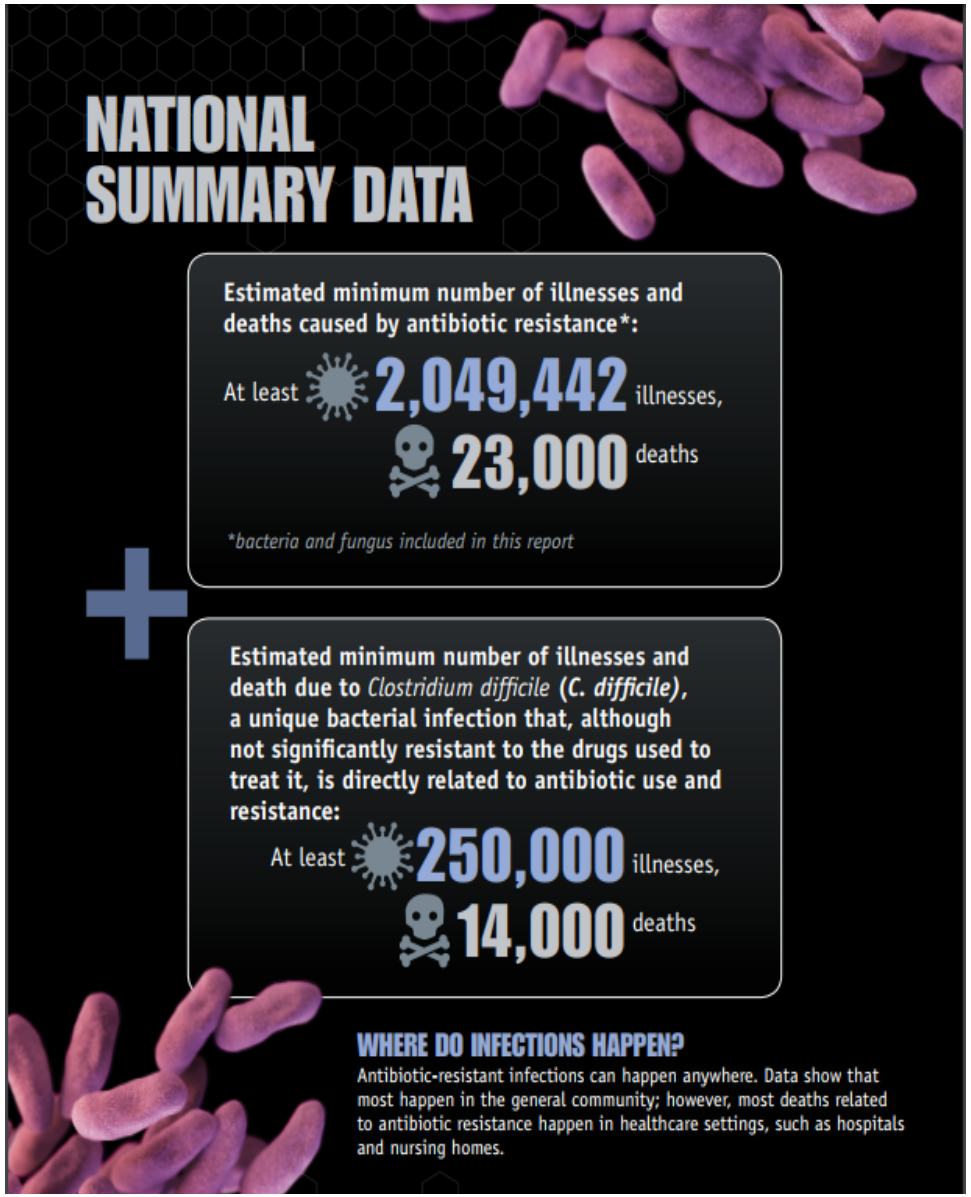
Design of RNA probe for rapid profiling of antibiotic resistant bacteria

