

## SHAPE as a way to evaluate our ability to predict the effects of SNPs on RNA

---

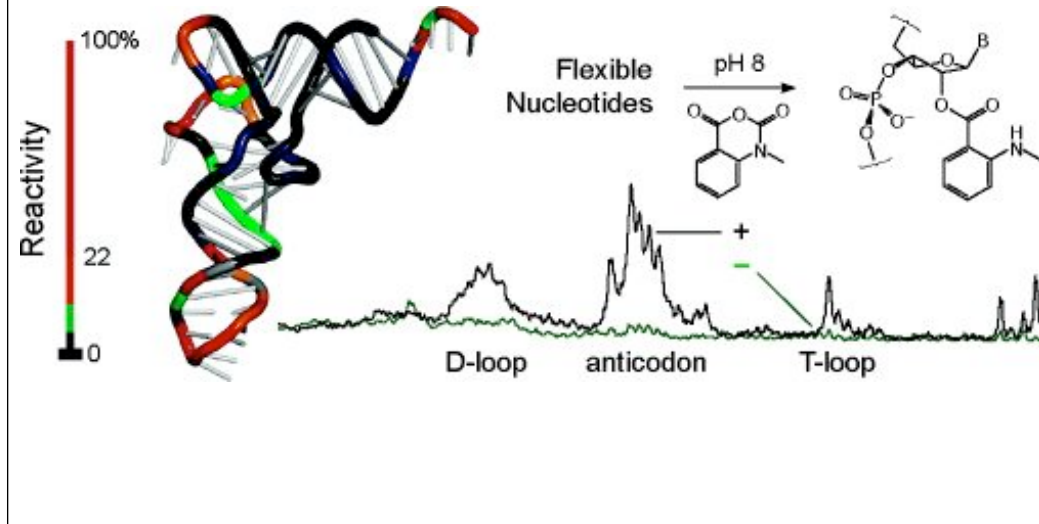
Laederach Lab

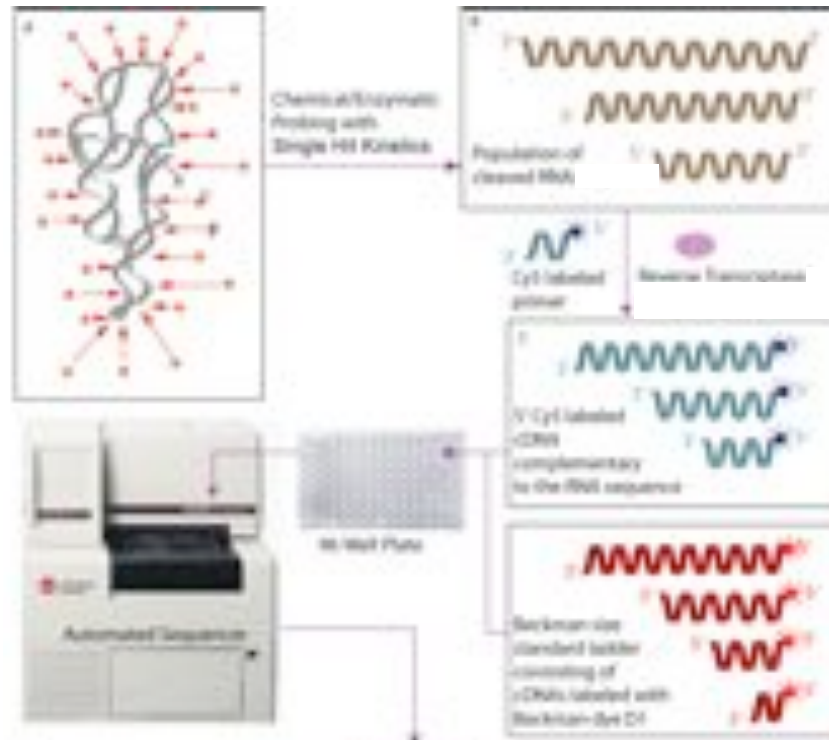
UNC Chapel Hill, Department of Biology

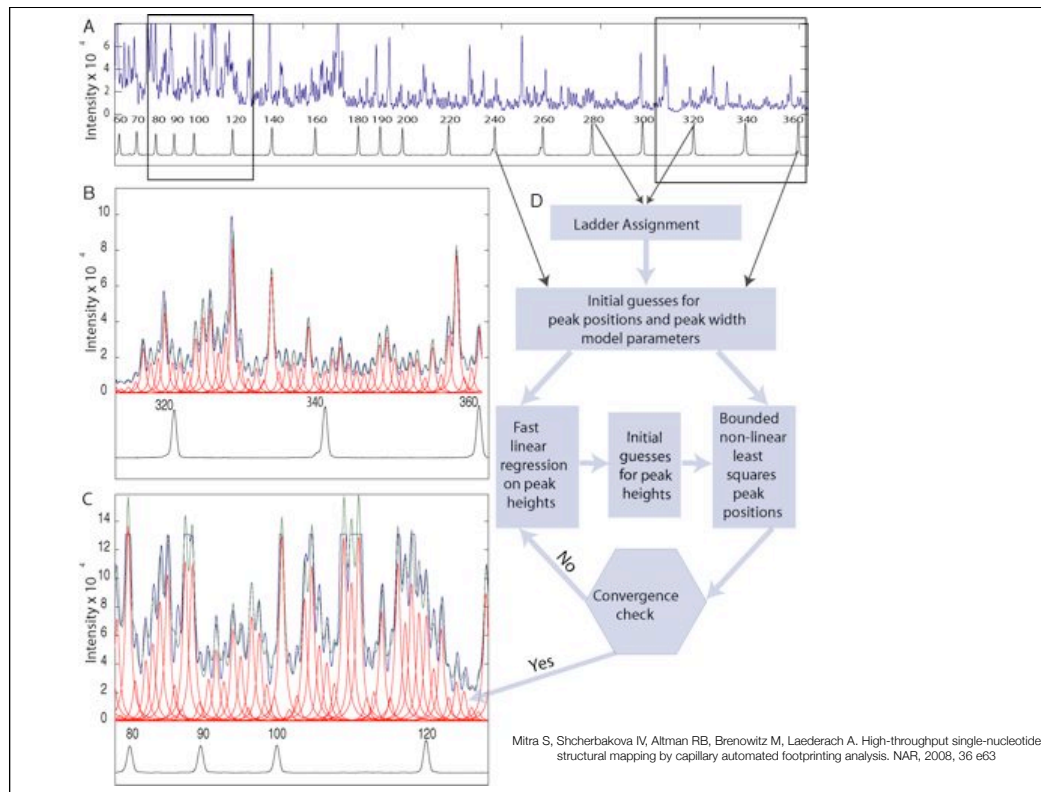
[alain@unc.edu](mailto:alain@unc.edu)

