

Deciphering the Structure-Function Relationships in RNA Editing

Jin Billy Li lab: Billy, Xin, Tao

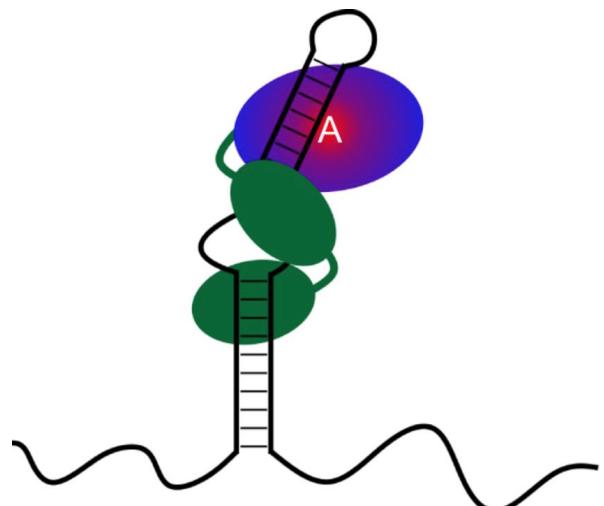
Anshul Kundaje lab: Anshul, Avanti

Noon, Nov. 17, 2017

Room: Lane 301.

Outline:

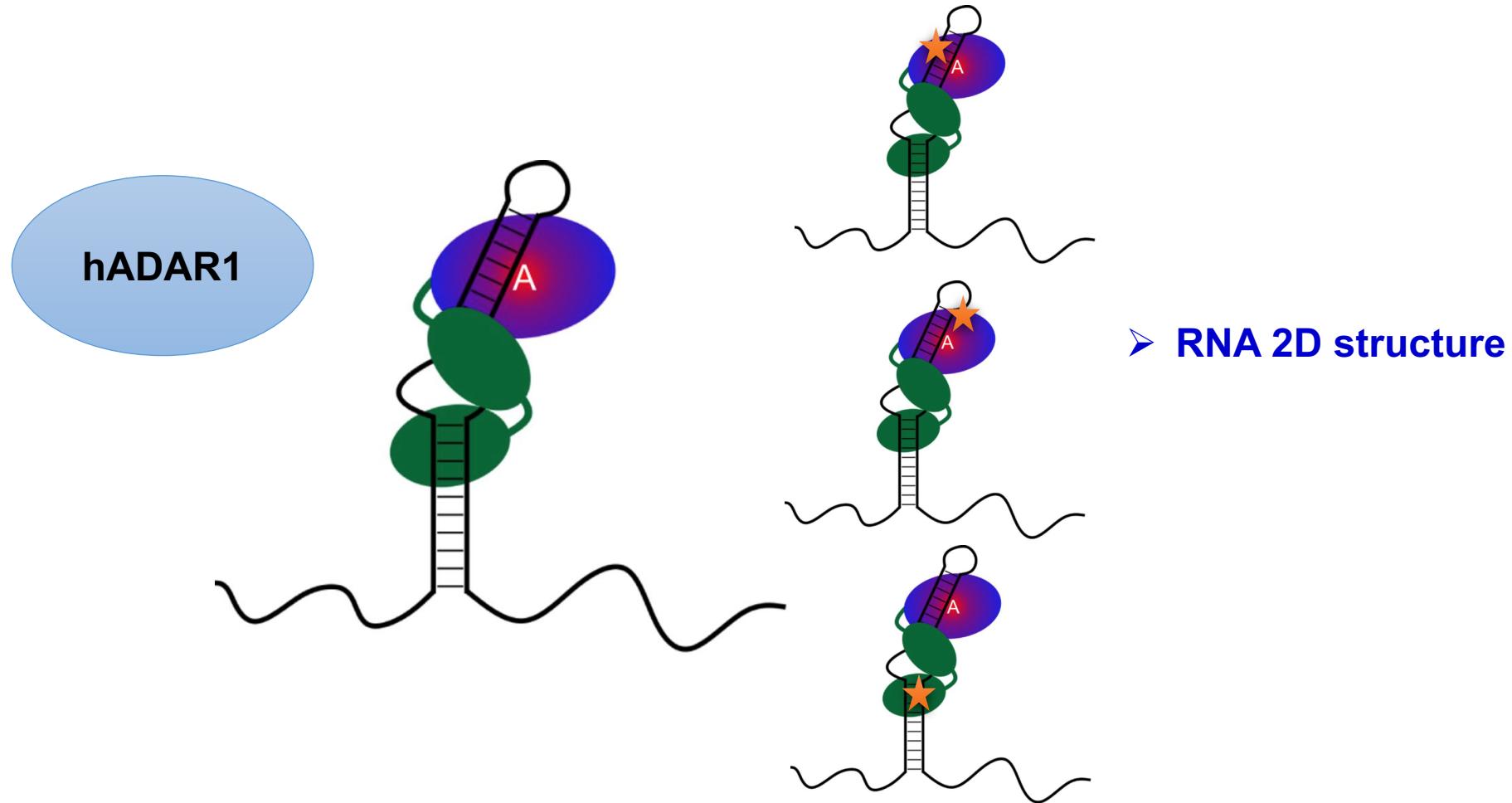
1. Goals/overarching question.
2. Current data
3. Open questions seeking advanced computational and machine learning methods



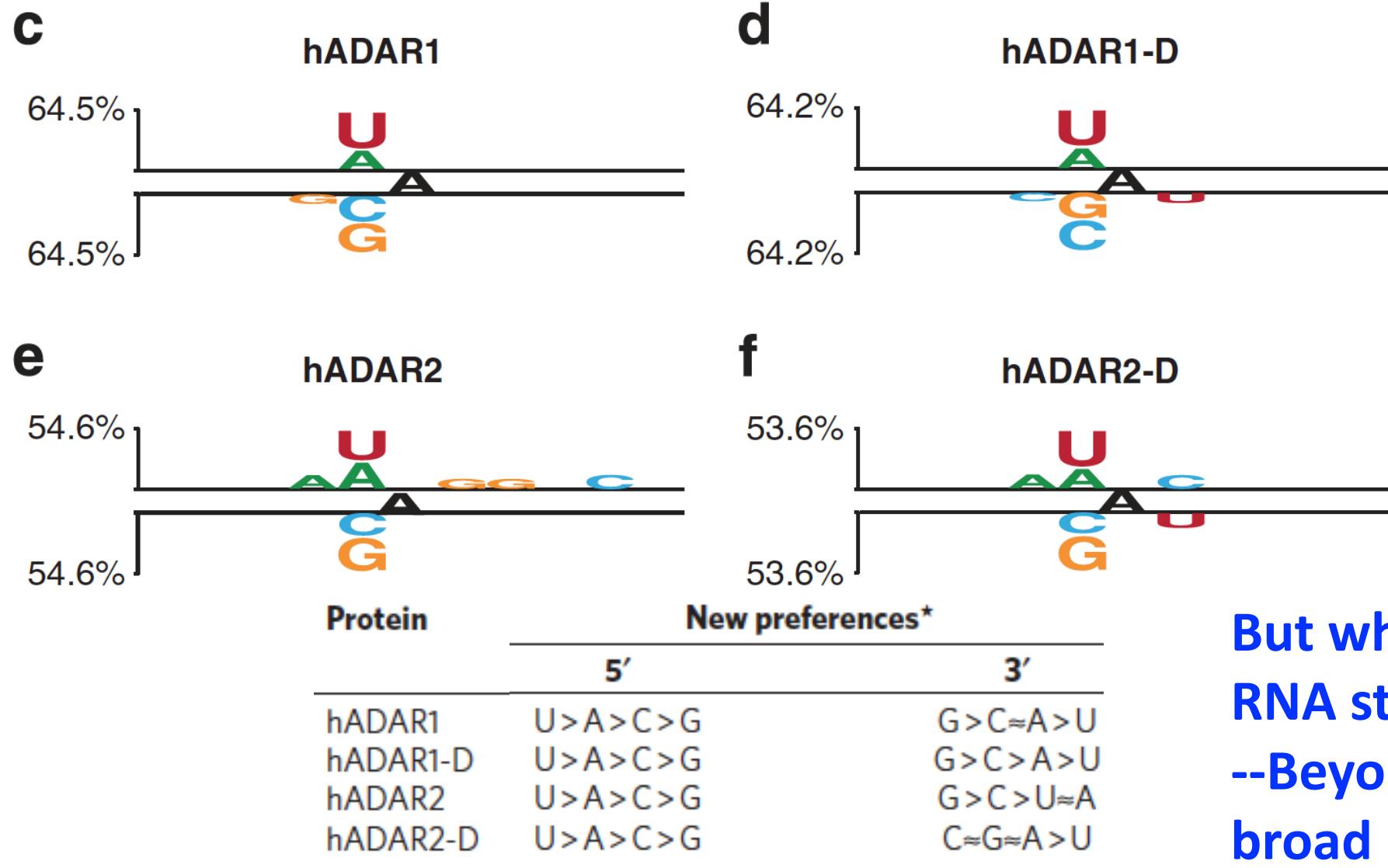
Question:

cis regulation of A-to-I RNA editing

How primary sequence and secondary structure of RNA affect ADAR editing?

