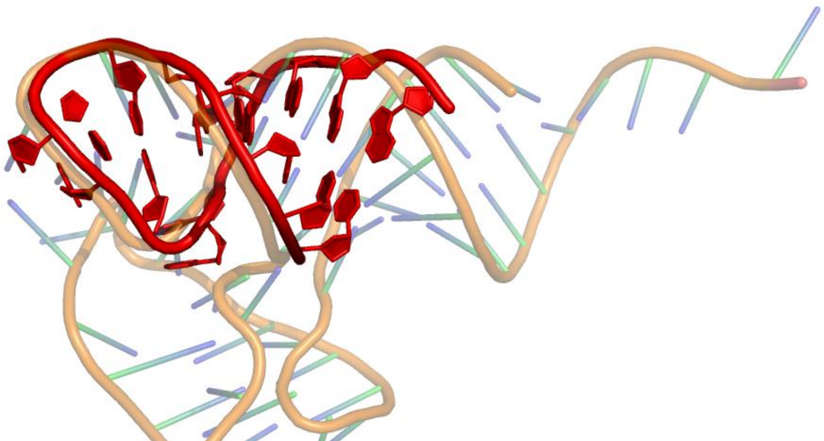
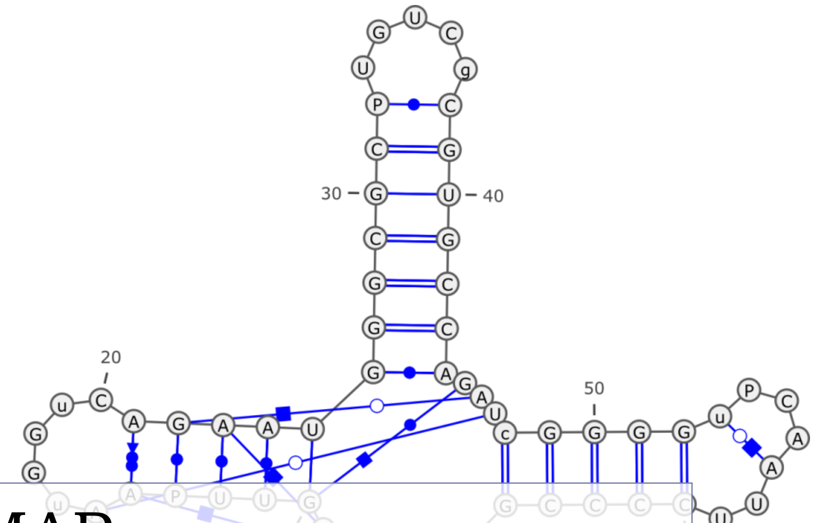




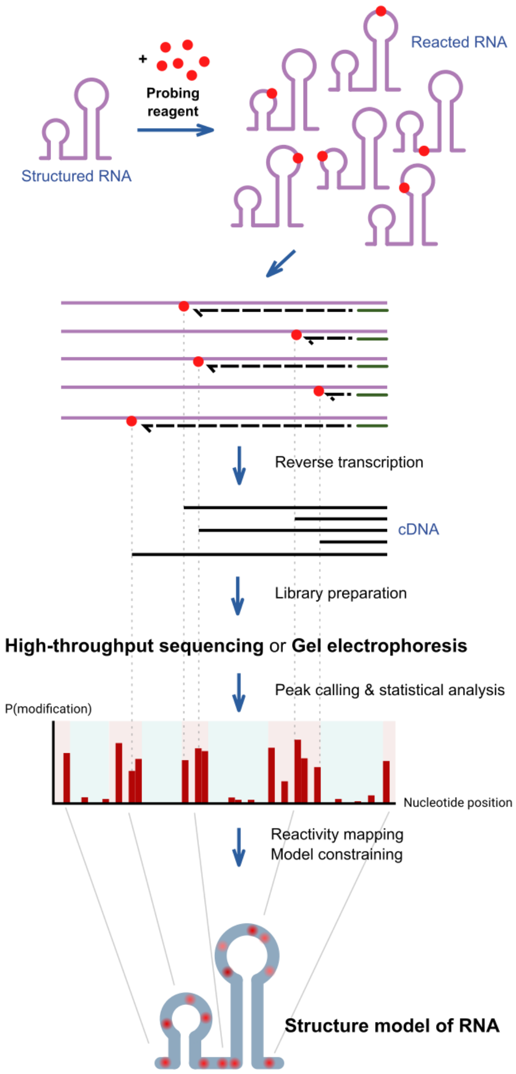
IPANEMAP

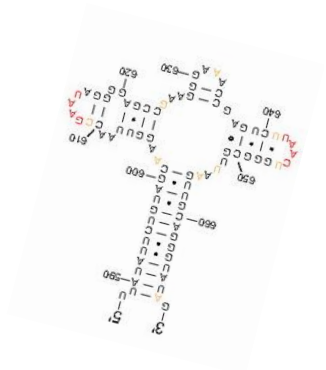
Integrative **P**robing **A**nalysis of **N**ucleic Acids **E**mpowered by **M**ultiple **A**ccessibility **P**rofiles

Afaf SAAIDI & Yann PONTY
CNRS/Ecole Polytechnique

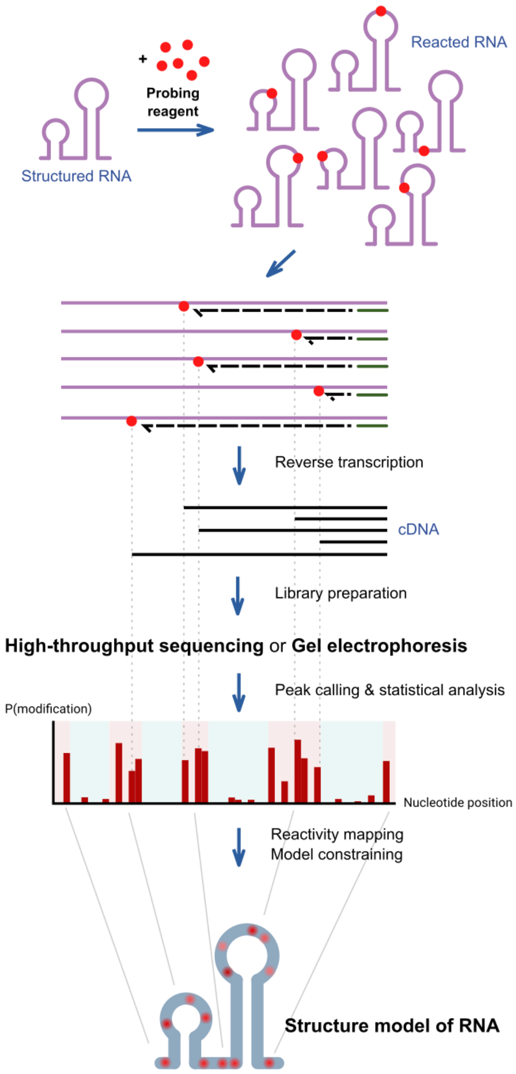
Bruno SARGUEIL & Delphine ALLOUCHE
Univ. Paris Descartes