Abdul Bhuiya, Antonio Ortega, Matthew Ning,

Mohammed Muzamil Khan

Antibiotic resistance

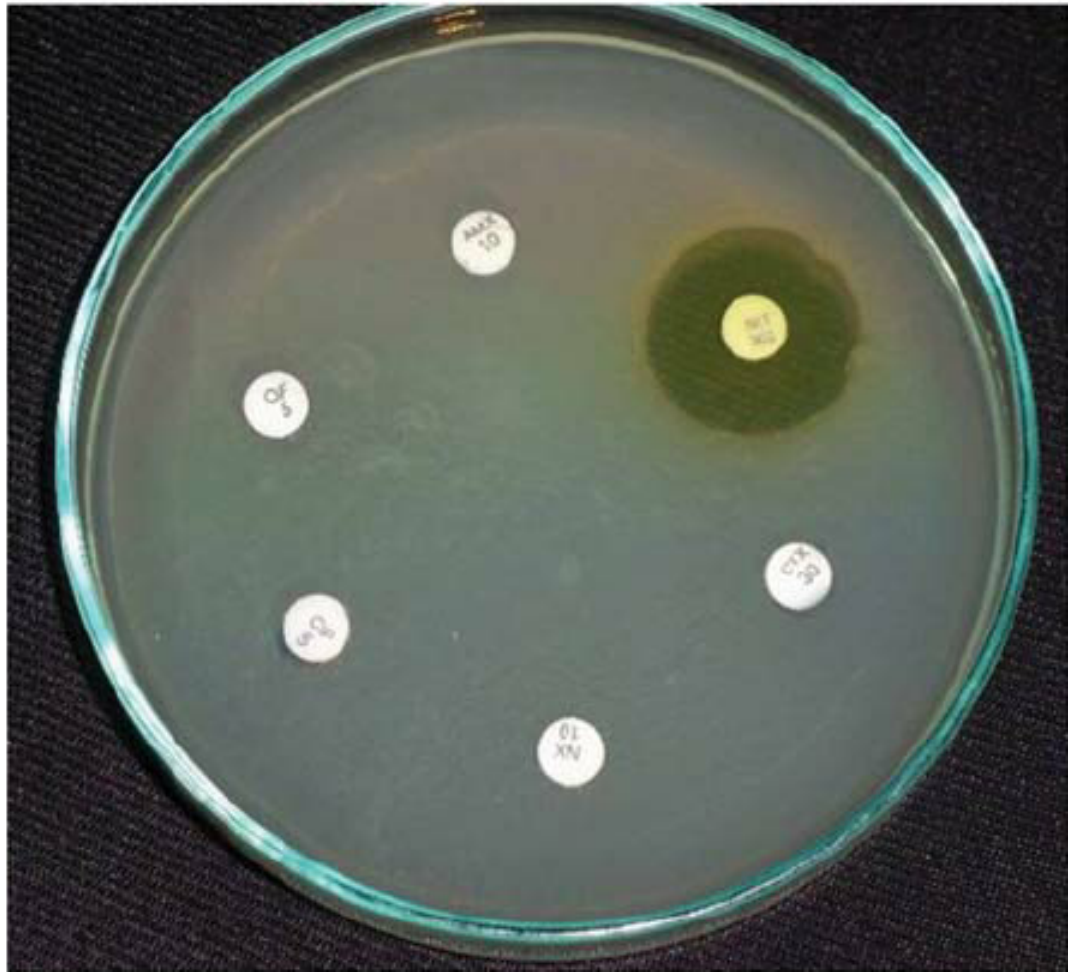
- More than two million cases a year in the US
- At least 23,000 deaths a year
- Annual impact- $20 billion in excess health care costs and 8 million additional hospital days
- Better assays are needed to take better and quicker clinical decisions



https://www.cdc.gov/drugresistance/about.html

Antibiotic susceptibility test

- Current gold standard is disk diffusion method (pictured) or dilution method
- Measures ability of antibiotic to inhibit bacterial growth
- Takes a couple of days to get results back
- Need faster diagnostic tests to prevent having to prescribe broad-spectrum antibiotics



https://www.researchgate.net/figure/236130281_fig2_Fig-2-Antibiotic-Susceptibility-Test-of-MDR-E-coli-in-Mueller-Hinton-Agar

RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities

Amy K. Barczak[a,b,c,d,1], James E. Gomez[a,c,1], Benjamin B. Kaufmann[a,b,c], Ella R. Hinson[a,b,c], Lisa Cosimi[a,e], Mark L. Borowsky[b], Andrew B. Onderdonk[f], Sarah A. Stanley[a,b,c], Devinder Kaur[g], Kevin F. Bryant[c], David M. Knipe[c], Alexander Sloutsky[g], and Deborah T. Hung[a,b,c,d,e,2]

[a]The Broad Institute, Cambridge, MA 02142; Departments of [b]Molecular Biology and [d]Medicine, Massachusetts General Hospital, Boston, MA 02114; [c]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; Departments of [e]Medicine and [f]Pathology, Brigham and Women's Hospital, Boston, MA 02115; and [g]Massachusetts Supranational Tuberculosis Reference Laboratory, University of Massachusetts Medical School, Jamaica Plain, MA 02130

Edited* by Eric S. Lander, The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, and approved March 1, 2012 (received for review November 28, 2011)

