom-right: with RNA patches and heat elution. Each dot in the scatter plot represented a unique 8-mer, color coded by the alignment score with TAR. Pearson correlations of the *k*-mer profiles between each sample and their blank control were listed in the center of this image. A $y = x$ dash line is included in each scatter plot. (**B**) UpSet charts for selected DNA-binding features and features enriched via the binding of the small molecules to the magnetic beads (namely beads binder features) enriched in all four TAR samples. P+: RNA patches were used; P–: RNA patches were not used; C+: Ribocil-C Competitive elution; C–: heat elution.

to FMN riboswitch induces a conformational change in the expression platform and regulates *rib* expression coding for the riboflavin biosynthetic enzyme. FMN riboswitch serves as a regulatory motif coordinating riboflavin synthesis which is essential for bacterial growth, and disrupting riboflavin biosynthesis represents a promising strategy for anti-bacterial treatment and FMN riboswitch servers as a validated drug target. Ribocil-C is a reported small molecule specifically binds FMN Riboswitch and exhibits anti-bacterial growth effect (24), we further investigated DEL selection for this target in order to demonstrate the ability of the optimized selection strategy and find active compounds.

As the length of *E. coli* FMN Riboswitch is longer than 100 nt and chemical synthesis of the RNA held difficulty in accuracy of the sequence, we prepared the RNA by *in vitro* transcription and introduced a poly A tail the end for the ease of immobilization by Oligo dT beads. As demonstrated in the Figure 4A, activity of the riboswitch with and without poly A tail was comparable, suggesting introduction of poly A tail into the riboswitch did not affect its structural arrangement and activity. We further explored the immobilization and confirmed Oligo dT beads could successfully

capture the RNA (Figure 4B). The immobilized activity was also validated under the selection condition by confirming the binding of FMN with the RNA applying Ribocil-C competitive elution (Figure 4C). To validate the effectiveness of RNA patches in reducing non-specific binding of DNA tags, we first confirmed the RNA patches didn't affect the target RNA structure and interfere FMN binding by activity test using FMN Riboswitch without (Figure 4D) and with poly A (Figure 4E), we then assessed the enrichment ratio of the DNA tag (without compound attached) on FMN Riboswitch with and without pre-blocking by RNA patches. The result suggested that DNA enrichment ratio was significantly reduced to 1/2 when compared with the group without RNA patch blocking (Figure 4F), indicating the RNA patches were effective in reducing RNA–DNA binding.

Encouraged by these findings, we then applied DEL selection of FMN Riboswitch with 46.3 billion member DEL incorporating the RNA patch blocking and Ribocil-C competitive elution. Importantly, we also included a group of heat elution for comparison. The selection afforded significant enrichment of DEL molecules. After ruling out computationally predicted DNA-binding features, the enriched

Table 1. A list of in-junction and in-single-codon motifs predicted for TAR RNA selection

Motif	Motif type	Representative feature from P−/C−	PolyO score					Multiple sequence alignment[b] 3′-CCGUCUCUCGAGGGUC CGAGUCUAGACGG-5′ (TAR)
			P−/C−	P+/C−	P−/C+	P+/C+	Blank[a]	
AGCTCCTAGGC	In-single-codon	DEL0951-0-541-0	478.564	2.234	12.909	0	0	- - - -AGCTCCTAGGC- - - -
AACTTGGCAGA	In-single-codon	DEL1228-0-0-265	15.867	11.859	1.94	0	0	AACTTGGCAGA- - - - - - - -
AGAGATTCCAG	In-single-codon	DEL1065-0-491-0	12.133	0	2.386	0	0	- - - -AGAGATTCCAG- - - - -
ATTCCAGGCTC	In-single-codon	DEL1244-526-0-0	11.883	4.071	0	0	0	- - - - -ATTCCAGGCTC- - -
CTTATGGCAGA	In-single-codon	DEL1228-0-0-1126	17.917	13.693	2.355	0	0	CTTATGGCAGA- - - - - - - -
AGAGTTCC	In-junction	DEL0934-26-0-0	41.684	1.191	6.501	0	0	- - - -AGAGTTCC- - - - - -
GGCAGAGAG	In-junction	DEL0937-0-259-0	15.155	0	7.063	0	0	- -GGCAGAGAG- - - - - - -
CAGGCTCA	In-junction	DEL0991-0-105-0	13.765	0	1.893	0	0	- - - - - -CAGGCTCA- - -