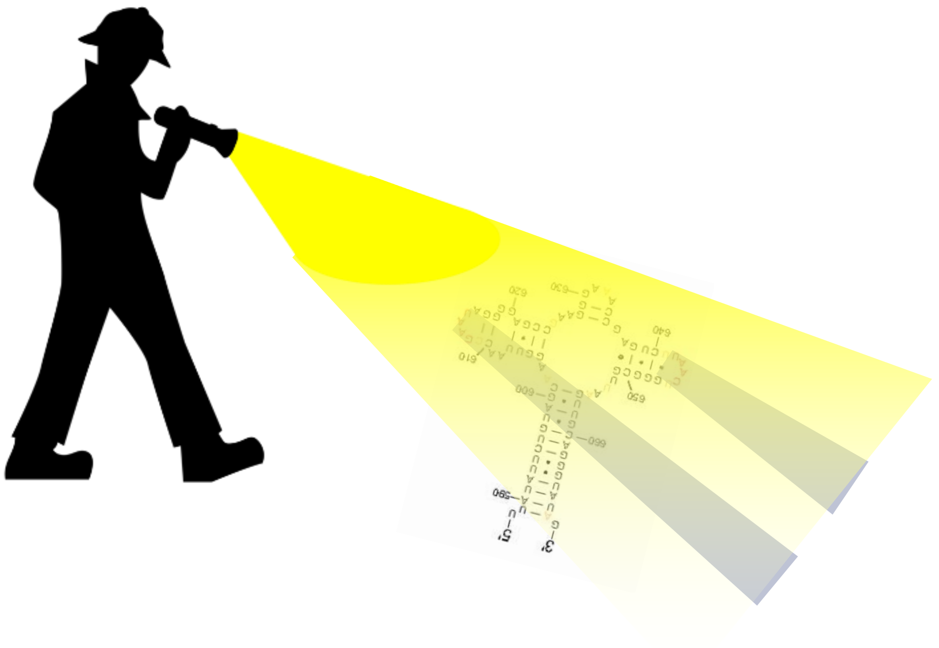
Structure probing in a nutshell



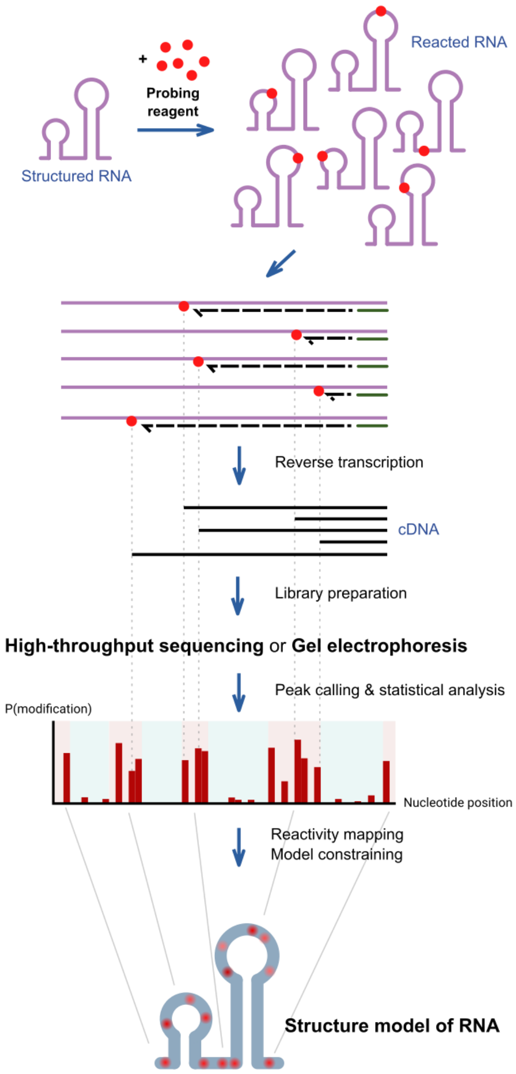


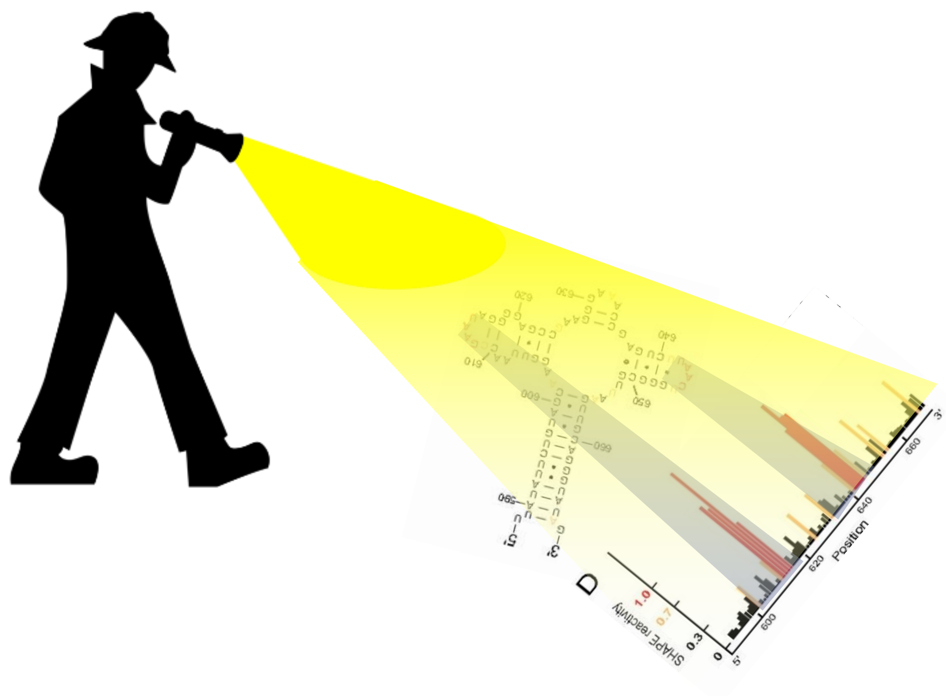
Structure probing in a nutshell



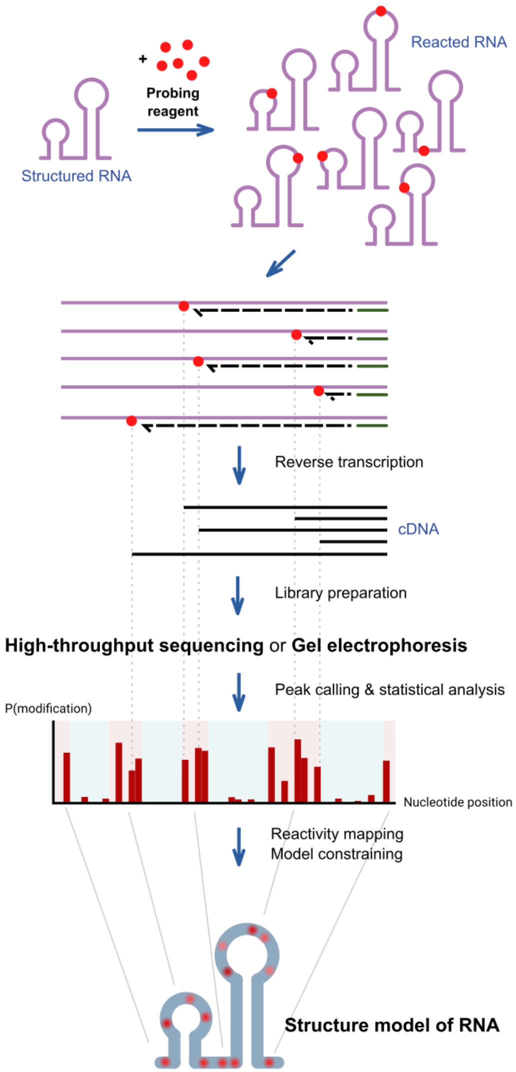


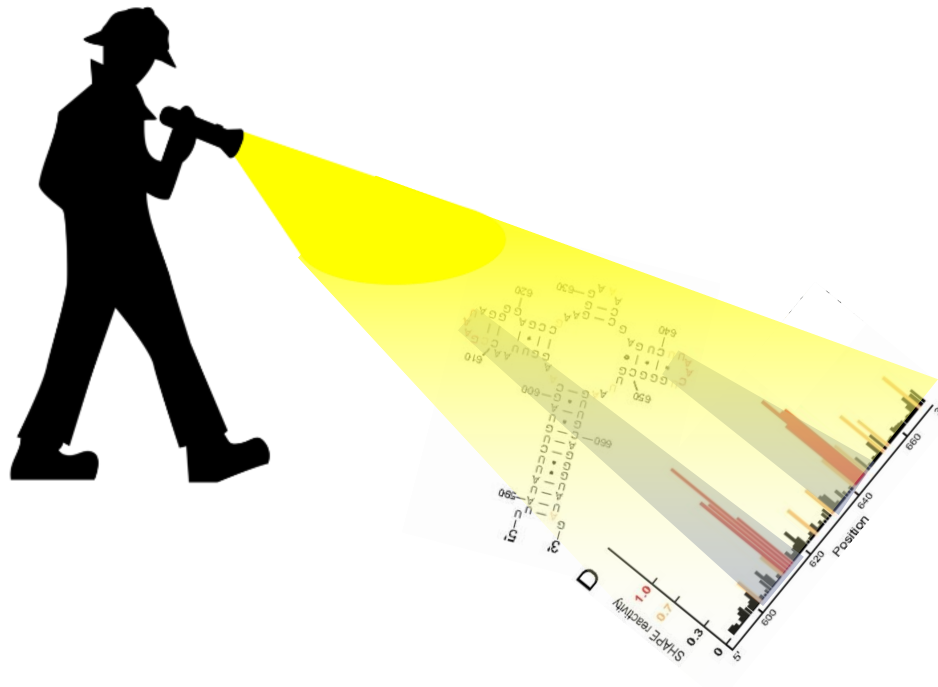
Structure probing in a nutshell





Structure probing in a nutshell

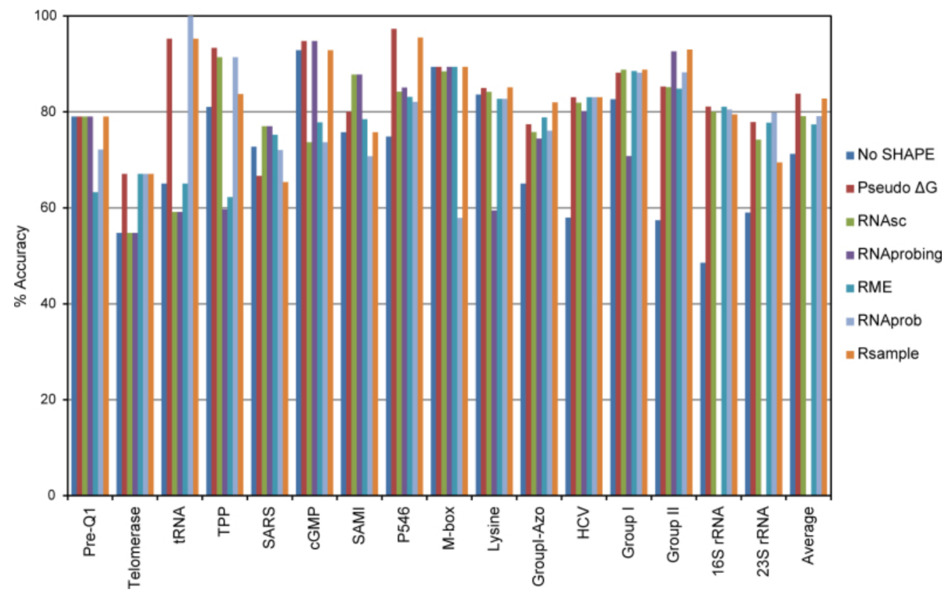




The SHAPE wars: 100% accuracy and beyond…

Now, wait a minute guys…!

David H Mathews



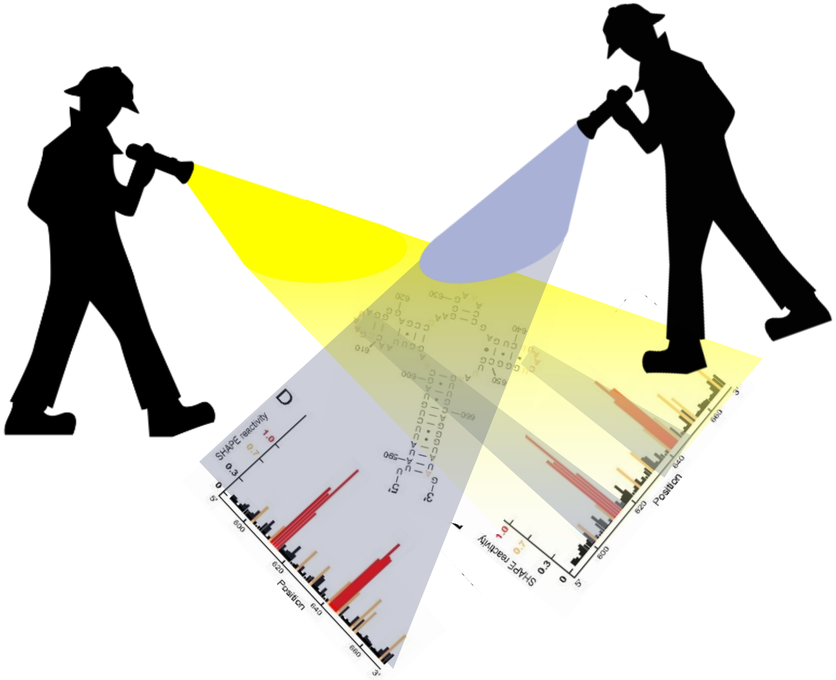
Spasic *et al*, NAR 2018

The practice of RNA modeling

Bruno Sargueil

Probing is tricky to perform and interpret.

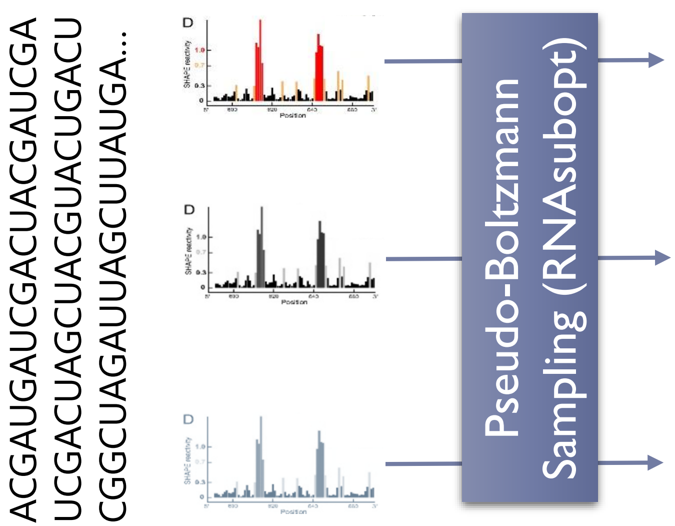
Beyond RNAPuzzle, modelers usually try different techniques and reagents, but integration is no gimme…



How to integrate multiple probing profiles?

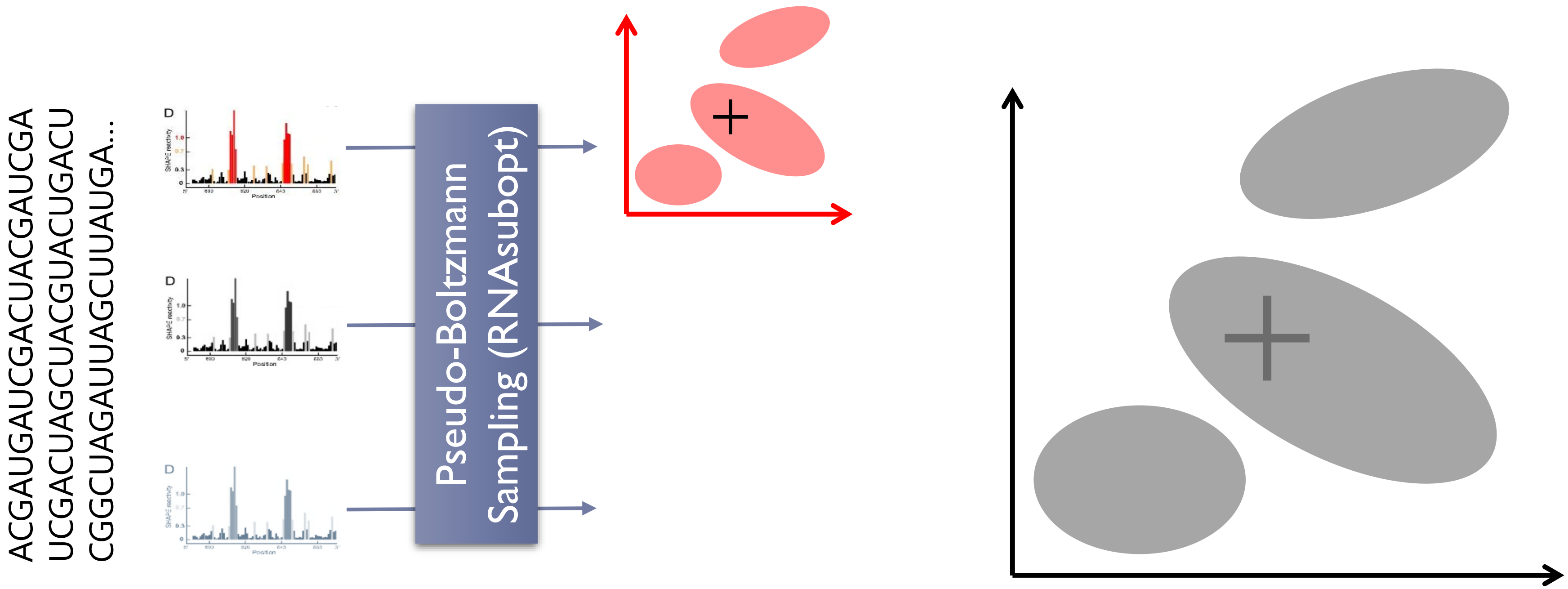
A KISS approach

- Reactivity profiles uniformly captured as pseudo potentials in prediction methods (RNAsubopt + Soft constraints)
- Native structure should be represented in each of the pseudo-Boltzmann ensemble (maybe not at the top)



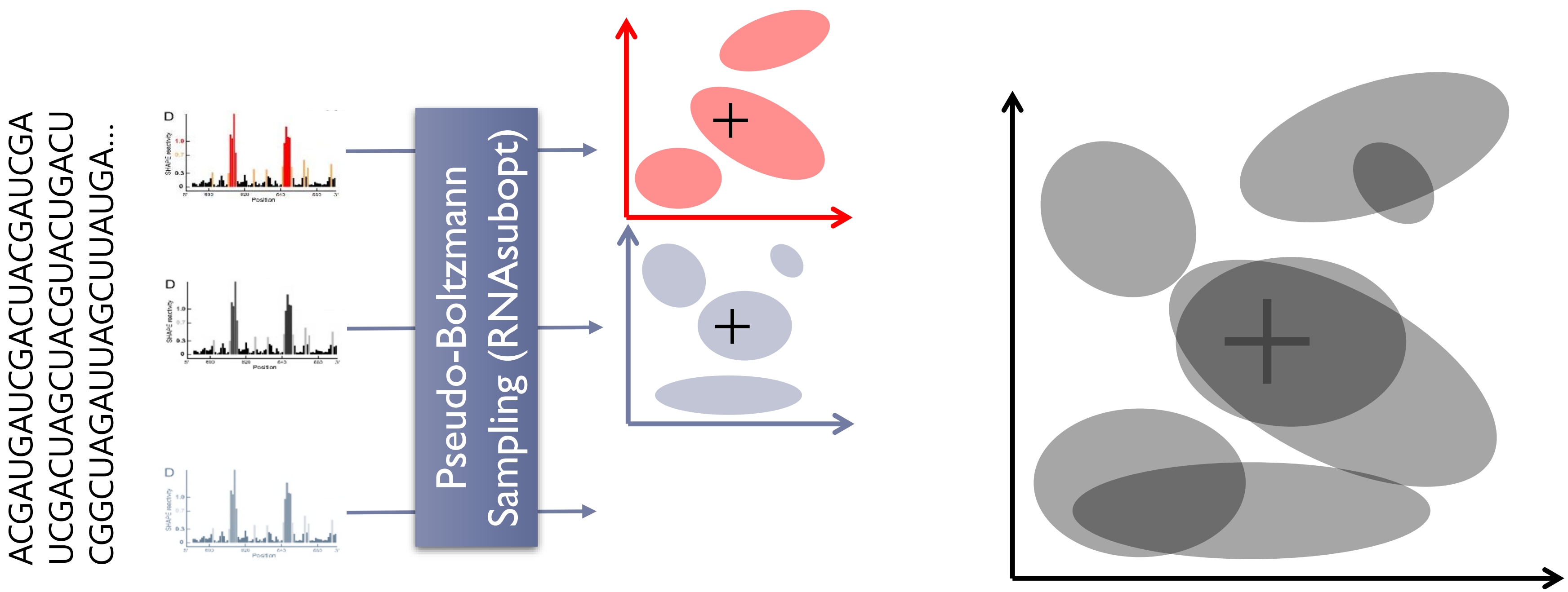
A KISS approach

- Reactivity profiles uniformly captured as pseudo potentials in prediction methods (RNAsubopt + Soft constraints)
- Native structure should be represented in each of the pseudo-Boltzmann ensemble (maybe not at the top)



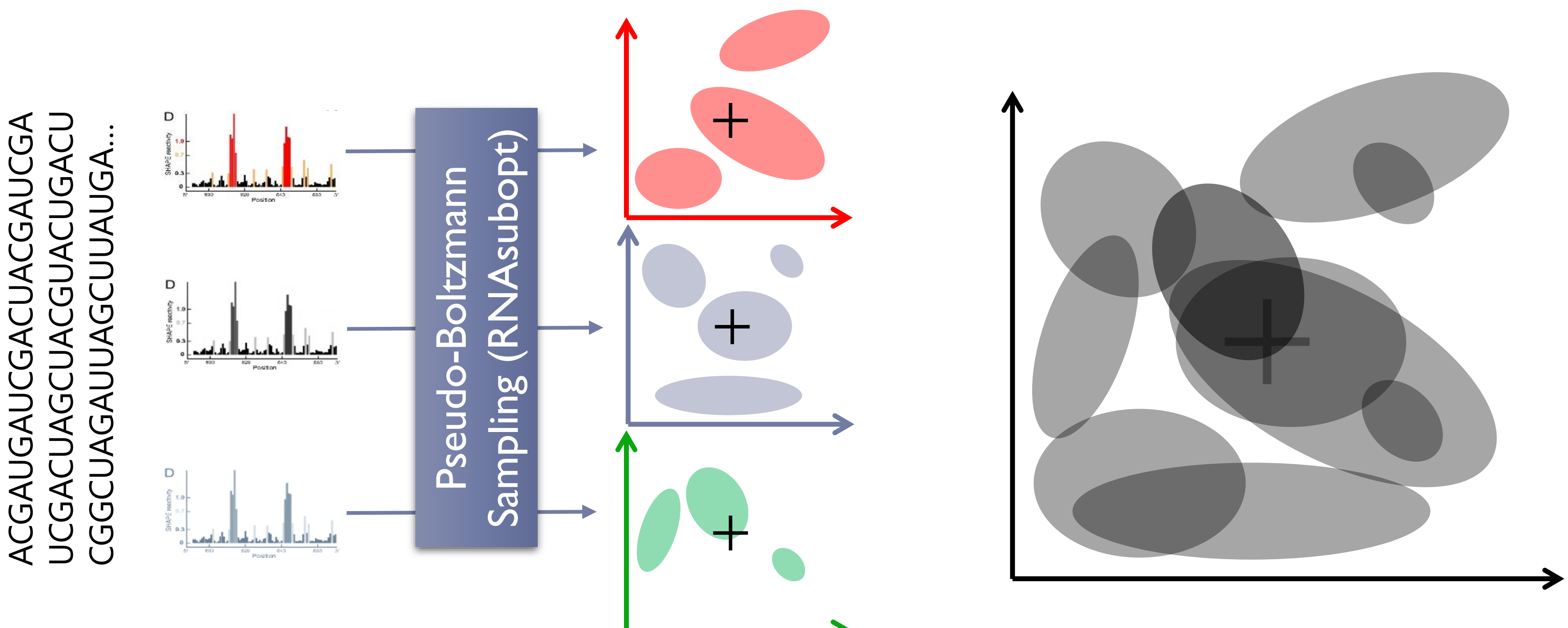
A KISS approach

- Reactivity profiles uniformly captured as pseudo potentials in prediction methods (RNAsubopt + Soft constraints)
- Native structure should be represented in each of the pseudo-Boltzmann ensemble (maybe not at the top)