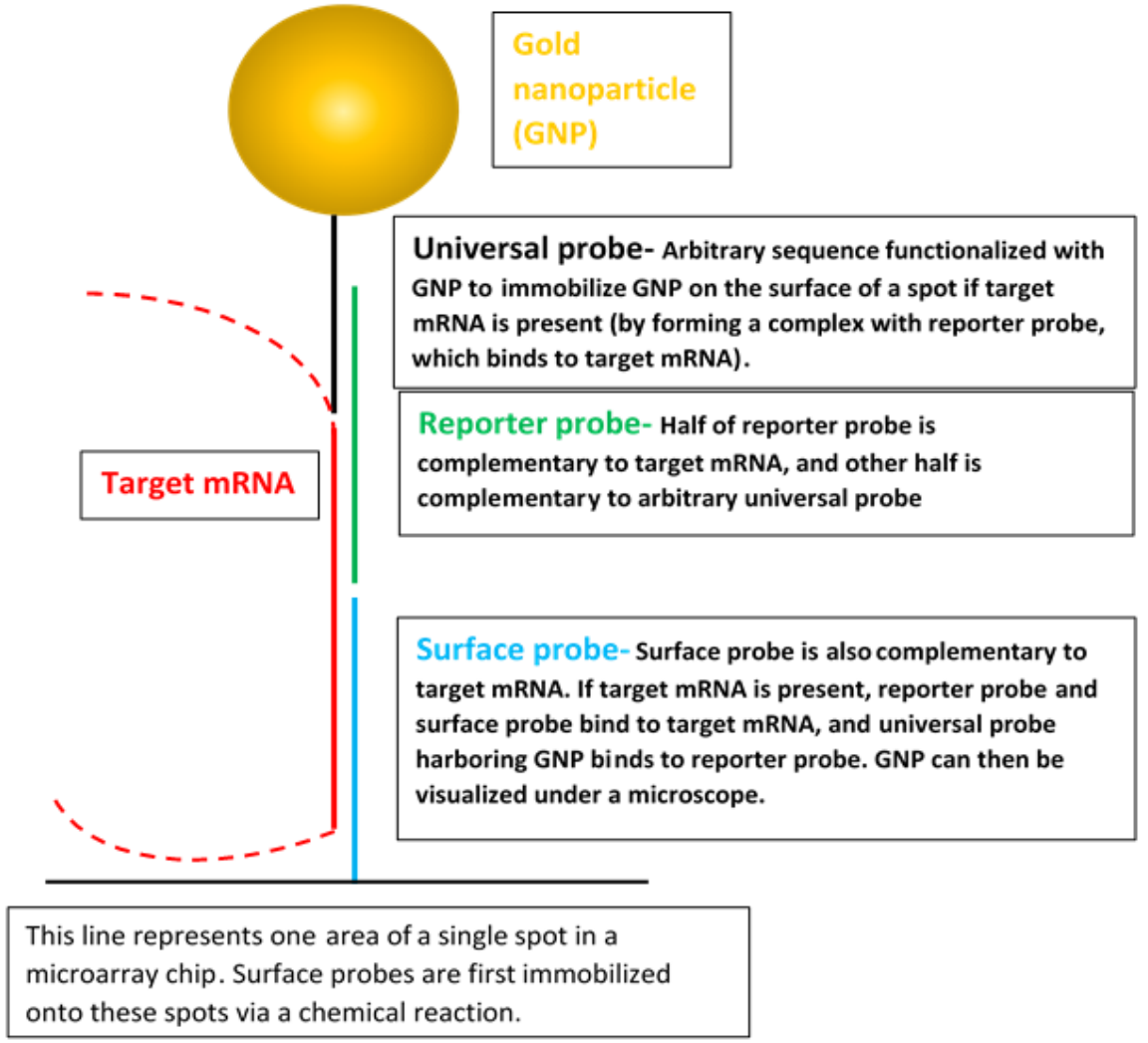
With rising rates of drug-resistant infections, there is a need for diagnostic methods that rapidly can detect the presence of pathogens and reveal their susceptibility to antibiotics. Here we propose an approach to diagnosing the presence and drug-susceptibility of infectious diseases based on direct detection of RNA from clinical samples. We demonstrate that species-specific RNA signatures can be used to identify a broad spectrum of infectious agents, including bacteria, viruses, yeast, and parasites. Moreover, we show that the behavior of a small set of bacterial transcripts after a brief antibiotic pulse can rapidly differentiate drug-susceptible and -resistant organisms and that these measurements can be made directly from clinical materials. Thus, transcriptional signa-

Here we explore an alternative approach: using RNA detection to obtain both genotypic and phenotypic information. Like DNA, RNA contains abundant genomic information to allow accurate identification of pathogens. However, unlike DNA, the RNA transcriptome also provides critical dynamic phenotypic information. Brief antibiotic exposure can trigger transcriptional responses in susceptible, but not in resistant, microbes within a few minutes (14), providing a means to couple pathogen identification directly with antibiotic-susceptibility testing.

In this work, we provide a proof-of-principle for RNA-based detection and drug-resistance testing of pathogens. Specifically, we demonstrate that detection of a set of pathogen RNA tran-

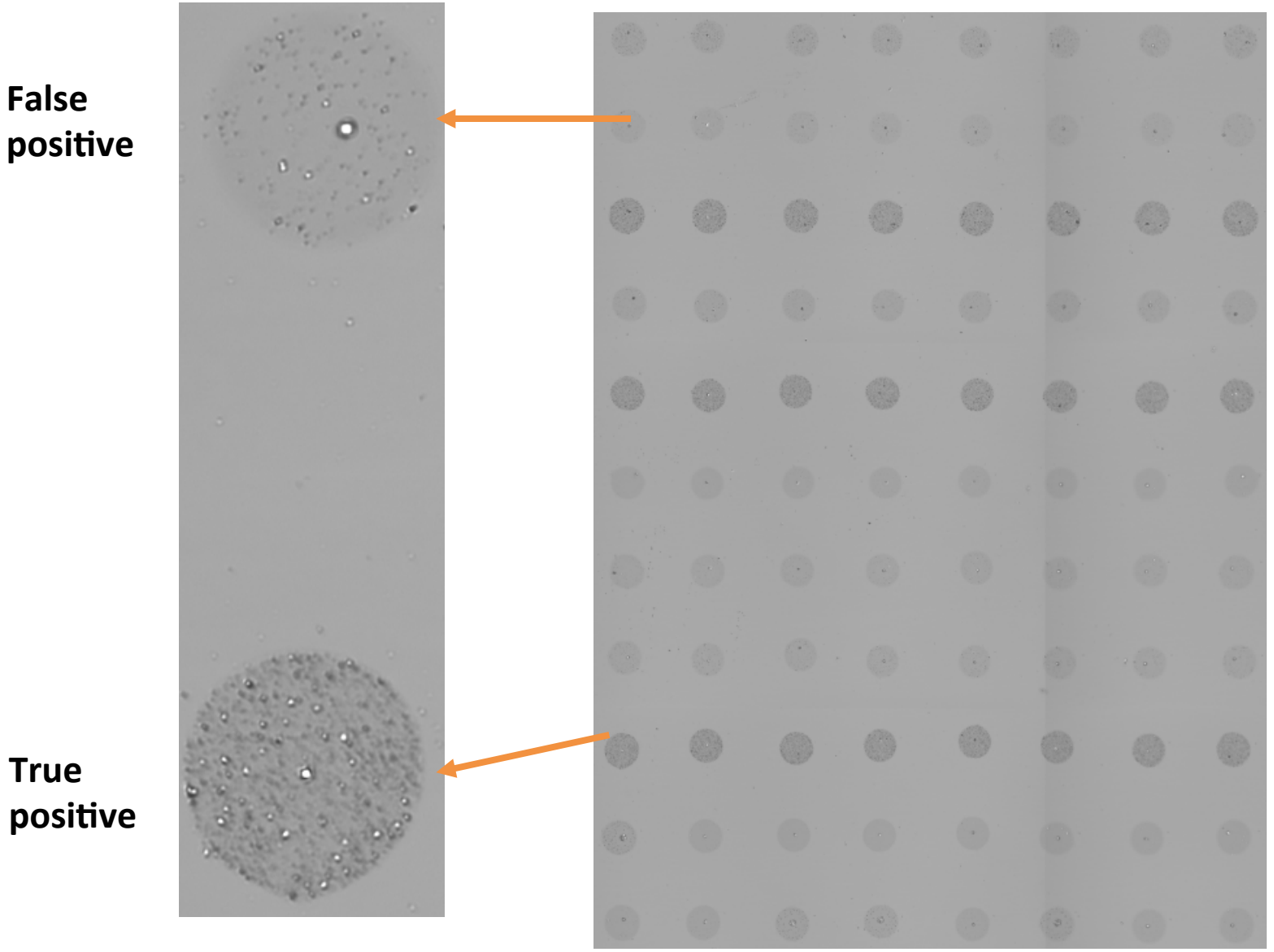
Single particle interferometric reflectance imaging sensor (SP-IRIS)