[a]Two blank control samples with different elution methods (heat elution and competitive elution) were included in the TAR selection experiment. DNA-binding features were not enriched in both samples. Therefore, only one column was used to represent the feature enrichment in the blank samples.
[b]The TAR sequence in the multiple sequence alignment column was written in 3′ to 5′ direction to ensure the sequences listed in the motif column are aligned by their original sequences.
In total, eight motifs were predicted, with three of them are in-junction motifs. Representative features from the TAR P−/C− sample were used to demonstrate the enrichment score changes across the other samples. The representative features exhibited significant enrichment on P−/C− sample and no enrichment on P+/C + and the blank control samples. Using RNA patches or competitive elution along could not completely eliminate the DNA-binding features. Multiple sequence alignment suggested that the first 8 bases near 5' did not contribute to significant DNA binding features.
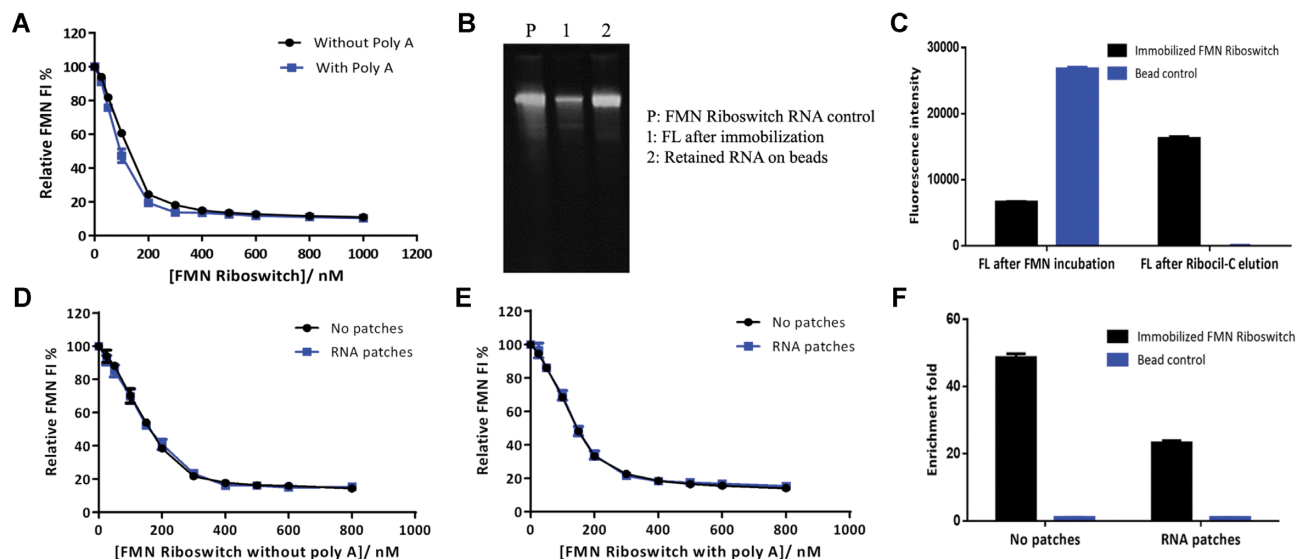

Figure 4. FMN riboswitch RNA validation before DEL selection. (**A**) Activity of FMN Riboswitch was confirmed by binding with FMN in fluorescence-based assay, poly A did not affect activity as RNA with and without poly A tail exhibited similar activity. (**B**) Gel electrophoresis showed FMN Riboswitch RNA with poly A could be successfully immobilized by Oligo d(T)25 beads. 20 pmol RNA tested with 10 μl beads in the assay. (**C**) FMN binding assay revealed that immobilized FMN Riboswitch was active under DEL selection condition with significant higher amount of FMN binding to RNA when compared with blank beads in Ribocil-C elution, the flow through (FL) after FMN incubation was lower, also indicated more FMN bound with immobilized RNA than with blank beads. (**D**, **E**) RNA patches didn't affect FMN Riboswitch binding with FMN, incubation of equal amount of RNA patches with FMN Riboswitch without (D) and with poly A (E) indicated similar FMN binding when compared with the no RNA patch group. (**F**) RNA patches significantly reduced the binding of control DNA conjugate (no compound attached) introduced by RNA–DNA interaction in qPCR assay. Values were expressed as mean ± SD ($n = 4$ for A, D, E, F, $n = 2$ for C).

DEL signals were visualized in DataWarrior cubes with each axis representing one cycle of the DEL synthesis. As shown in Figure 5A, heat elution and Ribocil-C competitive elution revealed similar enrichment with feature intensity stronger in competitive elution groups, suggesting the robustness of the enriched signal and the reliability of competitive elution. The 29.27 million sub-library (DEL1307) was constructed using a DNA recorded split and pool strategy (30) in 3 synthesis cycles consisting of 776, 328 and 115 building blocks (BBs) at cycle 1, 2 and 3, respectively (Figure 5B). We further looked into the details of the enriched feature, this feature exhibited as a line in the cube, suggesting two building blocks of the structures were fixed and the

first building block (R1) remained variable. Representative structures on this feature were chosen for off-DNA synthesis as illustrated in Figure 5C. The synthesis of these two compounds is detailed in supplementary method 2 and Figure S4.

