



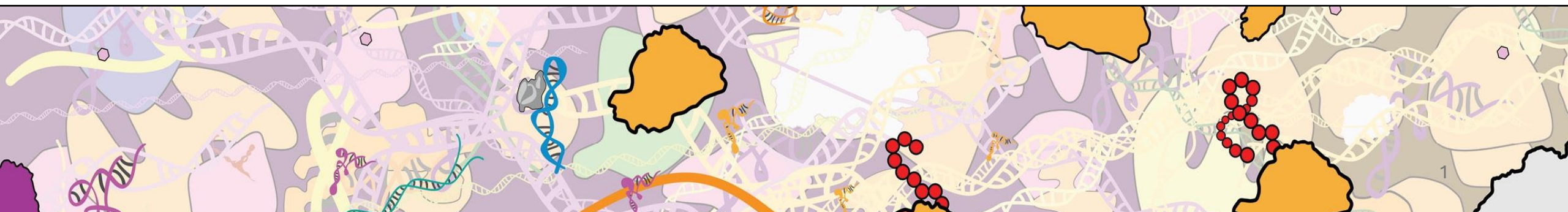
Expanding covariation analysis for riboswitches and gene editing RNA-protein systems

Laura Marie Hertz, Ph.D.

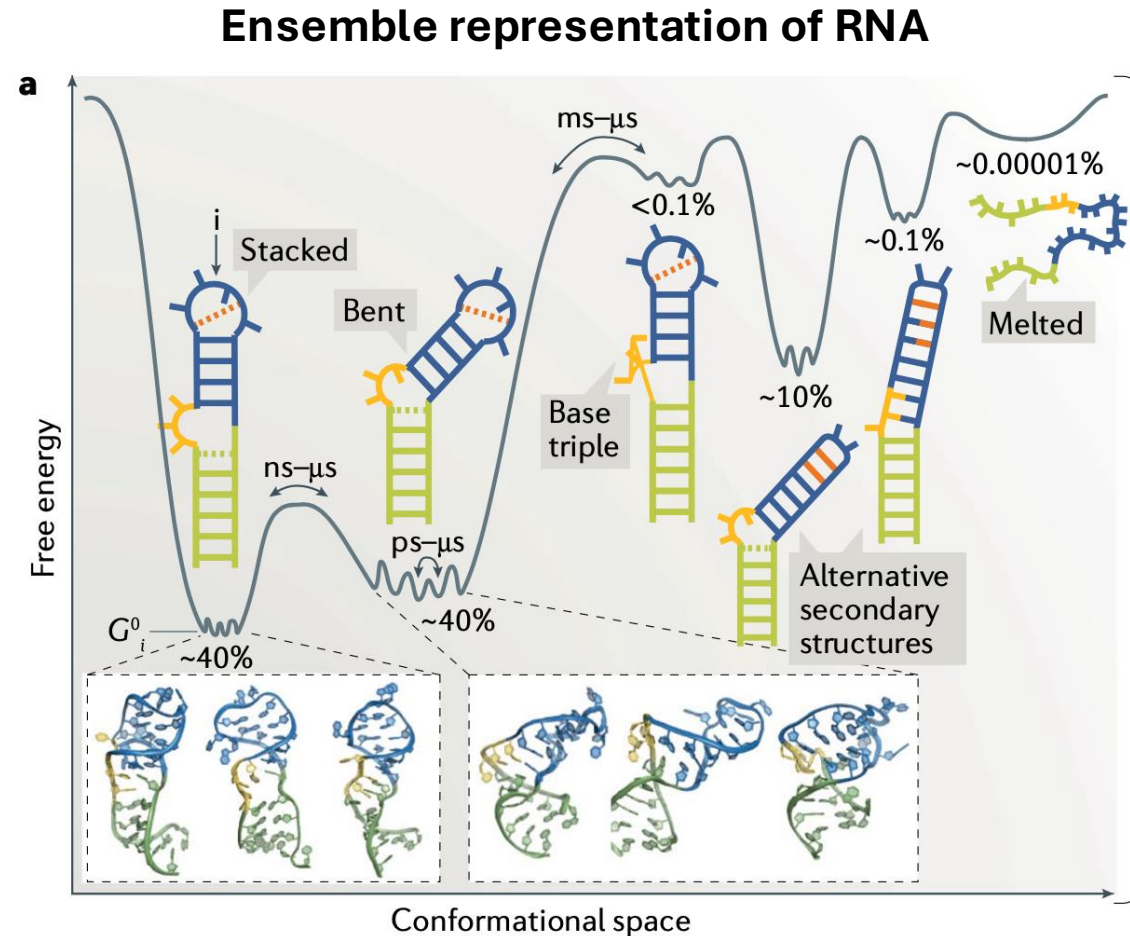
Julius Lucks' lab (PhD) → Eric Chan's lab (Research Fellow)

16 July 2026

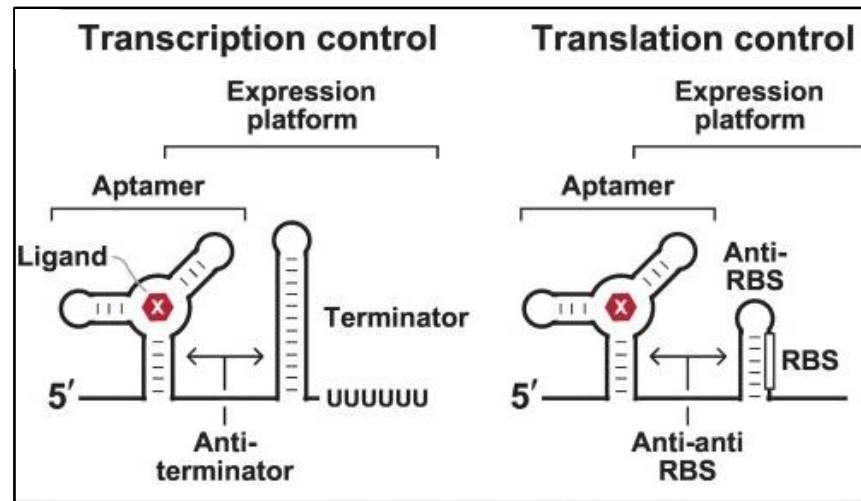
RNA Benasque 2026



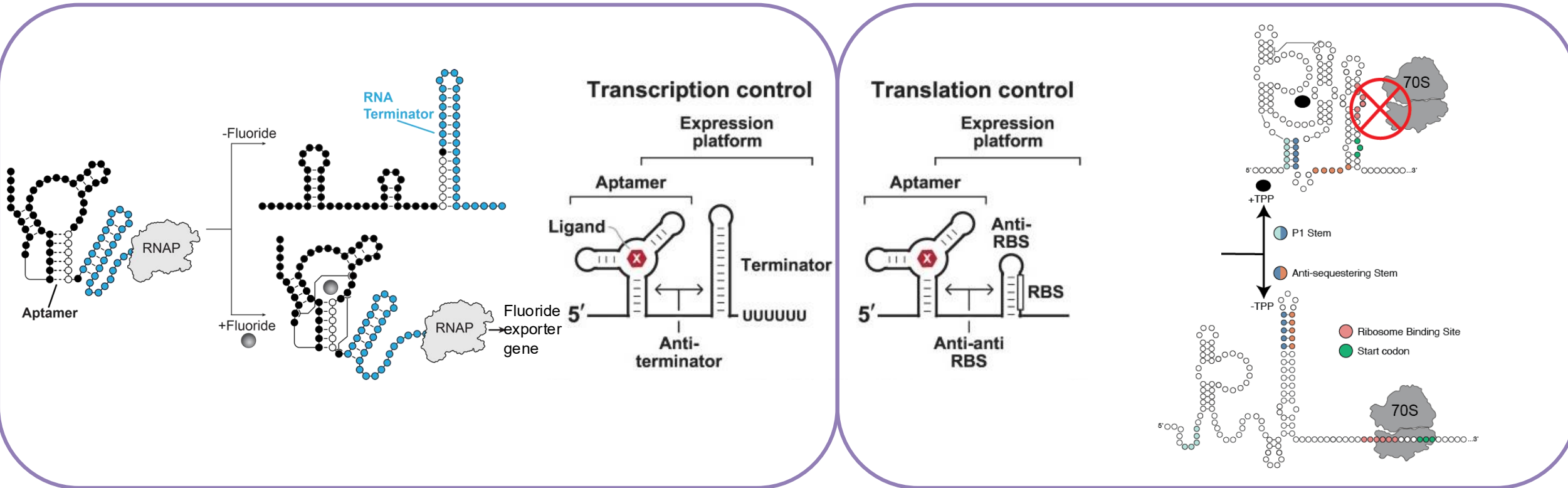
Ground truth: RNA and complex and fun



Riboswitches exploit RNA dynamics to regulate gene expression



Riboswitches exploit RNA dynamics to regulate gene expression



2020 | Riboswitches + Rfam = Problem?

Rfam HOME SEARCH BROWSE FTP BLOG HELP Rfam's 20th COVID-19 Search Rfam Q Search

Family: Fluoride (RF01734)
Description: Fluoride riboswitch (*crcB*)

2382 sequences 1643 species 6 structures

Summary

Wikipedia annotation [Edit Wikipedia article](#)

The Rfam group coordinates the annotation of Rfam families in [Wikipedia](#). This family is described by a Wikipedia entry [CrcB RNA motif](#). [More...](#)

The **fluoride riboswitch** (formerly called the ***crcB* RNA motif**) is a conserved RNA structure identified by [bioinformatics](#) in a wide variety of [bacteria](#) and [archaea](#). These RNAs were later shown to function as [riboswitches](#) that sense [fluoride ions](#). These "fluoride riboswitches" increase expression of downstream genes when fluoride levels are elevated, and the genes are proposed to help mitigate the toxic effects of very high levels of fluoride.