<http://ribosnitch.bio.unc.edu>

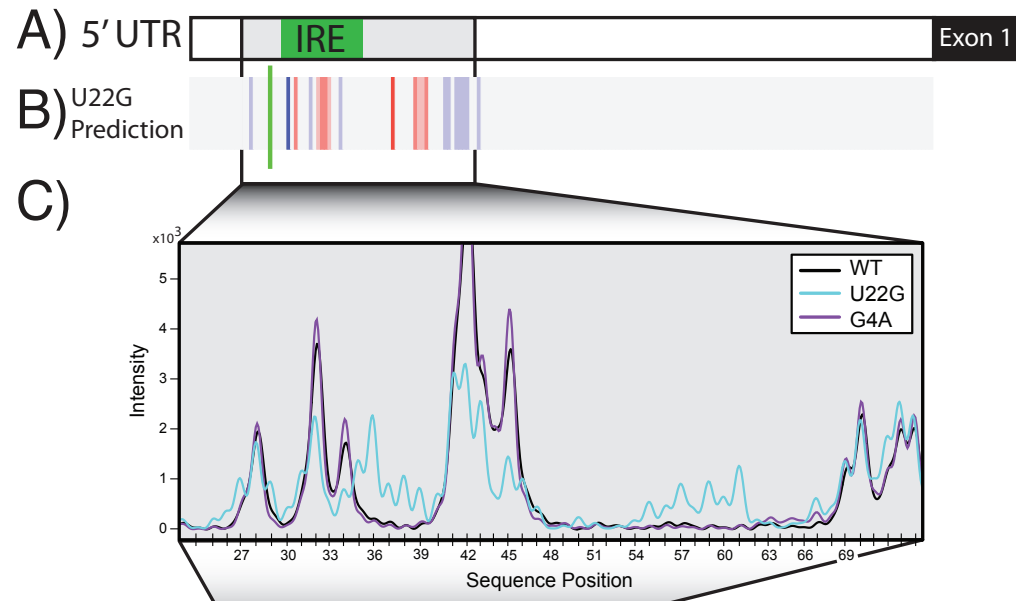
# SHAPE Chemical Mapping







## SHAPE Analysis of FTL 5' UTR



\*In our CAFA analysis, peaks are proportional to the amount of accessibility-dependent labeling in folded RNA at different positions

