Transcriptome-wide *in cellulo*, *in vitro* and ΔSHAPE data from the K562 cell line

Introduction

Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) is a chemical probing method that measures RNA flexibility at single nucleotide resolution. NAI adducts, formed during probing, induce a higher rate of mutations by reverse transcriptase at the modified positions during the generation of cDNA, compared to DMSO control (Fig. 1). SHAPE technology measures local RNA flexibility both in cellular *(in cellulo)* conditions and in purified, deproteinized RNA *(in vitro)*, which can aid in the revealing of RNA secondary structure.

In vitro SHAPE reveals flexibility information across the RNA strand, based on Watson-Crick base pairing alone. *In vivo* SHAPE reveals flexibility information about an RNA molecule whose folding is affected by the binding of RBPs and other factors in the cell. The combination of *in vivo* and *in vitro* mutation rates, ΔSHAPE, can identify RNA positions that directly interact with proteins.

Database Construction

eSHAPE *db-K562*

eSHAPE *db-K562* is a deep sequencing database containing *in cellulo* and *in vitro* SHAPE data, as well as ΔSHAPE data from the K562 cell line. The eSHAPE *db-K562* is based on 6 technical replicates of polyA-selected RNA and is delivered as a data package for a single gene of interest, sets of genes, or the entire (covered) transcriptome.

Data Deliverables

.bam	SHAPE *db-K562* processed and merged reads aligned to the genome for DMSO and NAI treated	*in cellulo*	*in vitro*	*ΔSHAPE*
.bedgraph	Reactivity score at each position in the human hg38 genome with >300x coverage	*in cellulo*	*in vitro*	*ΔSHAPE*
.py	Simple python script to extract SHAPE data for regions of interest from the reactivity bedgraph			
Plots	*Optional* per gene coverage, mutation rate, and reactivity line plots and box plots (.svg)			

eSHAPE *db-K562* Workflow

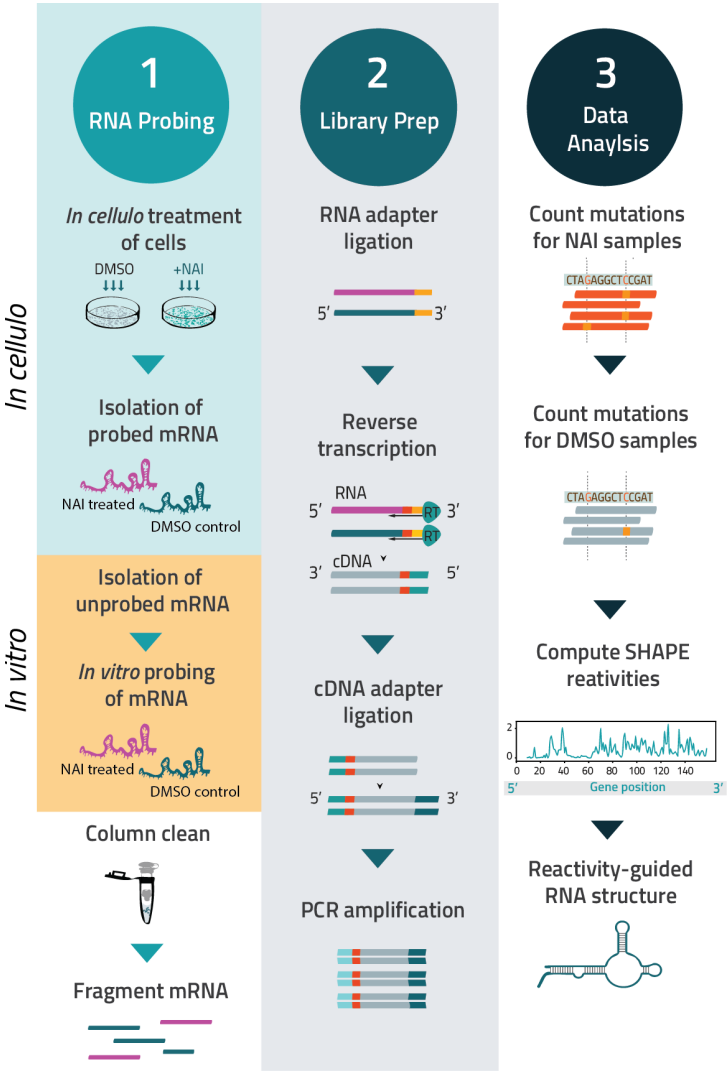

Figure 1. K562 cells were probed and harvested according to the schematic in Figure 1, with 6 replicates for each condition. For *in vitro* SHAPE, mRNA was isolated, folded, and probed (highlighted in purple). Once sequencing reads are aligned, mutations are counted for each position. SHAPE reactivities are computed from NAI mutation rates normalized to DMSO mutation rates. Computing the difference in mutation rates between NAI *in cellulo* and NAI *in vitro* yields ΔSHAPE reactivities.

More information about eSHAPE *db-k562* online at **eclipsebio.com** or contact us at **info@eclipsebio.com.**

Deep coverage of K562 transcripts

eSHAPE *db-K562* contains data for six technical replicates from each of the chemical probing conditions (NAI treated and control DMSO treated, for both *in cellulo* and *in vitro*). Each replicate was sequenced at ~130M reads. Reads were aligned to the hg38 reference genome, UMI deduplicated and merged, yielding a deep coverage SHAPE dataset on K562 transcripts.