Many genes are presumed to be regulated by these fluoride riboswitches. Two of the most common encode proteins that are proposed to function by removing fluoride from the cell. These proteins are CrcB proteins and a fluoride-specific subtype of chloride channels referred to as EriC or ClC. ClC proteins have been shown to function as fluoride-specific fluoride/proton antiporters.

The three-dimensional structure of a fluoride riboswitch has been solved at atomic resolution by X-ray crystallography.

Fluoride riboswitches are found in many organisms within the domains [bacteria](#) and [archaea](#), indicating that many of these organisms sometimes encounter elevated levels of fluoride. Of particular interest is *Streptococcus mutans*, a major cause of [dental caries](#). It has been shown that sodium fluoride has inhibited the growth rate of *S. mutans* using glucose as an energy and carbon source. However, it is also noteworthy that many organisms that do not encounter fluoride in the human mouth carry fluoride riboswitches or resistance genes.

Discovery of the fluoride riboswitch

The identity of fluoride as the riboswitch ligand was accidentally discovered when a compound contaminated with fluoride caused significant conformational changes to the non-coding *crcB* RNA motif during an in-line probing experiment. In-line probing was used to illuminate the

crcB RNA motif

base pair annotations
— covering mutations
— compatible mutations
— no mutations observed

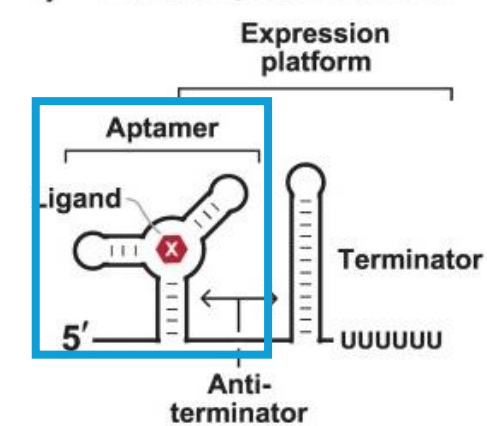
variable length region
variable length stem loop
modular sub-structure

nucleotide identity
● 97% ● 75% ● 50% ● 25% ● 10% ● 5% ● 1% ● 0%

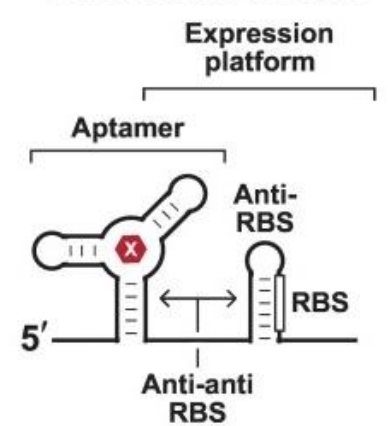
R = A or G, Y = C or U, "nt" = nucleotides, "PS" = pairing element 1.

Consensus secondary structure of *crcB* RNAs

A) Transcription control



Translation control



Aptamer != Riboswitch

So what? Any methods using Rfam to train for riboswitches, are operating under an incorrect assumption

Identification of potential riboswitch elements in *Homo sapiens* mRNA 5'UTR sequences using positive-unlabeled machine learning

William S Raymond , Jacob DeRoo, Brian Munsky

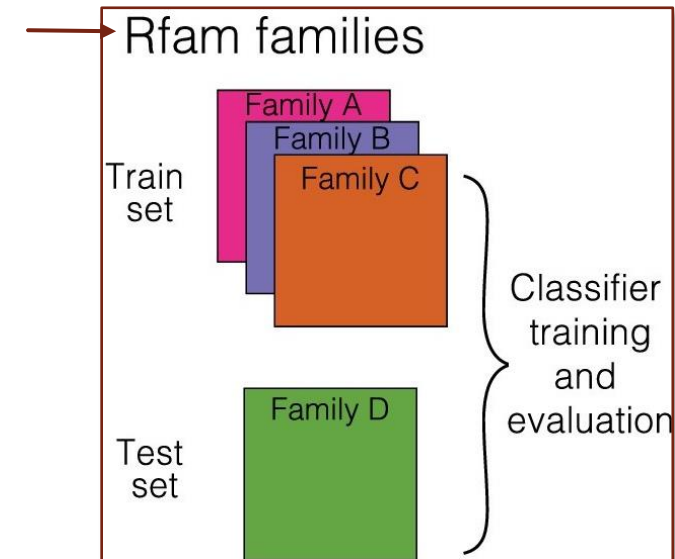
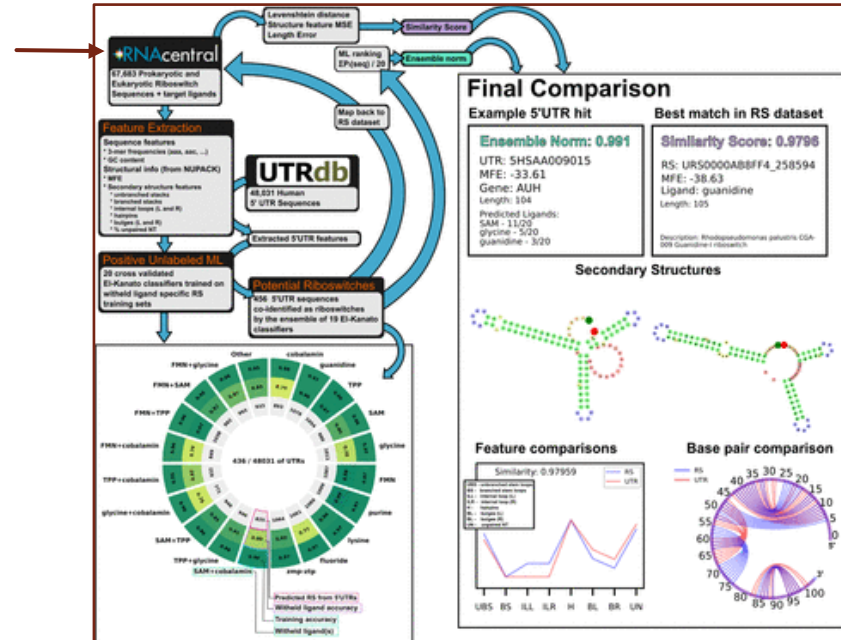
Published: April 24, 2025 • <https://doi.org/10.1371/journal.pone.0320282>

Article | [Open access](#) | Published: 16 July 2024

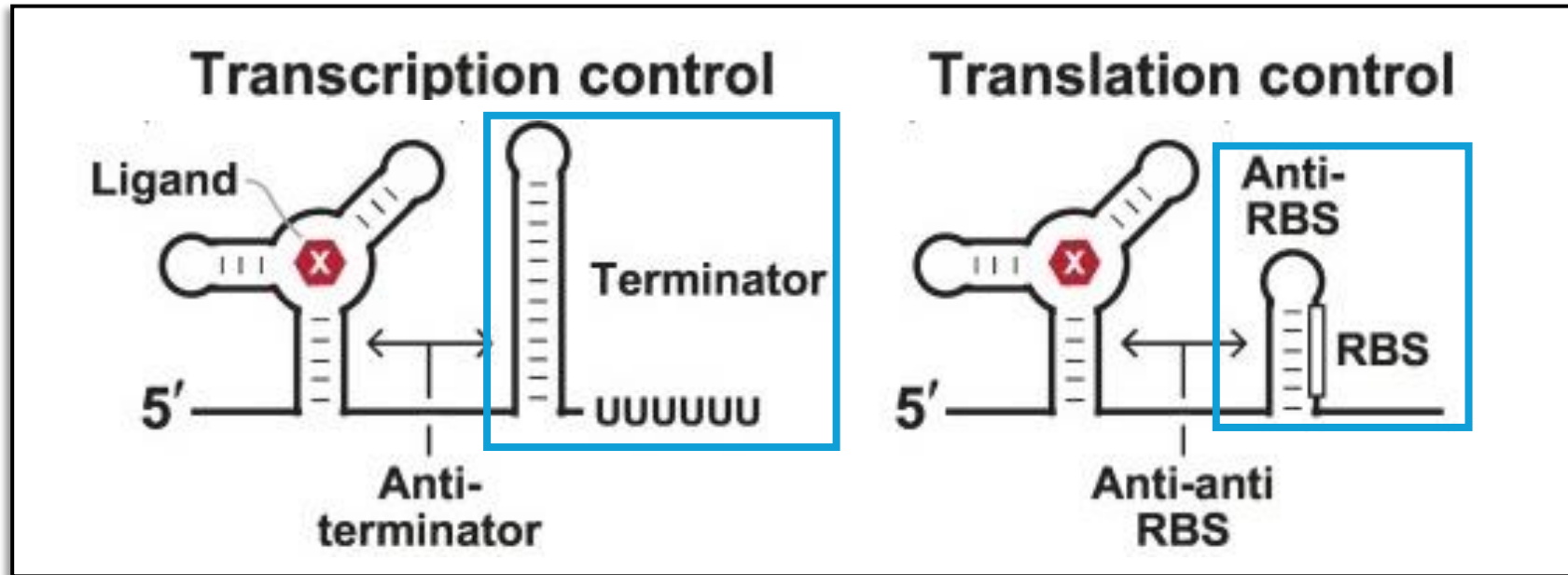
A systematic search for RNA structural switches across the human transcriptome