Compound activity against FMN riboswitch

HGC-1 and HGC-2 were synthesized off-DNA and tested in biophysical assays to validate the performance of DEL selection. As shown in Figure 6A, in the ITC assay, the two compounds exhibited dissociation constant K_D value of 11.38 and 15.03 nM, respectively, which is comparable to
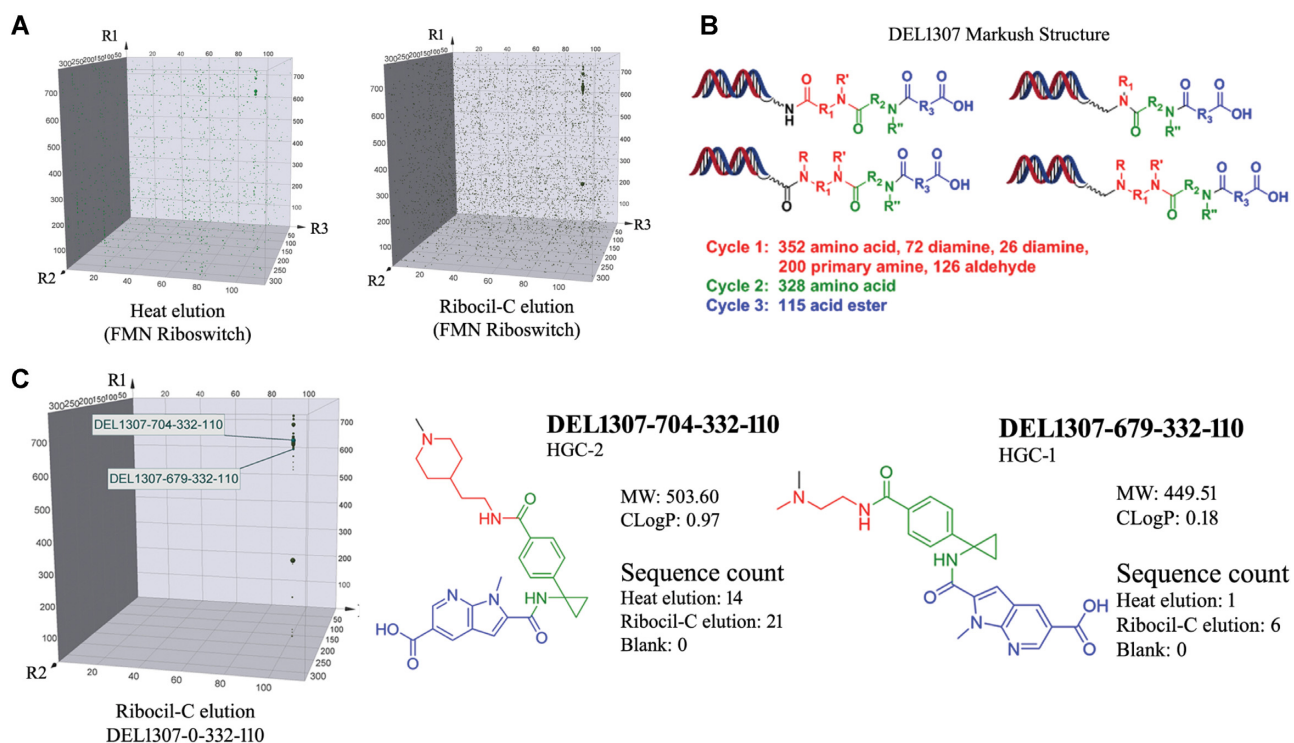
Figure 5. The general construction of DEL 1307 and DataWarrior cubic view showing the enriched feature against FMN Riboswitch. (**A**) Heat elution and Ribocil-C competitive elution revealed similar feature, indicating the specificity and robustness of the enriched signal. (**B**) DEL 1307 is a 3-cycle library consisting of 776, 328 and 115 BBs at cycle 1, 2 and 3, respectively, comprising 29.27 million individual compounds on DNA. (**C**) Zoom-in of the feature details, the line feature suggested R2 and R3 were essential for binding activity, variation in R1 was illustrated with compound structures depicted, the black methyl group in the red moiety indicated the DNA attachment point. Two representative structures (HGC-2 and HGC-1) were chosen for off-DNA synthesis. The structures, properties and sequence count across tested samples were listed.
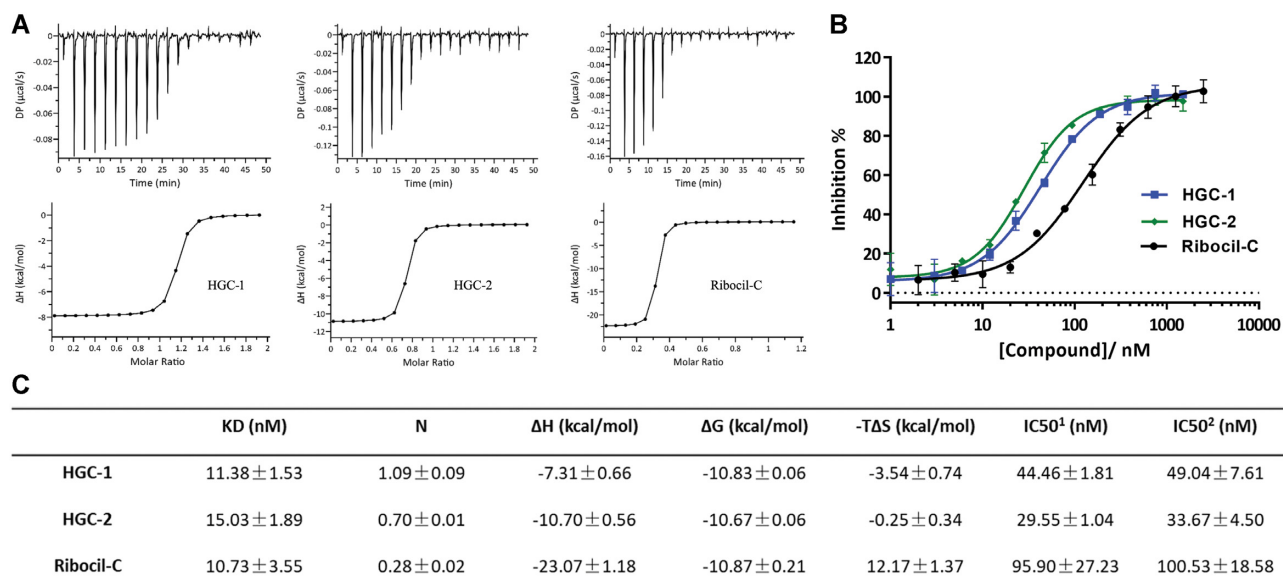