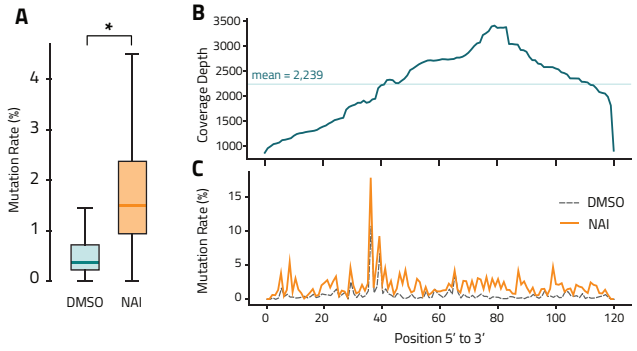

Figure 2. *In vitro* 5S rRNA (A) mutation rate box plot (*p-value of 10-24 by KS-test) (B) coverage line plot and (C) mutation rate line plot for merged datasets.

300X coverage yields reactivities with high sensitivity and specificity

Mutations generated by the NAI treatment yield valuable information on the paired or unpaired status of nucleotides, requiring a minimal coverage of approximately 300X for maximal RNA folding sensitivity and specificity.

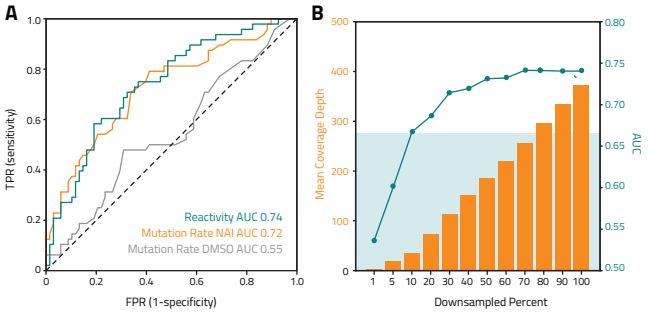

Figure 3. (A) The 5S rRNA mutation rates and reactivity scores for the eSHAPE *db-K562* dataset were compared with the paired/unpaired status of each nucleotide in the accepted reference 5S rRNA structure using Receiver Operating Characteristic (ROC) analysis and area under the curve (AUC) metric. Increased AUCs are observed for the SHAPE reactivities above the DMSO control. (B) The 5S rRNA *in vitro* SHAPE replicates were downsampled and mean coverage (orange, left axis) and reactivity AUC (teal, right axis) was computed for each downsampling. Error bars are the standard error across the six replicates. The AUC remains the same once coverage reaches approximately 300X.

SHAPE reactivities guide accurate RNA fold

In vitro SHAPE reactivities produced at each nucleotide are used as input to RNA folding algorithms such as RNAStructure to guide the folding of the RNA molecule toward an accurate secondary structure.

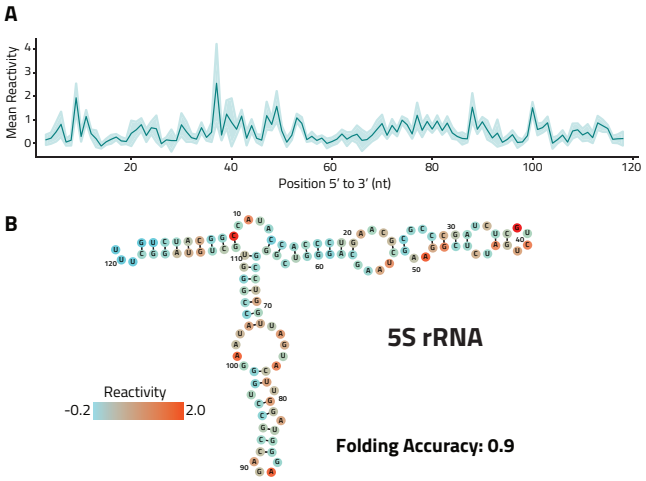

Figure 4. *In vitro* 5S rRNA (A) reactivity profile with shaded standard error across the 6 replicates and (B) SHAPE guided RNA fold. Folding accuracy for the 5S rRNA is 0.9 compared to the annotated fold in RNACentral.

ΔSHAPE predicts RBP binding to RNA

ΔSHAPE assesses the difference in mutation rates between *in cellulo* and *in vitro* NAI probing conditions, revealing positions along the RNA that interact with RNA binding proteins (RBPs).

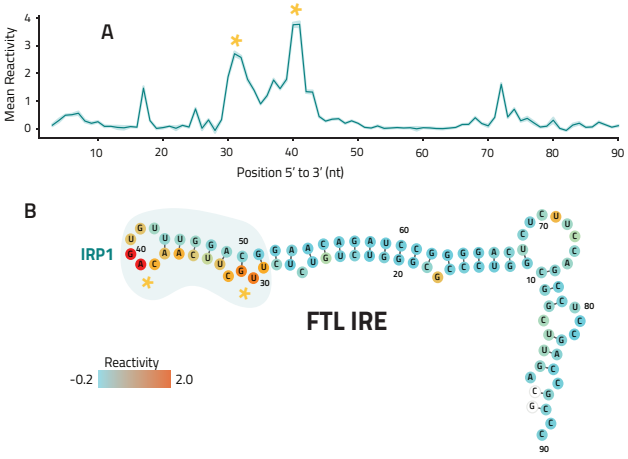

Figure 5. FTL gene ΔSHAPE (A) reactivity profile and (B) RNAStructure produced secondary structure with ΔSHAPE reactivities overlayed. Teal shading shows where IRP1 binding, and yellow asterisks marks bound nucleotides.