[Matvei Khoroshkin](#), [Daniel Asarnow](#), [Shaopu Zhou](#), [Albertas Navickas](#), [Aidan Winters](#), [Jackson Goudreau](#), [Simon K. Zhou](#), [Johnny Yu](#), [Christina Palka](#), [Lisa Fish](#), [Ashir Borah](#), [Kian Yousefi](#), [Christopher Carpenter](#), [K. Mark Ansel](#), [Yifan Cheng](#), [Luke A. Gilbert](#) & [Hani Goodarzi](#) 

Nature Methods 21, 1634–1645 (2024) | [Cite this article](#)



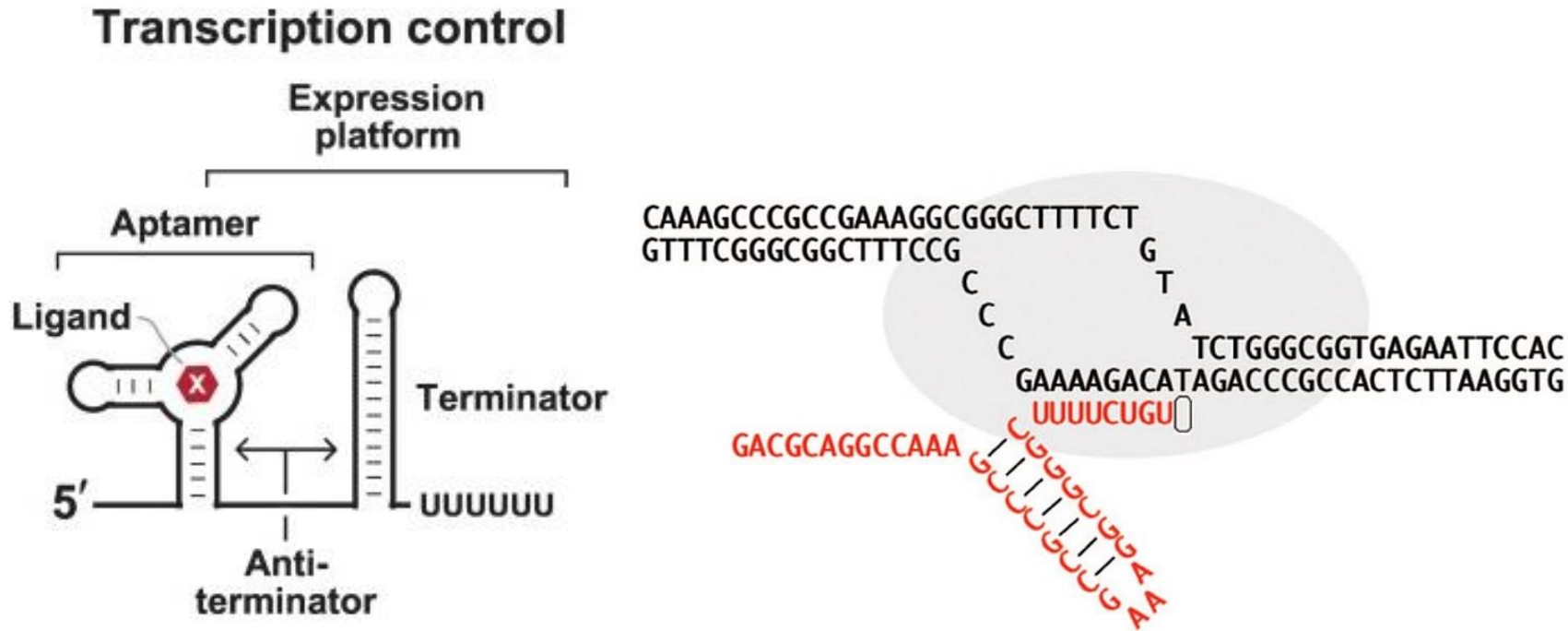
Why? Lack of uniformed function



This function variation interrupts the sequence-structure-function relationship that covariation detects

Sequence > Structure > Function

Sequence < Structure < Function



Riboswitch
covariation solution:
Restrict riboswitch
variants to those that
undergo intrinsic
termination.

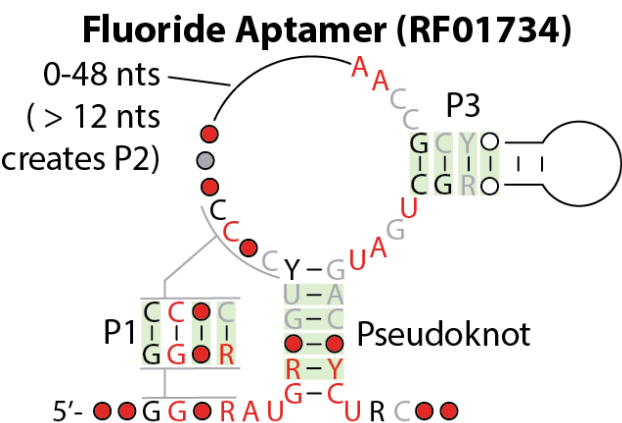
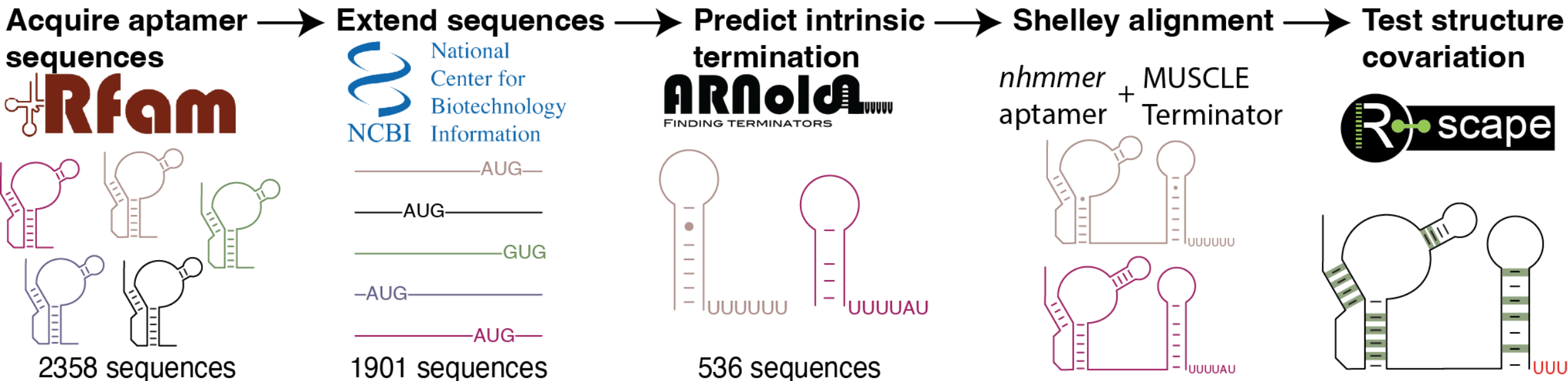
Kavita and Breaker (2023) *Trends in Biochemical Sciences*

Roberts (2019) *JMB*

Pipeline: The Shelley Alignment



Prof. Elena Rivas



Naville, M., et. al (2011) *RNA biology*

A portrait of a man with dark hair, smiling, wearing a black blazer over a black shirt. He is seated at a table, with his hands clasped in front of him. The background is a textured, greyish-blue wall.

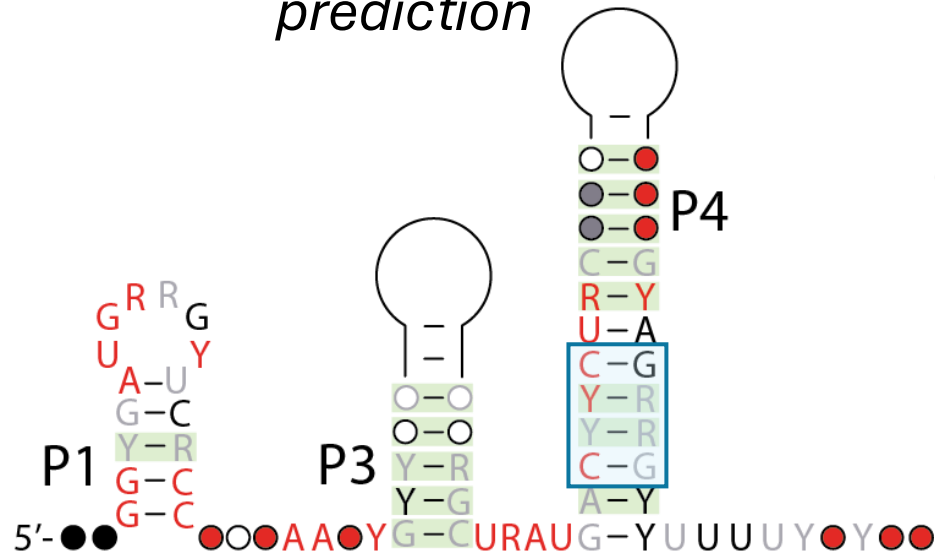
Acquire aptamer sequences **Rfam** → Extend sequences **NCBI National Center for Biotechnology Information** → Predict intrinsic termination **ARNold** → Shelley alignment *nhmmer* aptamer + MUSCLE Terminator → Test structure covariation **R-scape**



Additional successful approaches:

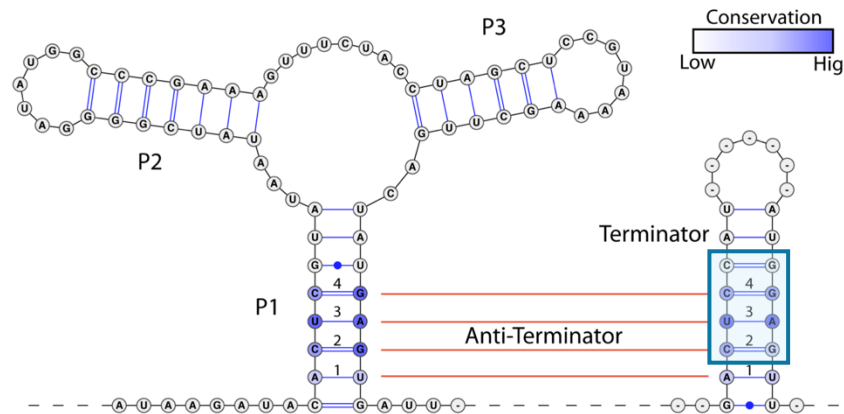
Fluoride

High-throughput prediction



Purine

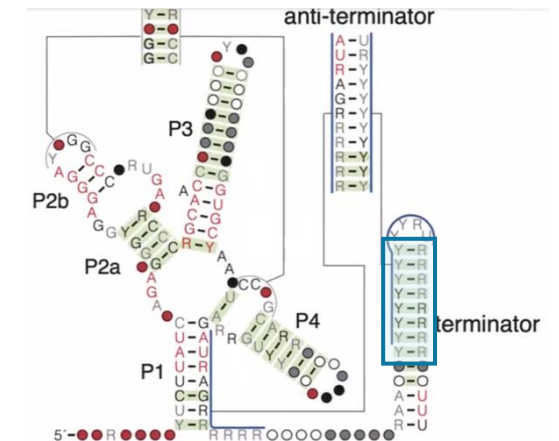
*Curated based
on crystal structure*



Ritz, et. al., (2013) *PLOS Comp Bio*

SAM

Curated



Zhu & Meyer (2015) *RNA Biology*
Specific representation provided by
Prof. Elena Rivas

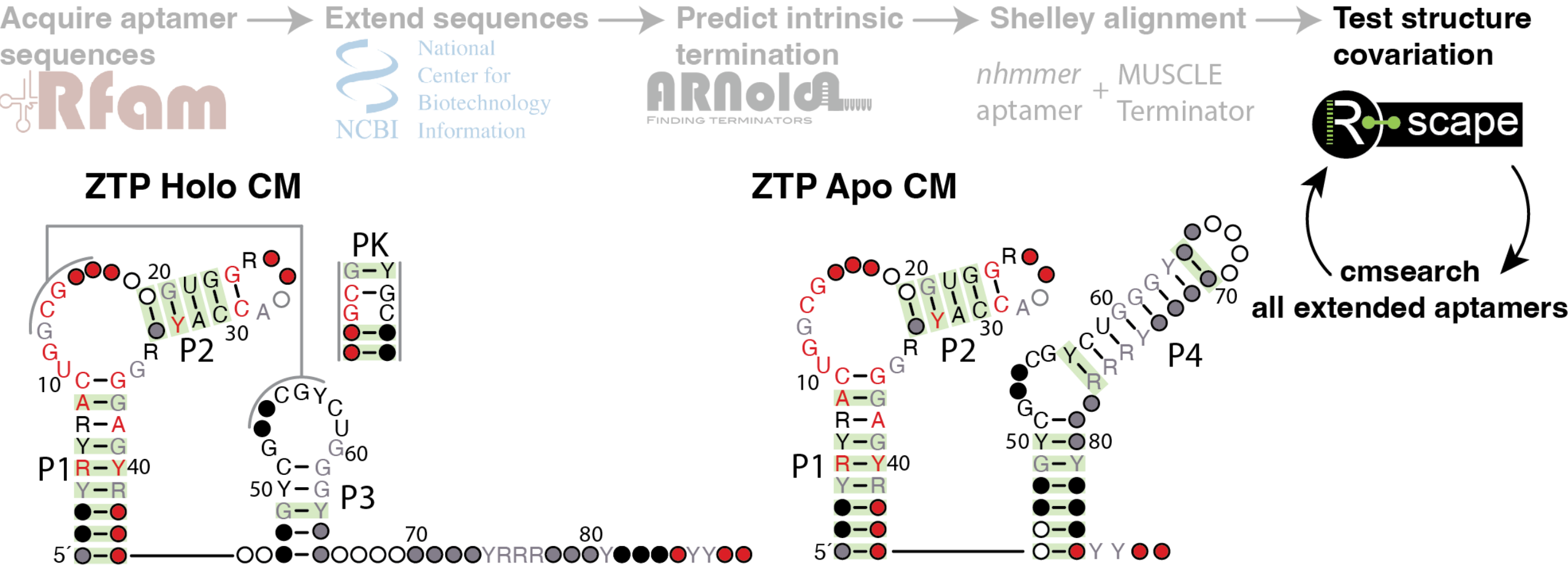
Y-R motif in terminator atop polyU

Not observed with general terminators

Iteration of the method

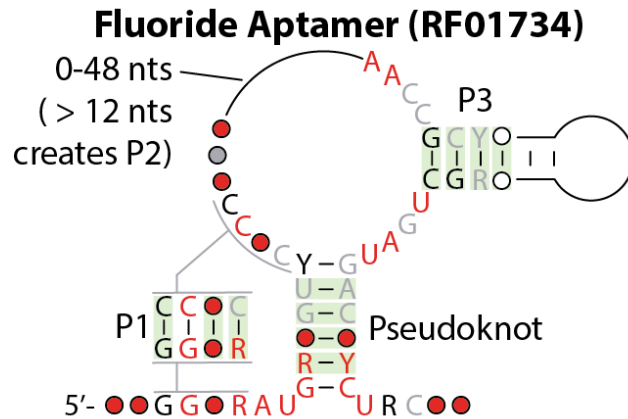


Prof. Elena Rivas





What Did Not work:



RM00022 or RM00023

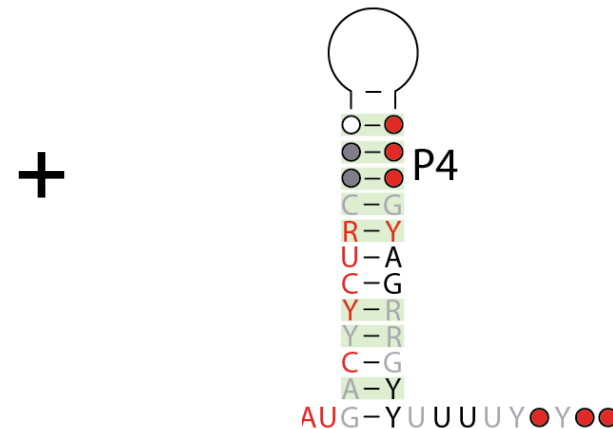
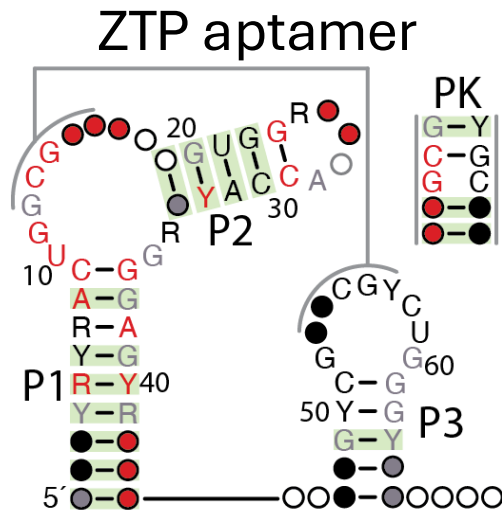
```

# STOCKHOLM 1.0

#=#G ID Terminator2
#=#G DE Rho independent terminator 2
#=#G PT Intrinsic_terminator; RIT
#=#G AU Gardner PP
#=#G SE Published; PMID:21478170
#=#G SS Published; PMID:21478170
#=#G GA 18.00
#=#G DR URL:http://github.com/ppgardne/RNIE
#=#G RN [1]
#=#G RM 21478170
#=#G RT RNIE: genome-wide prediction of bacterial intrinsic terminators.
#=#G RA Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z
#=#G RA Nucleic Acids Res. 2011;39:5045-52.
#=#G WK Intrinsic_termination
#=#G FT TP Stem Loop

rnies.7 -----CGCA-GU...GAA.....UUC...-GCGCC-G-GGU.....GGUU,UUUU...
rnies.8 -----CGCA-GG...UG.....UCA.....CCA...-CCGCG-C-UUU.....UU-UUUUA...
rnies.9 -----CGCA-GG...UG.....UCA.....CCA...-CCGCG-C-UUU.....UU-UUUUA...

```



2024 | Riboswitches + Rfam = Solved?