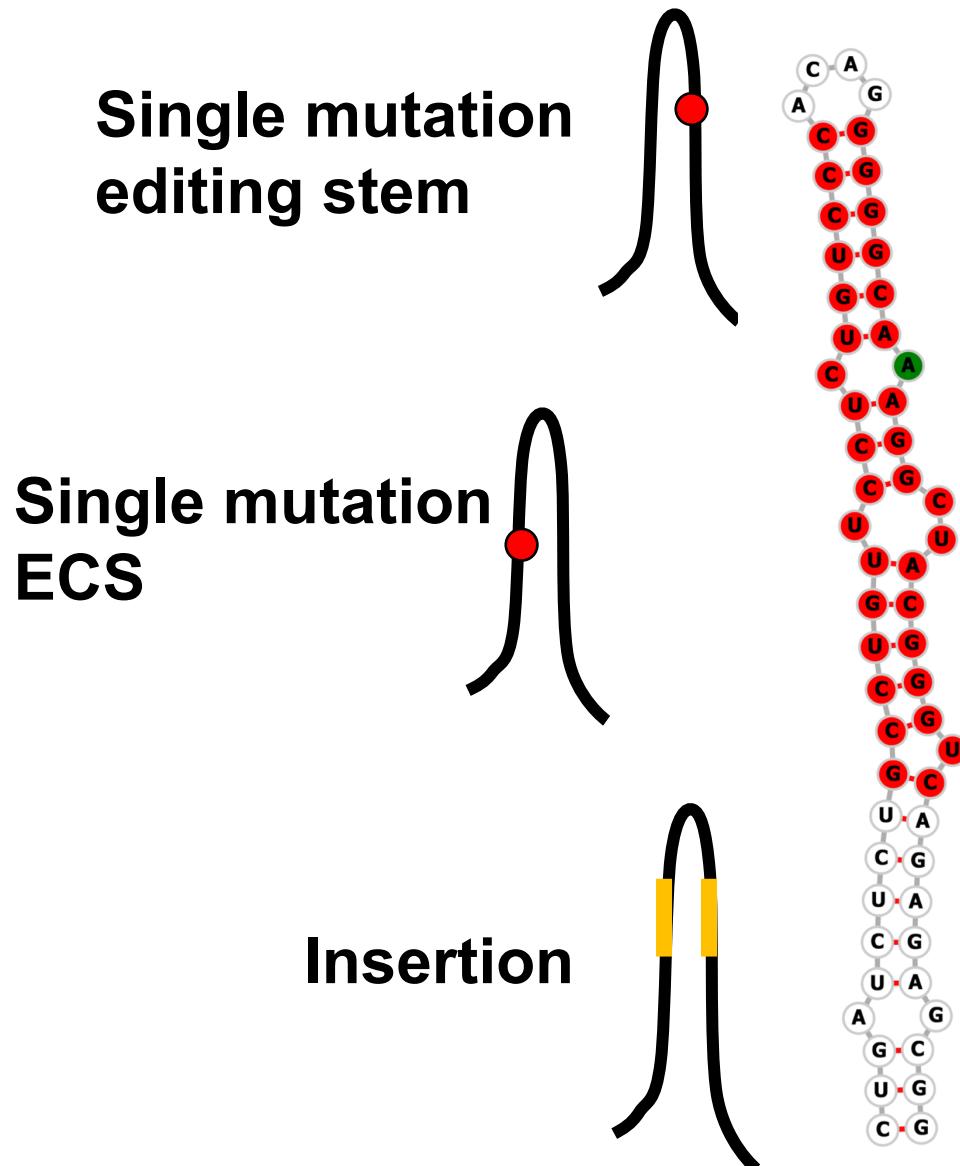


Known sequence preference by ADARs

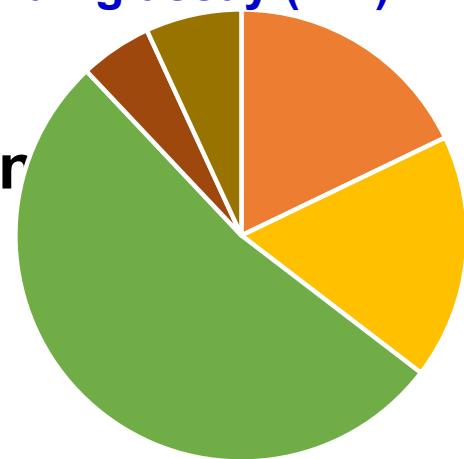


But what about
RNA structure?
--Beyond just
broad “dsRNA”?

Our approach: high-throughput assays: Example: Neil1 RNA variant library



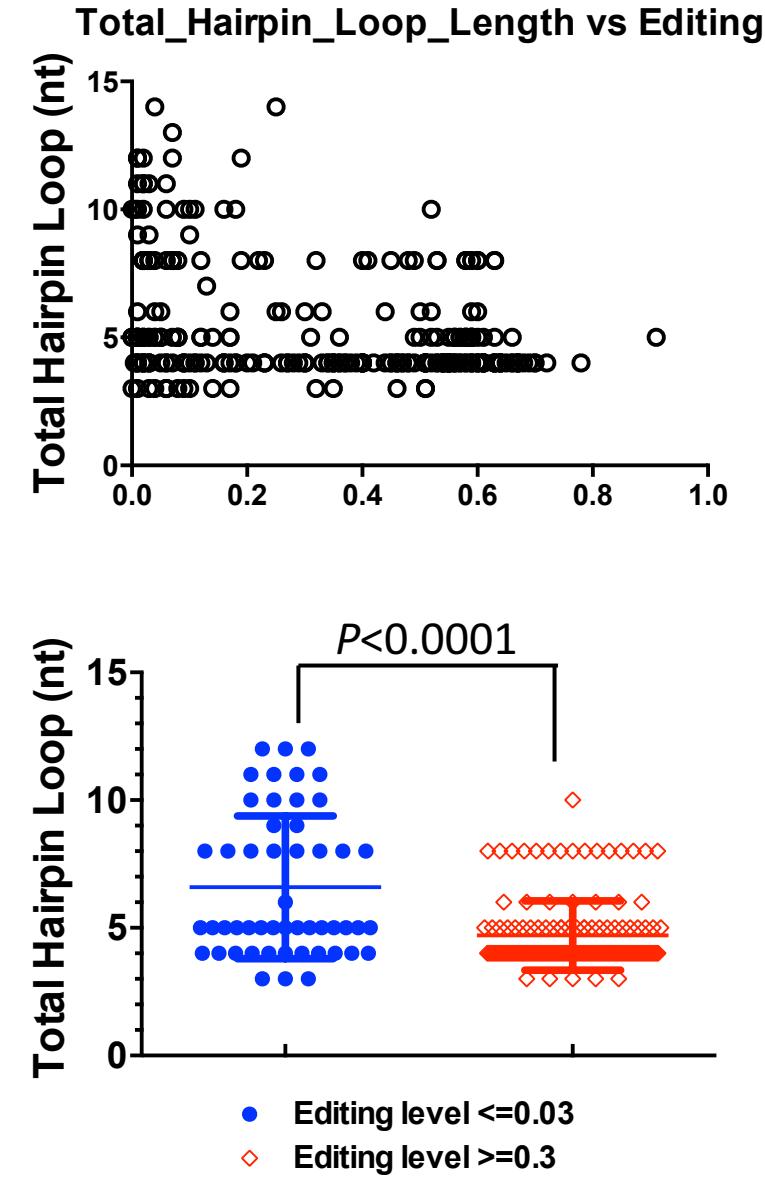
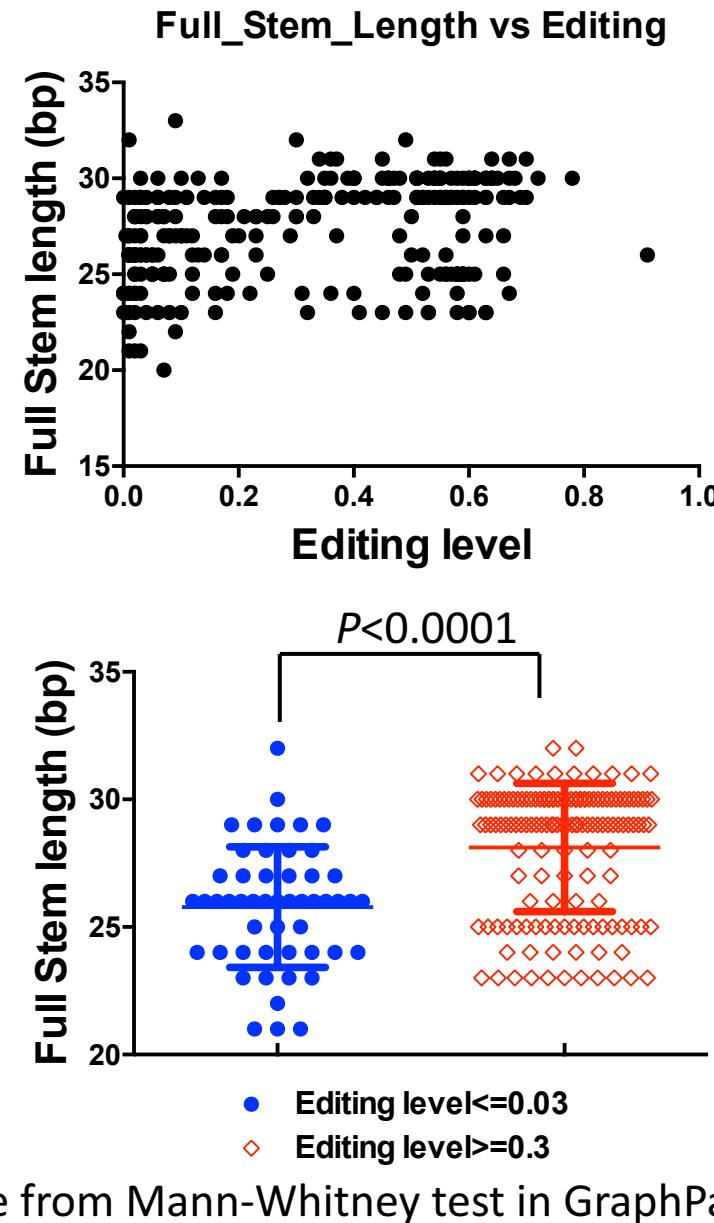
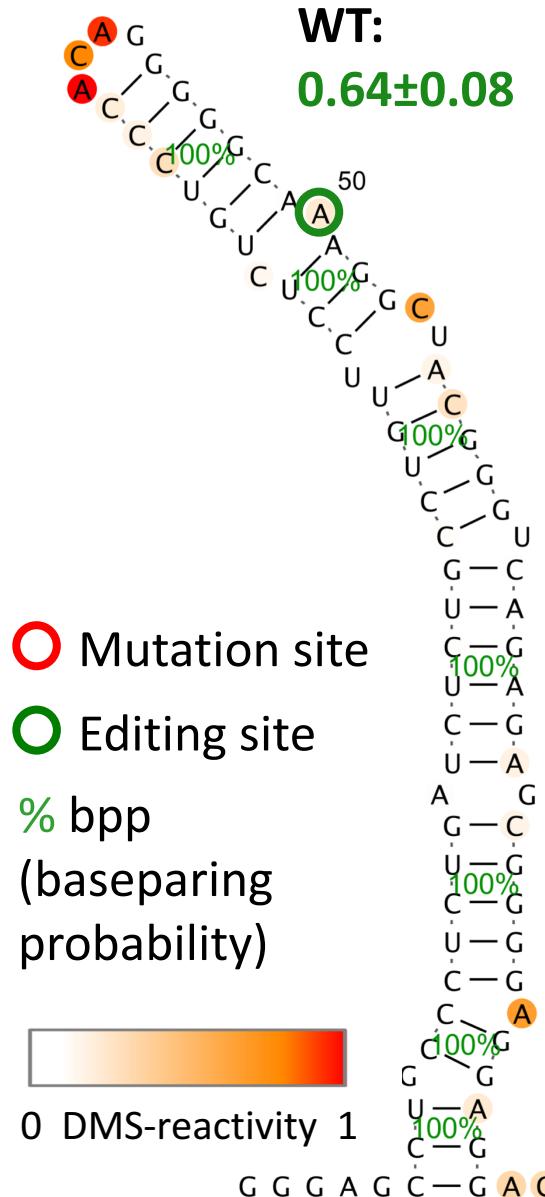
- ✓ RNA variants by CRISPR in cells (Tao)
- ✓ Chemical mapping of RNA 2D structure *in vitro* (Xin)
- ✓ On-going: RNA-protein binding assay (Xin)