- Panel of signature mRNA transcripts are used to identify antibiotic-resistant bacteria
- If signature mRNAs are present, gold nanorod forms complex with surface probe
- Visualized under microscope to quantify concentration of target mRNA by looking at number of bound gold nanorods
- Get results within hours



Microscope images of bound gold nanorods in our sensor

- Image showing true and false positives
- Suspect universal probe binding to surface probe in the absence of target mRNA

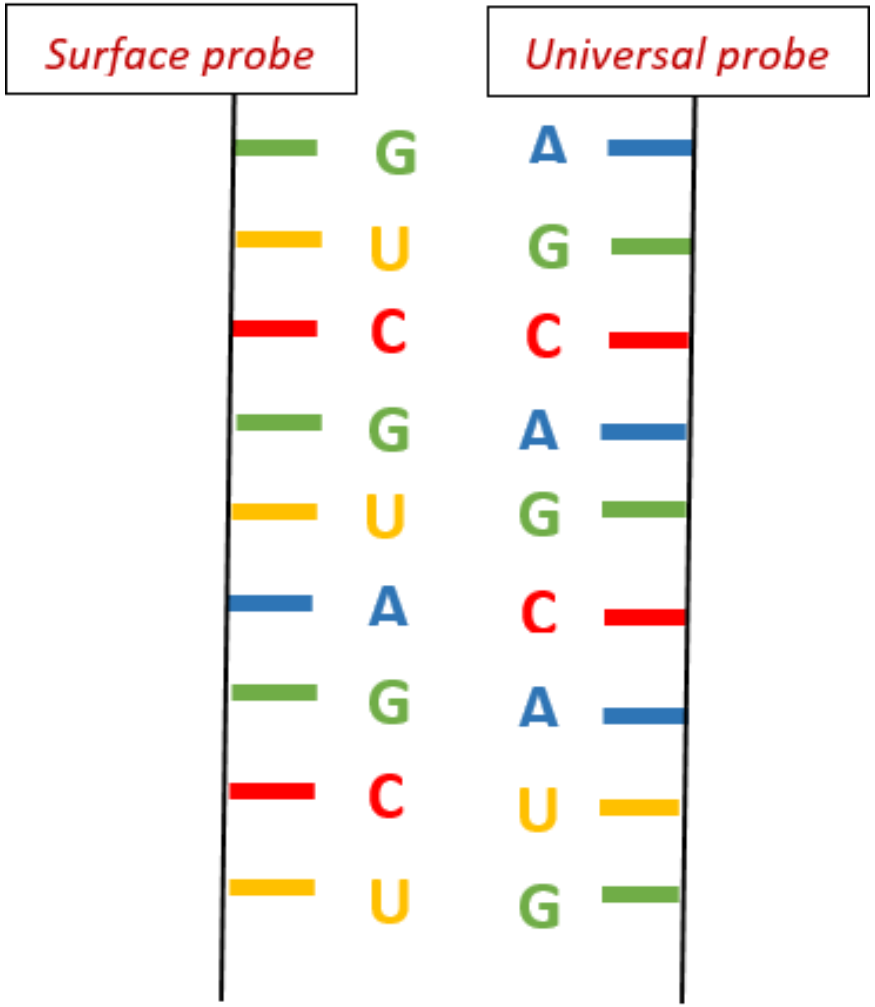


Objective

- To design an arbitrary 30 bases long RNA sequence for the Universal Probe (UP) that minimizes undesired binding to Surface Probe (SP)