2020 | Previously



Rfam HOME SEARCH BROWSE FTP BLOG HELP Rfam's 20th COVID-19 Search Rfam Q Search

Family: Fluoride (RF01734)
Description: Fluoride riboswitch (crcB)

2382 sequences 1643 species 6 structures

Summary

Sequences

Alignment

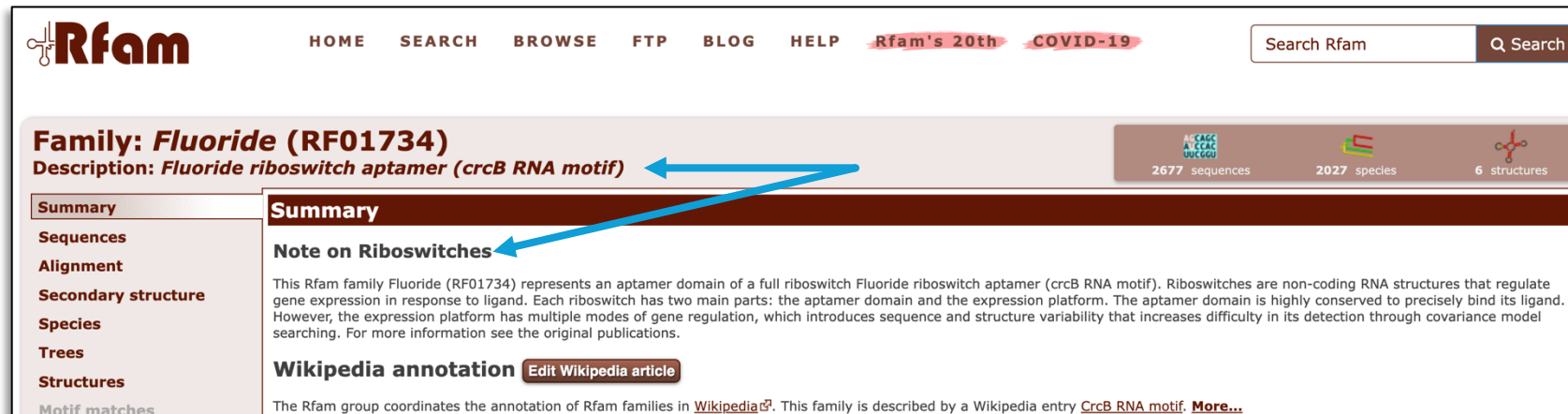
Secondary structure

Summary

Wikipedia annotation [Edit Wikipedia article](#)

The Rfam group coordinates the annotation of Rfam families in [Wikipedia](#). This family is described by a Wikipedia entry [CrcB RNA motif](#). [More...](#)

Now



Rfam HOME SEARCH BROWSE FTP BLOG HELP Rfam's 20th COVID-19 Search Rfam Q Search

Family: Fluoride (RF01734)
Description: Fluoride riboswitch aptamer (crcB RNA motif)

2677 sequences 2027 species 6 structures

Summary

Sequences

Alignment

Secondary structure

Species

Trees

Structures

Motif matches

Summary

Note on Riboswitches

This Rfam family Fluoride (RF01734) represents an aptamer domain of a full riboswitch Fluoride riboswitch aptamer (crcB RNA motif). Riboswitches are non-coding RNA structures that regulate gene expression in response to ligand. Each riboswitch has two main parts: the aptamer domain and the expression platform. The aptamer domain is highly conserved to precisely bind its ligand. However, the expression platform has multiple modes of gene regulation, which introduces sequence and structure variability that increases difficulty in its detection through covariance model searching. For more information see the original publications.

Wikipedia annotation [Edit Wikipedia article](#)

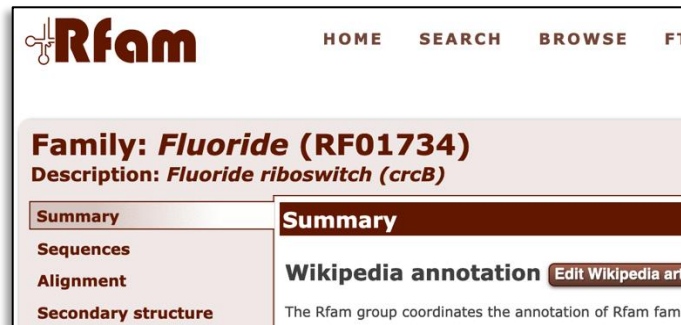
The Rfam group coordinates the annotation of Rfam families in [Wikipedia](#). This family is described by a Wikipedia entry [CrcB RNA motif](#). [More...](#)



Dr. Blake Sweeney

Supplying terminator alignments to Rfam

2020 | Previously



Rfam HOME SEARCH BROWSE FT

Family: Fluoride (RF01734)
Description: Fluoride riboswitch (*crcB*)

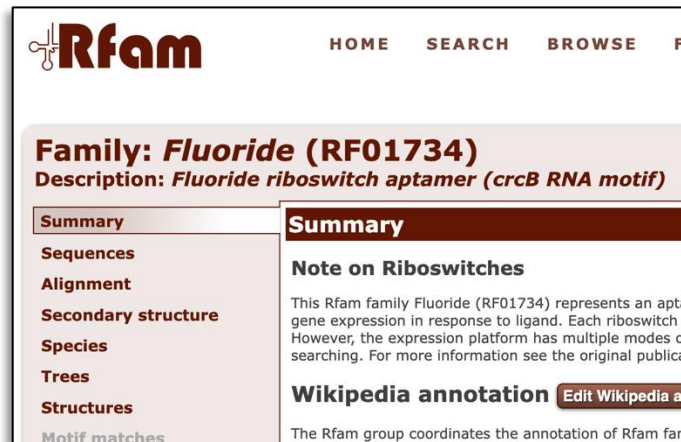
Summary Sequences Alignment Secondary structure

Summary

Wikipedia annotation [Edit Wikipedia a](#)

The Rfam group coordinates the annotation of Rfam fam

Now



Rfam HOME SEARCH BROWSE FT

Family: Fluoride (RF01734)
Description: Fluoride riboswitch aptamer (*crcB* RNA motif)

Summary Sequences Alignment Secondary structure Species Trees Structures Motif matches

Summary

Note on Riboswitches

This Rfam family Fluoride (RF01734) represents an apt gene expression in response to ligand. Each riboswitch. However, the expression platform has multiple modes o searching. For more information see the original publica

Wikipedia annotation [Edit Wikipedia a](#)

The Rfam group coordinates the annotation of Rfam fam

Generated transcriptional motifs for 20/49 Rfam aptamer entries

ID	Description	Alignment File
RF00050	FMN riboswitch aptamer (RFN element)	RF00050_Transcriptional_Riboswitch.sto
RF00059	TPP riboswitch aptamer (THI element)	RF00059_Transcriptional_Riboswitch.sto
RF00080	Manganese Mn ²⁺ riboswitch aptamer (yybP-ykoY)	RF00080_Transcriptional_Riboswitch.sto
RF00162	SAM riboswitch aptamer (S box leader)	RF00162_Transcriptional_Riboswitch.sto
RF00167	Purine riboswitch aptamer	RF00167_Transcriptional_Riboswitch.sto
RF00168	Lysine riboswitch aptamer	RF00168_Transcriptional_Riboswitch.sto
RF00174	Cobalamin riboswitch aptamer	RF00174_Transcriptional_Riboswitch.sto
RF00234	glmS glucosamine-6-phosphate activated ribozyme ap	No Terminators
RF00379	c-di-AMP riboswitch aptamer (ydaO/yuaA leader)	RF00379_Transcriptional_Riboswitch.sto
RF00380	M-box riboswitch aptamer (ykoK leader)	RF00380_Transcriptional_Riboswitch.sto

Nancy currently working on
[current feedback is to update the genomic coordinates]