- single mutation in editing stem
- single mutation in ECS
- double mutation in editing stem
- double compensatory mutations
- indel

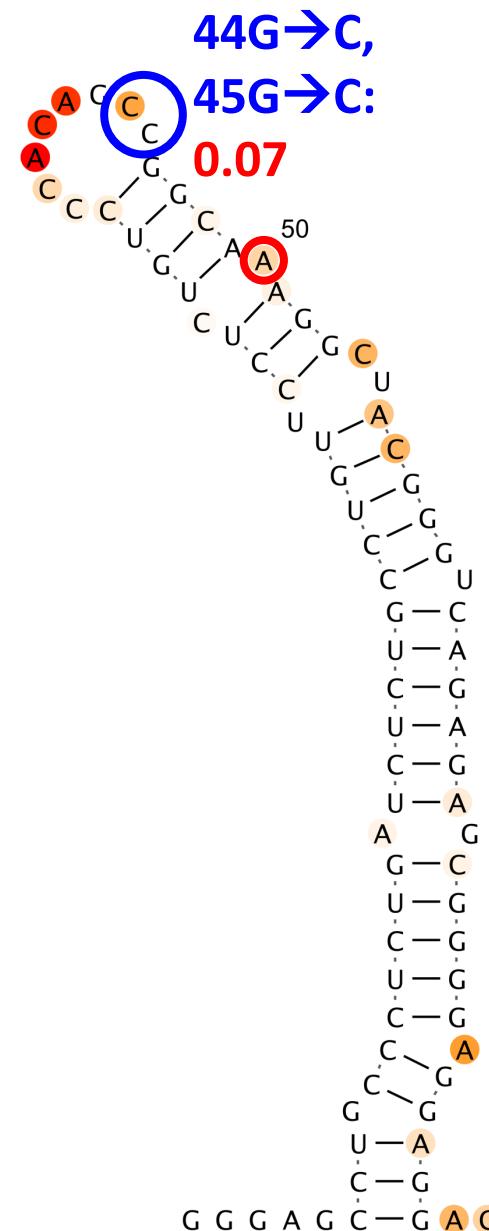
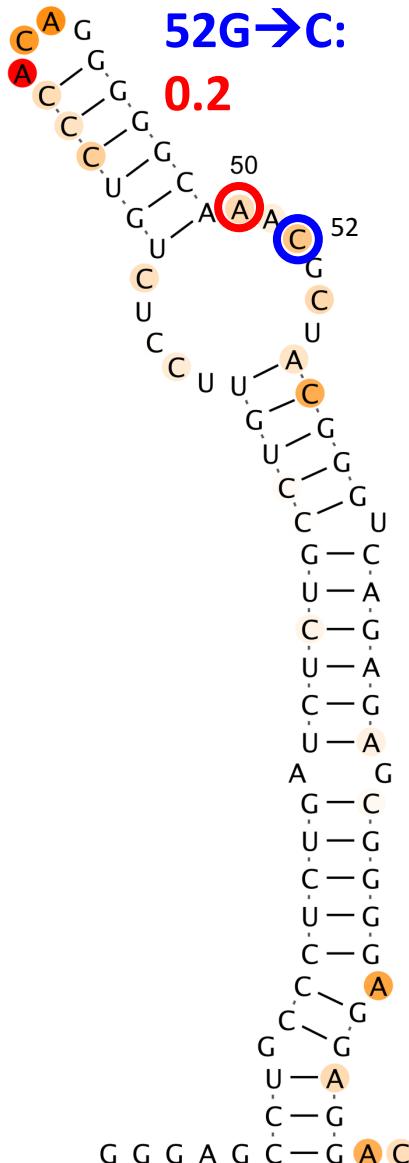
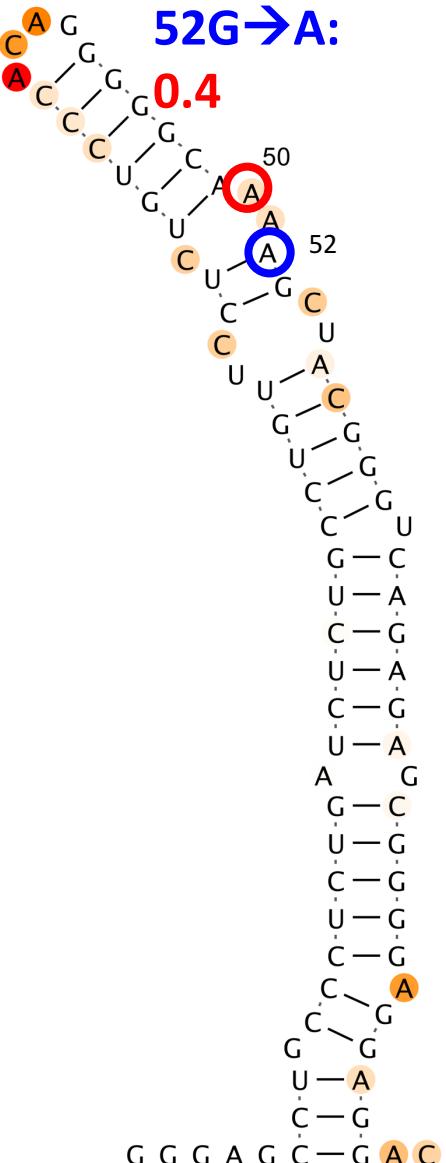
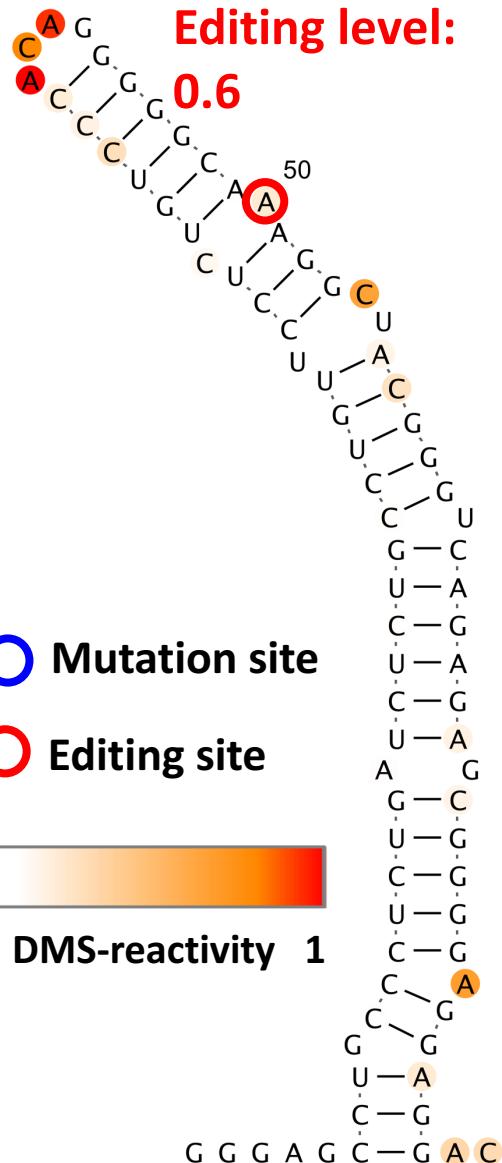
Tao

More stable dsRNA structures tend to have higher editing



More details in RNA secondary structure and Editing level by CRISPR

WT:



○ Mutation site

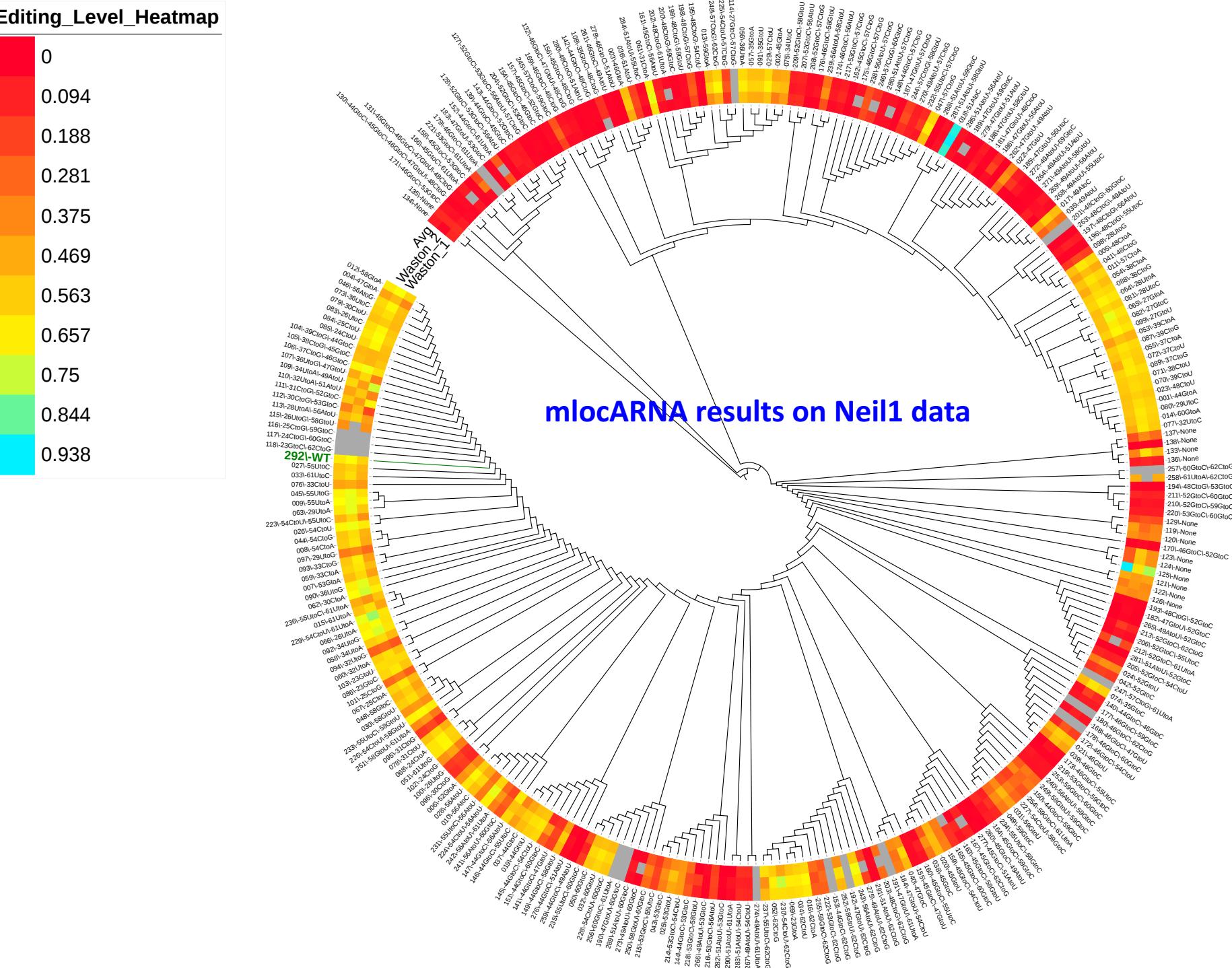
○ Editing site

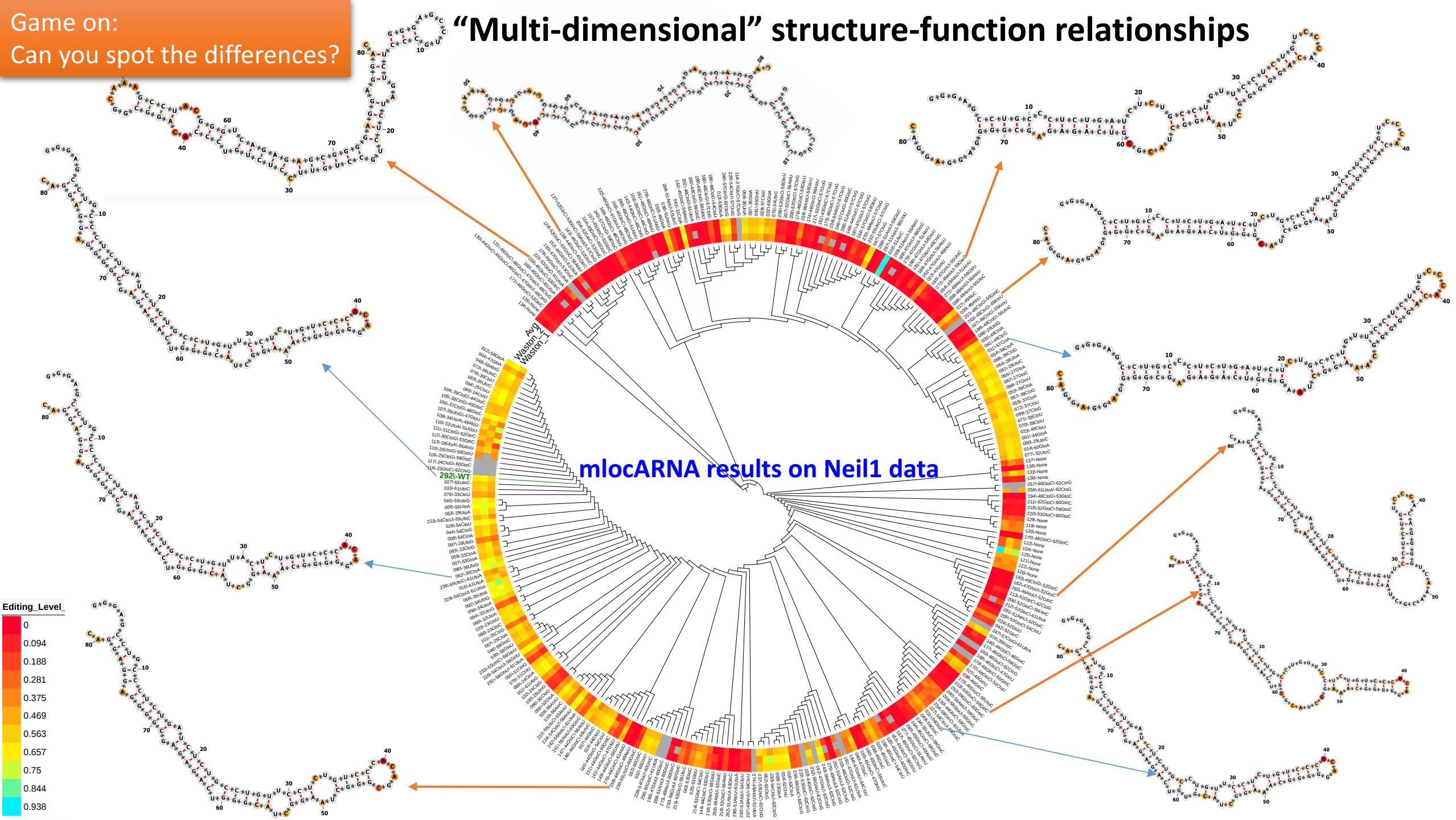


Tao

In search for a better way to visualize/quantify our data for structure-function analysis:

- Tried several ways:
 - RNAdistance,
 - RNA-TVcurve,
 - Web-Beagle,
 - locARNA
**(mlocARNA,
RNACLUST)**