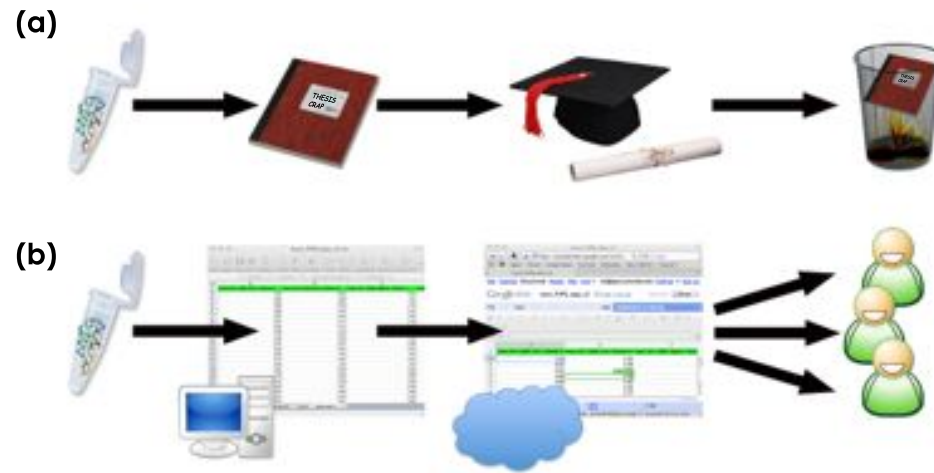
\*Peaks here represent unfiltered fluorescence intensities of the three variants of FTL 5'UTR that we will focus on today (WT, A196G and U22G)

\*The area that is being focused on here is representative of the IRE in the secondary structure of 5'UTR

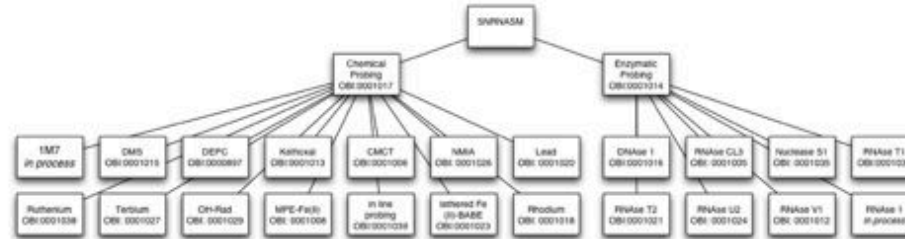
\*We see that even without filtration of datapoints, not only do WT and A196G capillary traces closely follow one another, but also U22G deviates significantly from the WT patterns of peaks shown

How do we get a lot more probing data?

---



# SNRNASM (Single Nucleotide Resolution Nucleic Acid Structure Mapping)



Listed below are some example data sets shared by different labs in the ISATab format. You can also obtain a blank ISATab file [here](#) and find the ontological references from the tutorial [here](#).

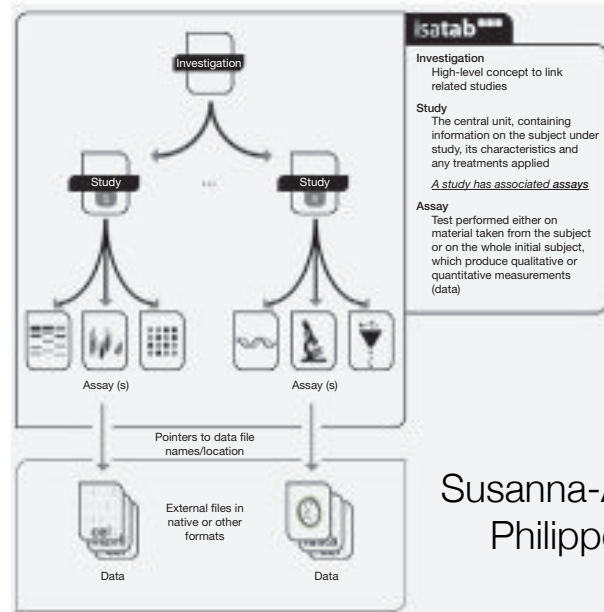
[Evaluation of the information content of RNA structure mapping data for secondary structure prediction](#). DMS and T1 structure probing of the Twort and Tetrahymena group I introns under low and high salt conditions.

