CEITEC

Central European Institute of Technology

BRNO | CZECH REPUBLIC

MUNI

Using attribution sequence alignment to interpret Deep Learning models for miRNA binding site prediction

Katarína Grešová

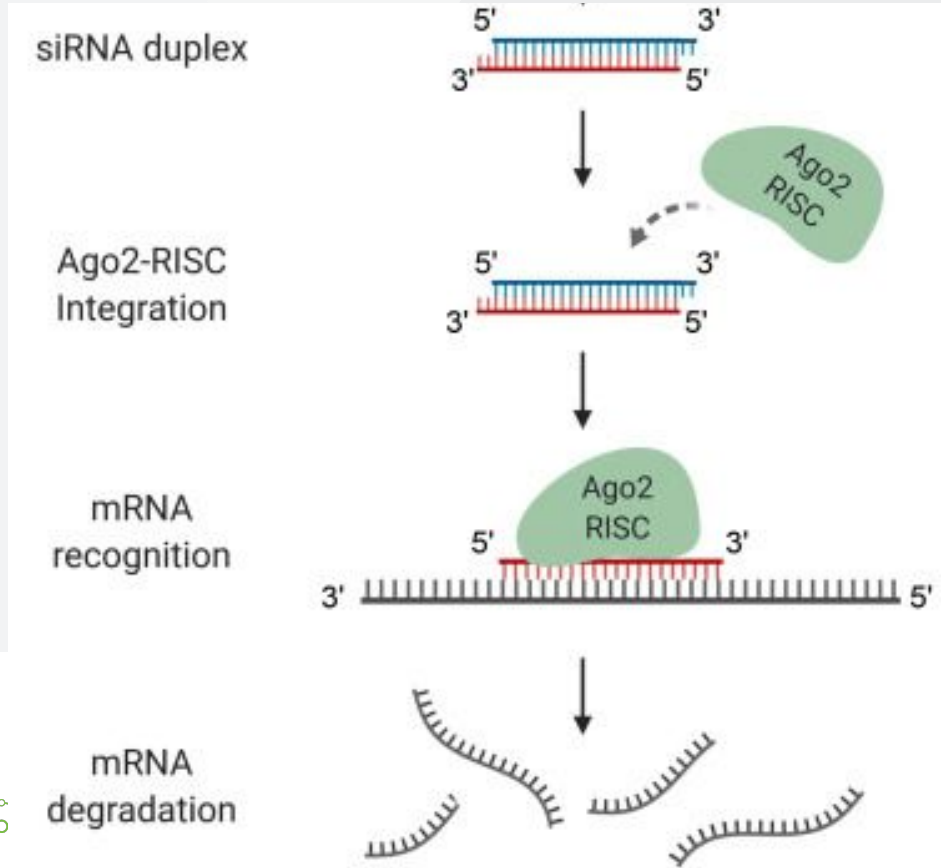
Panagiotis Alexiou Research Group, CEITEC-MU

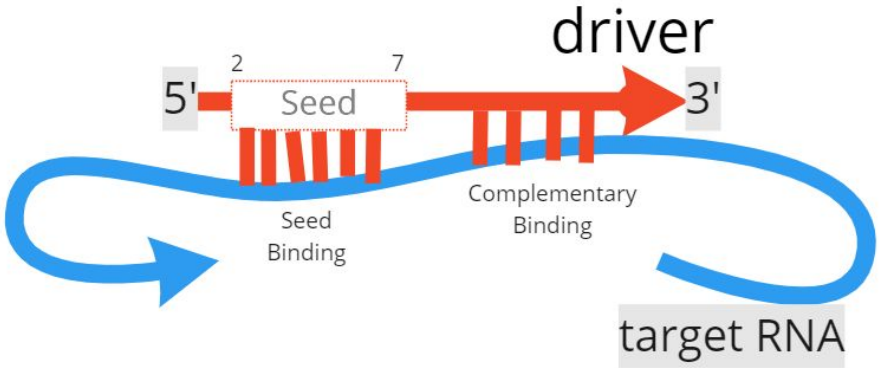
Faculty of Science, NCBR, Masaryk University