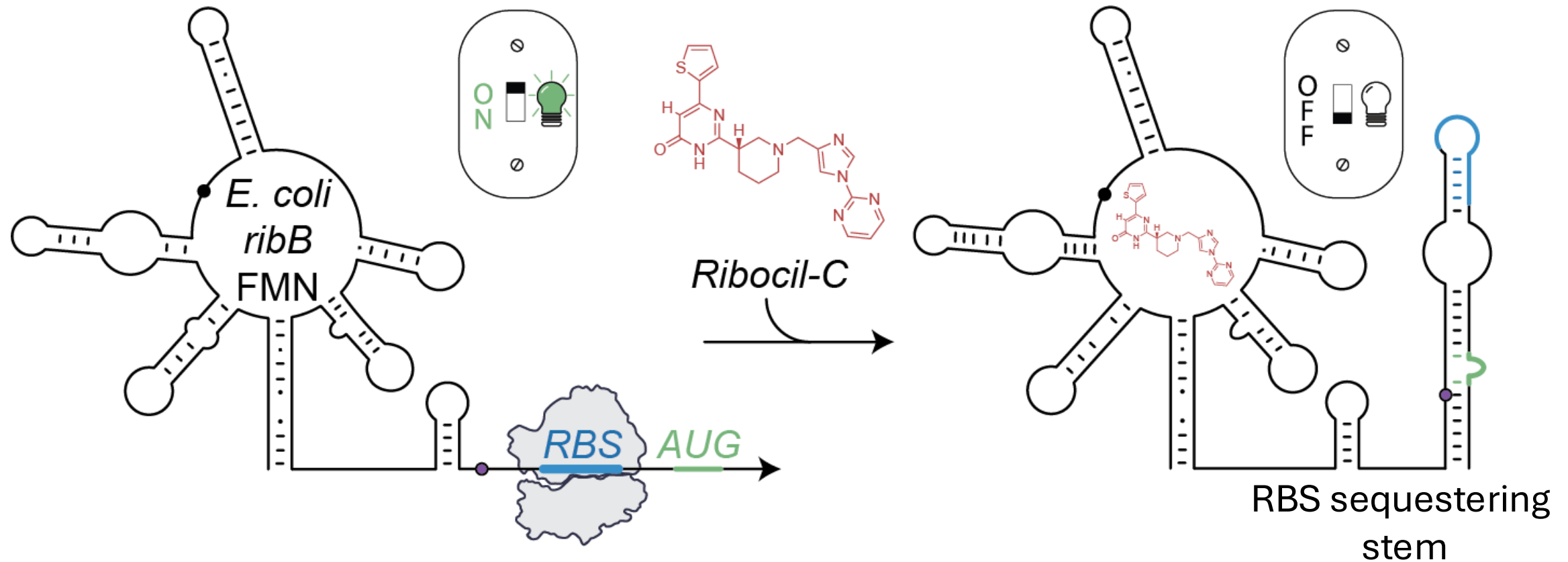
Nancy Ontiveros



Dr. Anibal Arce

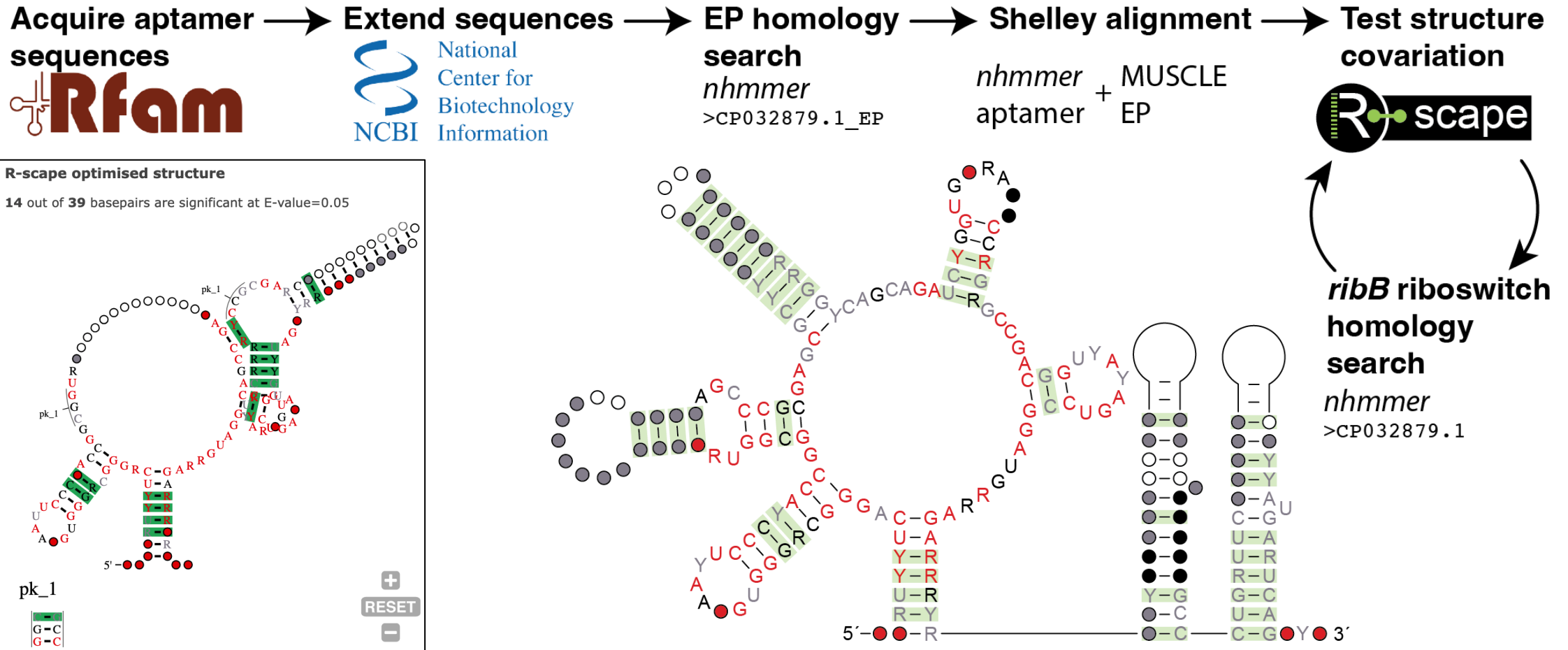
What about a **translational** riboswitch?

Case study: *E. coli* *ribB* FMN riboswitch



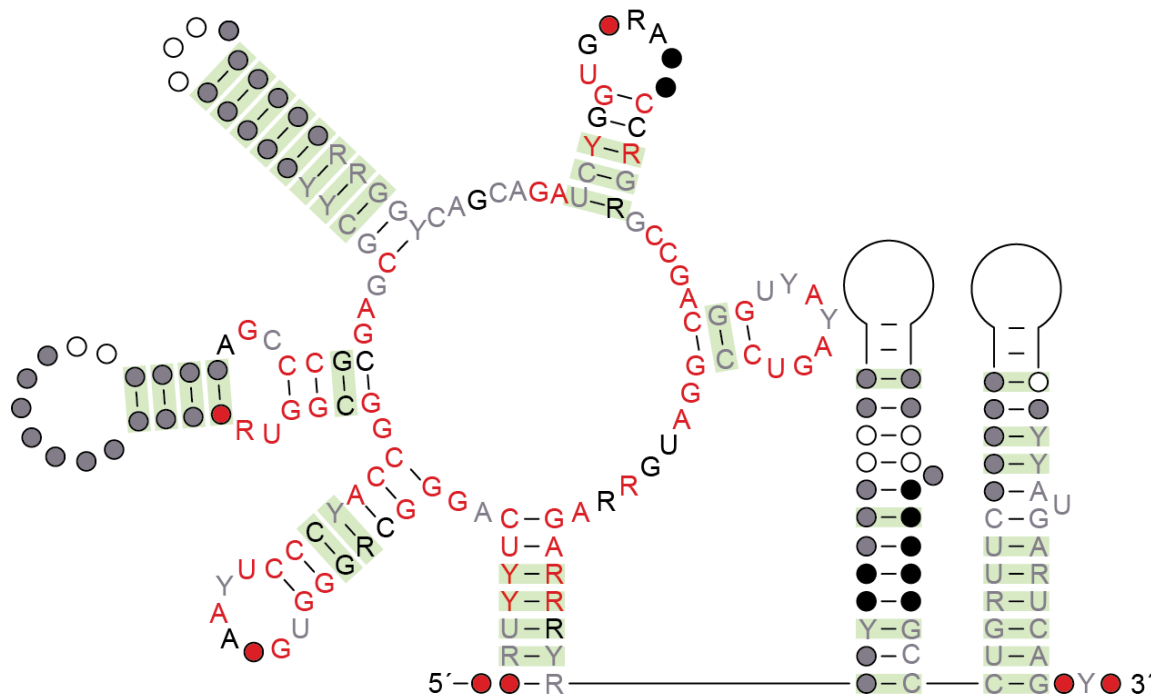
Adapting the Shelley Method for translation

Case study: *E. coli* *ribB* FMN riboswitch



Translational motif less informative of dynamics

Case study: *E. coli ribB* FMN riboswitch



- 1) No covariation information regarding switching mechanism
- 2) Conserved “U” budge in the start codon
- 3) Intermediate helix between aptamer and translational sequestering stem
- 4) Hairpin covariation of the first two codons

Riboswitches



Gene editing
therapeutics



Transition Time

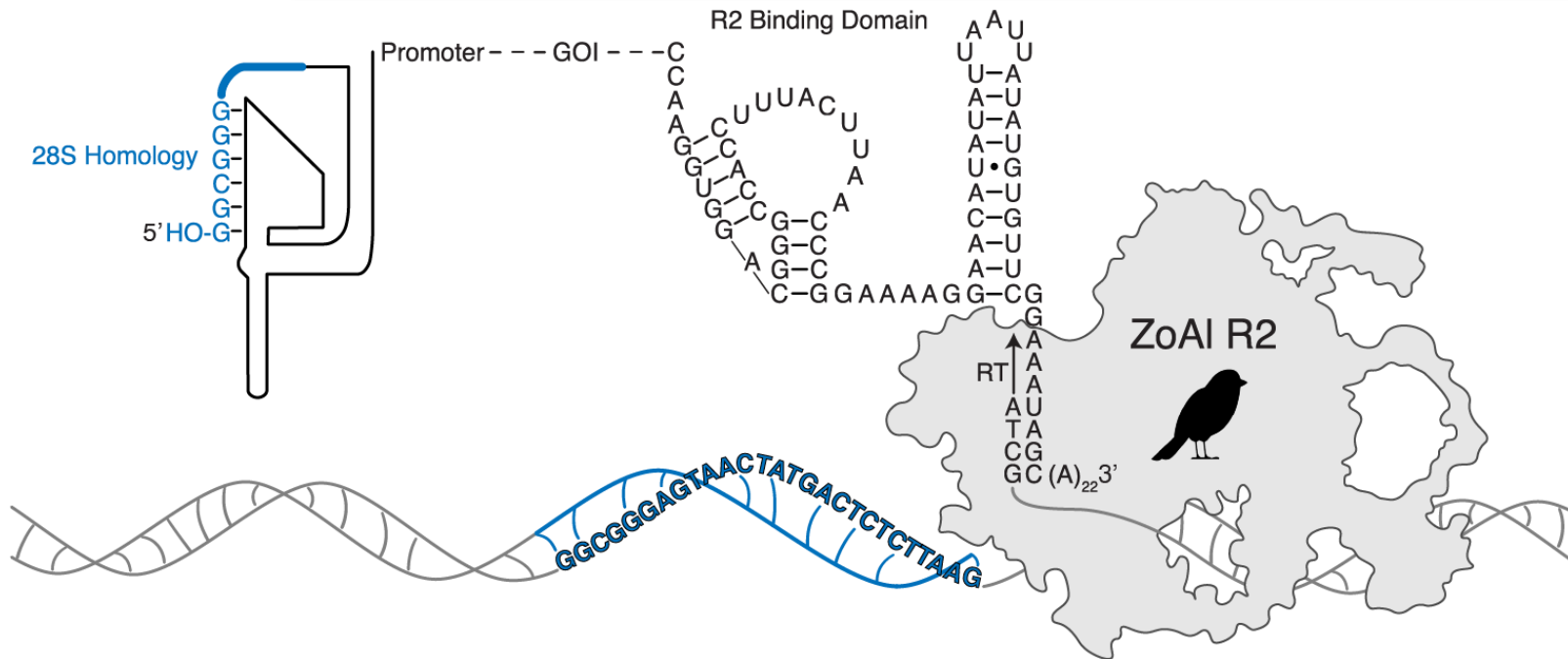
PRINT: Precise RNA-mediated INsertion of Transgenes