	KD (nM)	N	ΔH (kcal/mol)	ΔG (kcal/mol)	-TΔS (kcal/mol)	IC50[1] (nM)	IC50[2] (nM)
HGC-1	11.38±1.53	1.09±0.09	-7.31±0.66	-10.83±0.06	-3.54±0.74	44.46±1.81	49.04±7.61
HGC-2	15.03±1.89	0.70±0.01	-10.70±0.56	-10.67±0.06	-0.25±0.34	29.55±1.04	33.67±4.50
Ribocil-C	10.73±3.55	0.28±0.02	-23.07±1.18	-10.87±0.21	12.17±1.37	95.90±27.23	100.53±18.58

Figure 6. Synthesized off-DNA compounds HGC-1 and HGC-2 showed activity on FMN Riboswitch. Representative figures of HGC-1 and HGC-2 binding on FMN Riboswitch without poly A by ITC (**A**) and competition with FMN binding in fluorescence-based assay (**B**, mean ± SD from quadruplicates). Both compounds showed similar activity, the effect of Ribocil-C was presented for comparison. (**C**) Summary of the binding affinity K_D, stoichiometry N and thermodynamic parameters (ΔH, ΔG, $-T\Delta S$) from ITC, as well as IC$_{50}$ values from fluorescence-based assay for FMN riboswitch without (IC50[1]) and with poly A (IC50[2]), values are mean ± SD from three independent experiments.

that of Ribocil-C. The observed stoichiometry and thermodynamic parameters differ for the three ligands (Figure 6C), similar ΔG (change in free energy) indicated similar K_D of the compounds, the favorable change in enthalpy $-\Delta H$ is attributed to hydrogen bond formation during ligand binding for Ribocil-C over HGC-1 and HGC-2, the unfavorable $-T\Delta S$ (change in entropy ΔS multiplied by the absolute temperature T) suggested possible conformational rearrangements in the RNA upon Ribocil-C binding. In most cases, ligand and macromolecule binding stoichiometry is 1, with the exception that the macromolecule is not 100% active or the existence of multiple binding sites or moieties. In our case, the stoichiometry for Ribocil-C and HGC-2 was still not 1 when multiple assay conditions including different RNA annealing procedures, concentrations and buffer recipes were tried (data not shown). This could be possibly explained that the binding behavior changed under different conditions and some ligands induced conformational change or bound to a heterogeneous RNA population as shown in the literatures (31,32), the detailed mechanism remains to be further explored.

Per confirmation of binding affinity, we speculated the compounds would exhibit activity towards the target in the functional assay by competition with FMN binding. As demonstrated in Figure 6B, HGC-1 and HGC-2 showed competition effect, with IC50 value of 44.46 and 29.55 nM, respectively, which are both slightly better than that of Ribocil-C. The presented results were against the FMN Riboswitch without poly A, a parallel test with this RNA with poly A tails also revealed similar values as demonstrated in Figure 6C.

With the carboxylic acid at the end, we suspect the compounds may exhibit difficulty entering the bacteria, especially for Gram-negative pathogens as in the case of Ribocil-C (33), and further optimization may be needed to develop these compounds as anti-bacterial reagents. However, these compounds can serve as a good starting point for anti-bacterial drug discovery as Ribocil-C was also identified from the screening on MB5746 *E. coli* strain, which is outer membrane hyper-permeable and efflux deficient (24).

DNA patch versus RNA patch

Originally, only RNA patches were tested in the TAR RNA and FMN Riboswitch screening experiments because we thought the interactions between the RNA target and the DNA tags could be more accurately represented by that between RNA patches (when compared to DNA patches) and the DNA tags. Nevertheless, in a context of affinity-based selection, a more stable triplex formed by patches with double-strand DNA could be more efficient to block the interactions of DNA tags with the RNA target. Thermodynamic parameters of DNA, RNA and DNA–RNA hybrid duplexes were reported in multiple studies (34,35); however, few compared the stability of RNA–dsDNA triplex with ssDNA–dsDNA triplex. Since DNA patches are often more accessible and cost-effective, we conducted additional validation experiments to test whether DNA patches could be a better replacement for RNA patches.

To validate the effectiveness of DNA patches, after we confirmed the DNA patches did not affect FMN Ri-

boswitch RNA binding activity with FMN (Figure S6A, B), we compared the enrichment of the DNA tag (without compound attached) on FMN Riboswitch with RNA/DNA patch blocking and without blocking. The result indicated that DNA tag enrichment ratio was significantly reduced to 4/5 when compared with the group without blocking (Figure S6C). However, DNA patches were less effective as the RNA patches reduced the enrichment ratio to 1/2.