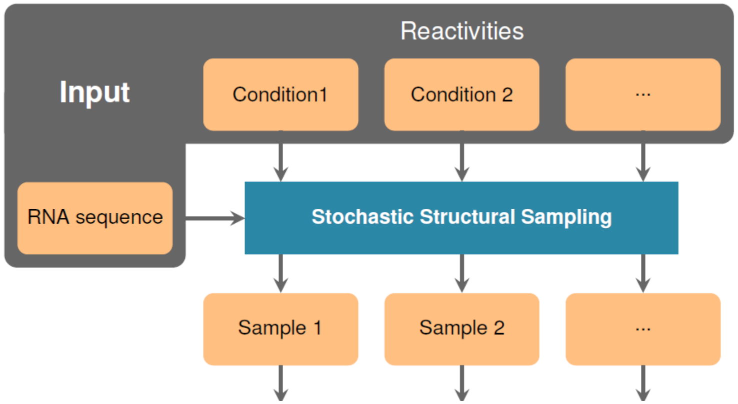


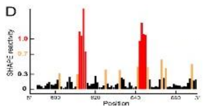
A KISS approach

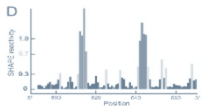
- Reactivity profiles uniformly captured as pseudo potentials in prediction methods (RNAsubopt + Soft constraints)
- Native structure should be represented in each of the pseudo-Boltzmann ensemble (maybe not at the top)