[Atomic accuracy in predicting and designing noncanonical RNA structure](#). DMS probing of a synthetic RNA and three mutants under multiple Mg solution conditions.

[RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting](#). Time Resolved Hydroxyl Radical Footprinting of the Tetrahymena group I intron when folding in 10mM Mg.

[Nonhierarchical Ribonucleoprotein Assembly Suggests a Strain-Preparation Model for Protein-Facilitated RNA Folding](#). SHAPE probing data on maturase bound RNA (Mrs-1-bound RNA) in the presence and absence of Maturase.

# ISA-Tab is a Standard



Susanna-Assunta Sansone  
Philippe Rocca-Serra



# SNRNASM is a link farm to Google Spreadsheets

The screenshot shows a Google Spreadsheet titled "DMS-data.xls" with the following data:

Source Name	Characteristics (Nucleotide Sequence)	Charact. Type	Characteristics (RN Production)	Protocol REF
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_WT	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_C8U	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay
SRP-Domain-IV_C8U	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAA	RNA	in vitro synthesis	DMS structure mapping assay

# Providing open access to the data and meta-data.

The screenshot shows a Google Sheets spreadsheet titled "DMS-data.xls" with a URL: <https://docs.google.com/spreadsheets/cc?key=0AvCayBYdTclldJocQJotbWESRGx0VzdobmVjK20>. The spreadsheet contains a table with 19 rows and 19 columns. The first column is labeled "SRP-Domain-IV\_WT\_MgO2\_BmM" and the subsequent columns are labeled "SRP-Domain-IV\_WT\_MgO2\_BmM" through "SRP-Domain-IV\_WT\_MgO2\_BmM". The data is numerical, ranging from approximately 0.1632 to 2.7775. The spreadsheet interface includes a menu bar (File, Edit, View, Insert, Format, Data, Tools, Help, View only) and a toolbar with various editing tools. A "Share" button is visible in the top right corner.

SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM	SRP-Domain-IV_WT_MgO2_BmM
3.1632	2.6468	2.7917	2.8995	3.3488	3.4161	3.0297	4.4494	4.3895	3.3443	3.9589	3.6496	3.7952	3.5325	3.7031	4.2629			
0.7752	0.6625	0.6668	0.7139	0.7587	0.9356	2.0238	1.0417	1.0783	2.4965	1.5426	2.43	2.1564	1.7485	2.426	0.9652			
0.8715	0.6542	0.6594	0.6222	0.6457	0.6079	0.16	0.806	0.8706	0.3093	0.5344	0.1635	0.2873	0.2875	0.0606	0.7531			
0.6956	0.599	0.5986	0.5436	0.5671	0.5447	0.4625	0.7302	0.7161	0.5068	0.7365	0.4865	0.5957	0.5536	0.3926	0.6964			
1.8694	1.628	1.6021	1.6027	1.5791	1.5904	1.6023	1.7694	1.7981	1.673	1.5415	1.4349	1.402	1.1963	1.2652	1.3482			
0.5771	0.5581	0.5442	0.5203	0.5279	0.5415	0.5615	0.6245	0.6806	0.6048	0.5942	0.5366	0.5657	0.4826	0.4761	0.5146			
0.2189	0.1901	0.1989	0.1811	0.1928	0.2153	0.228	0.2507	0.2696	0.2461	0.2509	0.2541	0.2602	0.2194	0.2128	0.2219			
1.7793	1.5389	1.5182	1.4979	1.5209	1.6041	1.733	2.0141	2.1061	1.9912	2.1186	2.0432	2.0747	1.9658	1.8557	1.795			
1.9935	1.7228	1.7534	1.7279	1.7632	1.8105	1.9047	2.2399	2.3541	2.1165	2.1329	2.059	2.058	1.9091	1.9125	1.8441			
1.6205	1.4371	1.436	1.4261	1.4325	1.465	1.564	1.8443	1.8927	1.697	1.768	1.6721	1.6922	1.6252	1.6876	1.6561			
0.348	0.2546	0.2566	0.241	0.2586	0.2543	0.2636	0.3023	0.3736	0.338	0.3167	0.3208	0.3567	0.297	0.2903	0.3007			
0.3793	0.3132	0.3329	0.3026	0.3227	0.3535	0.3648	0.403	0.4447	0.4018	0.3993	0.4083	0.3897	0.3469	0.3429	0.3958			
1.7865	1.5499	1.5654	1.5429	1.5428	1.5849	1.655	1.918	1.9783	1.7541	1.7435	1.6059	1.6062	1.4915	1.5818	1.7448			
2.1472	1.9354	1.9688	1.949	1.9602	2.0358	2.1434	2.5652	2.7523	2.4213	2.4485	2.3517	2.3812	2.247	2.274	2.3596			
2.3146	2.1586	2.174	2.2254	2.3059	2.3564	2.4721	3.0049	3.2379	2.8967	2.9665	2.8722	2.9058	2.7629	2.7237	2.6805			
2.3057	2.1642	2.1695	2.2276	2.3404	2.3601	2.4776	2.9496	3.1753	2.8787	2.9357	2.8298	2.8402	2.7842	2.7779	2.7530			
2.3587	2.3444	2.2992	2.3864	2.4692	2.5977	2.7739	3.0853	3.0206	2.9743	3.0992	2.7514	2.6781	2.5445	2.284	2.3131			
2.2744	2.2077	2.2179	2.2913	2.3872	2.4768	2.6002	3.0535	3.1762	2.9694	3.0967	2.8831	2.8542	2.8196	2.7775	2.7127			
0.4705	0.4176	0.4364	0.4445	0.4644	0.4844	0.5044	0.5244	0.5444	0.5644	0.5844	0.6044	0.6244	0.6444	0.6644	0.6844			