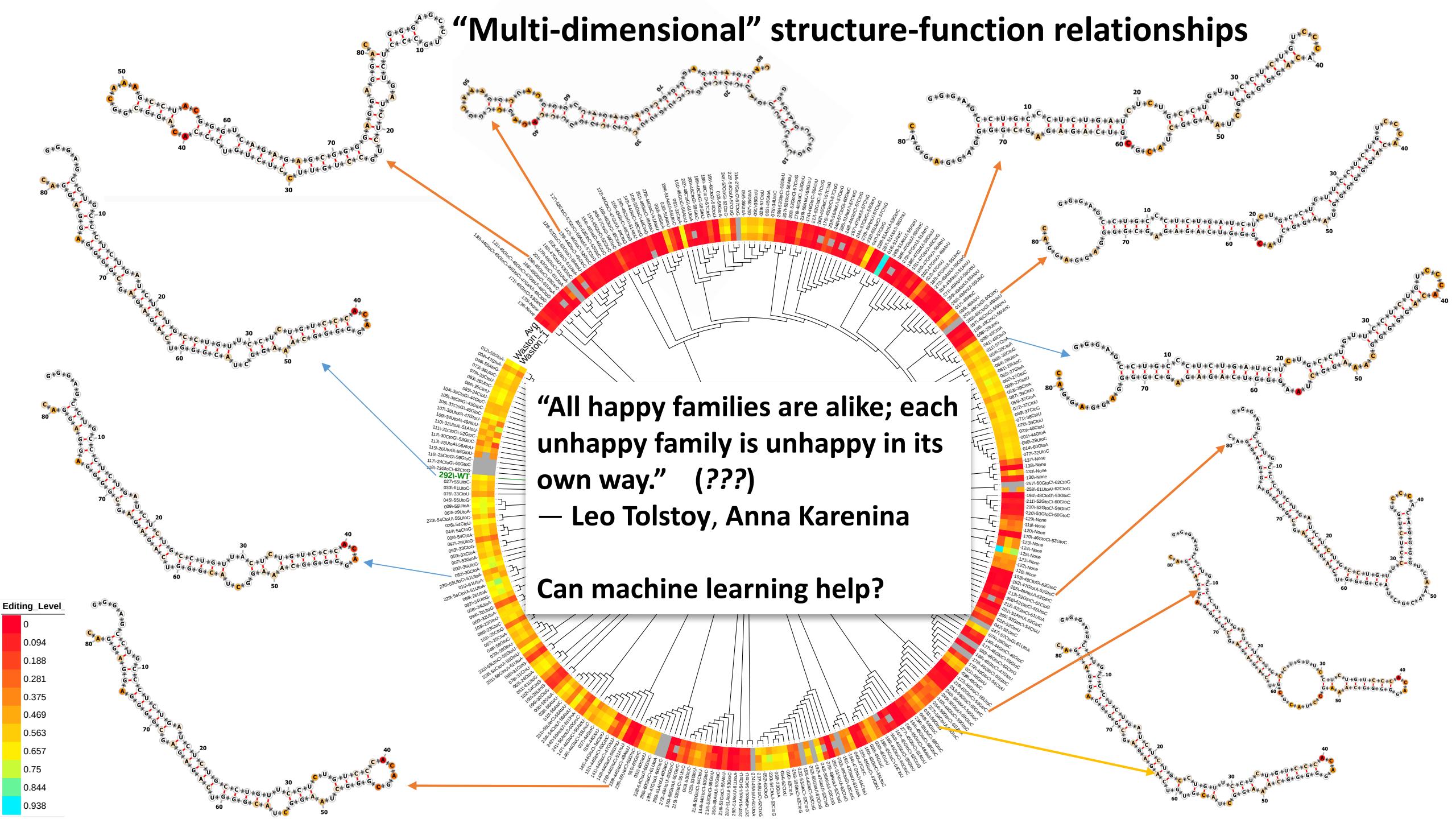


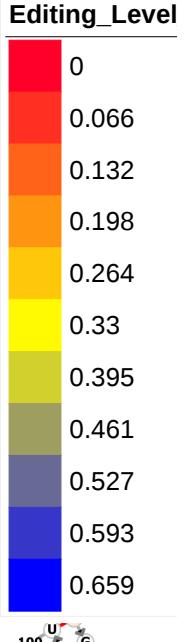


“Multi-dimensional” structure-function relationships

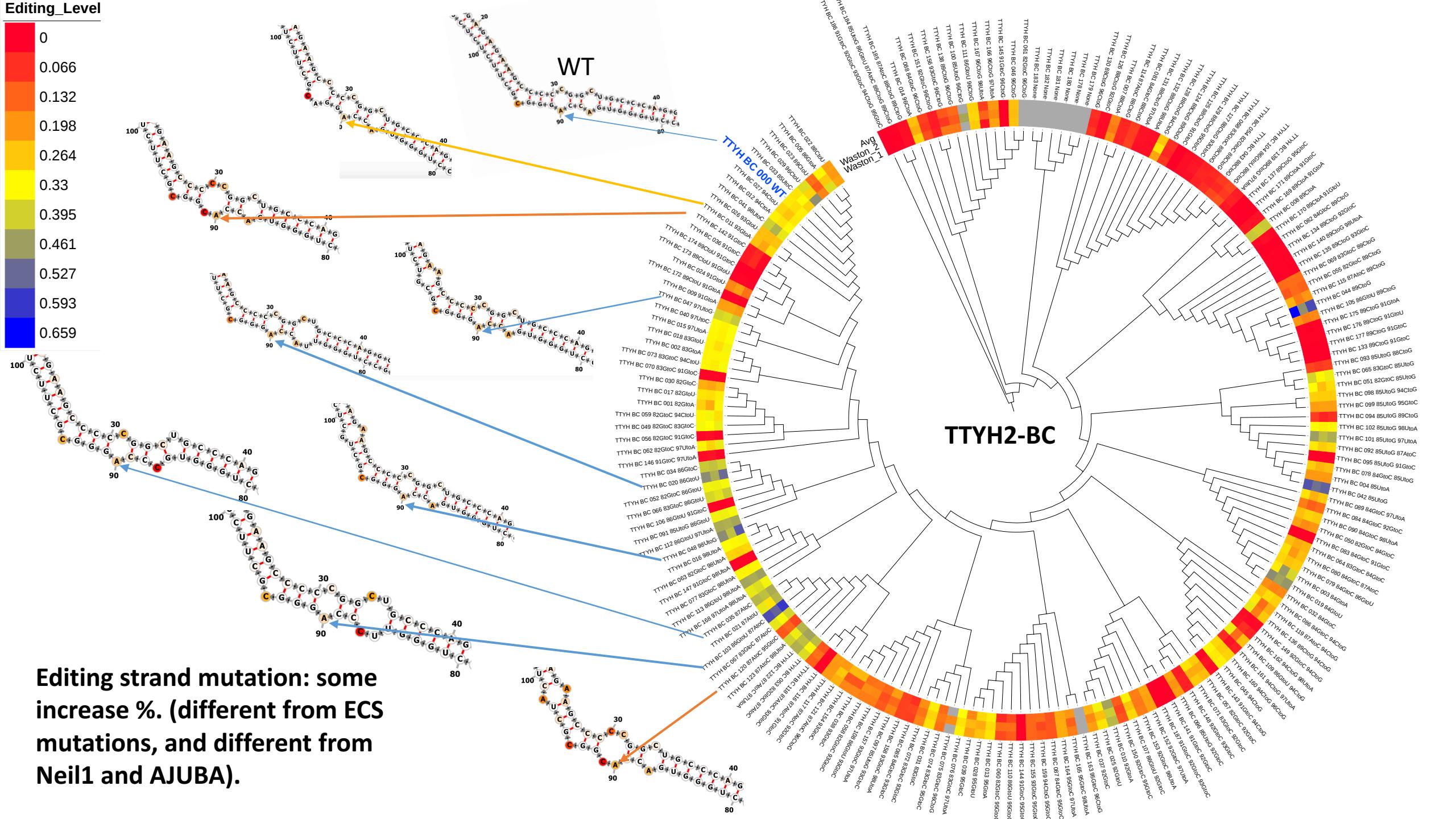
“All happy families are alike; each unhappy family is unhappy in its own way.” (???)
— Leo Tolstoy, Anna Karenina

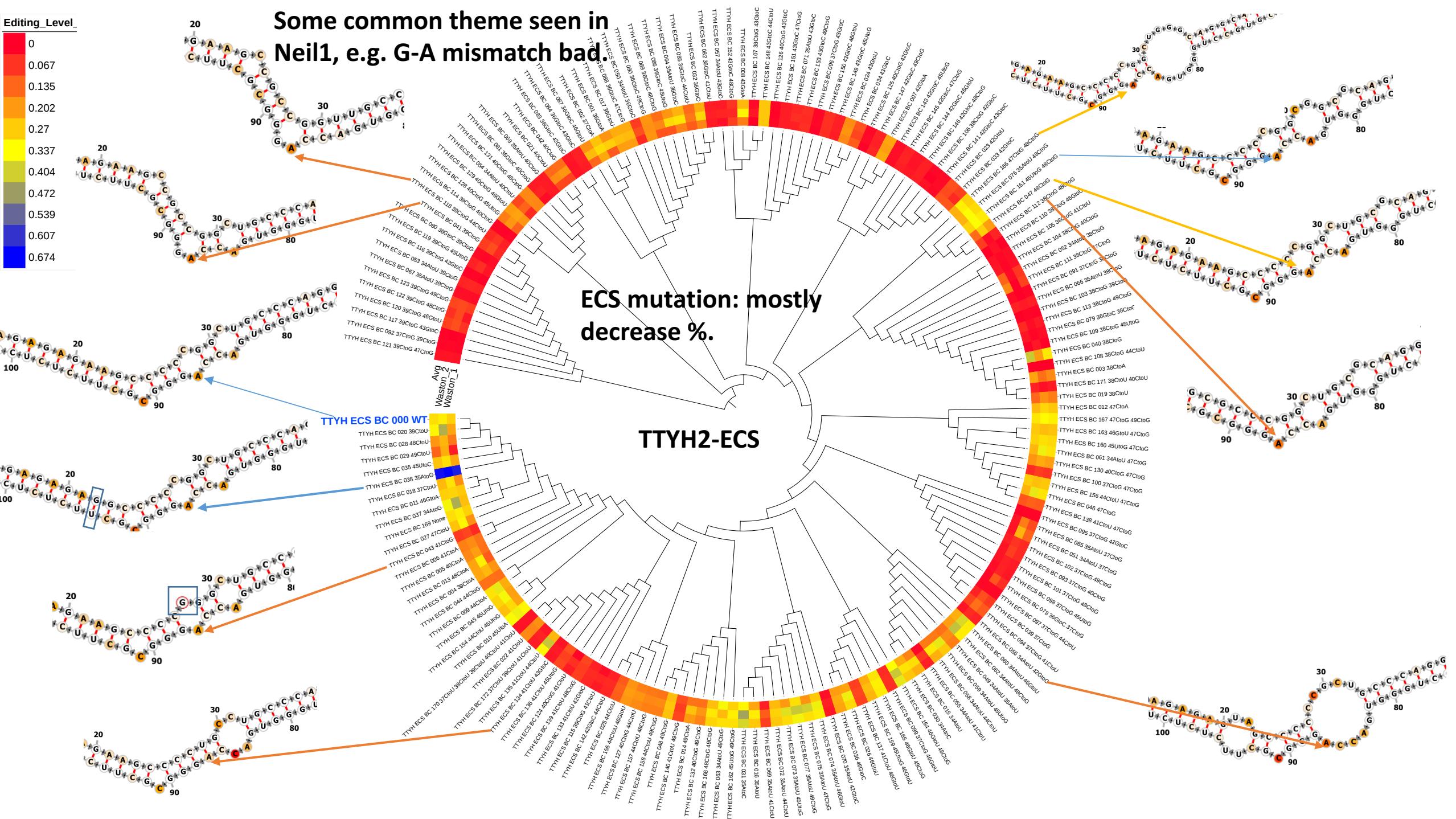
Can machine learning help?





Editing strand mutation: some increase %. (different from ECS mutations, and different from Neil1 and AJUBA).





Cracking the *cis* code for RNA editing:

- Summary of total amount of RNA structures we measured so far:

Library	Number of RNA variants
Neil1*	293
TTYH2-Editing strand	186
TTYH2-ECS	173
AJUBA-Editing strand	271
AJUBA-ECS**	285

Total **923** (maybe 1208 if AJUBA-ECS CRISPR results better).

Neil1*: Tao is mining data for editing on other Adenosines in the sequence (which could increase number of data for machine learning).

AJUBA-ECS**: challenging, missing many of the editing level CRISPR results.