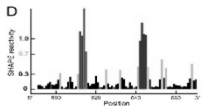


IPANEMAP Pipeline

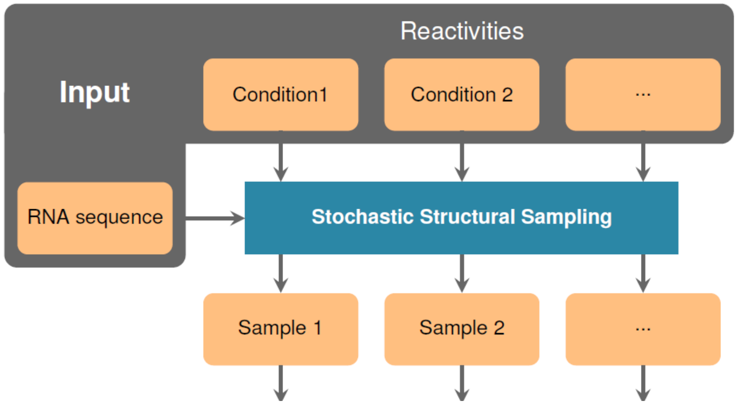


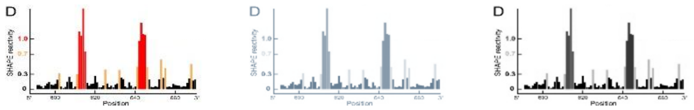


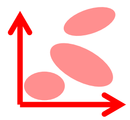




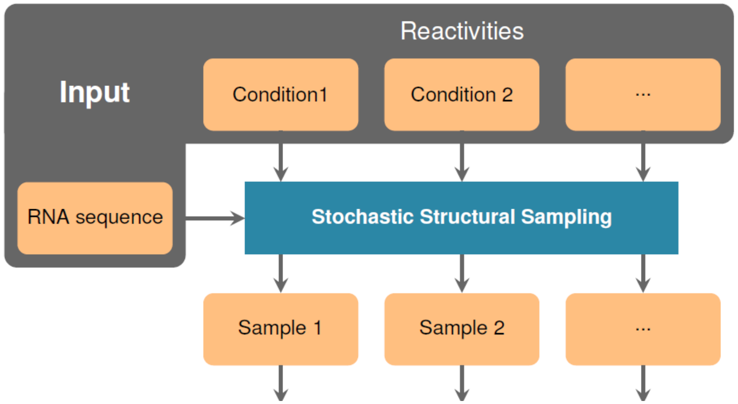
IPANEMAP Pipeline

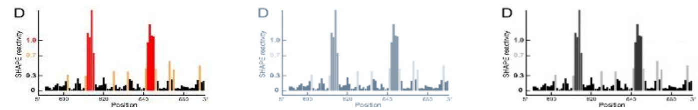




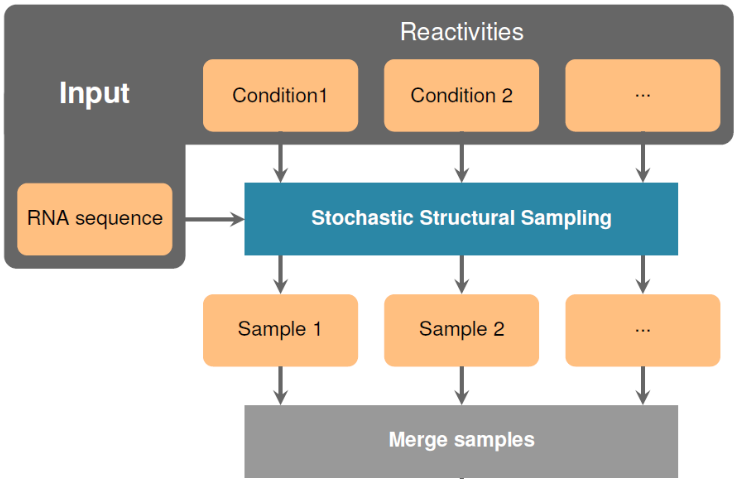


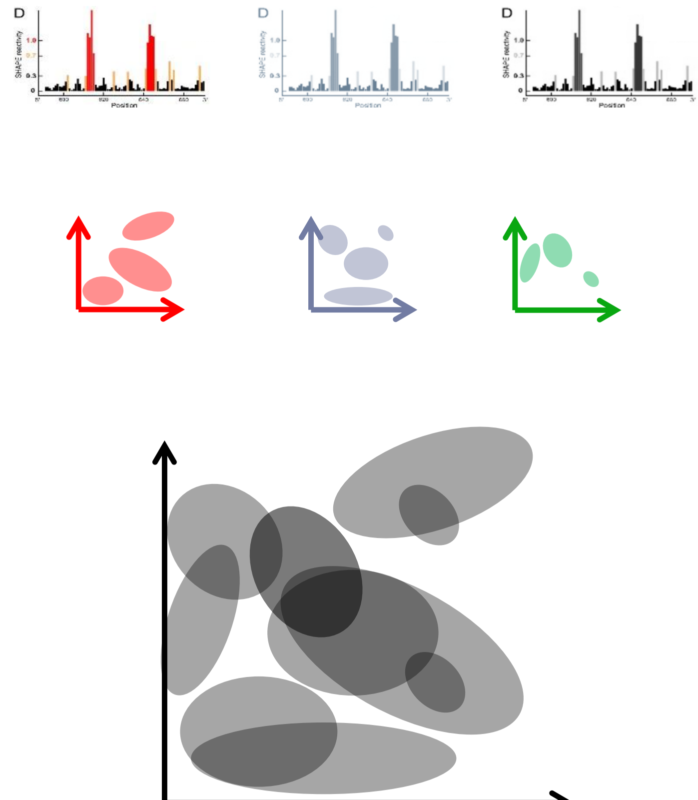
IPANEMAP Pipeline



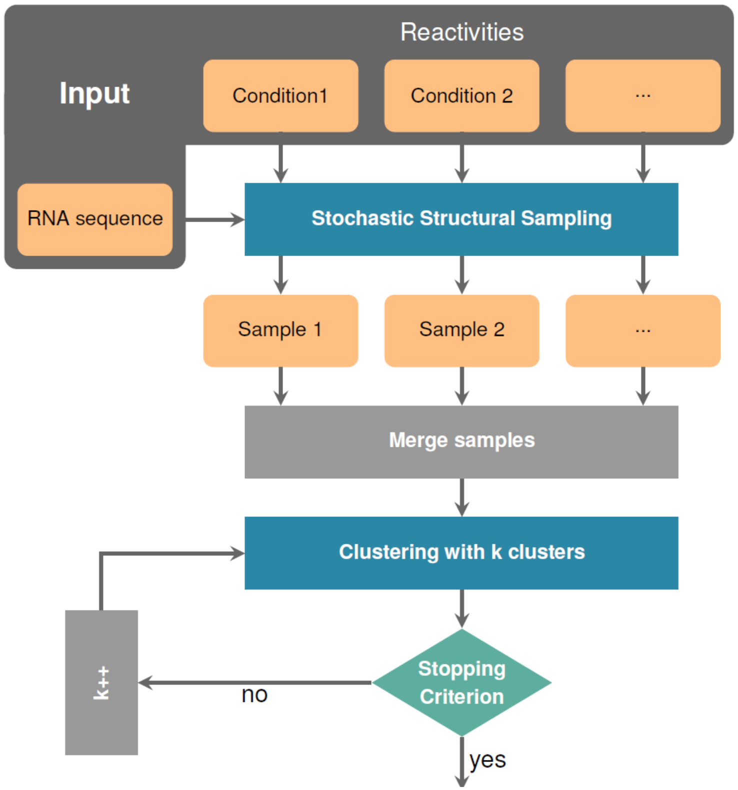


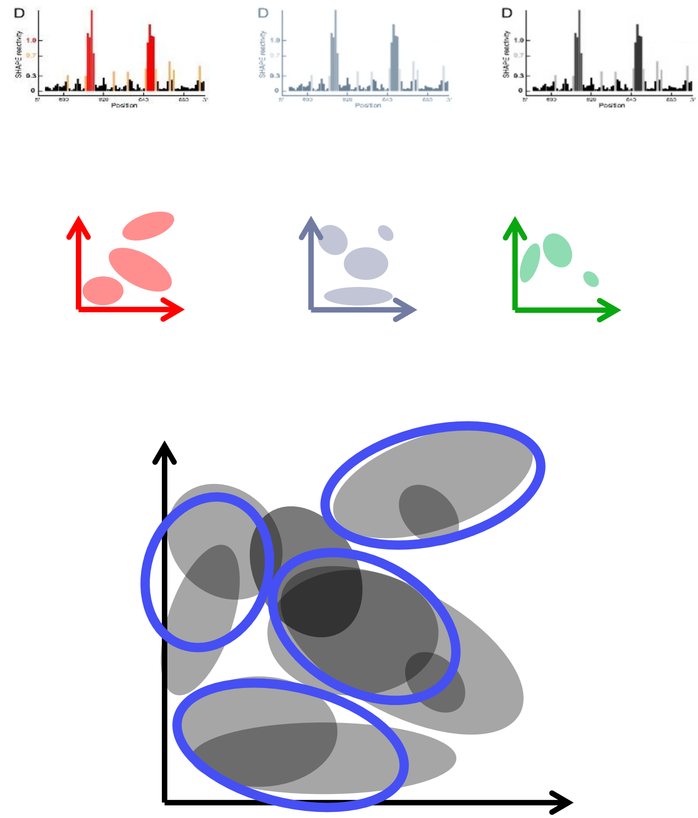
IPANEMAP Pipeline



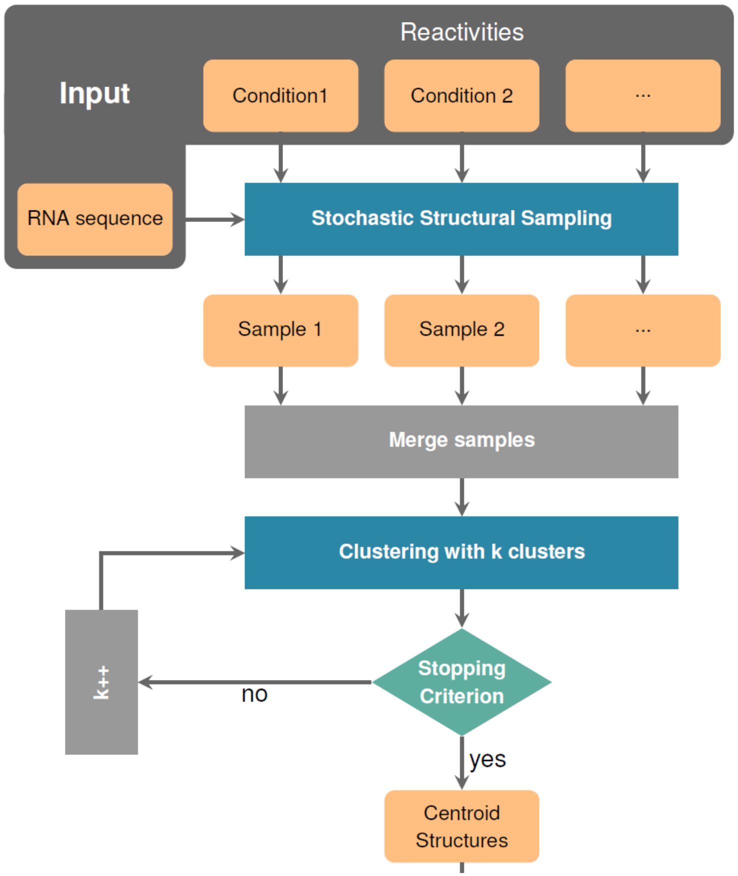


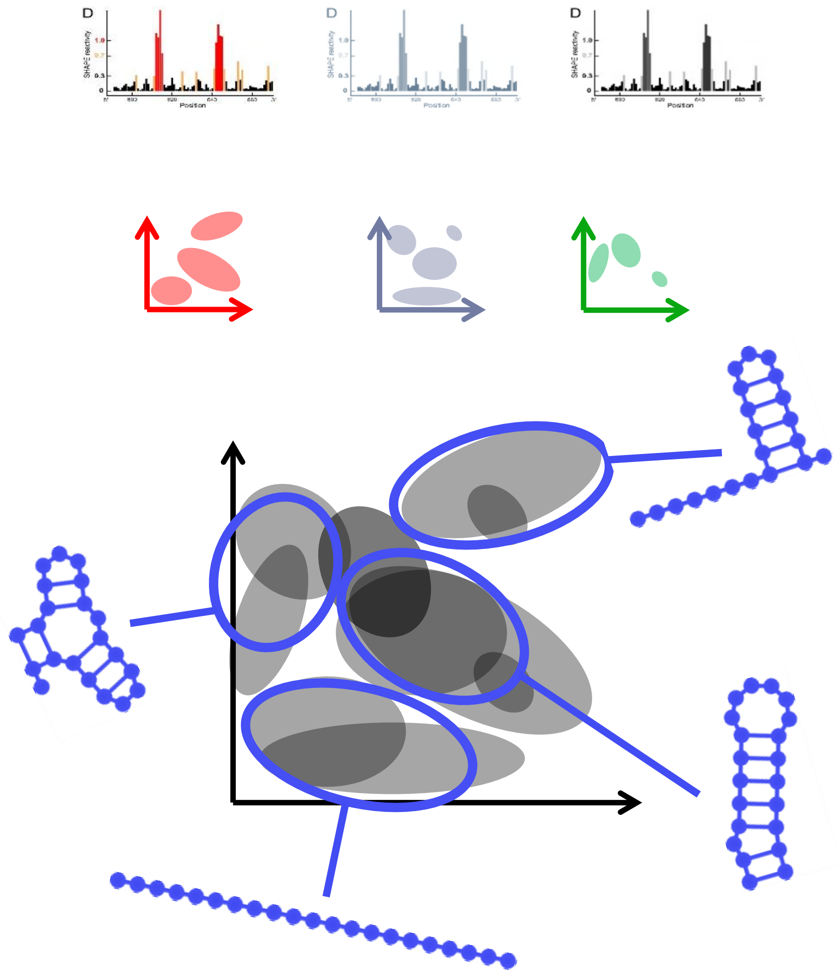
IPANEMAP Pipeline



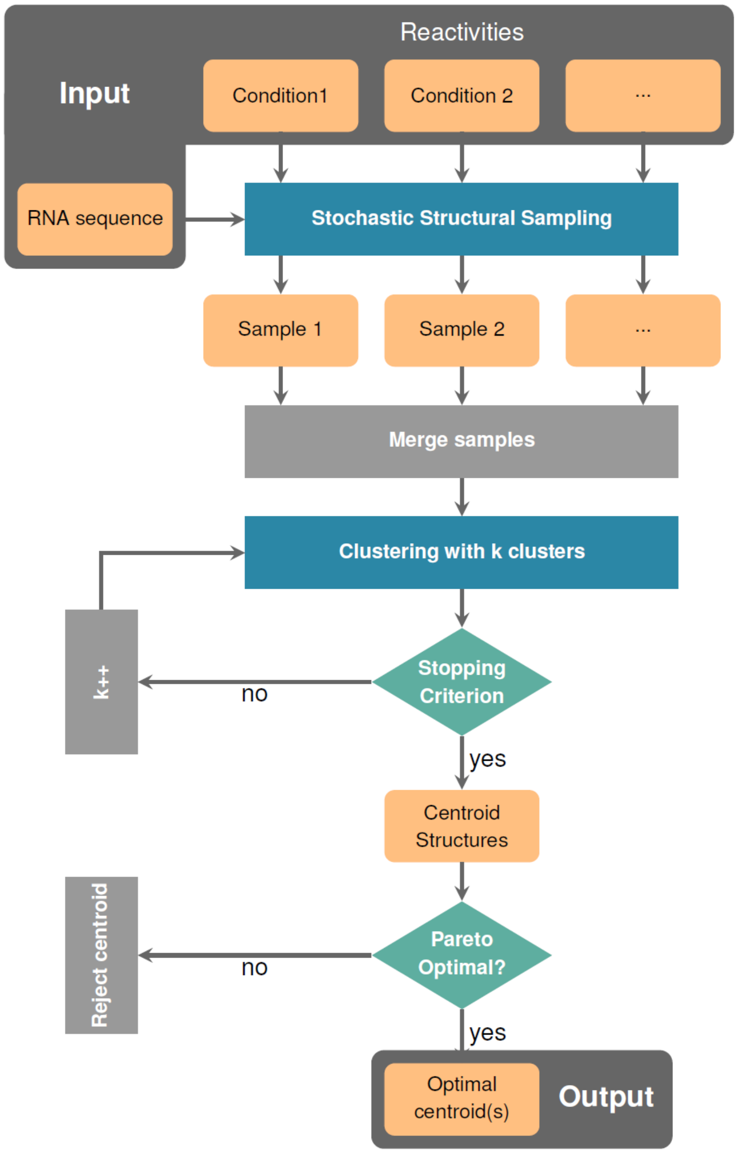


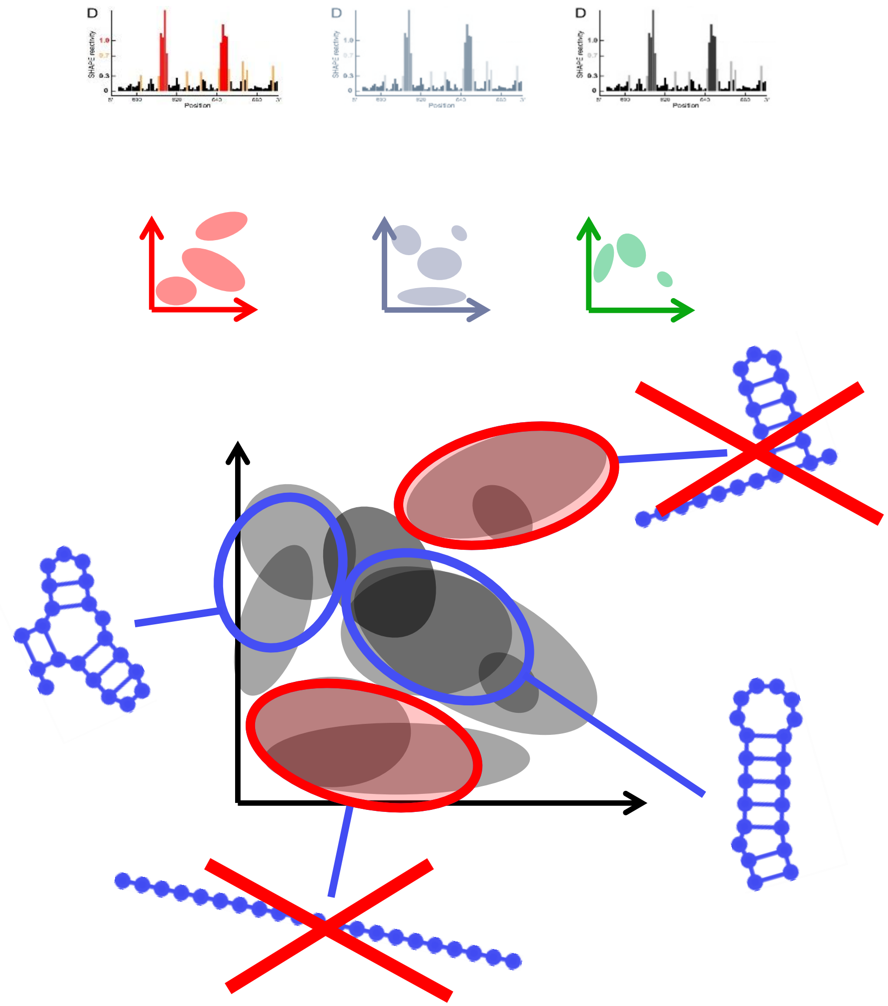
IPANEMAP Pipeline



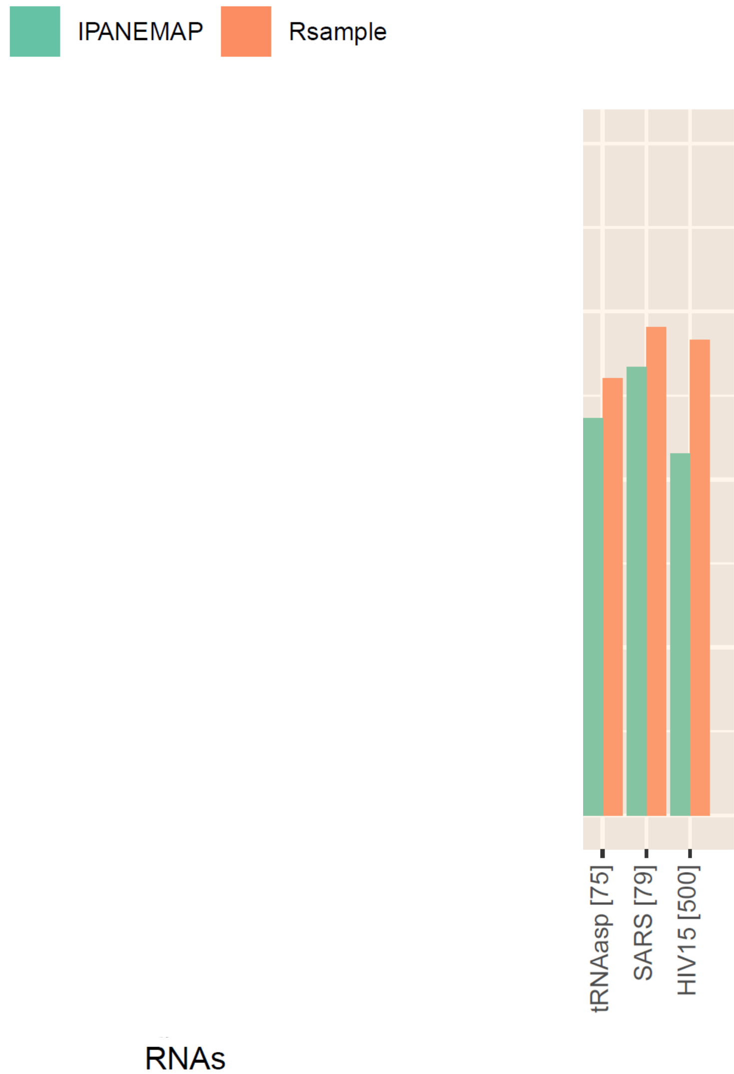


IPANEMAP Pipeline





Single condition/single structure dataset



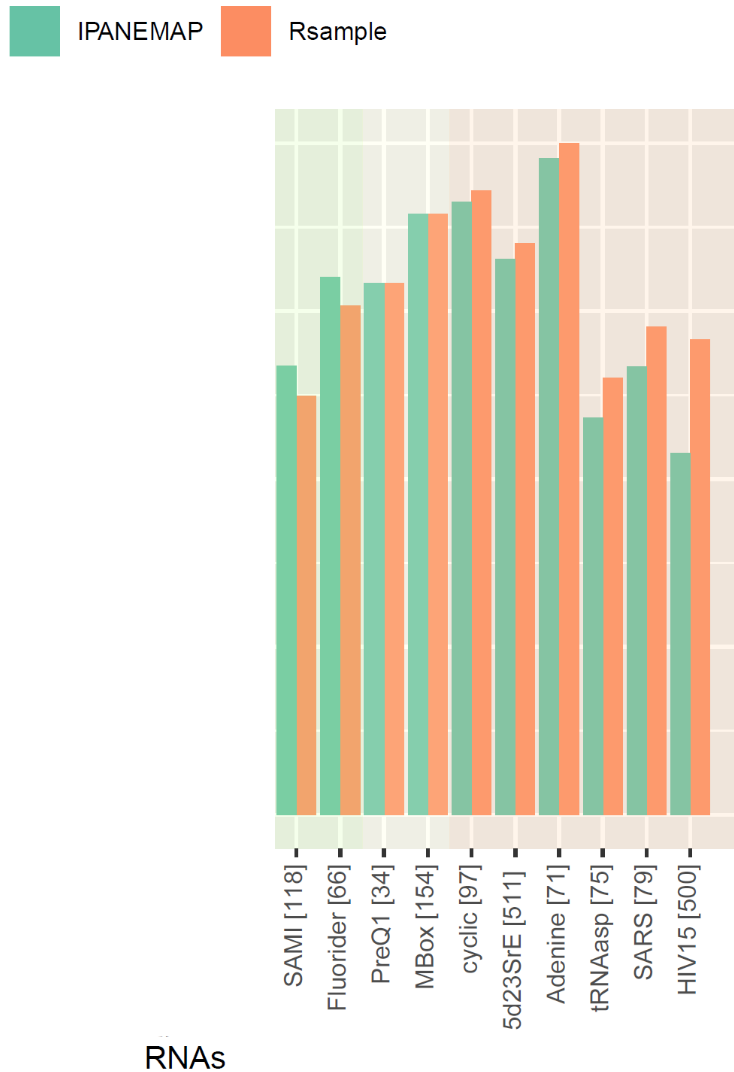
- IPANEMAP vs Rsample

  [Spasic *et al*, NAR 2018]

- Hajdin *et al* dataset
  SHAPE 1M7
  50-500nts RNAs

Geometric **M**ean $\mathrm{GM}(S \mid R) = \sqrt{\mathrm{Sens}(S \mid R) \times \mathrm{PPV}(S \mid R)}$

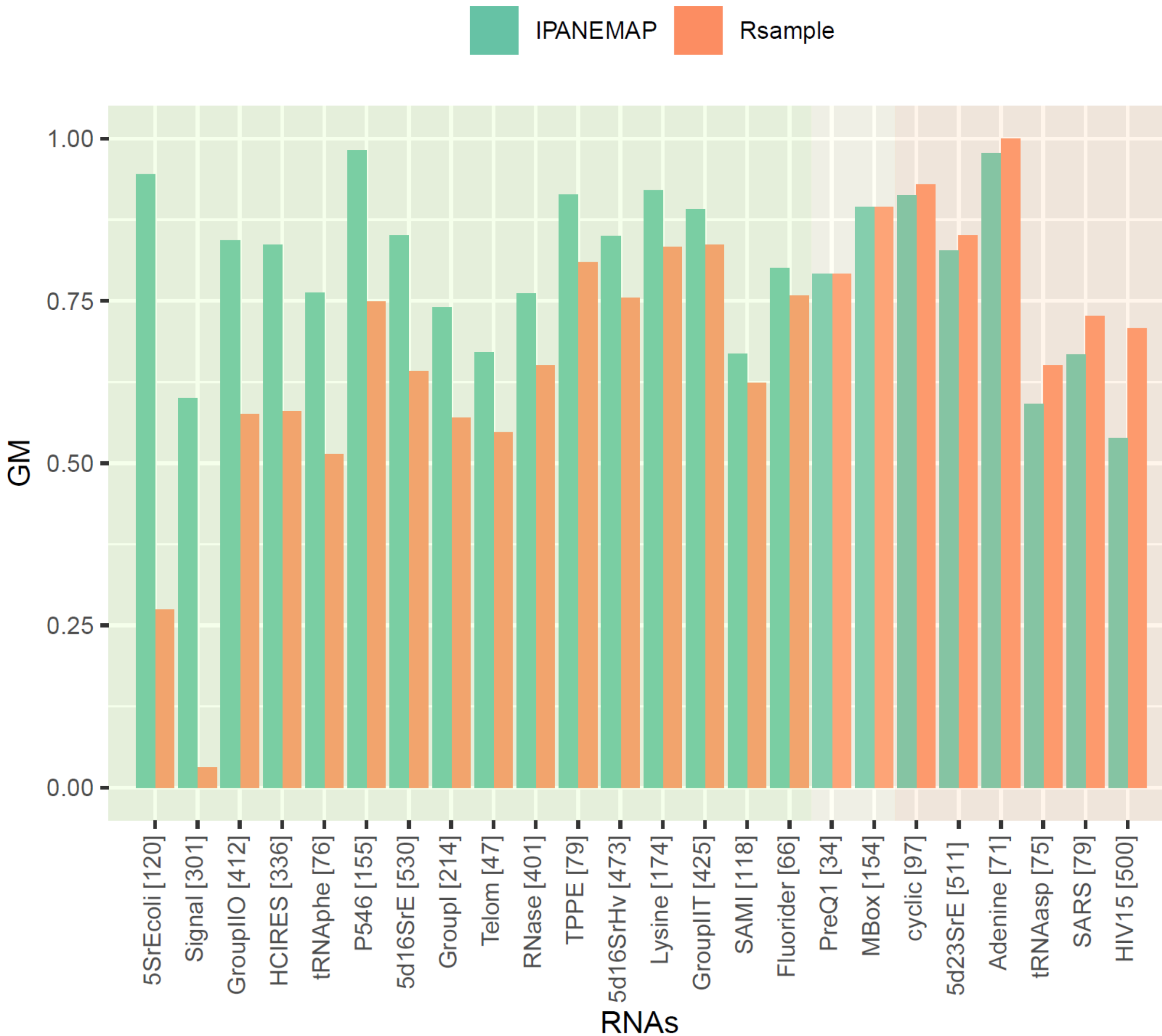
Single condition/single structure dataset



- IPANEMAP vs Rsample
  [Spasic *et al*, NAR 2018]
- Hajdin *et al* dataset
  SHAPE 1M7
  50-500nts RNAs

Geometric **M**ean $\mathrm{GM}(S \mid R) = \sqrt{\mathrm{Sens}(S \mid R) \times \mathrm{PPV}(S \mid R)}$

Single condition/single structure dataset



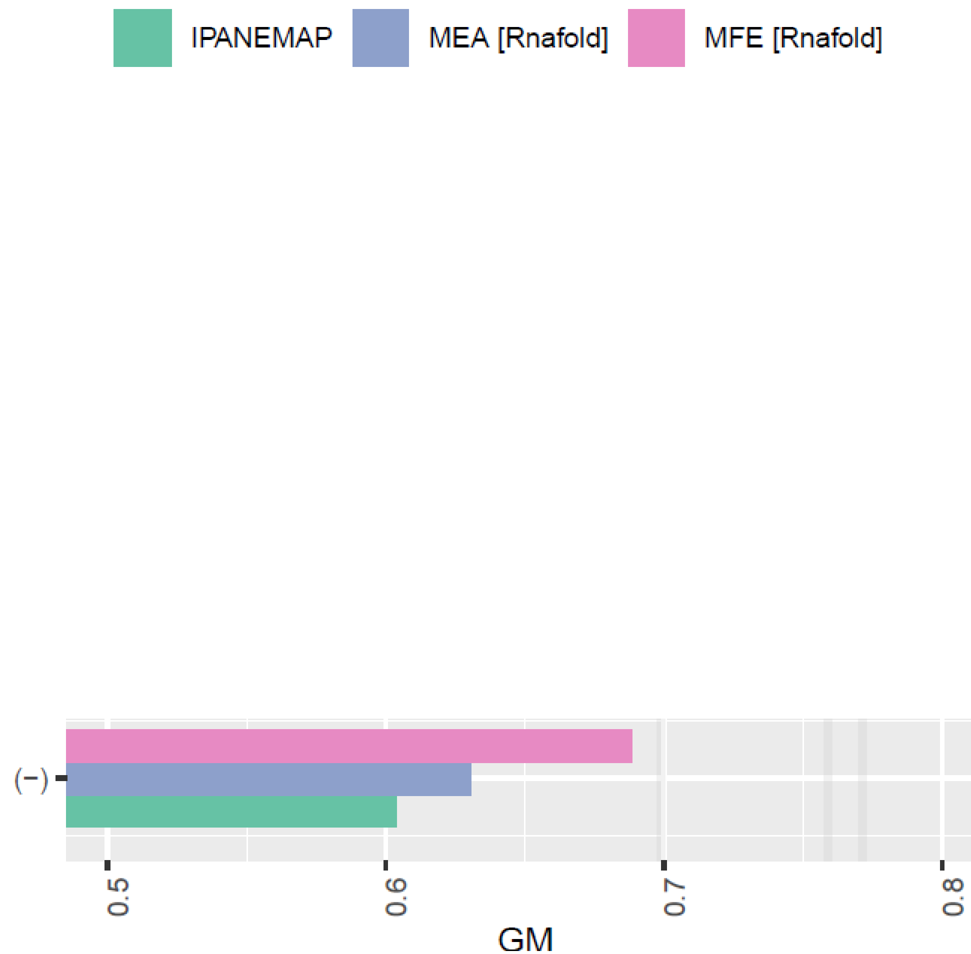
- IPANEMAP vs Rsample
  [Spasic *et al*, NAR 2018]
- Hajdin *et al* dataset
  SHAPE 1M7
  50-500nts RNAs
- 80% GM IPANEMAP
  vs 63% GM for Rsample
- Reason: Rsample ≈ MEA

Geometric **M**ean $\mathrm{GM}(S \mid R) = \sqrt{\mathrm{Sens}(S \mid R) \times \mathrm{PPV}(S \mid R)}$

Multiple conditions improve predictions?