# A community effort

BIOINFORMATICS

## Sharing and archiving nucleic acid structure mapping data

PHILIPPE ROCCA-SERRA,<sup>1</sup> STANISLAV BELLAOUSOV,<sup>2</sup> AMANDA BIRMINGHAM,<sup>3</sup> CHUNXIA CHEN,<sup>4</sup>  
PABLO CORDERO,<sup>5</sup> RHIJU DAS,<sup>5</sup> LAUREN DAVIS-NEULANDER,<sup>4</sup> CAIA DUNCAN,<sup>6</sup> MATTHEW HALVORSEN,<sup>4</sup>  
ROB KNIGHT,<sup>7</sup> NEOCLES LEONTIS,<sup>8</sup> DAVID H. MATHEWS,<sup>2</sup> JUSTIN RITZ,<sup>4</sup> JESSE STOMBAUGH,<sup>7</sup>  
KEVIN M. WEEKS,<sup>6</sup> CRAIG L. ZIRBEL,<sup>9</sup> and ALAIN LAEDERACH<sup>4,10</sup>

<sup>1</sup>Oxford e-Research Center, University of Oxford, OX1 3QG, Oxford, United Kingdom

<sup>2</sup>Department of Biochemistry and Biophysics and Center for RNA Biology, University of Rochester, Rochester, New York 14642, USA

<sup>3</sup>Thermo Fisher Scientific, Lafayette, Colorado 80026, USA

<sup>4</sup>Biology Department, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA

<sup>5</sup>Biochemistry Department, Stanford University, Stanford, California 94305, USA