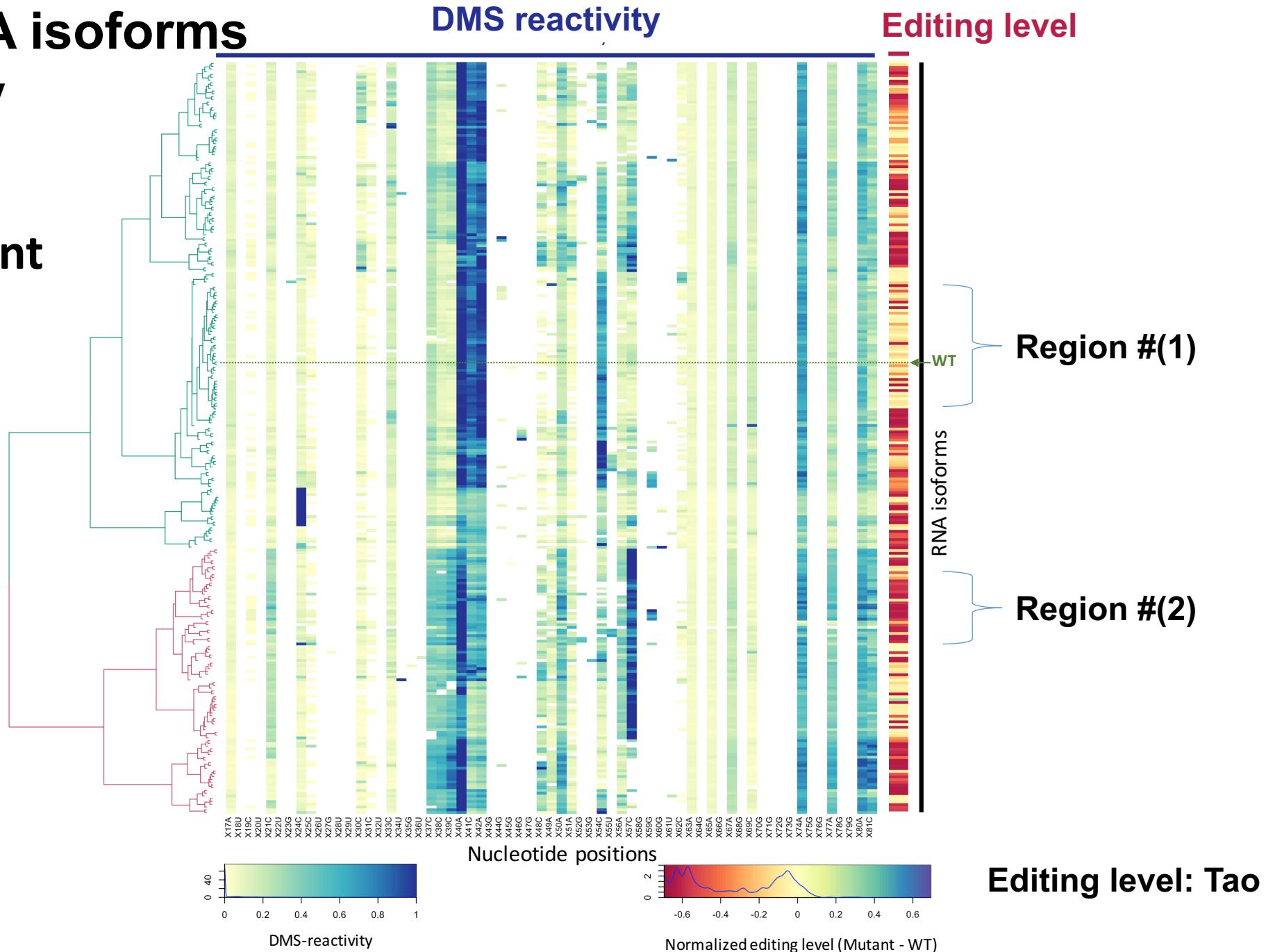
Open questions seeking help from experts in Kundaje lab

- Can we use
 1. Graphical computational tools to better present the data?
 2. Machine Learning methods to find the rules in explaining Structure-Function (structure vs editing level)?
- Suggestions/thoughts?

Backup slides:

Clustering of RNA isoforms by DMS-reactivity

looking for “dominant structures” in each RNA structure “clusters”



Nomenclatures in Neil1 structure:

#292

WT

P: helix

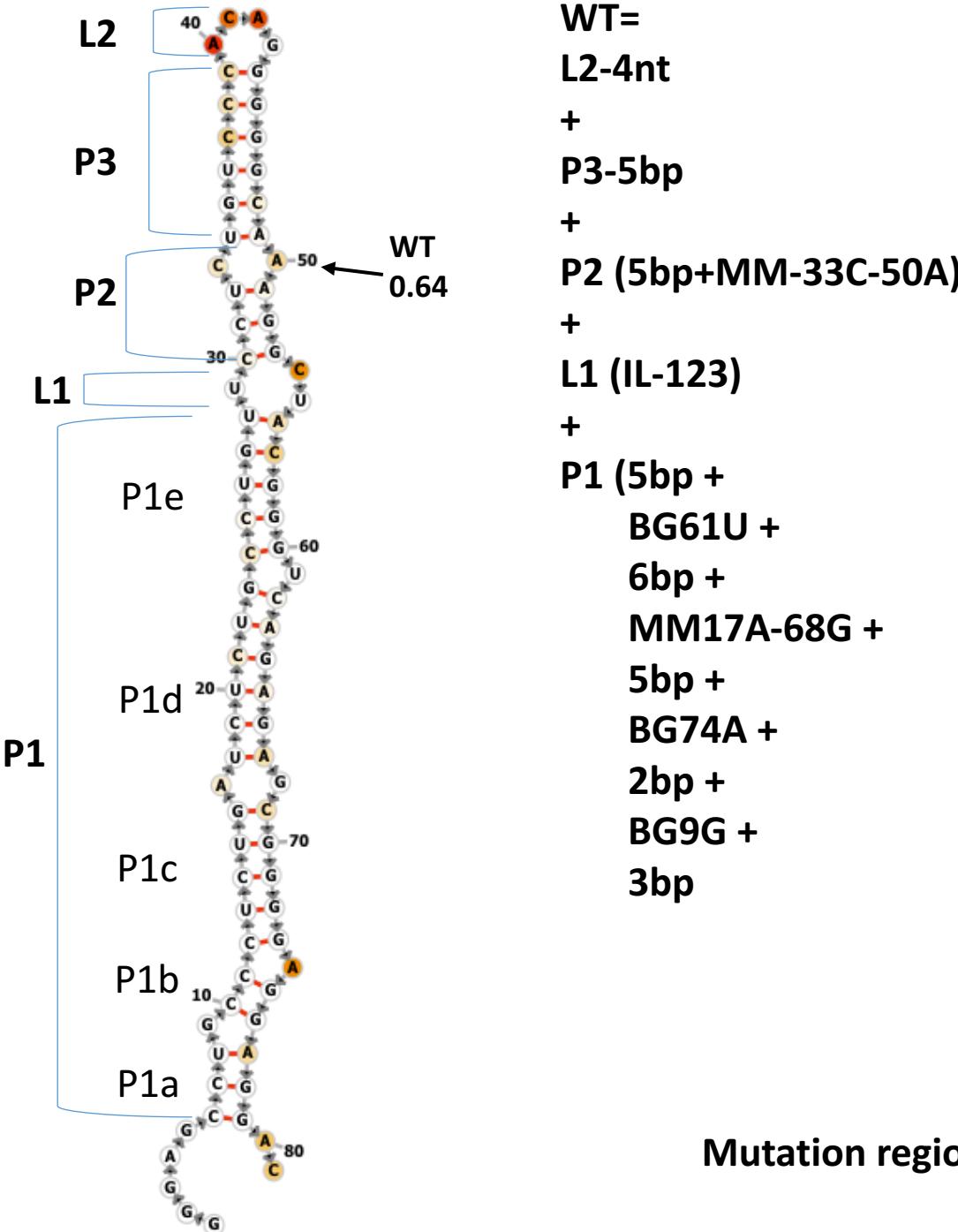
BG: Bulge

MM: Mis-Match

L: Loop

IL: Internal Loop

looking for “dominant structures” in each RNA structure “clusters”



WT=

L2-4nt

+

P3-5bp

+

P2 (5bp+MM-33C-50A)

+

L1 (IL-123)