To evaluate whether DNA patches are effective at reducing the overall DNA binding event, we performed an additional DEL selection and applied the UMI profiling method on three FMN Riboswitch samples, one without the use of patch sequences (denoted as P−), one with DNA patches (P^{DNA}) and one with RNA patches (P^{RNA}), all of which used Ribocil-C elution (described as competitive elution in general, labeled as C+). A separate sample with no patches under heat elution condition (denoted as P−/C−) was also included. The 8-mer profiles from these four samples (i.e. P−/C−, P−/C+, P^{RNA}/C+ and P^{DNA}/C+) were compared with their corresponding blank controls. Results of the UMI profiles in these samples correlated with the observations from the TAR RNA selection. Particularly, the Pearson correlation increased in the order P−/C− < P−/C+ < P^{RNA}/C+ . There is only a slight increase of the correlation score (+0.001) when DNA patches was used, as compared to P−/C+, suggesting that DNA patches may not be as effective as RNA patches at blocking the interactions between the RNA target and the single strand DNA region. Irrespective of different optimization methods, samples with patches or competitive elution all seemed to reduce the throughput occupied by DNA binding event, with the majority of species with low alignment scores distributed on the diagonal region (Figure 7A).

To test if the validated compounds/feature can be identified in these samples, we compared the PolyO score and the ranking of the true positive feature (namely TP feature, with a feature index of DEL1307-0-332-110) (Figure 7B). It has been shown that the ranking of the TP feature decreased in the order P^{RNA}/C+ > P^{DNA}/C+ > P−/C+ >> P−/C−. Nevertheless, the ranking difference between P^{RNA}/C+ and P^{DNA}/C+ should not be interpreted as significant. Moreover, it should be noted that the TP feature is still detectable in no patch/heat elution sample, ranking at 271 out of all 651 significantly enriched features, though the PolyO score is very close to the significant feature cut-off (PolyO ≥ 4).

DISCUSSION

It is well accepted that interactions between RNA targets and the DNA tags was one of the major obstacles that prevents the wide application of DEL screening on RNA targets. By analyzing DNA motifs enriched from each DEL screening experiment, we found out that the majority of DNA–RNA interactions still follow base pairing rules. We believe that, similar to the formation of LncRNA-DNA triplex in Eukaryotes (36), interactions between DEL DNA tags and the RNA target are likely to happen on the Hoogsteen-edge (37) when canonical Watson-Crick edges are occupied (i.e. DNA bases in the double-strand region or RNA bases in stem region).
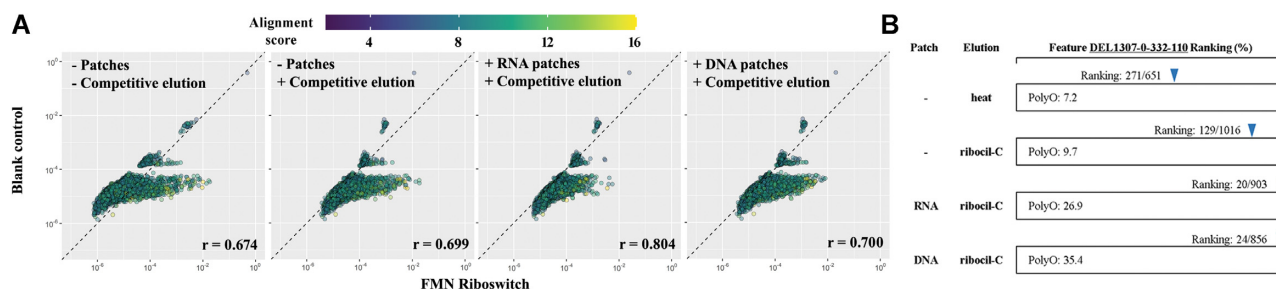
Figure 7. Levels of DNA binding measured by UMI *k*-mer profiling and the ranking of the feature with validated hits for FMN Riboswitch experiment. (**A**) The scatter plots of the UMI *k*-mer profiles for the target sample (x-axis) versus the corresponding blank control (y-axis). From left to right: no patches/heat elution, no patches/competitive elution, RNA patches/competitive elution, DNA patches/competitive elution. Pearson correlation and a $y = x$ dash line is labelled for each scatter plot. (**B**) The PolyO score and the ranking of DEL1307-0-332-110 in four FMN Riboswitch samples. For each sample, the PolyO score of the feature is listed inside the box, with the ranking among all significantly enriched features labelled on top of the box.