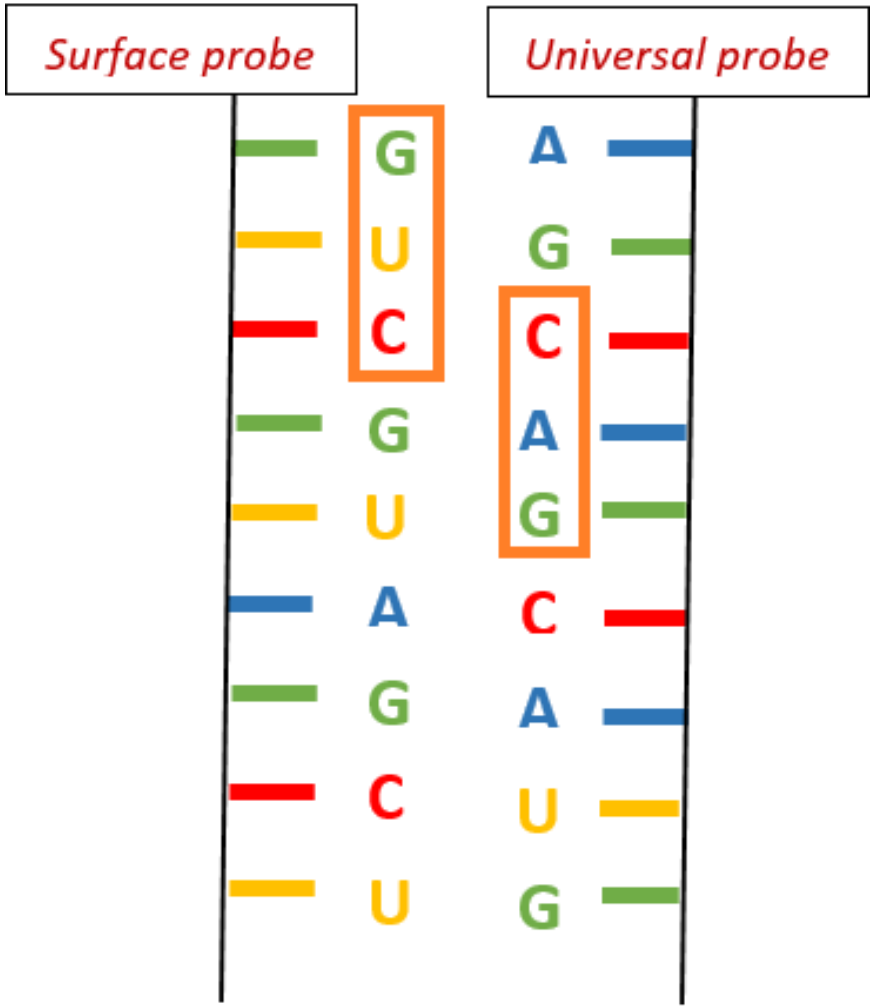
Partial matches

- Even if the sequences are different…



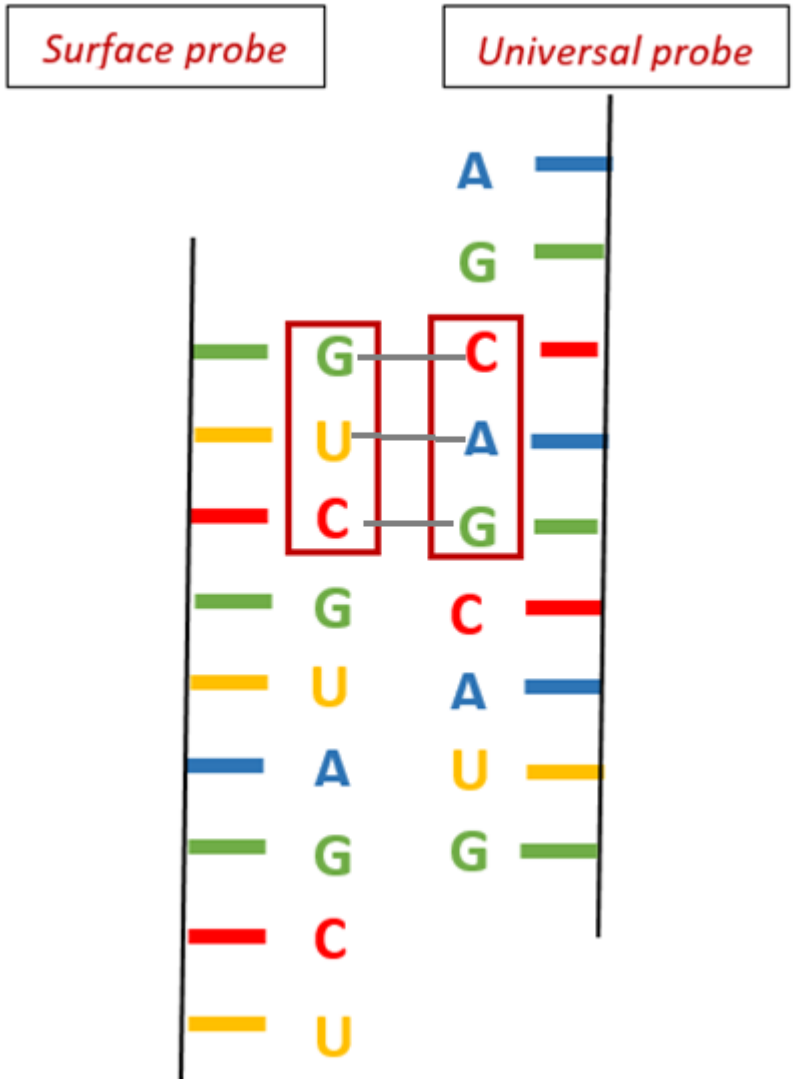
Partial matches

- Even if the sequences are different…
- Partial matches are possible

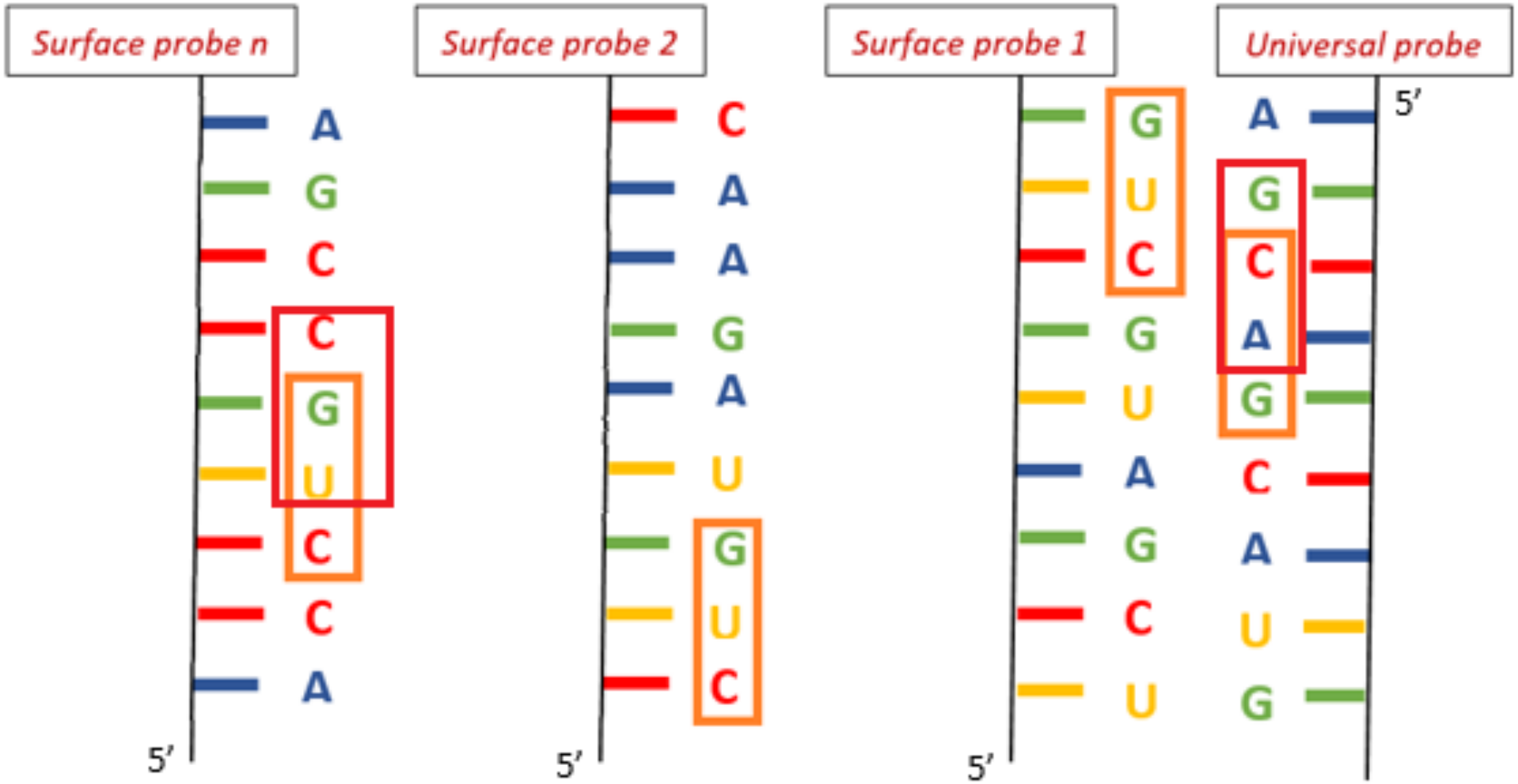


Model of match and mismatch