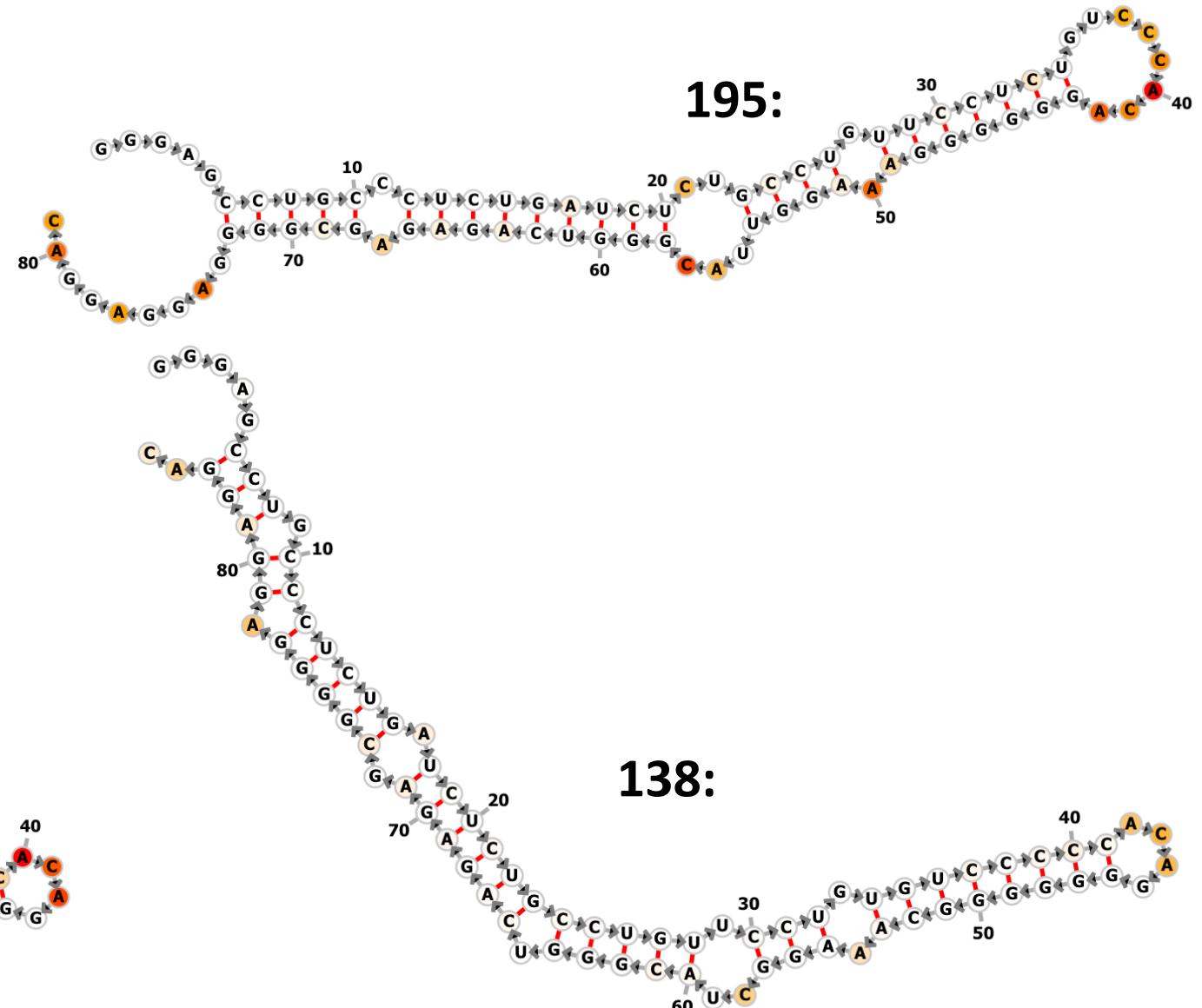
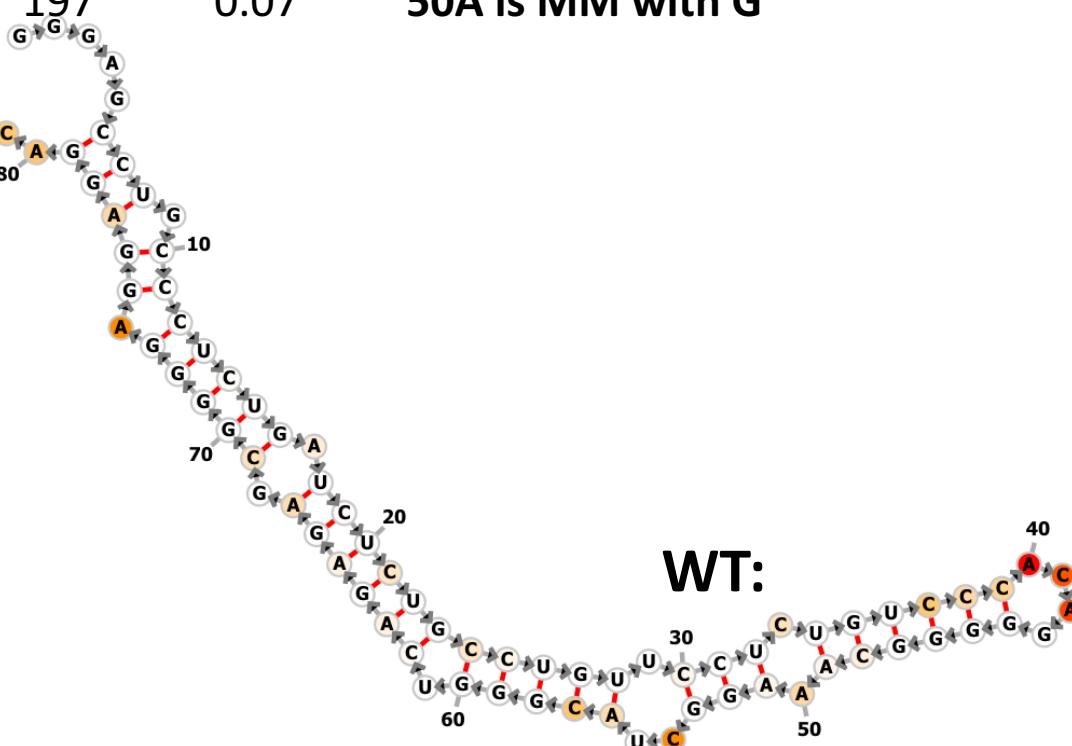
+

P1 (5bp +
BG61U +
6bp +
MM17A-68G +
5bp +
BG74A +
2bp +
BG9G +
3bp

Mutation region: G23 to C62

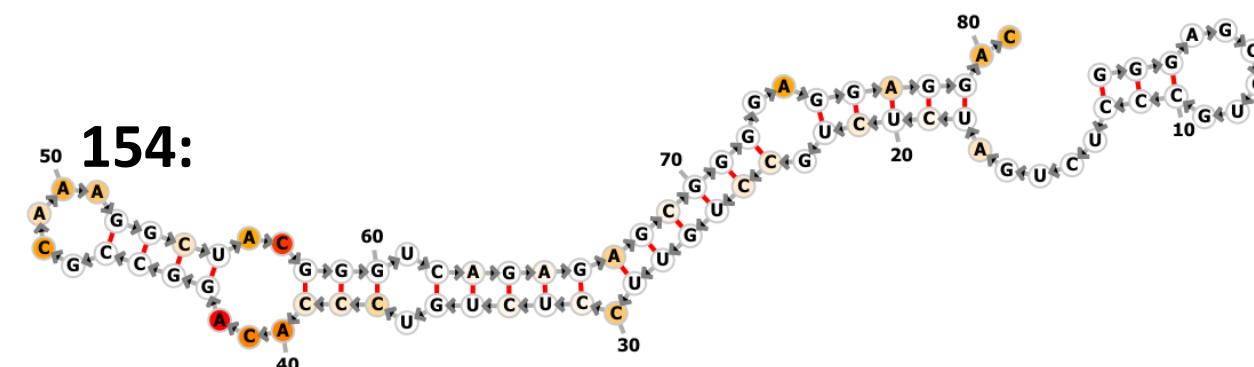
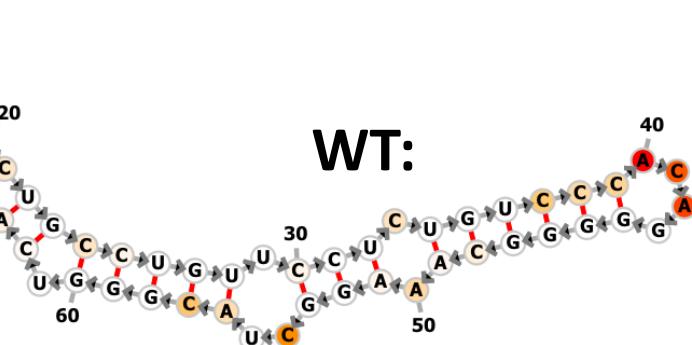
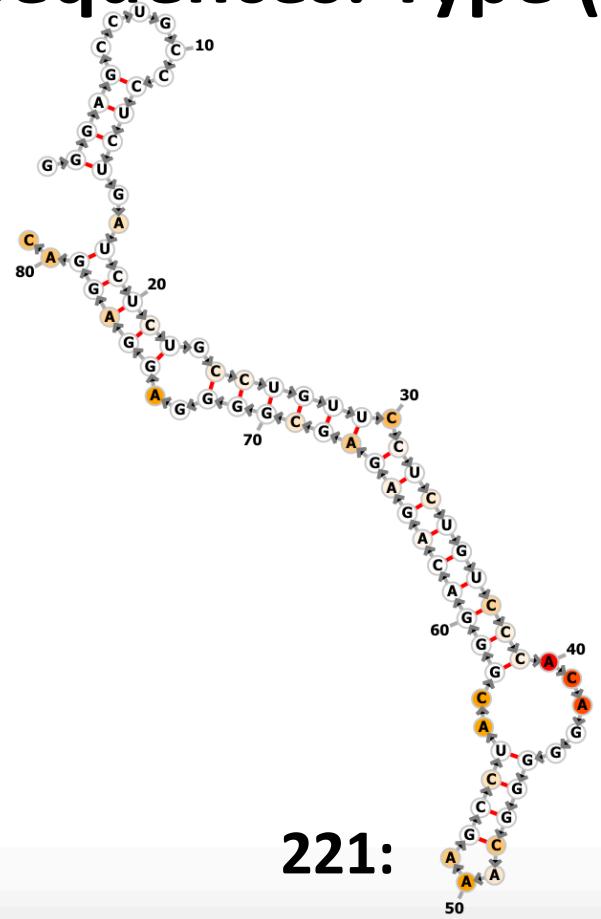
“Dominant structures” in low Editing level sequences: Type (1)

rna_id	editing_level	Structure_Feature
138	0.01	50A is MM with G
195	0.02	50A is MM with G
196	0.02	50A is MM with G
200	0.02	50A is MM with G
199	0.03	50A is MM with G
202	0.03	50A is MM with G
197	0.07	50A is MM with G