Harnessing eukaryotic retroelement proteins for transgene insertion into human safe-harbor loci

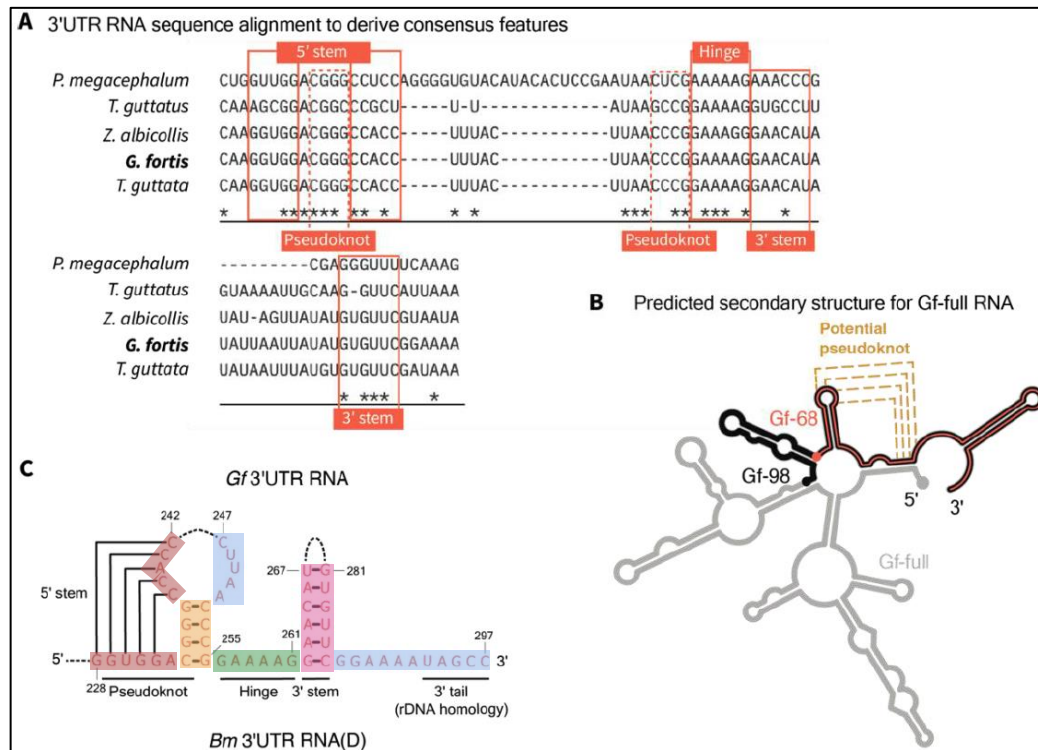
Received: 14 December 2022

Accepted: 10 January 2024

**Xiaozhu Zhang^{1,2}, Briana Van Treeck^{1,2}, Connor A. Horton¹,
Jeremy J. R. McIntyre¹, Sarah M. Palm¹, Justin L. Shumate¹ &
Kathleen Collins¹**  

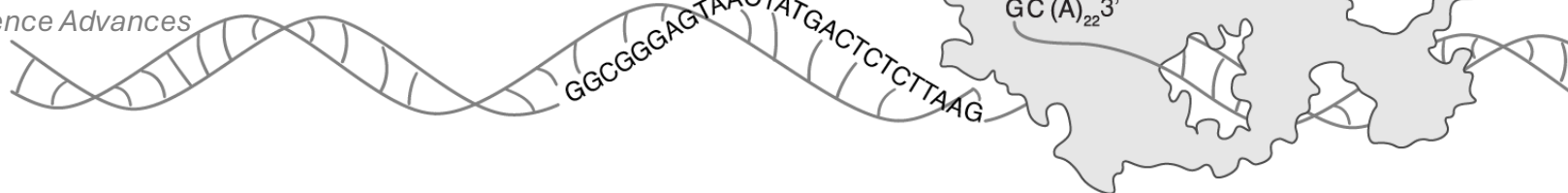


MSA of 3' UTR RNA Structure?



```
# STOCKHOLM 1.0
#=GF From 2025 PRINT SciAdv SI
QXTE01005434.1/4231-4305
NW_010576467.1/4400-4337
NW_027428349.1/170655-170589
NW_005054324.1/2492118-2492051
NC_133062.1/2957923-2957856
#=GC SS_cons
..((((((.....))))).(((.....)))).....
CUGGUUGGACGGGCCUCCAGGGGUGUACAUACACUCCGAAUACUCGAAAAAGAACCCG-----CGAGGGUUUCAAAG
CAAAGCGGACGGGCCGCU-----U-U-----AUAAGCCGAAAGGUGCCUUGUAAAAUUGCAAG-GUUCAUUAAA
CAAGGUGGACGGGCCACC-----UUUAC-----UUAACCCGAAAGGAACAUAUAGUUAUUGUGUUCGUAUA
CAAGGUGGACGGGCCACC-----UUUAC-----UUAACCCGAAAGGAACAUAUAGUUAUUGUGUUCGUAUA
CAAGGUGGACGGGCCACC-----UUUAC-----UUAACCCGAAAGGAACAUAUAGUUAUUGUGUUCGUAUA
..((((((.....))))).(((.....)))).....
```