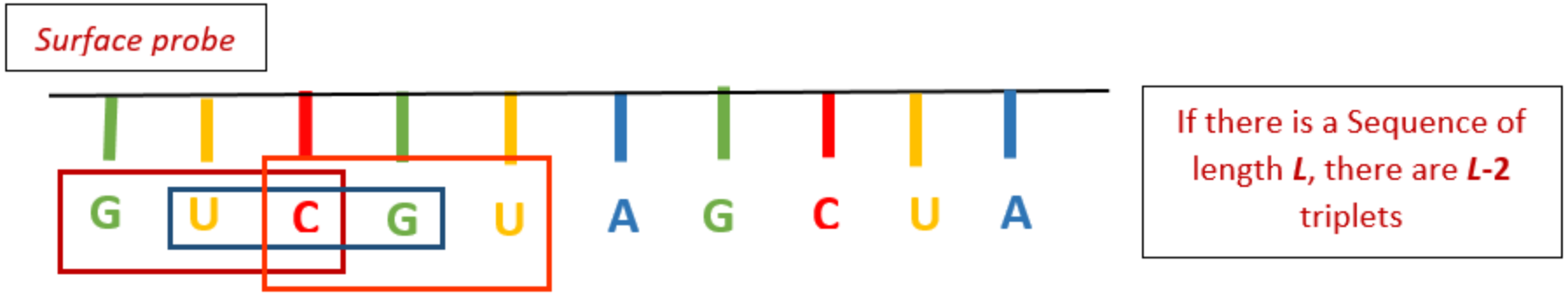
- Even if the sequences are different…
- Partial matches are possible
- Three nucleotide matches can be enough for binding



Even more with multiple surface probes!



Triplet counting



For N surface probes of length ***L***, there are N(***L*-2)** triplets.