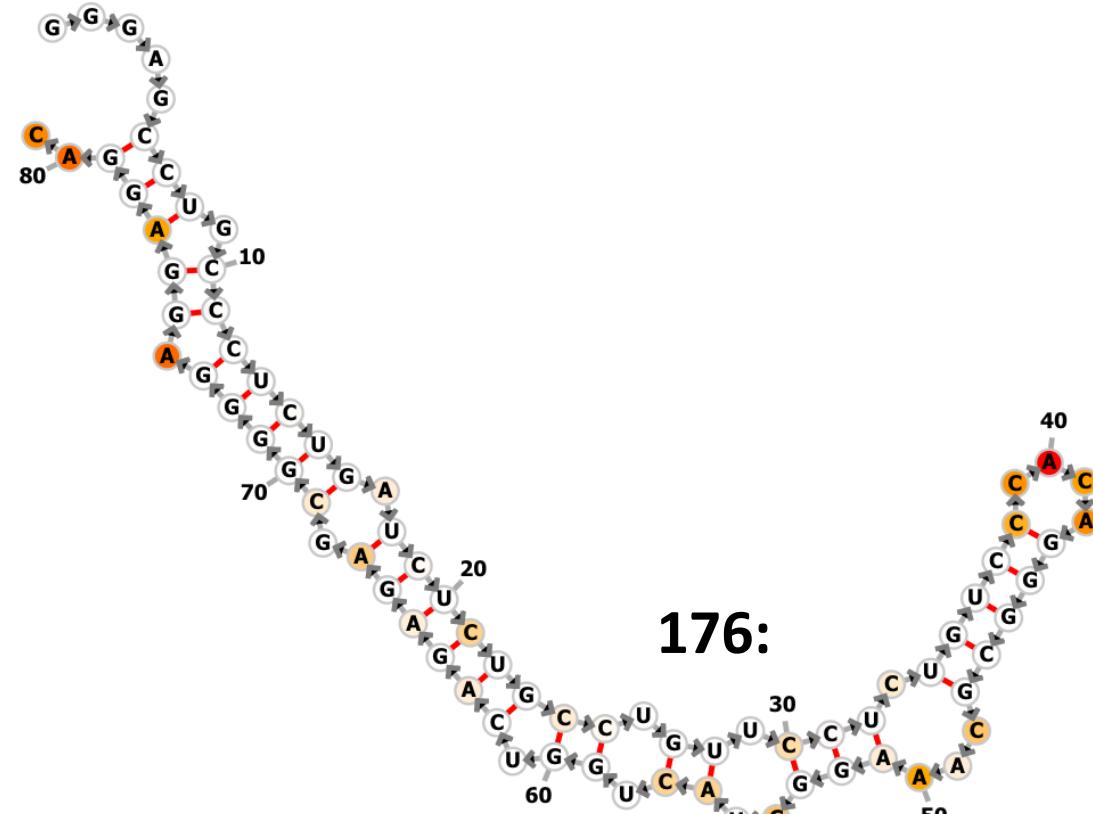
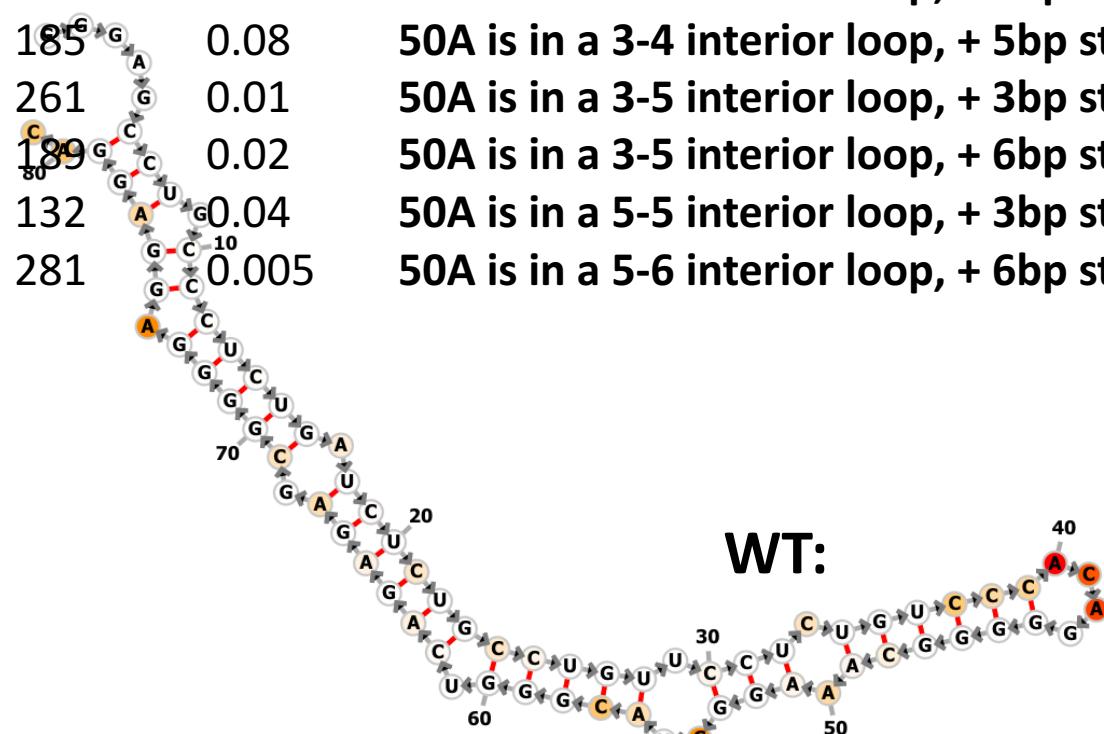
“Dominant structures” in low Editing level sequences: Type (2)

rna_id	editing_level	Structure_Feature
221	0.19	50A is in the 3nt hairpin loop, 4bp stem
144	0.25	50A is in the 3nt hairpin loop, 4bp stem
131	0.03	50A is in the 3nt hairpin loop, 5bp stem
216	0.18	50A is in the 3nt hairpin loop, 6bp stem
170	0.02	50A is in the 4nt hairpin loop
154	0.01	50A is in the 5nt hairpin loop
127	0.02	50A is in the 5nt hairpin loop
204	0.02	50A is in the 5nt hairpin loop
157	0.03	50A is in the 5nt hairpin loop
171	0.06	50A is in the 5nt hairpin loop
128	0.01	50A is in the 6nt hairpin loop
158	0.01	50A is in the 6nt hairpin loop
143	0.07	50A is in the 6nt hairpin loop
130	0.02	50A is in the 6nt hairpin loop and around U20 a 9nt loop



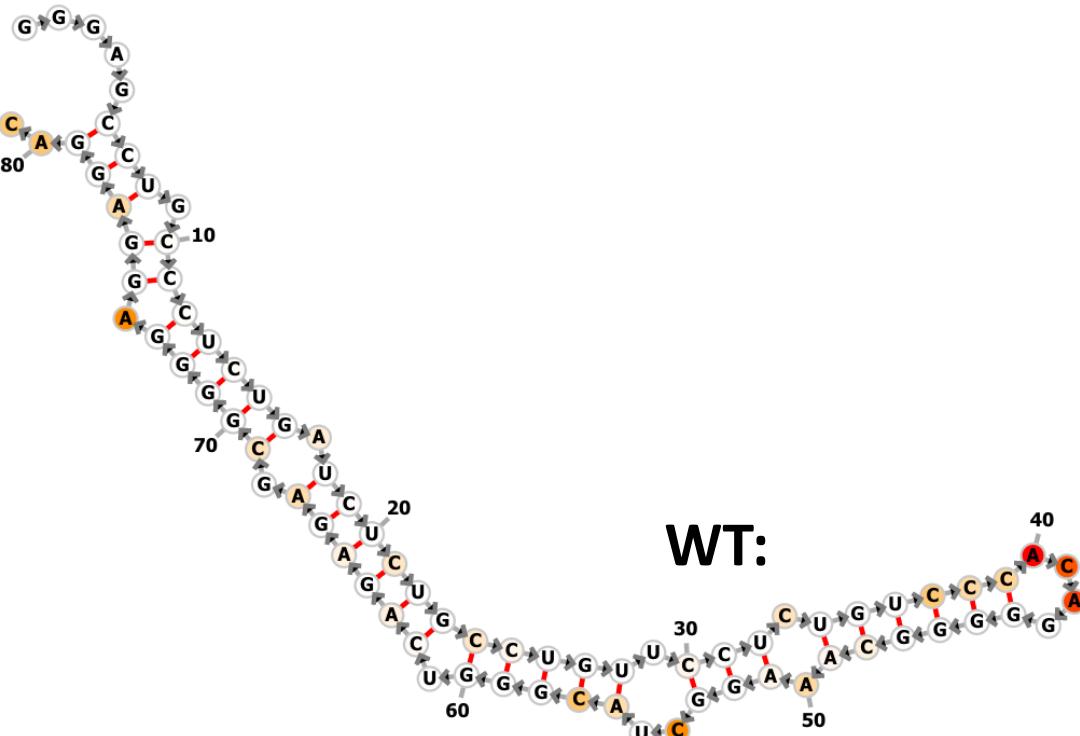
“Dominant structures” in low Editing level sequences: Type (3)

rna_id	editing_level	Structure_Feature
155	0.03	50A is in a 1-3 interior loop + 2bp upstream of 50A
134	0.07	50A is in a 1-3 interior loop, + 4bp stem+3-3 interior loop+2bp
176	0.04	50A is in a 1-3 interior loop, + 5bp stem
282	0.004	50A is in a 1-3 interior loop, + 7bp stem
259	0	50A is in a 3-3 interior loop
272	0.03	50A is in a 3-3 interior loop, + 5bp stem
193	0.02	50A is in a 3-3 interior loop, + 7bp stem
187	0.02	50A is in a 3-4 interior loop, + 5bp stem
185	0.08	50A is in a 3-4 interior loop, + 5bp stem
261	0.01	50A is in a 3-5 interior loop, + 3bp stem
189	0.02	50A is in a 3-5 interior loop, + 6bp stem
132	0.04	50A is in a 5-5 interior loop, + 3bp stem
281	0.005	50A is in a 5-6 interior loop, + 6bp stem



Structure don't always correlate with editing level

rna_id	editing_level	Structure_Feature
151	0.26	a U BG on ECS 2bp upstream of 50A + 3bp stem
150	0.33	a U BG on ECS 2bp upstream of 50A + 3bp stem
19	0.57	a U BG on ECS 2bp upstream of 50A + 3bp stem
147	0.59	a U BG on ECS 2bp upstream of 50A + 3bp stem
145	0.6	a U BG on ECS 2bp upstream of 50A + 3bp stem
217	0.08	a U BG on ECS 2bp upstream of 50A, + 7bp stem.



rna_id	editing_level	Structure_Feature
278	0	50A in a 2 BG
288	0	50A in a 2 BG
284	0.03	50A in a 2 BG