Thawani, et al. (2025) *Science Advances*



Many cmsearch in various genomes iterations

Pseudoknot not found



Base pair with statistically significant covariation in R-scape

Nucleotide Present

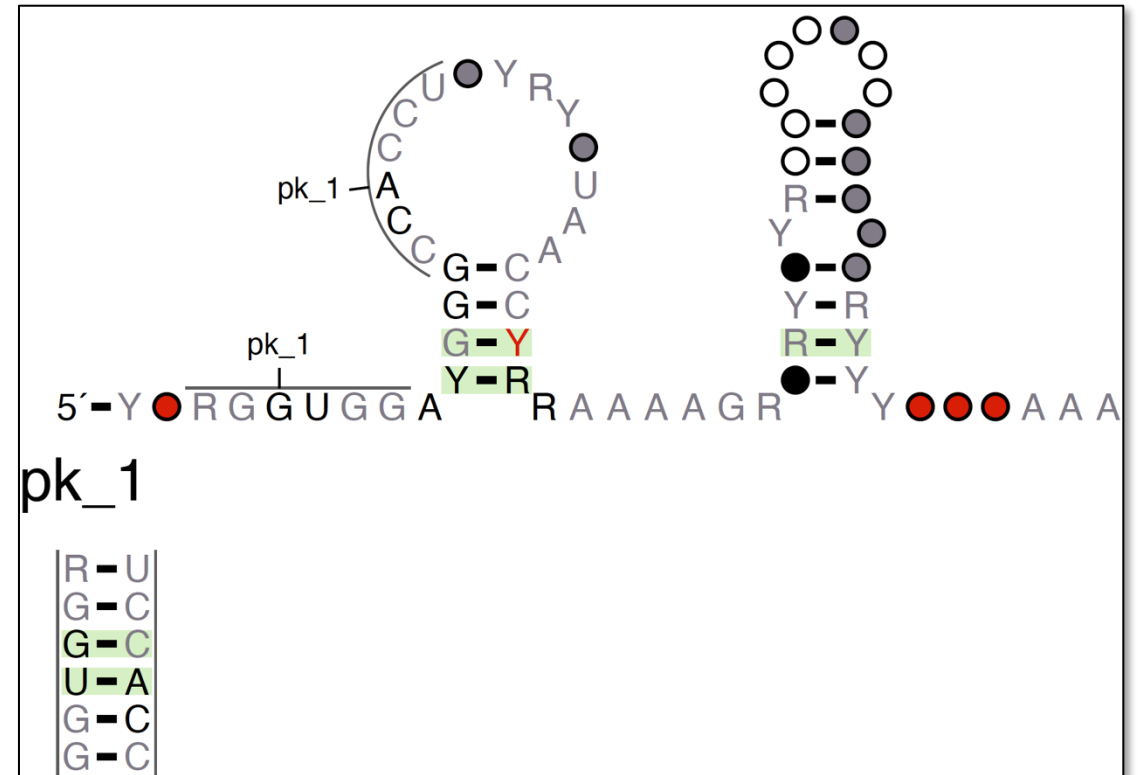
● 97% ● 90% ● 75% ○ 50%

Identity

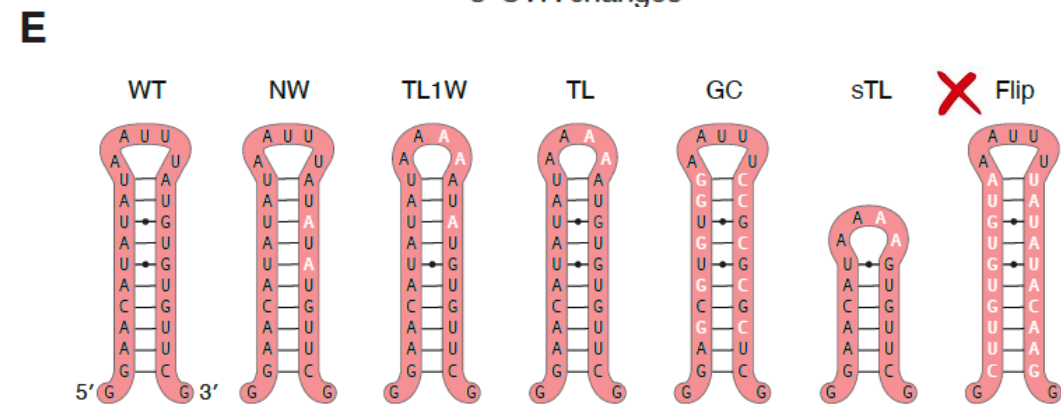
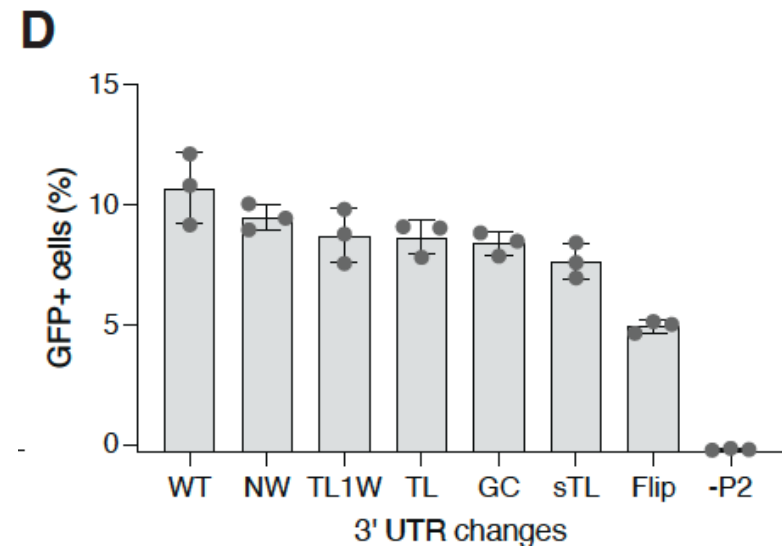
N 97% N 90% N 75%

+ more genomes

cacofold with -s (structure) option



Principles with RNA Engineering

Edmonds, et. al. (2025) *Nature Comms*

Future looking:

In (hopefully) 15-18 months...

I *should* have a dataset of >1000 RNAs purified with a specific function and predominately interaction with two specific proteins.

I expect 50-90% mRNA.

Likely some frog mitochondriaRNA.

If this dataset sounds interesting to you, please talk to me. Let's exchange contacts and keep in touch!

Acknowledgements

Lucks Lab

Dr. Julius LUCKS

Dr. Anibal ARCE

...

Covariation Collaborators:

Dr. Elena RIVAS

Dr. Blake SWEENEY

Nancy ONTIVEROS

Chan Lab:

Dr. Eric CHAN

Xingyu ZHU

...



Funding



DMR-2308691



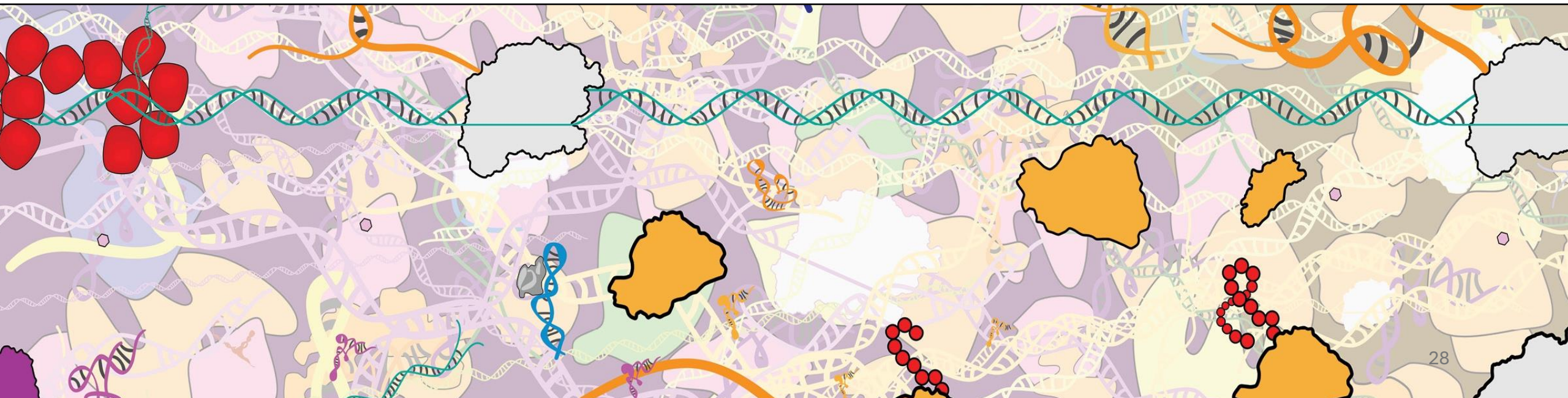
T32GM008449



DMM26022578



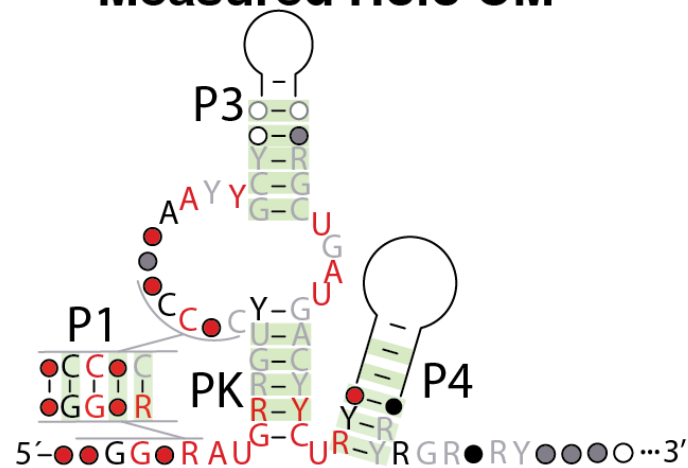
Thank you for your attention



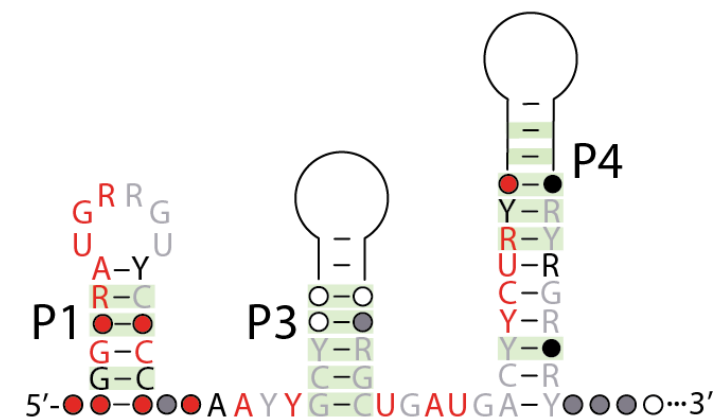
Similar motif with NGS data



Measured Holo CM



Measured Apo CM



Family: *R2_retro_el* (RF00524)

Description: *R2 RNA element*

Summary

Sequences

Alignment

Secondary structure

Species

Trees

Structures

Motif matches

Database references

Curation

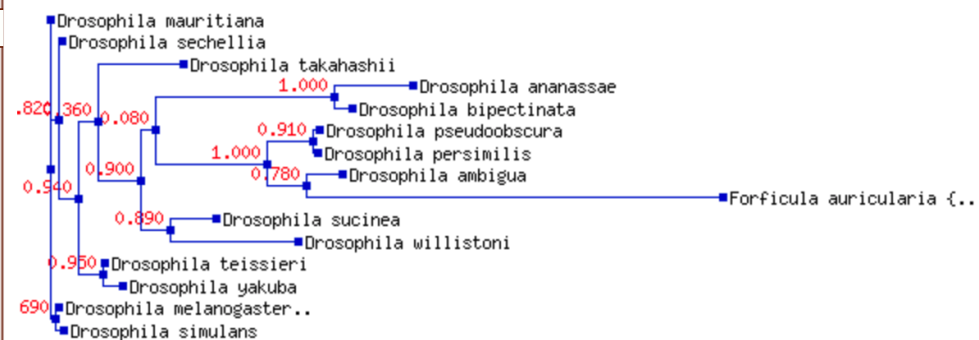
Publications

Trees

This page displays the predicted phylogenetic tree for the alignment. [More...](#)

This tree was built using the *fasttree* method.

Label the trees using ☒ species names or ☐ accessions.



R-scape optimised structure

4 out of 43 basepairs are significant at E-value=0.05