- 6 RMDB RNAs
  - 5s RNA *E. coli*
  - Glycine Riboswitch *F. Nucleatum*
  - cidGMP riboswitch *V. Cholerae*
  - P4 - P6 domain Tetrahymena ribozyme
  - *add* Adenine Riboswitch
  - tRNA phenylalanine yeast

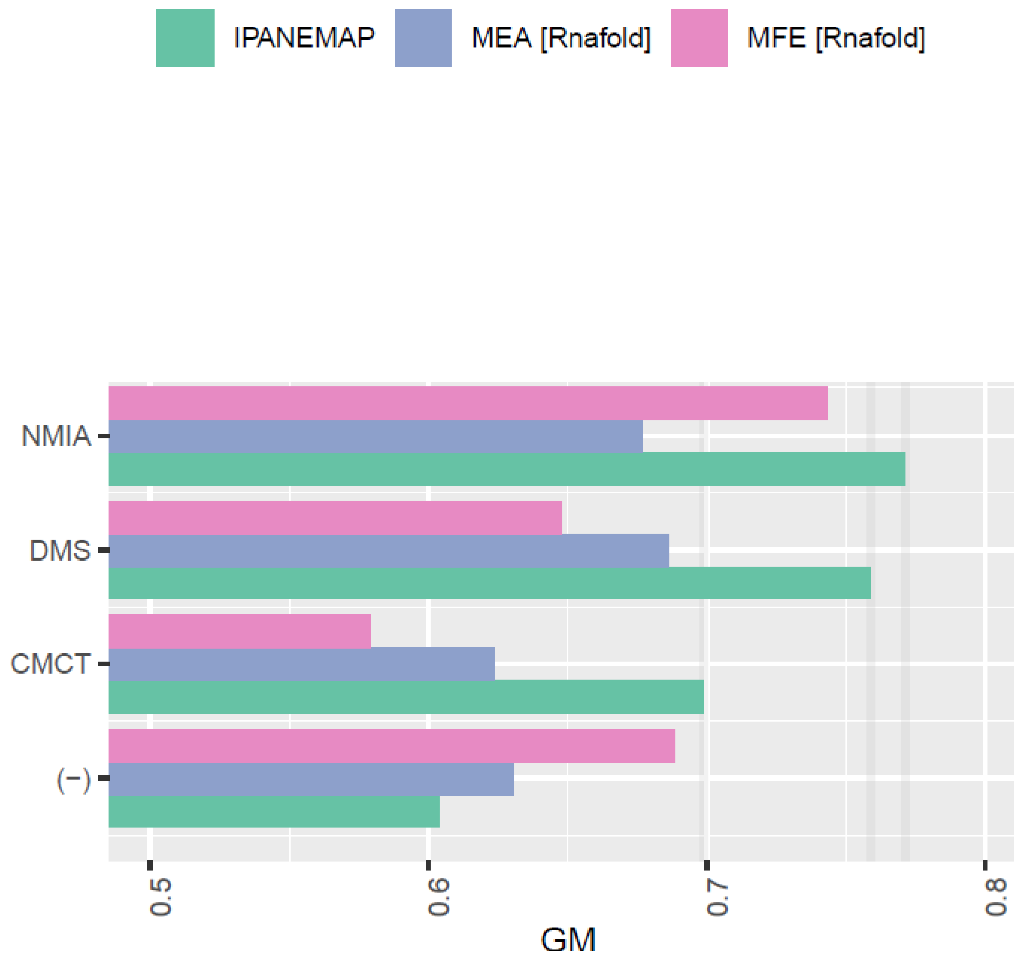
- 3 conditions
  - SHAPE (NMIA)
  - DMS
  - CMCT



Multiple conditions improve predictions?

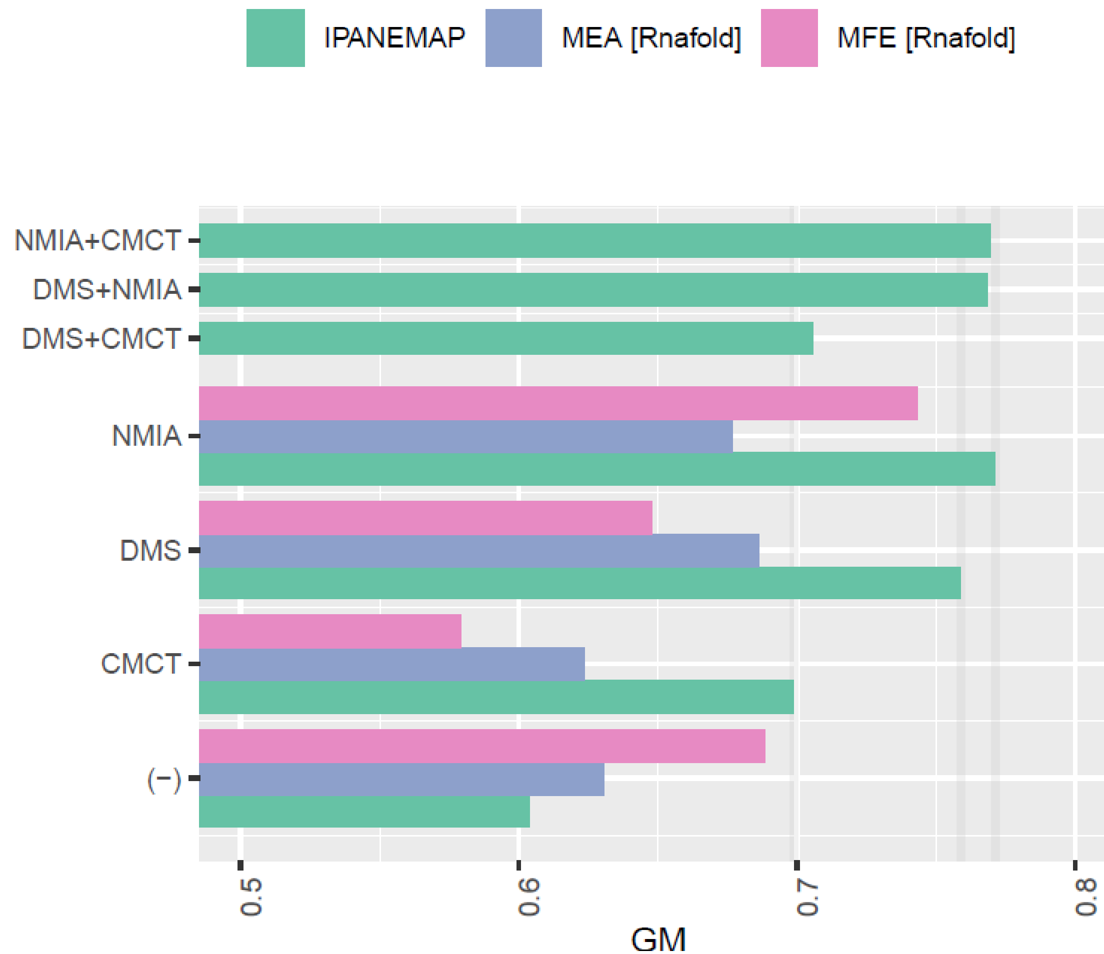
- 6 RMDB RNAs
  - 5s RNA *E. coli*
  - Glycine Riboswitch *F. Nucleatum*
  - cidGMP riboswitch *V. Cholerae*
  - P4 - P6 domain Tetrahymena ribozyme
  - *add* Adenine Riboswitch
  - tRNA phenylalanine yeast

- 3 conditions
  - SHAPE (NMIA)
  - DMS
  - CMCT



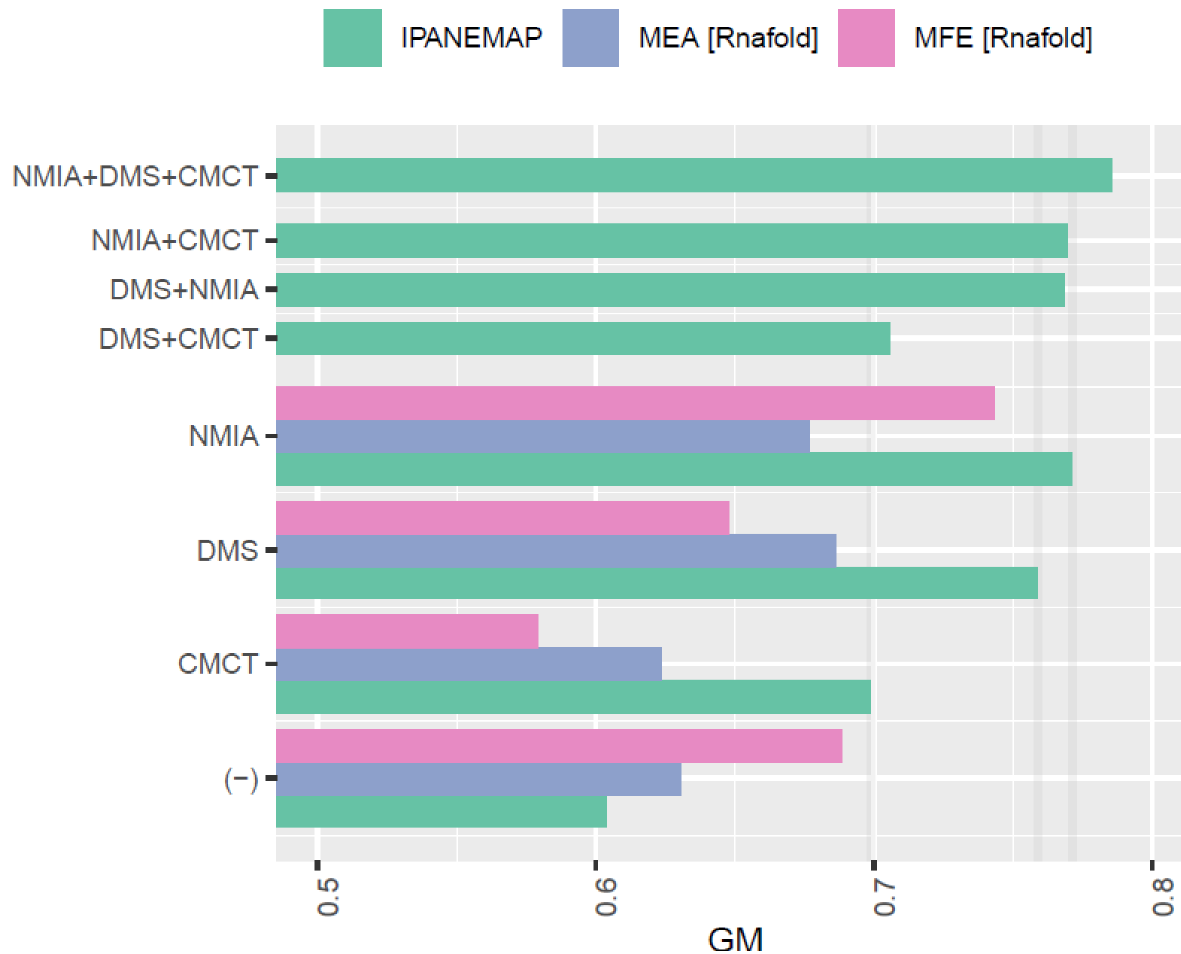
Multiple conditions improve predictions?

- 6 RMDB RNAs
  - 5s RNA *E. coli*
  - Glycine Riboswitch *F. Nucleatum*
  - cidGMP riboswitch *V. Cholerae*
  - P4 - P6 domain Tetrahymena ribozyme
  - *add* Adenine Riboswitch
  - tRNA phenylalanine yeast
- 3 conditions
  - SHAPE (NMIA)
  - DMS
  - CMCT



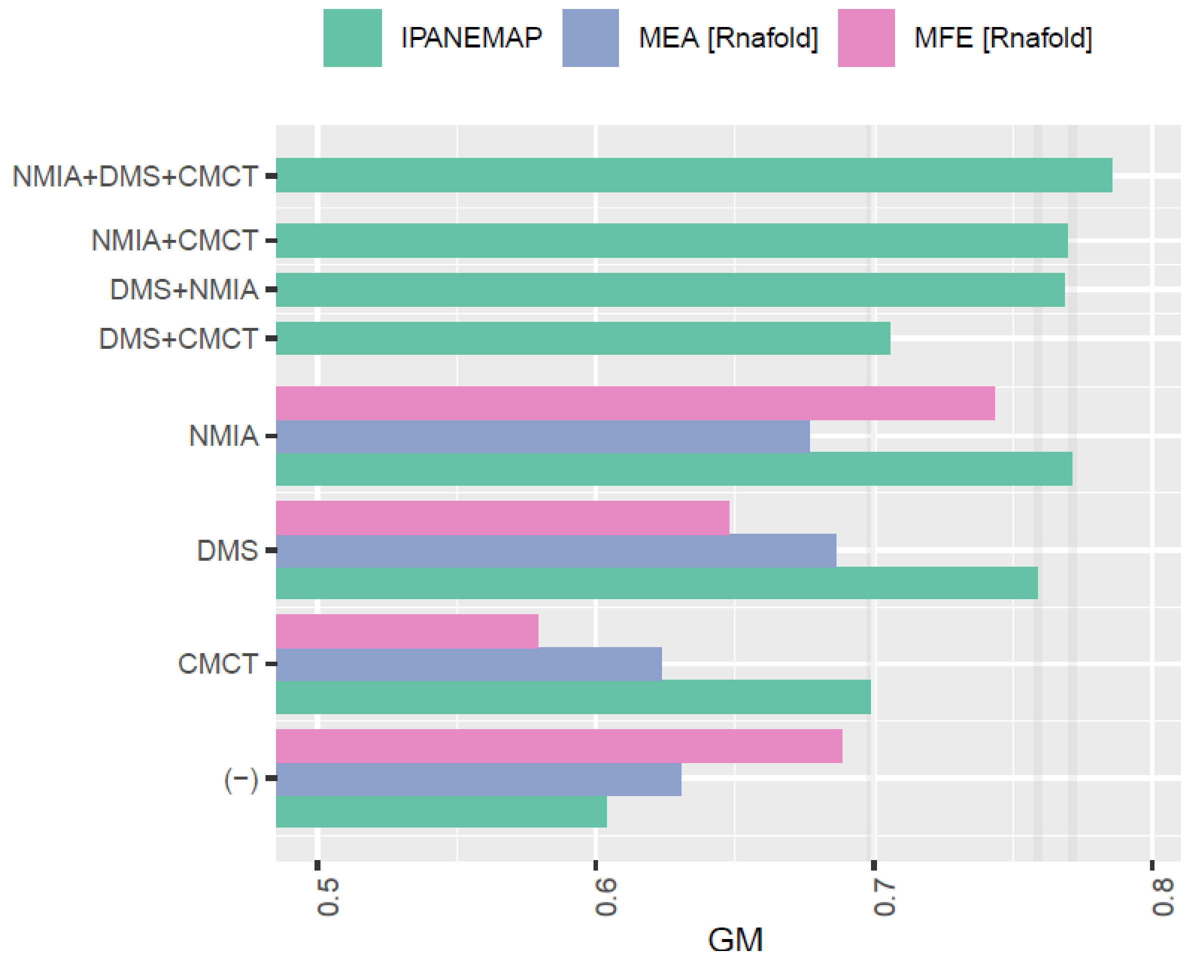
Multiple conditions improve predictions?