<sup>6</sup>Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599, USA

<sup>7</sup>Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

<sup>8</sup>Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, USA

<sup>9</sup>Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403, USA

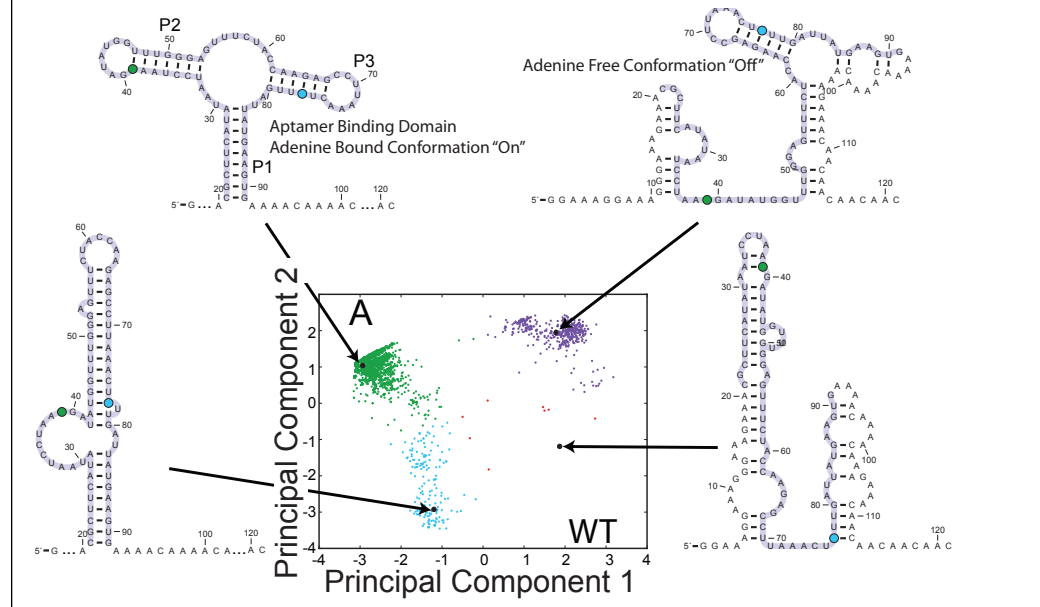
### 15 ABSTRACT

**ABSTRACT** Nucleic acids are particularly amenable to structural characterization using chemical and enzymatic probes. Each individual structure mapping experiment reveals specific information about the structure and/or dynamics of the nucleic acid. Currently, there is no simple approach for making these data publicly available in a standardized format. We therefore developed a standard for reporting the results of single nucleotide resolution nucleic acid structure mapping experiments, or SNRNASMs. We propose a schema for sharing nucleic acid chemical probing data that uses generic public servers for storing, retrieving, and searching the data. We have also developed a consistent nomenclature (ontology) within the Ontology of Biomedical Investigations (OBI), which provides unique identifiers (termed persistent URLs, or PURLs) for classifying the data. Links to standardized data sets shared using our proposed format along with a tutorial and links to templates can be found at <http://snrnasm.bio.unc.edu>.

**25 Keywords:** RNA structure; chemical mapping; secondary structure

RNA (2011), 17:00–00. Published by Cold Spring Harbor Laboratory Press. Copyright © 2011 RNA Society.

# On/off conformations in Boltzmann sampling



# SNitching the Switch

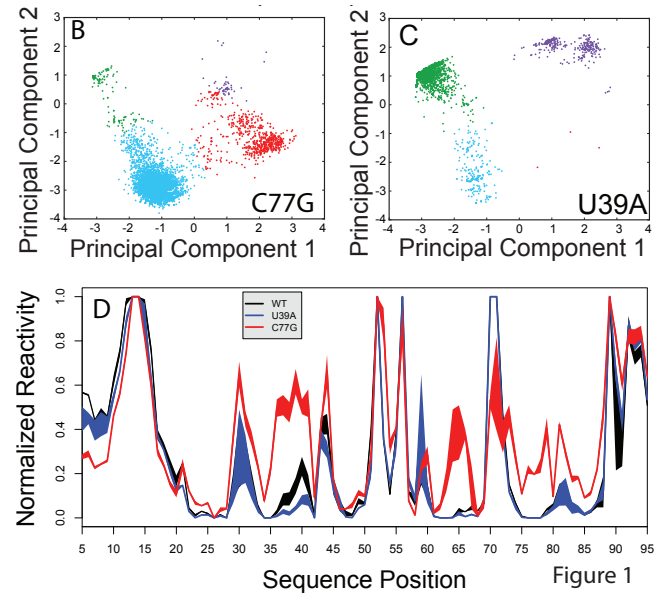
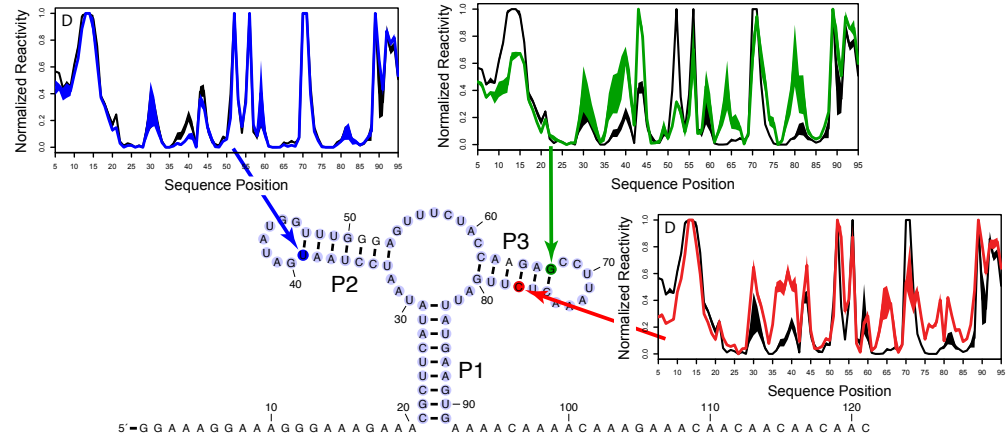