Features enriched by DNA–RNA interaction usually contain common DNA binding motifs in the coding region. To date, no motif identification methods for DEL screening data were reported. Motif discovery methods designed for genomic-related sequencing data (38,39) are not directly applicable for DEL for several reasons. Firstly, a regular DEL molecule contains at least two different moieties, i.e. a small molecule and a DNA tag. For efficient binding motif prediction, sequences enriched by interactions between small molecules and the target should be excluded. However, DNA-binding features will no longer be a concern if features enriched by small molecules can be accurately identified. To solve this issue, one can think of including DNA-only libraries, either as a separate sample, or as spike-in libraries pooled with the rest of DELs. The total combinations of DNA codons have to be large enough to be comparable to that of the actual DEL pools. It is uncertain whether the use of large spike-in DNA-only libraries affects the enrichment of true positive features. Therefore, a target-by-target validation is required. It is worth mentioning that DEL screening is an affinity selection process that is highly sensitive to weak interactions. Specific motifs and their variants with mismatches are enriched and mixed together. Considering that compounds on a same feature also contain the same conserved DNA codons, these variables make the motif discovery even more challenging. Lastly, we also explored the feasibility of aligning the RNA sequence to consensus sequences for DNA-binding feature prediction. False positives (regions that do not contribute to DNA–RNA interaction) and false negatives (motifs with poor alignment due to mismatches and gaps) were introduced using this approach.

Although the algorithm can differentiate the DNA–RNA binding signals, experimental procedure to reduce the false positive signals is of paramount importance. With the application of the RNA patches and competitive elution, binding mediated by DNA tag-RNA target interaction was reduced and exhibited lower recovery of DEL molecules during selection, this on the one hand saved sequencing capacity as the DNA–RNA binding signals would occupy the sequencing depth and lead to huge waste, on the other hand, lower recovery resulted in less round of selection, which saved the effort of performing additional round of selection, as additional round of selection would bring down

the absolute signals while increasing the signal-to-noise ratio (40).

So far, one case of DEL selection on the DNA target G-quartets was reported (19), in which the binding introduced by DNA tags was not mentioned. A possible explanation is that G-quartet DNA contains guanine rich repeats in its sequences, and DEL DNA tags rarely contain such specific sequences, thus interaction introduced by binding with the DNA tags may not be significant. However, such case is not applicable to conventional nucleic acid targets. For the two RNA targets investigated in this paper, there is no specific sequence thus the interactions introduced by DNA tags are observed. In addition, G-quartets is remarkably stable in structure, which further alleviates the possible DEL tag binding effect and facilities a successful screen.

The utilization of RNA patches in DEL selection has not been reported. Fragmenting the RNA target into small segments (RNA patches) and employing them in the DEL selection may reduce the interference of RNA target binding to DNA tags while effectively recognize the interaction of RNA target with small molecules. It should be noted that the patches should not be too long to form secondary structures.

Most DEL selection adopts heat elution to fully denature the target and elute all molecules binding to the target. This is the most efficient method and all binders will be non-specifically eluted off. Competitive elution utilizes the effect of a known binder, by incubation of the binder at high concentration with target-DEL complex, the DEL molecules that bind at the same site as the known binder does will be competitively eluted off. This increases the specificity of elution and facilitates selection performance especially for targets with known binding pocket and preferred mechanism of action. In our case, competitive elution not only increases the chance of finding DEL molecules with known binding site, but also helps alleviate the false positive signals introduced by DNA–RNA interaction. The G-quartet DEL selection paper (19) utilized NaOH elution to denature the target, we tried this strategy for TAR RNA but found NaOH didn't fully release FAM-tat peptide. In addition, we suspect NaOH elution would not help decrease RNA-DNA binding signals as NaOH breaks base pairing between strands. For the FMN Riboswitch case, elution efficiency was facilitated by preparing the competitor Ribocil-

C in buffer without Mg^{2+} as Mg^{2+} helps the folding of FMN Riboswitch RNA (41), the tertiary structure of FMN Riboswitch is impaired under low Mg^{2+} condition, it further helps the release of bound small molecules without disrupting the DNA–RNA interaction. One may argue competitive elution only works for the RNA targets with known binders, for those without reported binders, the application of this strategy is limited. We tested the binding of the non-specific RNA binder neomycin to some RNA targets, and are currently investigating using neomycin as a competitive elution binder for RNA targets without known binder reported. Even without competitive elution, the RNA patches and heat elution in combination with our algorithm are still able to identify the real binding signals as shown in the case of FMN Riboswitch.

TAR RNA serves as an anti-viral target for HIV-1 infection and we took this target for the demonstration of DEL selection. Even the activity under selection condition was properly validated prior to DEL selection, the result also surprised us with most of the signals identified as DNA–RNA binding, while the recovery of DEL molecules during selection was not alarmingly high when using the conventional selection procedure. Our algorithm and RNA patches in combination with competitive elution successfully identified the RNA–DNA binding signals and reduced these features in selection. The developed strategy was further applied on FMN Riboswitch selection, a strong feature was identified from DEL1307 and the robustness of the signal was also confirmed in the heat elution group screened in parallel. The compounds synthesized from the feature were validated as active hits with double digit nanomolar binding affinity, in addition, competition with FMN binding to the target was also observed. Compared to DNA patches, RNA patches seem to be more effective at reducing the overall DNA binding event near the double strand and single strand DNA. Based on the enrichment of DEL1307-0-332-110 from the FMN Riboswitch experiment, DNA patches could be used as an alternative reagent to block the interactions between double strand DNA with the RNA target. The structural bases of why DNA and RNA patches behave different requires further investigation.

In summary, we presented the first report on DEL selection with RNA targets, with experimental strategies including RNA patches in combination with competitive elution and analytical approach incorporating motif identification. This methodology ensures robust selection result and shall have broad application on DEL selection with other RNA targets to expand the utilization of DEL selection and fuel drug discovery.