- 6 RMDB RNAs
  - 5s RNA *E. coli*
  - Glycine Riboswitch *F. Nucleatum*
  - cidGMP riboswitch *V. Cholerae*
  - P4 - P6 domain Tetrahymena ribozyme
  - *add* Adenine Riboswitch
  - tRNA phenylalanine yeast
- 3 conditions
  - SHAPE (NMIA)
  - DMS
  - CMCT

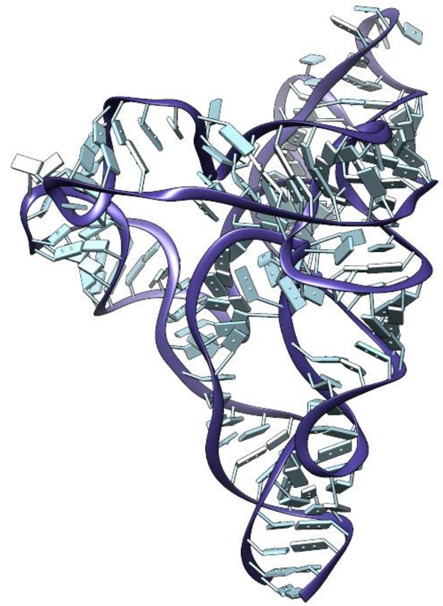


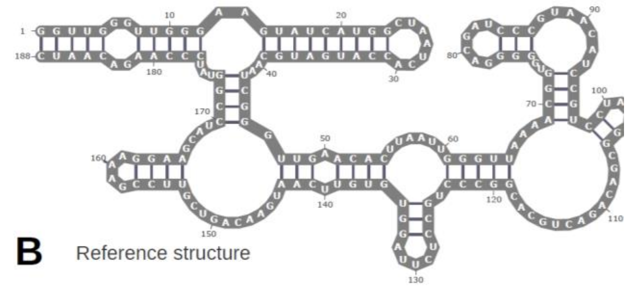
Multiple conditions improve predictions?

- **6 RMDB RNAs**
  - 5s RNA *E. coli*
  - Glycine Riboswitch *F. Nucleatum*
  - cidGMP riboswitch *V. Cholerae*
  - P4 - P6 domain Tetrahymena ribozyme
  - *add* Adenine Riboswitch
  - tRNA phenylalanine yeast
- **3 conditions**
  - SHAPE (NMIA)
  - DMS
  - CMCT



Back to the lab for even more data

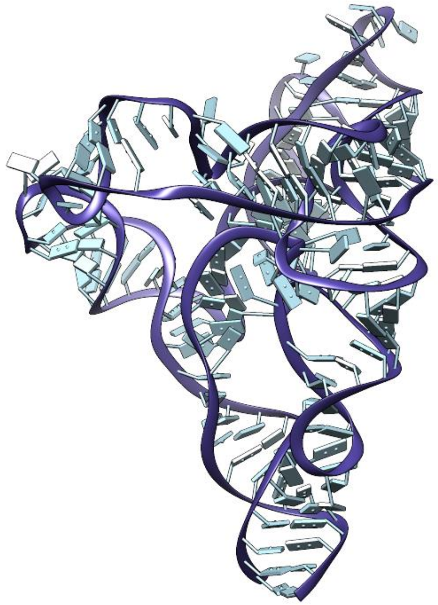


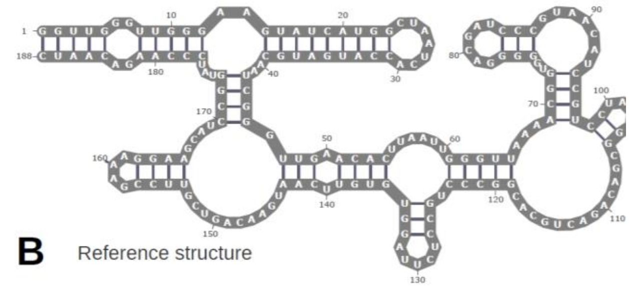


B Reference structure

- Lariat Capping Ribozyme (PDB 4P8Z) from Didymium iridis

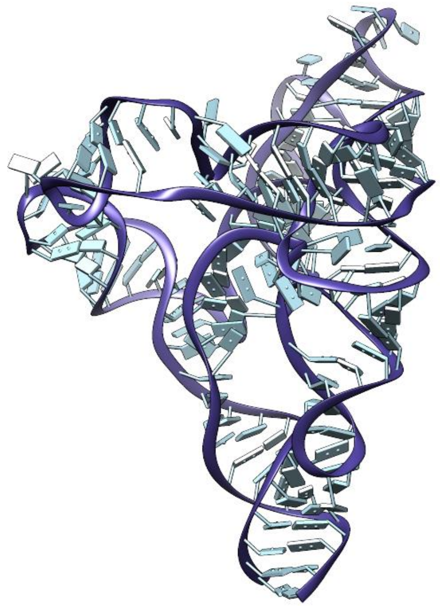
Back to the lab for even more data

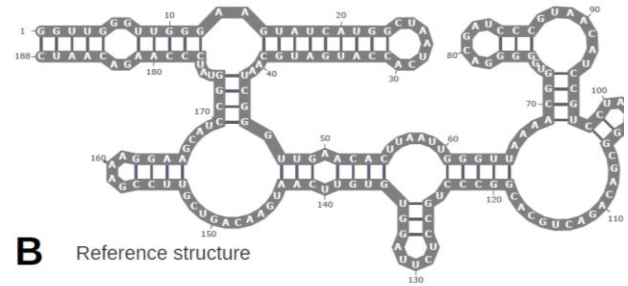




- Lariat Capping Ribozyme (PDB 4P8Z) from Didymium iridis
- Probed under 14 different conditions:
  - Technique: SHAPE, DMS, CMCT
  - Reagents: 1M7, NMIA, DMS, CMCT, NAI, BzCN
  - Stop based and Mutation based readouts
  - Presence/absence Magnesium

Back to the lab for even more data

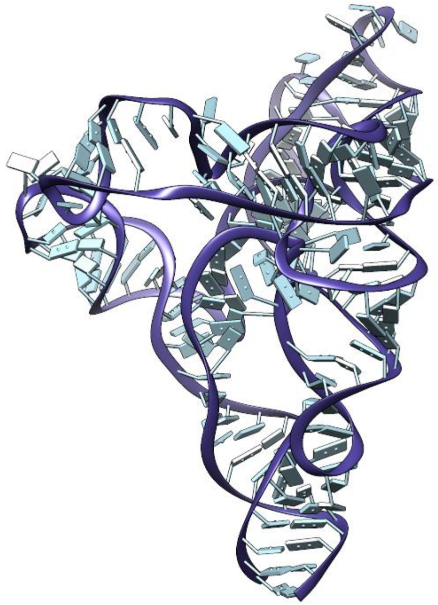


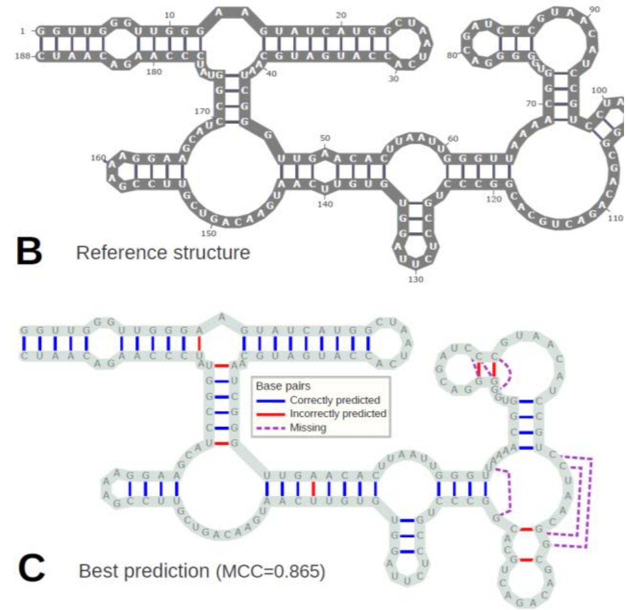


B Reference structure

- Lariat Capping Ribozyme (PDB 4P8Z) from Didymium iridis
- Probed under 14 different conditions:
  - Technique: SHAPE, DMS, CMCT
  - Reagents: 1M7, NMIA, DMS, CMCT, NAI, BzCN
  - Stop based and Mutation based readouts
  - Presence/absence Magnesium

Back to the lab for even more data





- Lariat Capping Ribozyme (PDB 4P8Z) from Didymium iridis
- Probed under 14 different conditions:
  - Technique: SHAPE, DMS, CMCT
  - Reagents: 1M7, NMIA, DMS, CMCT, NAI, BzCN
  - Stop based and Mutation based readouts
  - Presence/absence Magnesium

Mono probing analysis

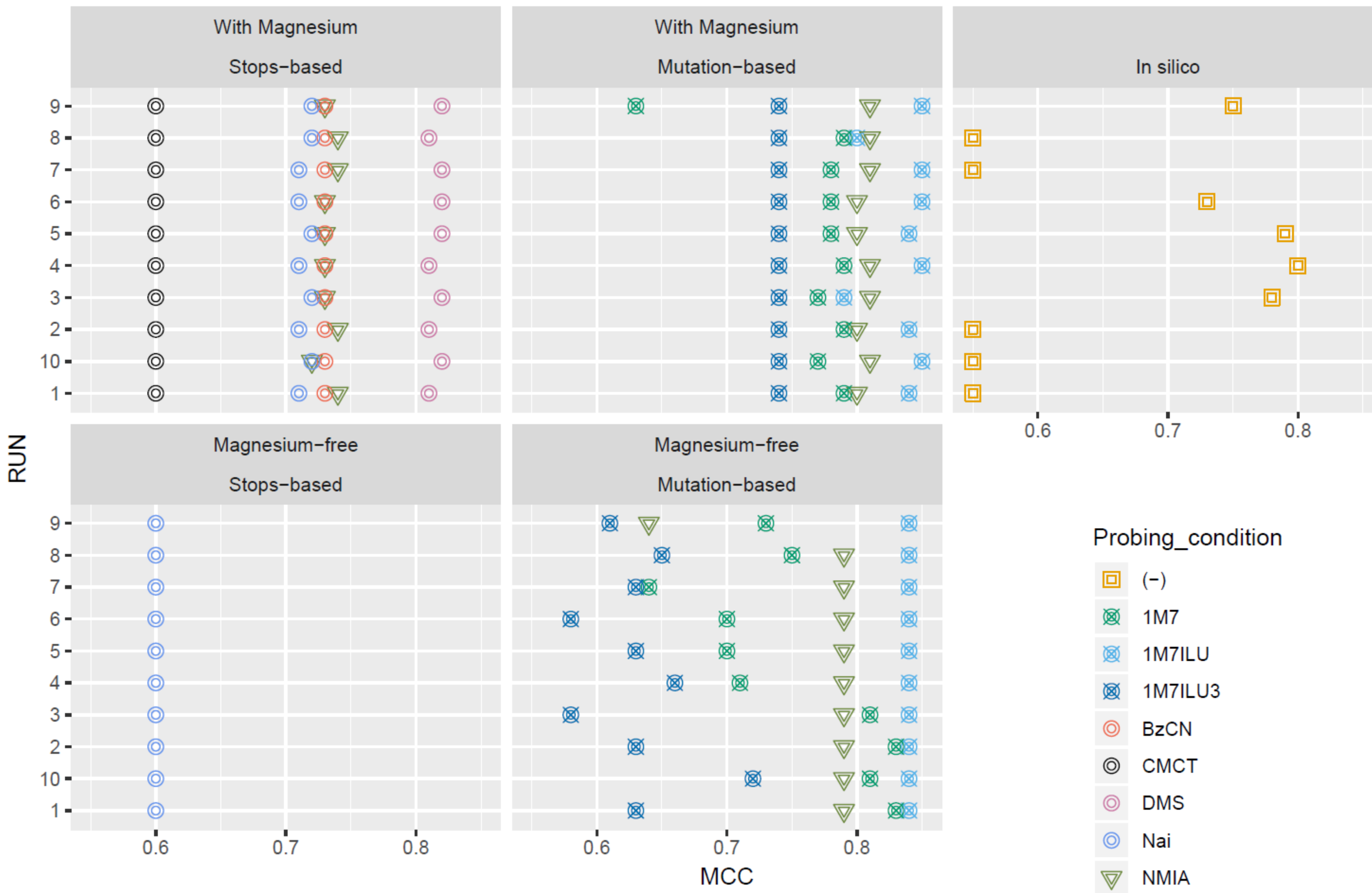
Condition name	Tech.	+/- Mg^{2+}	MCC IPANEMAP	MFE	MEA
1M7ILU	Mut	○	.84	.82	.85
1M7ILUMg	Mut	●	.83	.82	.8
DMSMg	Stop	●	.82	.41	.86
NMIAMg	Mut	●	.81	.8	.8
1M7ILU3Mg	Mut	●	.77	.75	.76
NMIAMgCE	Stop	●	.77	.675	.74
1M7Mg	Mut	●	.74	.735	.64
BzCNMg	Stop	●	.74	.72	.73
NMIA	Mut	○	.73	.69	.71
NaiMg	Stop	●	.73	.65	.71
1M7	Mut	○	.71	.51	.56
1M7ILU3	Mut	○	.62	.59	.67
CMCTMg	Stop	●	.6	.58	.59
Nai	Stop	○	.6	.59	.6
Avg Technology	Mut	–	.76	.71	.72
	Stop	–	.71	.61	.71
Avg +/- Mg^{2+}	–	○	.76	.68	.74
	–	●	.70	.64	.68
Average overall	–	–	.74	.67	.72

- Different performances
- No clear factor
- Some outliers (NAI & CMCTMg)

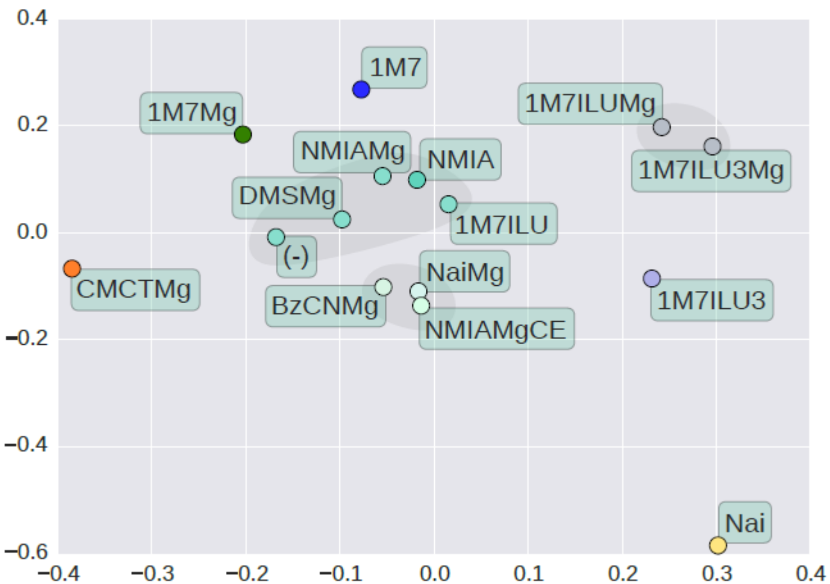
MCC:

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Reproducibility



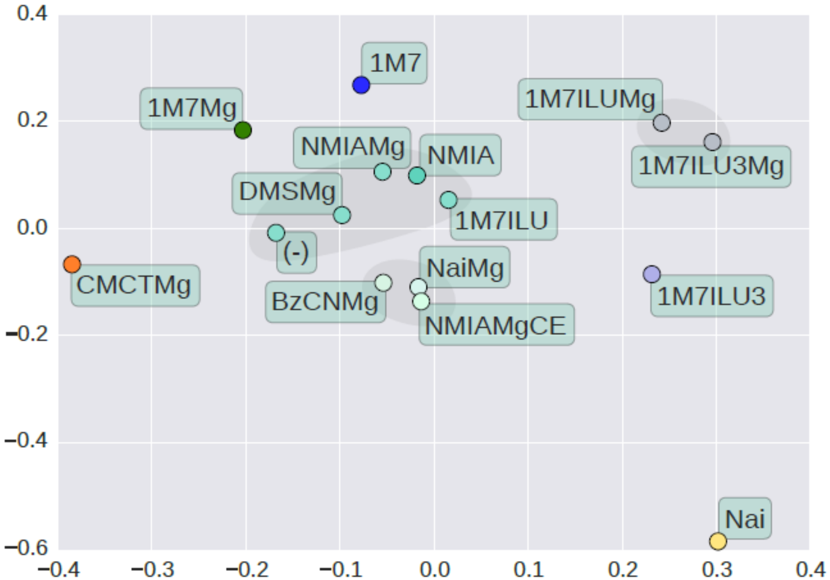
More mono probing analysis



(Squared) Euclidian distance
between conditions

$$\mathrm{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=x+1}^{n} \left(\mathbb{P}(i,j \mid d) - \mathbb{P}(i,j \mid d')\right)^2$$

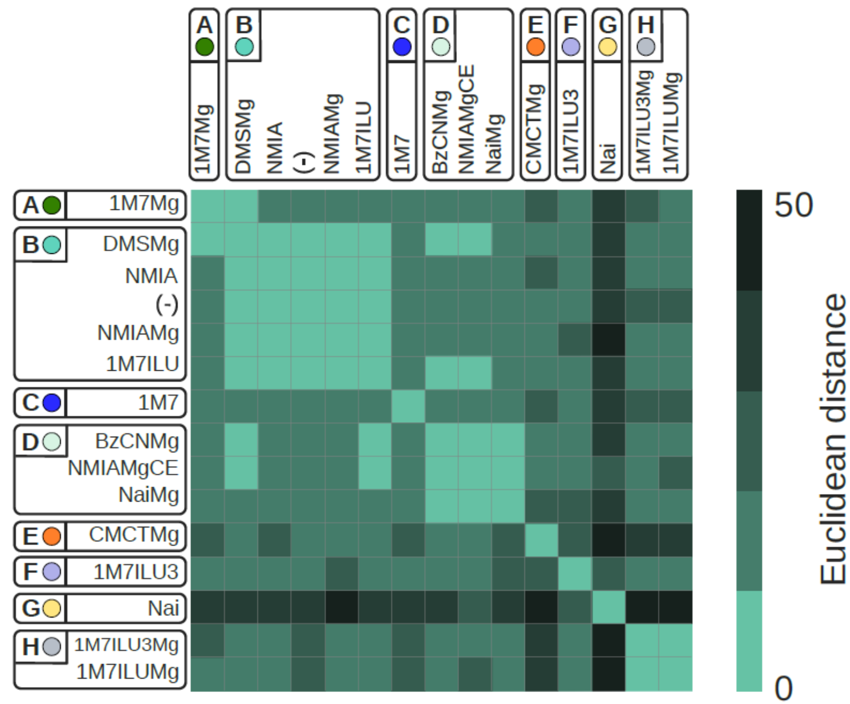
More mono probing analysis



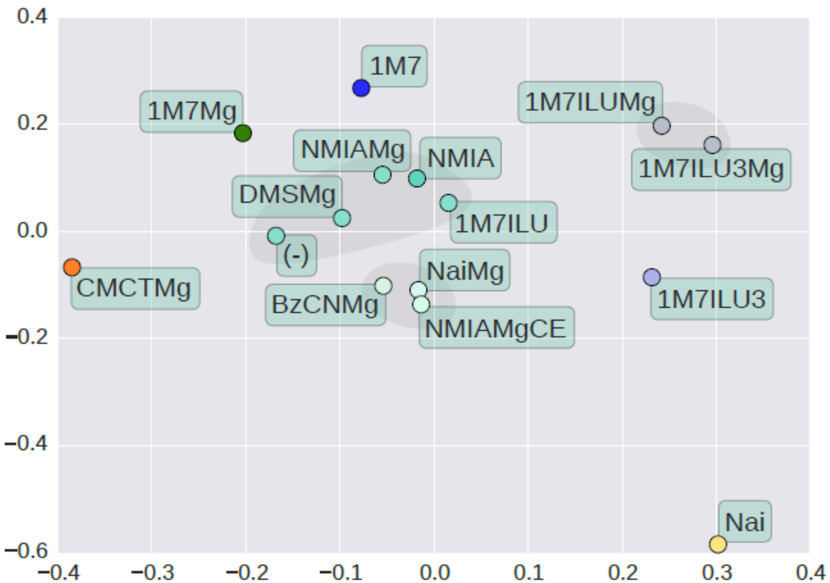
(Squared) Euclidian distance
between conditions

$$\text{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=x+1}^{n} \left(\mathbb{P}(i,j \mid d) - \mathbb{P}(i,j \mid d')\right)^2$$

- 14 conditions + no-probing (-)

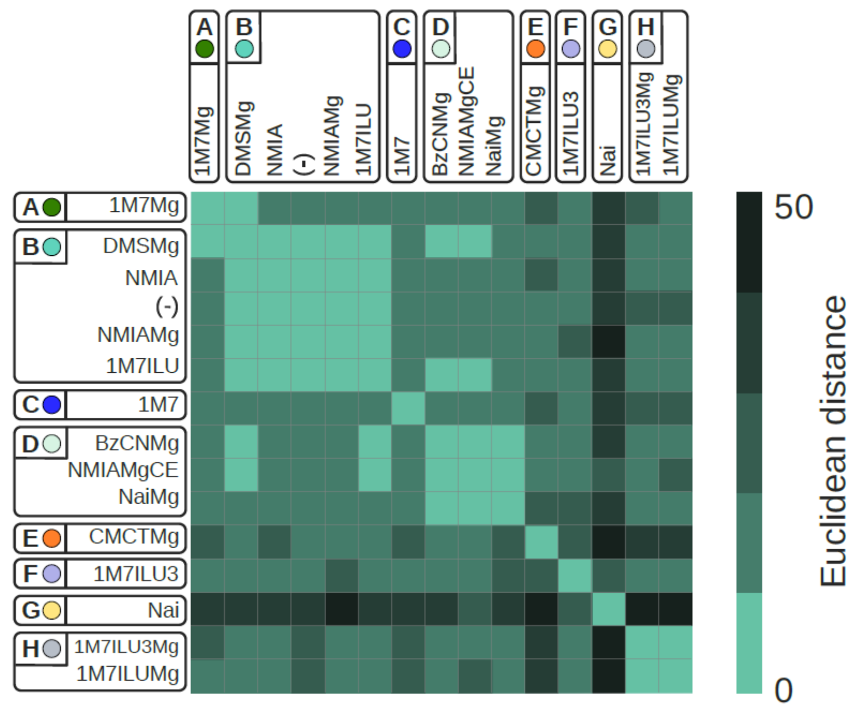


More mono probing analysis



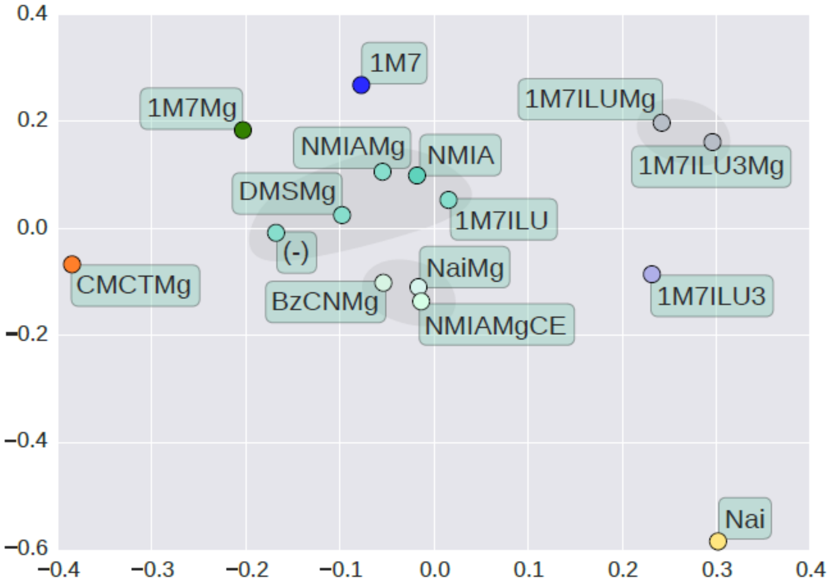
(Squared) Euclidian distance between conditions

$$\mathrm{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=x+1}^{n} \left(\mathbb{P}(i,j \mid d) - \mathbb{P}(i,j \mid d')\right)^2$$



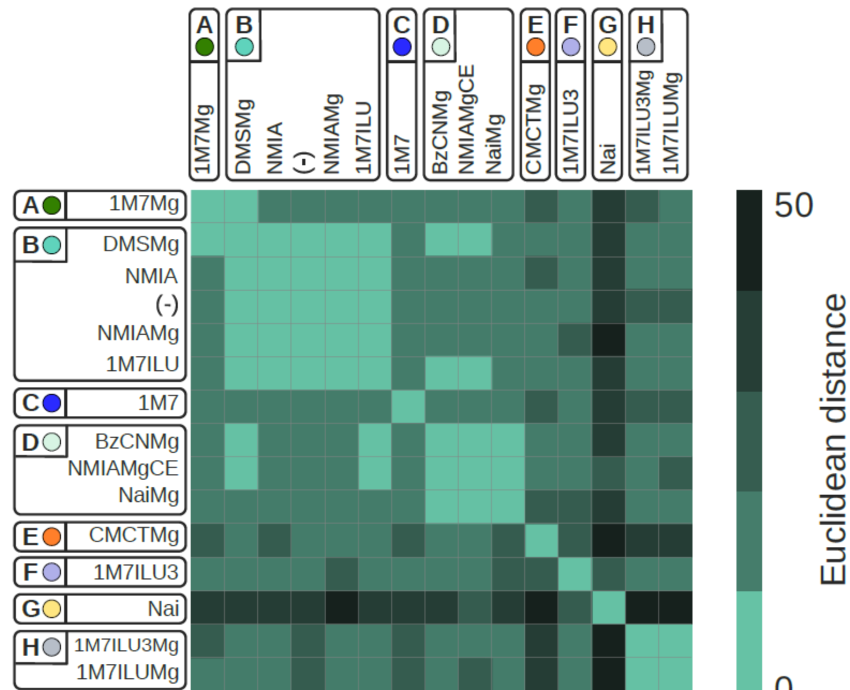
- 14 conditions + no-probing (-)
- Conditions cluster in 8 groups

More mono probing analysis



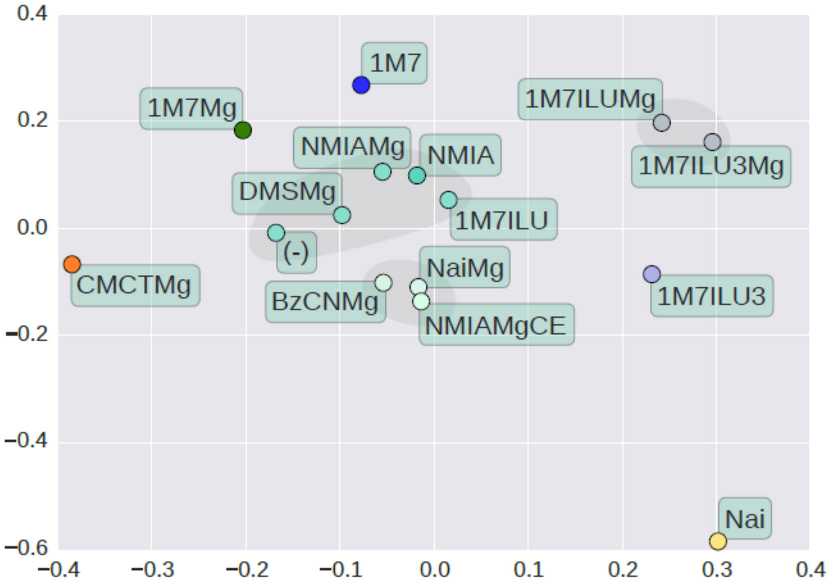
(Squared) Euclidian distance
between conditions

$$\mathrm{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=x+1}^{n} \left(\mathbb{P}(i,j \mid d) - \mathbb{P}(i,j \mid d')\right)^2$$



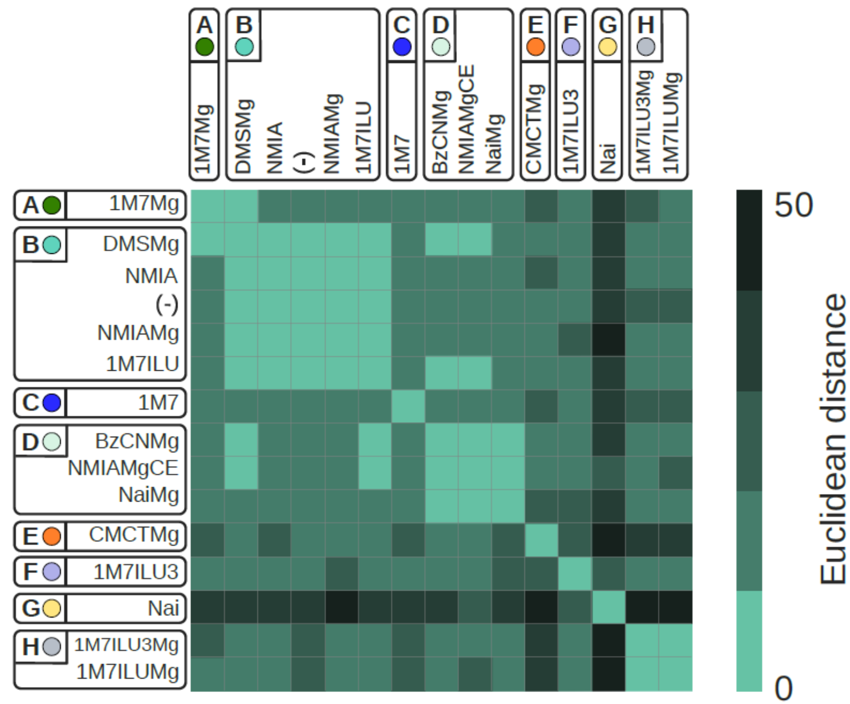
- 14 conditions + no-probing (-)
- Conditions cluster in 8 groups
- 5 singletons

More mono probing analysis



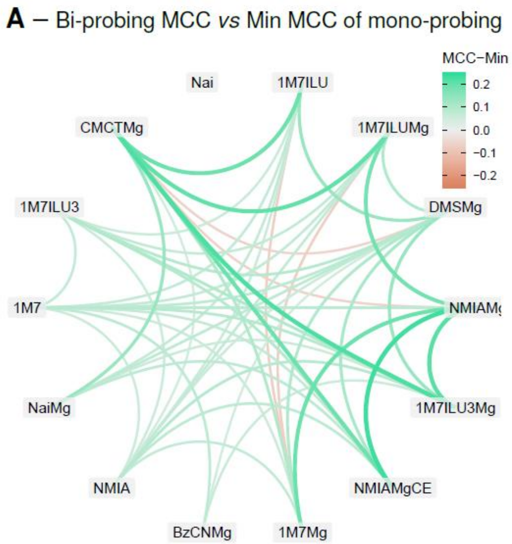
(Squared) Euclidian distance
between conditions

$$\mathrm{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=x+1}^{n} \left(\mathbb{P}(i,j \mid d) - \mathbb{P}(i,j \mid d')\right)^2$$



- 14 conditions + no-probing (-)
- Conditions cluster in 8 groups
- 5 singletons
- Outliers confirmed: NAI & CMCTMg