Figure 1

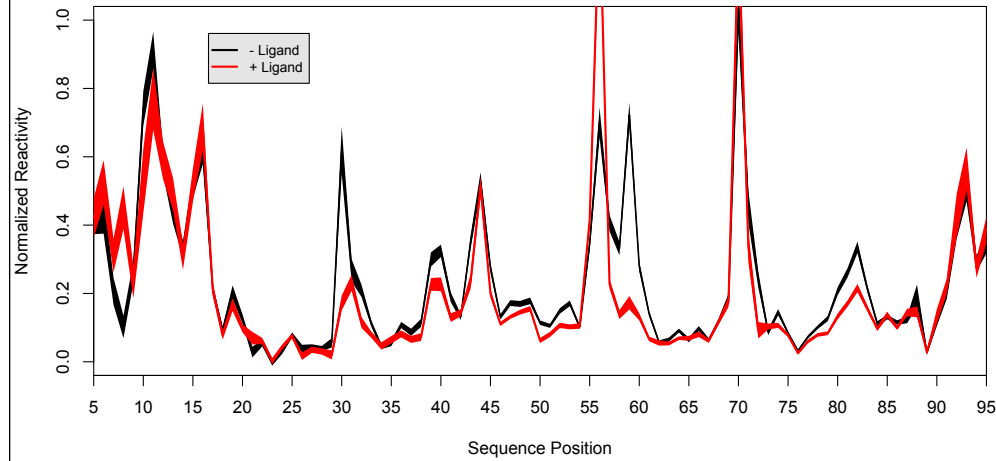
# SHAPE data for most single point mutations!



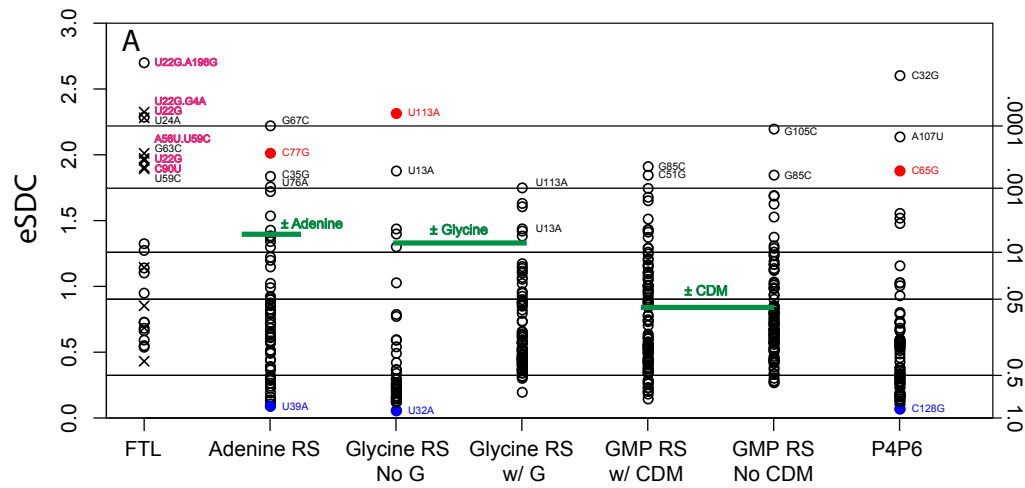
## Quantifying the structure change

$$eSDC = (1 - pCC) \times \sqrt{n}$$

eSDC = 1.45, Biological "significance."