miRNA targeting

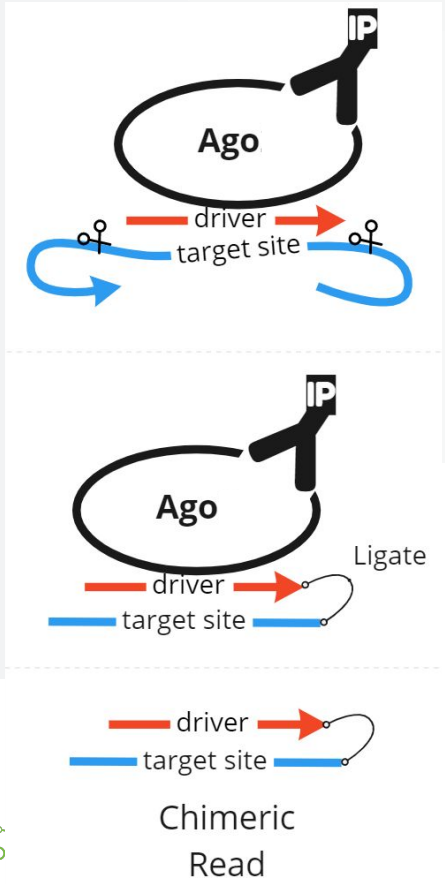
A–T

G–C

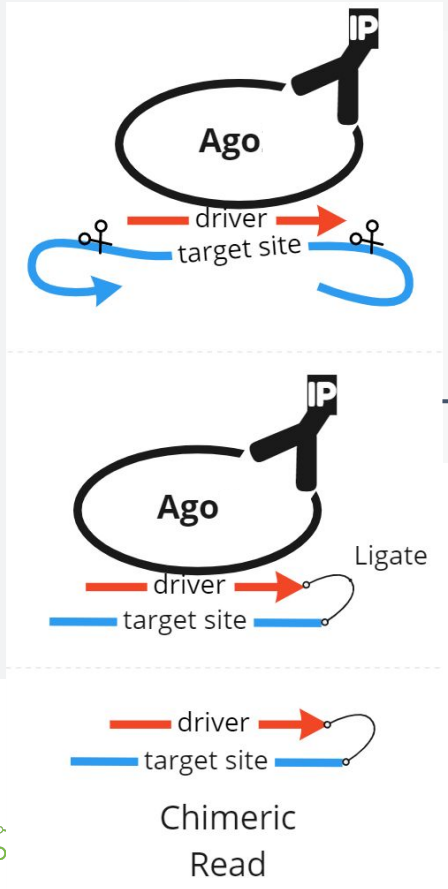




Biological experiment - CLASH



Biological experiment - CLASH



miRNA	gene	label
AACTGGCCCTCAAAGTCCCG	TGGAGAGCGGGCTTAAGAAGTGGCGGTTCGGCCGGAGGTTCCATCGTATC	1
ATCAGGGCTTGTGGAATGGG	CTCGCTGGCGTTCTCCGGGGTGGTTGGCATTGTGTCCTGGAAGCGGCCAT	0
TGGGGAGCTGAGGCTCTGGG	CTACACCTCAGCCCGGGGCTGCACTGCCACCCTGGGCAACTTCGCCAAGG	0
GTGAGGGCATGCAGGCCTGG	GTAAGGAGCTGGAGTCGCTGGTAGAGAACGAGGGCAGTGAGGTGCTGGCG	0
ATGCACCTGGGCAAGGATTC	GCATATGGGGGCCTTAAGGAATAACAGTGTGCGTGGTGGTGTGCAGGAGA	0
TGCACGGCACTGGGGACACG	TCAGGGTTTCTTGGGGGCTTATGAGTCTCACCGGTCAACCCAGGAGGCCT	0
AACTGGCCCTCAAAGTCCCG	ACCTCTTAATGGGCCAGTGAATAACACTCACTGCTGGCATTTAATGTGCA	1
TGGGTTCCTGGCATGCTGAT	CACCTGCTGCCCCTTCTACCCCAGCTCCACCACCTGCAGTCCCTAAAGAA	0
TCAGTGCATCACAGAACTTT	ACCCGCACAGCAAGCACCTGTACACGGCCGACATGTTCACGCACGGGATC	0
CTGGCCCTCTCTGCCCTTCC	CTGATTGTGGCAGAGGGGCCACTACCCAAGGTCTAGCTAGGCCCAAGACC	1
TGAGGTAGTAGGTTGTATAG	ATGACCCAACCTACCACCCTGTTTTTACATATCCAATTCCAGTAACTCTC	1
TAAAGTGCTTATAGTGCAGG	CAAAAGCATACCTACCTTCCCCTAGAGGTCTGTAACATTGTGGCTGGGCA	1
TGAGAACTGAATTCCATGGG	CCTGGGACCCCCAGGCGTGGAGGACAGTCAAGCCGTGGAGGCCGTGGAGG	0
TGAGGTAGTAGGTTGTATAG	CCCAACCTCAACCTCAACCTCCCAGCACCACACATCATGCCAGGGGTTGG	1
CTGTACAGGCCACTGCCTTG	GAAGGTAAAGAGGGTCATTGGGGTCGAGCTATGCCCAGAGGCTGTGGAGG	0
GTCCCTCTCCAAATGTGTCT	GCTGGCCAGCGGACTTCTGGAGTTAGCCTTTGCTTTTGGAGGACTGTGTG	0
TTAGGGCCCTGGCTCCATCT	ACACAGGAAGAGGAGCCAGGCCCTTGTACCTATGGGATTGGACAGGACTG	1
TAGGTAGTTTCATGTTGTTG	TCCGCCCTCTTTTGCCAGCCCAGCCCCTCCATGCACATTTGGACGCTGTC	0
TAAAGAGCCCTGTGGAGACA	TCCTGAGGCCTGGGGCACCTTTCGTCTGATGAGCCTCTGCATGGAGAGAG	0
GTGGGTACGGCCCAGTGGGG	CATCTTGTCCTCACAGCCCAGAGCATGTTCCAGATCCCAGAGTTTGAGCC	0