DATA AVAILABILITY

The DELTA toolkit used for conducting the DEL selection data analysis and motif prediction is available upon request from the authors. The manual for the DELTA toolkit can be found at https://github.com/LiYouBioinfo/DELTA-Toolkit.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Brenner,S. and Lerner,R.A. (1992) Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 5381–5383.
2. Gironda-Martinez,A., Donckele,E.J., Samain,F. and Neri,D. (2021) DNA-Encoded chemical libraries: a comprehensive review with succesful stories and future challenges. *ACS Pharmacol. Transl. Sci.*, **4**, 1265–1279.
3. Chen,Q., Cheng,X., Zhang,L., Li,X., Chen,P., Liu,J., Zhang,L., Wei,H., Li,Z. and Dou,D. (2020) Exploring the lower limit of individual DNA-encoded library molecules in selection. *SLAS Discov.*, **25**, 523–529.
4. Decurtins,W., Wichert,M., Franzini,R.M., Buller,F., Stravs,M.A., Zhang,Y., Neri,D. and Scheuermann,J. (2016) Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.*, **11**, 764–780.
5. Wu,Z., Graybill,T.L., Zeng,X., Platchek,M., Zhang,J., Bodmer,V.Q., Wisnoski,D.D., Deng,J., Coppo,F.T., Yao,G. *et al.* (2015) Cell-Based selection expands the utility of DNA-Encoded small-molecule library technology to cell surface drug targets: identification of novel antagonists of the NK3 tachykinin receptor. *ACS Combinatorial Sci.*, **17**, 722–731.
6. Huang,Y., Meng,L., Nie,Q., Zhou,Y., Chen,L., Yang,S., Fung,Y.M.E., Li,X., Huang,C., Cao,Y. *et al.* (2021) Selection of DNA-encoded chemical libraries against endogenous membrane proteins on live cells. *Nat. Chem.*, **13**, 77–88.
7. Oehler,S., Catalano,M., Scapozza,I., Bigatti,M., Bassi,G., Favalli,N., Mortensen,M.R., Samain,F., Scheuermann,J. and Neri,D. (2021) Affinity selections of DNA-encoded chemical libraries on carbonic anhydrase IX-Expressing tumor cells reveal a dependence on ligand valence. *Chemistry*, **27**, 8985–8993.
8. Cai,B., Kim,D., Akhand,S., Sun,Y., Cassell,R.J., Alpsoy,A., Dykhuizen,E.C., Van Rijn,R.M., Wendt,M.K. and Krusemark,C.J. (2019) Selection of DNA-encoded libraries to protein targets within and on living cells. *J. Am. Chem. Soc.*, **141**, 17057–17061.
9. Petersen,L.K., Christensen,A.B., Andersen,J., Folkesson,C.G., Kristensen,O., Andersen,C., Alzu,A., Slok,F.A., Blakskjaer,P., Madsen,D. *et al.* (2021) Screening of DNA-Encoded small molecule libraries inside a living cell. *J. Am. Chem. Soc.*, **143**, 2751–2756.
10. Warner,K.D., Hajdin,C.E. and Weeks,K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.*, **17**, 547–558.
11. Zhang,P., Park,H.J., Zhang,J., Junn,E., Andrews,R.J., Velagapudi,S.P., Abegg,D., Vishnu,K., Costales,M.G., Childs-Disney,J.L. *et al.* (2020) Translation of the intrinsically disordered protein alpha-synuclein is inhibited by a small molecule targeting its structured mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 1457–1467.
12. Ratni,H., Ebeling,M., Baird,J., Bendels,S., Bylund,J., Chen,K.S., Denk,N., Feng,Z., Green,L., Guerard,M. *et al.* (2018) Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene

splicing modifier for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.*, **61**, 6501–6517.

13. Palacino,J., Swalley,S.E., Song,C., Cheung,A.K., Shu,L., Zhang,X., Van Hoosear,M., Shin,Y., Chin,D.N., Keller,C.G. *et al.* (2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.*, **11**, 511–517.

14. Ren,Y., Wang,Y.F., Zhang,J., Wang,Q.X., Han,L., Mei,M. and Kang,C.S. (2019) Targeted design and identification of AC1NOD4Q to block activity of HOTAIR by abrogating the scaffold interaction with EZH2. *Clin. Epigenetics*, **11**, 29.

15. Lee,M.K., Bottini,A., Kim,M., Bardaro,M.F., Zhang,Z., Pellecchia,M., Choi,B.S. and Varani,G. (2014) A novel small-molecule binds to the influenza a virus RNA promoter and inhibits viral replication. *Chem. Commun.*, **50**, 368–370.

16. Velagapudi,S.P., Seedhouse,S.J. and Disney,M.D. (2010) Structure-activity relationships through sequencing (StARTS) defines optimal and suboptimal RNA motif targets for small molecules. *Angew. Chem.*, **49**, 3816–3818.

17. Velagapudi,S.P., Seedhouse,S.J., French,J. and Disney,M.D. (2011) Defining the RNA internal loops preferred by benzimidazole derivatives via 2D combinatorial screening and computational analysis. *J. Am. Chem. Soc.*, **133**, 10111–10118.