# eSDC, the experimental Structure Disruption Coefficient

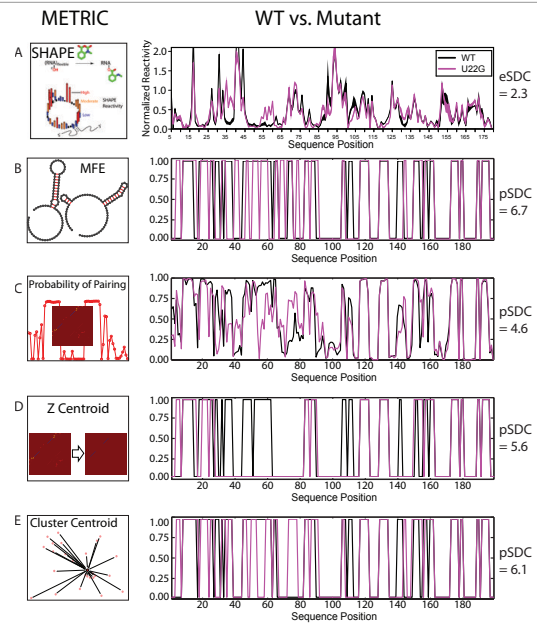




# Predicting SNitch inducing SNPs

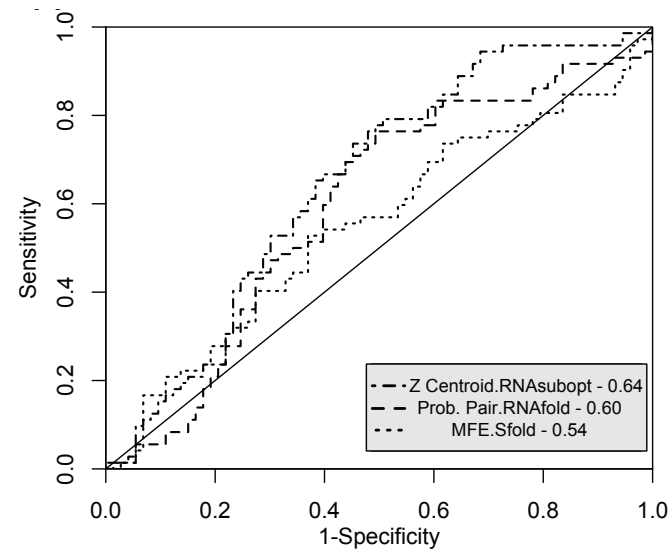
 <p>Minimal Free Energy Structure RNAfold -naPS &lt; seq.fa &gt; outfile.txt</p> <p>Stochastic Sampled Structures RNAsubopt -p 10000 -s &lt; seq.fa &gt; outfile.txt</p> <p>Suboptimal Structures RNAsubopt -e E -s &lt; seq.fa &gt; outfile.txt</p>	 <p>Minimal Free Energy Structure Fold seq.seq outfile.ct -m 1</p> <p>Stochastic Sampled Structures partition seq.seq outfile.pfs ProbabilityPlot outfile.pfs outfile.txt -t</p> <p>Suboptimal Structures AllSub seq.seq outfile.ct -a E</p>
 <p>Minimal Free Energy Structure sfold -o outdir seq.fa run 10 times find MFE structure</p> <p>Stochastic Sampled Structures sfold -o outdir seq.fa run 10 times</p>	 <p>Minimal Free Energy Structure RNAmutants -l ~/RNAmutants /lib/ -f seq.fa -mutants 1 -n 1 &gt; outfile.txt</p> <p>Mutant Structures RNAmutants -l ~/RNAmutants /lib/ -f seq.fa -mutants 1 -n 10000 &gt; outfile.txt</p>

# pSDC, or the predicted Structure Disruption Coefficient

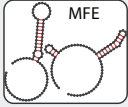

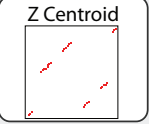
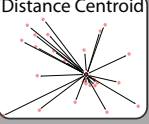


## PPV and AUC analysis of predictions

---



AUC reveals there is still work to be done.

Metric	Program	AUC +/- .03
 MFE	RNAfold	0.62
	RNAmutants	0.61
	RNAstructure	0.57
	sFold	0.54
	mFold	0.62
 Probability of Pairing	Allsub	0.59
	RNAfold	0.60
	RNAmutants	0.59
	RNAstructure	0.59
	RNAsubopt	0.61
 Z Centroid	Allsub	0.56
	RNAfold	0.62
	RNAmutants	0.58
	RNAstructure	0.56
	RNAsubopt	0.64
 Distance Centroid	Allsub	0.57
	RNAfold	0.63
	RNAmutants	0.59
	RNAstructure	0.58
	RNAsubopt	0.62
	sFold	0.55

## Open Questions

---

- What exactly is SHAPE telling us? Notice we only used it comparatively, i.e. does the structure change and how much?
- Is there information in partial SHAPE (or for that matter any probing) reactivities or do we just need to know the most reactive species?
- Why do we correctly predict that there are major structure disrupting SNPs but no one gets the right ones?
- Are hyper-reactive nucleotides even structurally important or just an artifact of ideal acylation geometry?

R00 GM079953 (NIGMS),  
R21 MH087336 (NIMH),  
R01 HL111527 (NHLBI),  
R01 GM101237 (NGMS).



Matt Halvorsen

Wes Sanders

Justin Ritz

Josh Martin

Chetna Gopinath

Gabriela Phillips