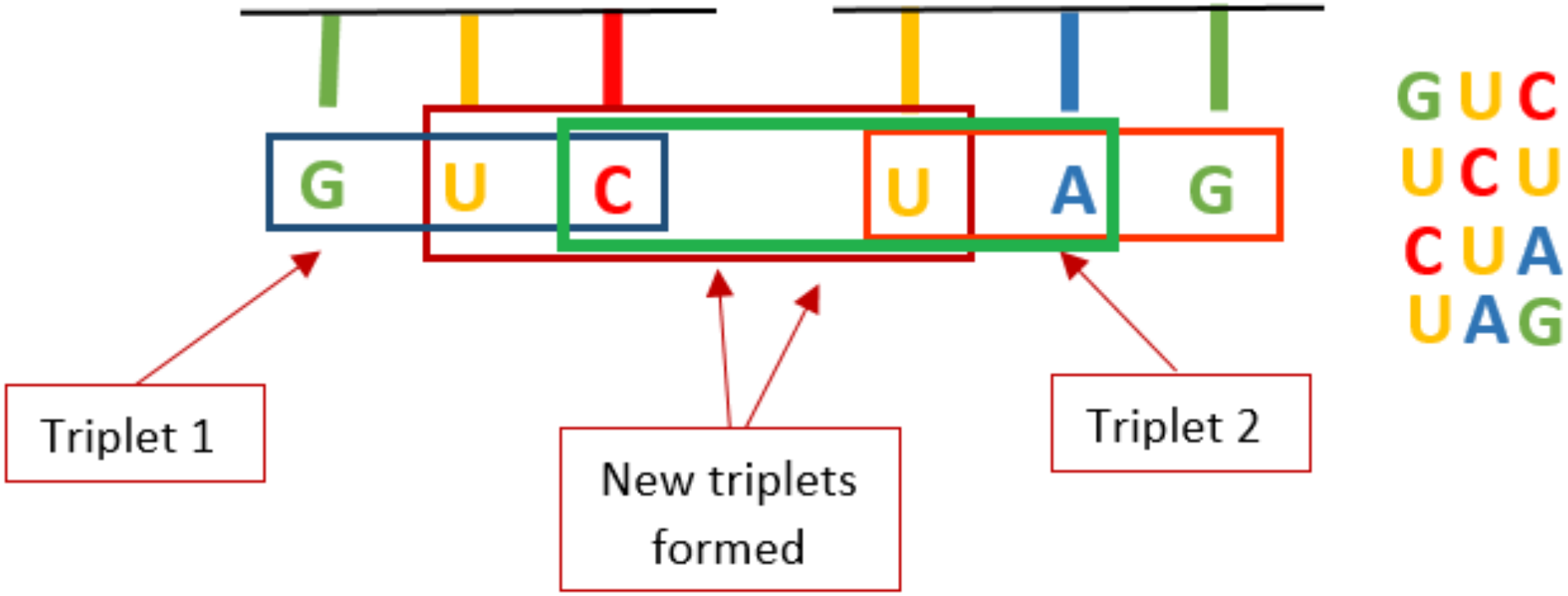
Frequency table

- We want to minimize the probability of three nucleotide matches
- We count how often they appear on the surface probes
- Build the UP from less common triplets!

Surface probe Triplets	Frequencies of SP triplet	Reverse complement
AGA	0.0357	UCU
GAC	0	GUC
CGU	0.0714	ACG
UAG	0.1071	CUA
CUA	.001	UAG
.	.	.
.	.	.

Building the universal probe sequence

- Even if GUC and UAG have low frequency, UCU or CUA could have high frequency, how to solve this?



Score of two triplets together

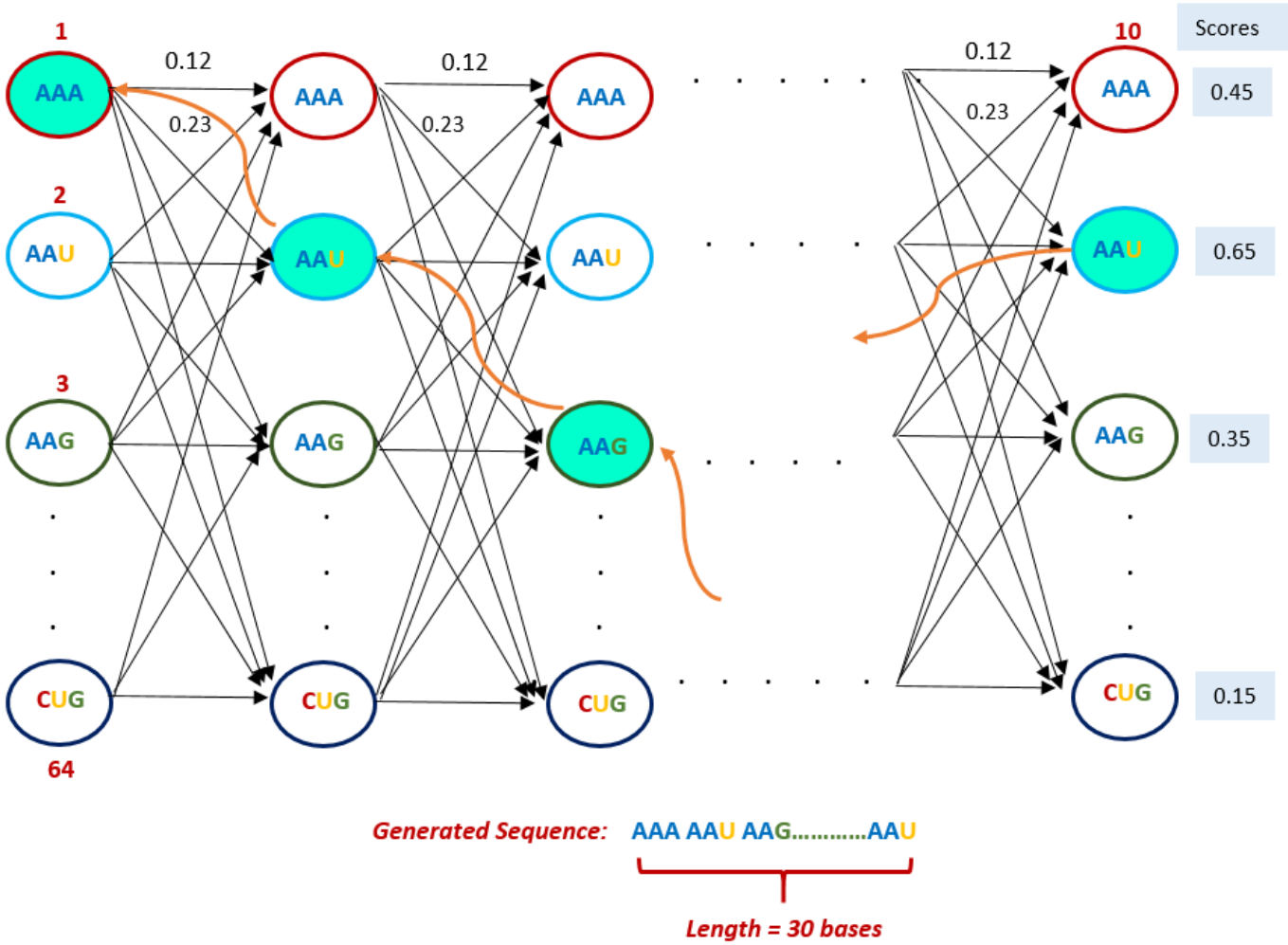
- If we form the sequence GUCUAG, we can calculate a score for this particular combination:

Triplet on UP	**Frequency of reverse complement on SP**	**Score (1-f)**
GUC	0	1
UCU	0.0357	0.9643
CUA	0.1071	0.8929
UAG	0.001	0.999
	Total score product:	0.86

- Choose triplets to maximize score of 30 nucleotide sequence!

Model as a trellis

- Transition cost fixed from specific triplet to specific triplet (transition matrix a_{jk})
- Find highest score by maximizing the best route at each step (dynamic programming)
- Final sequence calculated by trace-back at best final score



Thermodynamics

- Some triplets are more likely to bind than others, even if they have the same frequency.
- We define a thermodynamic score based on the Boltzmann weight from Gibbs free energy of hybridization of each specific three nucleotide sequence

$$1-e\uparrow-\Delta G\downarrow i\,/RT\ /\sum i\uparrow\blacksquare e\uparrow-\Delta G\downarrow i\,/RT$$

Example

Triplet	Rev Comp	ΔG	Frequency	Freq score	ΔG score	Score
'AAA'	'UUU'	14895.04	0.015714	0.984286	0.999983	0.984269
'AAC'	'GUU'	7531.2	0.012857	0.987143	0.999709	0.986856
'AAU'	'AUU'	14183.76	0.017143	0.982857	0.999978	0.982836
'AAG'	'CUU'	8200.64	0.022857	0.977143	0.999776	0.976924
'ACA'	'UGU'	4476.88	0.011429	0.988571	0.999048	0.98763
'ACC'	'GGU'	-2217.52	0.018571	0.981429	0.987206	0.968872
'ACU'	'AGU'	4602.4	0.012857	0.987143	0.999093	0.986248
'ACG'	'CGU'	1548.08	0.017143	0.982857	0.997033	0.979941
'AUA'	'UAU'	12510.16	0.01	0.99	0.999958	0.989958
…	…	…	…	…	…	…

Algorithm overview

- Input: RNA sequences of surface probes
- Count triplets on input
- Calculate reverse complements
- Calculate transition score matrix
- Maximize with recursion: $V\downarrow k\ (t)=\max\nolimits_{\top j}\ a\downarrow jk\ V\downarrow j\ (t-1)$
- Traceback

Results

- Applied the algorithm to random sequences and to previously used surface probes
- Score of known universal probes fits their performance (bad probe score: 0.13, good probe score: 0.53)

'AUGUUGGUUUGCAGCGCGCCUACUCAAUUG'
'UGCAAGUCUGUGAUAGUCGACCGCAUUUUC'
'GUAGACAAUAUGUAACAGGUCUAUCAGCCA'
'GGACAAGGGAAUUAAUUCCAAUUUACGAAG'
'UAGAAAGUGCUAGUACGUUACCACGUGGUC'
'UCGGGGCGGGACAUUAUUGUGGGACGGGAG'
'GCUAACCUUUACAUCUUGAUGAGGUUCUGC'
'CUAUGCGAGGCCCGUGGCAACGUCGCCUUU'
'CACGUUCGAAUACUGCGUCCUGCUUGGGUA'
'UACUAGACCUGACAGUCUACCCCGUGACUU'
'CCAUCCGUAUUGAAGAGUAAGUGCAUGGUC'
'UAUCUUCAUAUCGAGACCGAUCGGCGGUGC'
'AACUAACGUUUUCAGAACCCAACGGCCCAG'
….

CUAUAUCUAUAUCUAUAUCUACUAUAUAUC

Conclusions

- We implemented an algorithm to generate arbitrary sequences based on minimizing hybridization (by triplets or more) to input sequences
- Scoring scheme fits observed data qualitatively

Future work

- Explore secondary structure of Universal Probe RNA
- Validation on wet lab
- Integrating the algorithm to design pipeline for the rest of the probes in the assay

Thank You

Questions?