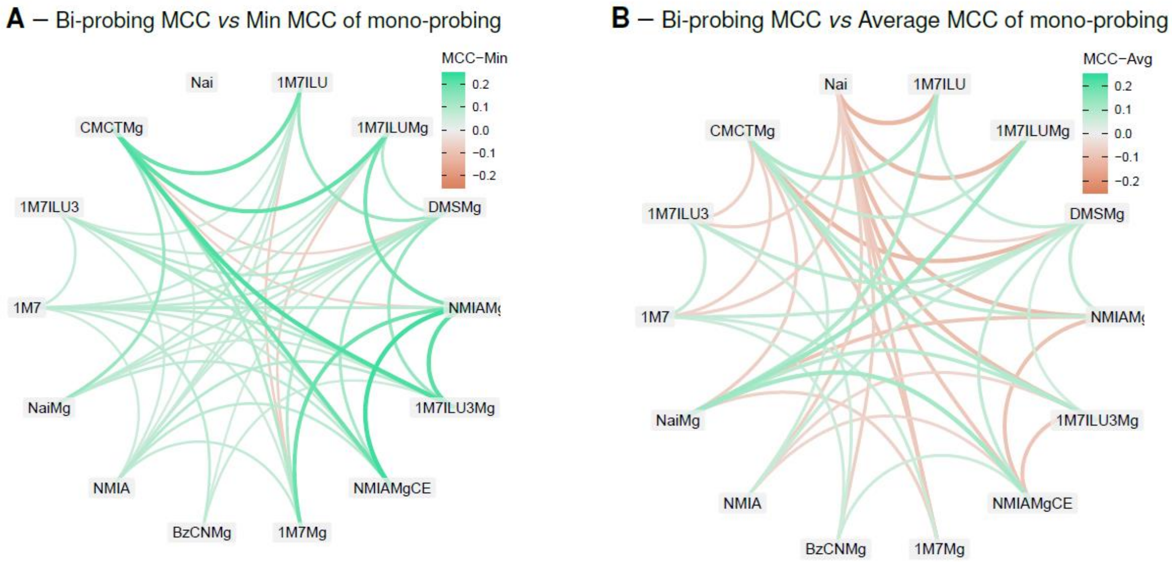
Dual probing analysis



Pairs of conditions do not improve predictions (+.2% MCC) but:

- Mitigates the risk of poor conditions:
  +5% MCC on average against worst condition

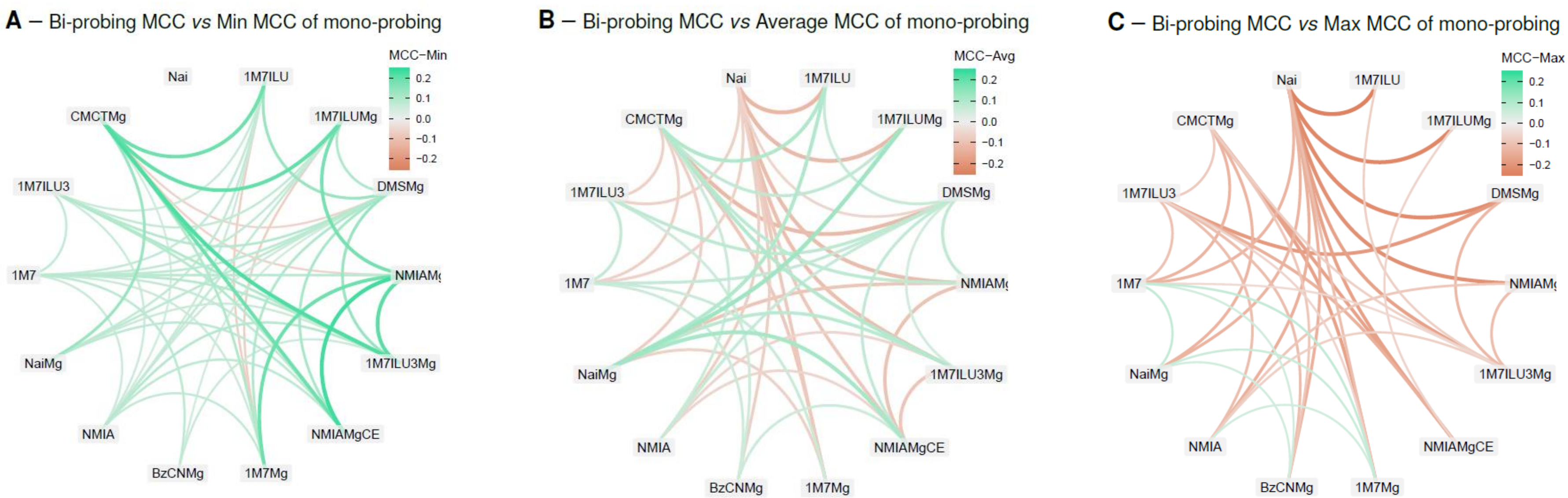
Dual probing analysis



Pairs of conditions do not improve predictions (+.2% MCC) but:

- Mitigates the risk of poor conditions:
  +5% MCC on average against worst condition
- Slightly increases expectation:
  +0.2% against mean MCC, +1% w/o NAI

Dual probing analysis



Pairs of conditions do not improve predictions (+.2% MCC) but:

- Mitigates the risk of poor conditions:
  +5% MCC on average against worst condition
- Slightly increases expectation:
  +0.2% against mean MCC, +1% w/o NAI
- Sensitive to contamination:
  -4.4% against best, -1.5% w/o NAI & CMCTMg

Triplets of conditions improve predictions

Considering triplets of conditions (364 runs) :

Triplets of conditions improve predictions

Considering triplets of conditions (364 runs) :

- Improves MCC by 10% over worst MCC

Triplets of conditions improve predictions

Considering triplets of conditions (364 runs) :

- Improves MCC by 10% over worst MCC
- Improves by 2.6% over mean MCC on avg (+3.4% median)

Triplets of conditions improve predictions

Considering triplets of conditions (364 runs) :

- Improves MCC by 10% over worst MCC
- Improves by 2.6% over mean MCC on avg (+3.4% median)
- Degrades by 4% over best MCC of triplet (2% median)

Triplets of conditions improve predictions

Considering triplets of conditions (364 runs) :

- Improves MCC by 10% over worst MCC
- Improves by 2.6% over mean MCC on avg (+3.4% median)
- Degrades by 4% over best MCC of triplet (2% median)
- Some triplets improve overall best MCCs
  - 1M7ILUMg + NMIAMgCE + 1M7ILU3
  - 1M7ILUMg + 1M7 + BzCNMg
  - 1M7ILUMg + 1M7ILU3 + 1M7

→ 85.3%MCC

Beyond three conditions?

The full monty: Multi probing

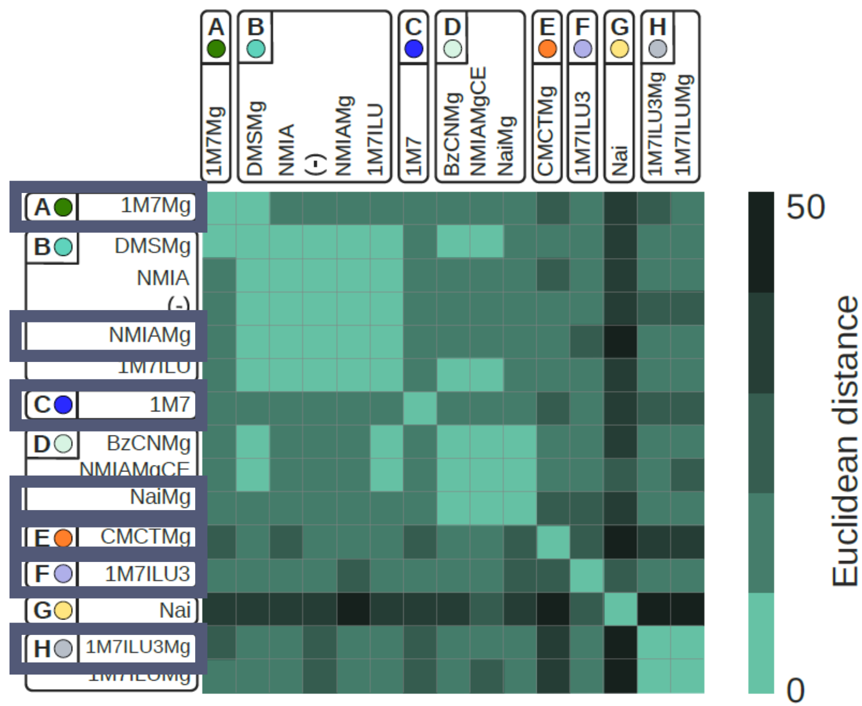
- All 14 conditions: 80% MCC (vs 73% Avg)

The full monty: Multi probing

- All 14 conditions: 80% MCC (vs 73% Avg)
- Within clustered conditions: Limited info/improvement

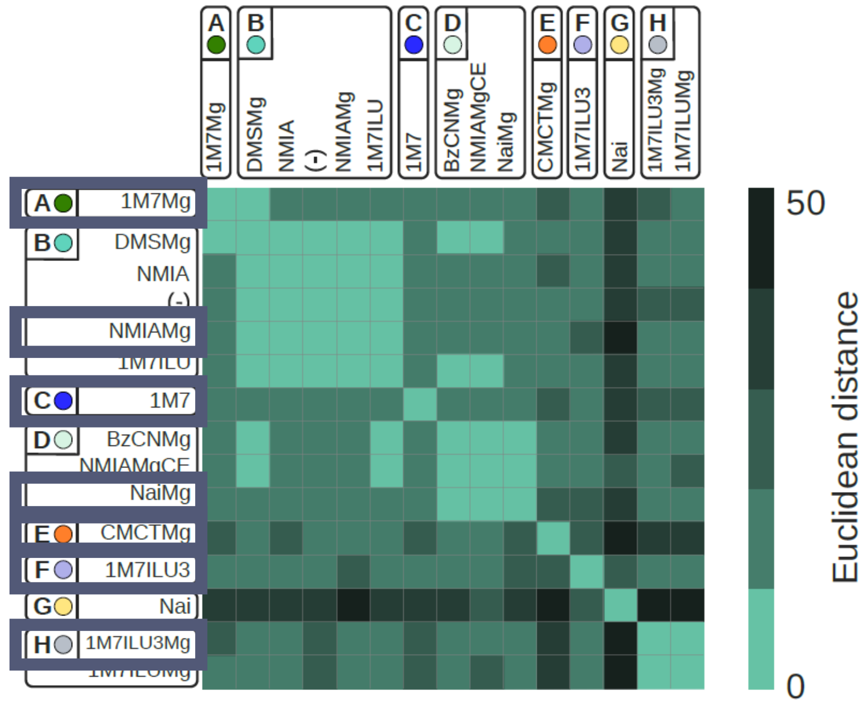
The full monty: Multi probing

- All 14 conditions: 80% MCC (vs 73% Avg)
- Within clustered conditions: Limited info/improvement
- Across clusters
  - One condition for each cluster
  - 79% MCC w/o NAI
  - 74% MCC with NAI
  - But best (86.5% MCC) has NAI



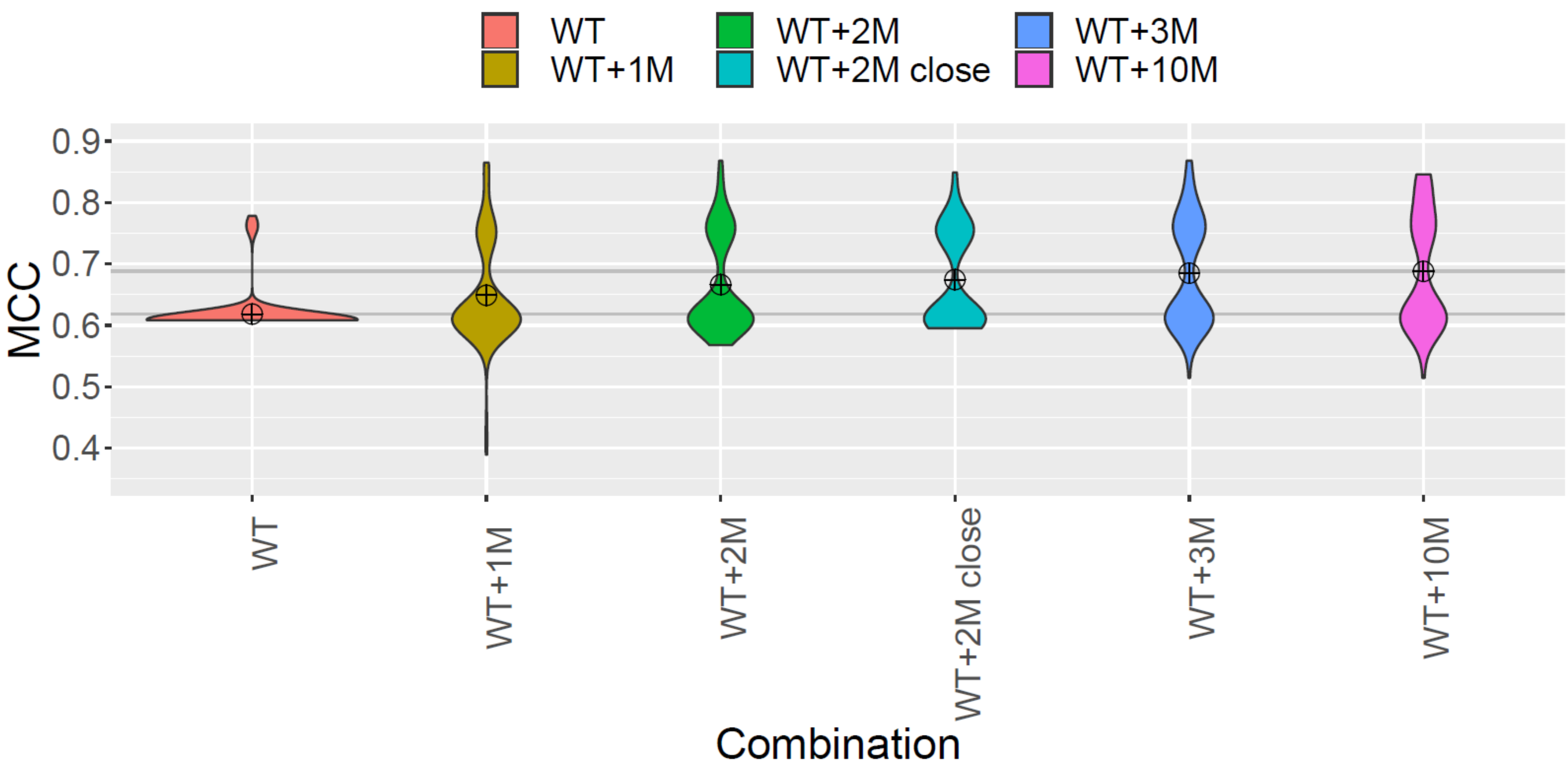
The full monty: Multi probing

- All 14 conditions: 80% MCC (vs 73% Avg)
- Within clustered conditions: Limited info/improvement
- Across clusters
  - One condition for each cluster
  - 79% MCC w/o NAI
  - 74% MCC with NAI
  - But best (86.5% MCC) has NAI
- Orthogonality?
- Complementarity?



Preliminary: Treating mutants as conditions

- Mutate-And-Map profiles produced by Das lab
- Systematic single-point mutants, usually similar structure



Conclusion

Conclusion

- IPANEMAP is way more than just a cool acronym

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

- Explore better pseudo-potentials

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

- Explore better pseudo-potentials
- Better clustering procedures (beyond majority rule)

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

- Explore better pseudo-potentials
- Better clustering procedures (beyond majority rule)
- Test on multi-stable RNAs

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

- Explore better pseudo-potentials
- Better clustering procedures (beyond majority rule)
- Test on multi-stable RNAs
- Optimal/iterative design of probing experiments

Conclusion

- IPANEMAP is way more than just a cool acronym
- Exploits the complementarity of probing experiments
- Paves the way for quantifying a notion of orthogonality

Take home message:

- Probing remains challenging (accuracy still slightly below 120%)
- Having >2 reactivity profiles (even bad ones) appears to help

The future needs:

- Explore better pseudo-potentials
- Better clustering procedures (beyond majority rule)
- Test on multi-stable RNAs
- Optimal/iterative design of probing experiments
- More data to build mechanical understanding of SHAPE

Thank you!

Afaf Saaidi
Postdoc?

Bruno Sargueil

Delphine Allouche

Ronny (*da main man*) for modular extensions to Vienna package

Soft constraints x Sampling = Awesomeness

Supported by Fondation pour la Recherche Médicale

FRM
FONDATION
RECHERCHE
MÉDICALE