18. Velagapudi,S.P., Gallo,S.M. and Disney,M.D. (2014) Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nature Chemical Biology*, **10**, 291–297.

19. Litovchick,A., Tian,X., Monteiro,M.I., Kennedy,K.M., Guie,M.A., Centrella,P., Zhang,Y., Clark,M.A. and Keefe,A.D. (2019) Novel nucleic acid binding small molecules discovered using DNA-encoded chemistry. *Molecules*, **24**, 2026.

20. Mukherjee,H., Blain,J.C., Vandivier,L.E., Chin,D.N., Friedman,J.E., Liu,F., Maillet,A., Fang,C., Kaplan,J.B., Li,J. *et al.* (2020) PEARL-seq: a photoaffinity platform for the analysis of small molecule-RNA interactions. *ACS Chem. Biol.*, **15**, 2374–2381.

21. Matsumoto,C., Hamasaki,K., Mihara,H. and Ueno,A. (2000) A high-throughput screening utilizing intramolecular fluorescence resonance energy transfer for the discovery of the molecules that bind HIV-1 TAR RNA specifically. *Bioorg. Med. Chem. Lett.*, **10**, 1857–1861.

22. Miner,B.E., Stoger,R.J., Burden,A.F., Laird,C.D. and Hansen,R.S. (2004) Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res.*, **32**, e135.

23. Shannon,C.E. (1997) The mathematical theory of communication. 1963. *M.D. Comput.: Comput. Med. Pract.*, **14**, 306–317.

24. Howe,J.A., Wang,H., Fischmann,T.O., Balibar,C.J., Xiao,L., Galgoci,A.M., Malinverni,J.C., Mayhood,T., Villafania,A., Nahvi,A. *et al.* (2015) Selective small-molecule inhibition of an RNA structural element. *Nature*, **526**, 672–677.

25. Murchie,A.I., Davis,B., Isel,C., Afshar,M., Drysdale,M.J., Bower,J., Potter,A.J., Starkey,I.D., Swarbrick,T.M., Mirza,S. *et al.* (2004) Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.*, **336**, 625–638.

26. Davidson,A., Leeper,T.C., Athanassiou,Z., Patora-Komisarska,K., Karn,J., Robinson,J.A. and Varani,G. (2009) Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of tat protein. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11931–11936.

27. Pearson,W.R. (2013) Selecting the right similarity-scoring matrix. *Curr. Protoc. Bioinformatics*, **43**, 351–359.

28. Gruber,A.R., Bernhart,S.H. and Lorenz,R. (2015) The ViennaRNA web services. *Methods Mol. Biol.*, **1269**, 307–326.

29. Conway,J.R., Lex,A. and Gehlenborg,N. (2017) UpSetR: an r package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.

30. Kung,P.P., Bingham,P., Burke,B.J., Chen,Q., Cheng,X., Deng,Y.L., Dou,D., Feng,J., Gallego,G.M., Gehring,M.R. *et al.* (2020) Characterization of specific N-alpha-Acetyltransferase 50 (Naa50) inhibitors identified using a DNA encoded library. *ACS Med. Chem. Lett.*, **11**, 1175–1184.

31. Rode,A.B., Endoh,T. and Sugimoto,N. (2018) Crowding shifts the FMN recognition mechanism of riboswitch aptamer from conformational selection to induced fit. *Angew. Chem.*, **57**, 6868–6872.

32. Sokoloski,J.E., Dombrowski,S.E. and Bevilacqua,P.C. (2012) Thermodynamics of ligand binding to a heterogeneous RNA population in the malachite green aptamer. *Biochemistry*, **51**, 565–572.

33. Motika,S.E., Ulrich,R.J., Geddes,E.J., Lee,H.Y., Lau,G.W. and Hergenrother,P.J. (2020) Gram-Negative antibiotic active through inhibition of an essential riboswitch. *J. Am. Chem. Soc.*, **142**, 10856–10862.

34. Lesnik,E.A. and Freier,S.M. (1995) Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry*, **34**, 10807–10815.

35. Ratmeyer,L., Vinayak,R., Zhong,Y.Y., Zon,G. and Wilson,W.D. (1994) Sequence specific thermodynamic and structural properties for DNA.RNA duplexes. *Biochemistry*, **33**, 5298–5304.

36. Deogharia,M. and Gurha,P. (2021) The "guiding" principles of noncoding RNA function. *Wiley Interdiscipl. Rev. RNA*, e1704.

37. Halder,S. and Bhattacharyya,D. (2013) RNA structure and dynamics: a base pairing perspective. *Progr. Biophys. Mol. Biol.*, **113**, 264–283.

38. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.

39. Yu,Q., Wei,D. and Huo,H. (2018) SamSelect: a sample sequence selection algorithm for quorum planted motif search on large DNA datasets. *BMC Bioinformatics*, **19**, 228.

40. Chen,Q., Hall,J., Foley,T.L., Wan,J., Li,Y. and Israel,D.I. (2020) A method for estimating binding affinity from primary DEL selection data. *Biochem. Biophys. Res. Commun.*, **533**, 249–255.

41. Ferre-D'Amare,A.R. and Winkler,W.C. (2011) The roles of metal ions in regulation by riboswitches. *Met. Ions Life Sci.*, **9**, 141–173.