Helwak et al., 2013 CLASH dataset - 30 785 miRNA:target site pairs

miRBind model

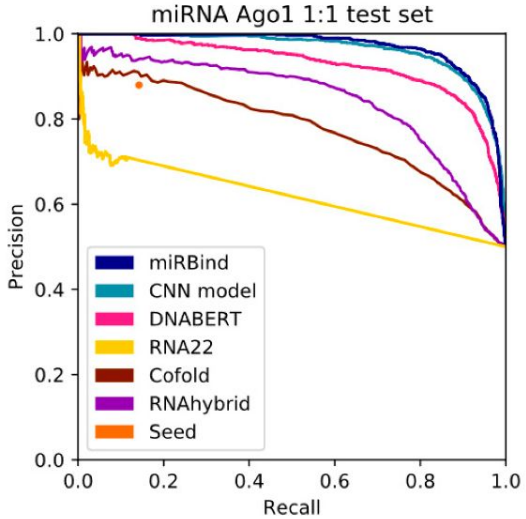
genes

MDPI

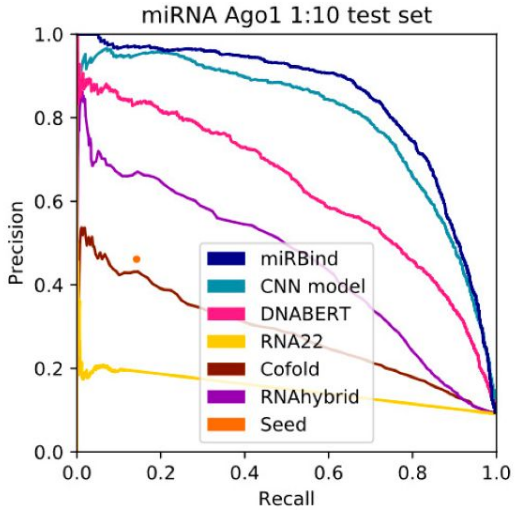
Article

miRBind: A Deep Learning Method for miRNA Binding Classification

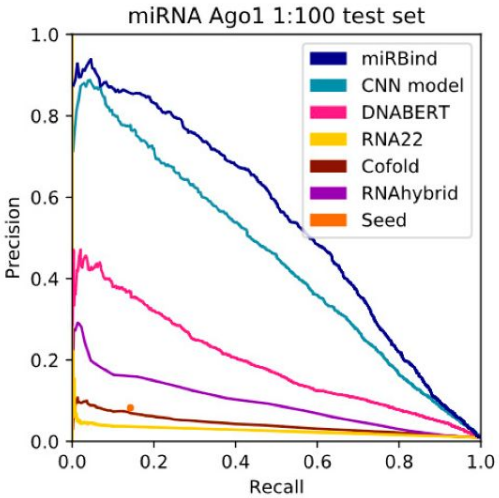
Eva Klimentová [1,†], Václav Hejret [1,2,†], Ján Krčmář [3], Katarína Grešová [1,2], Ilektra-Chara Giassa [1,*] and Panagiotis Alexiou [1]



(a)

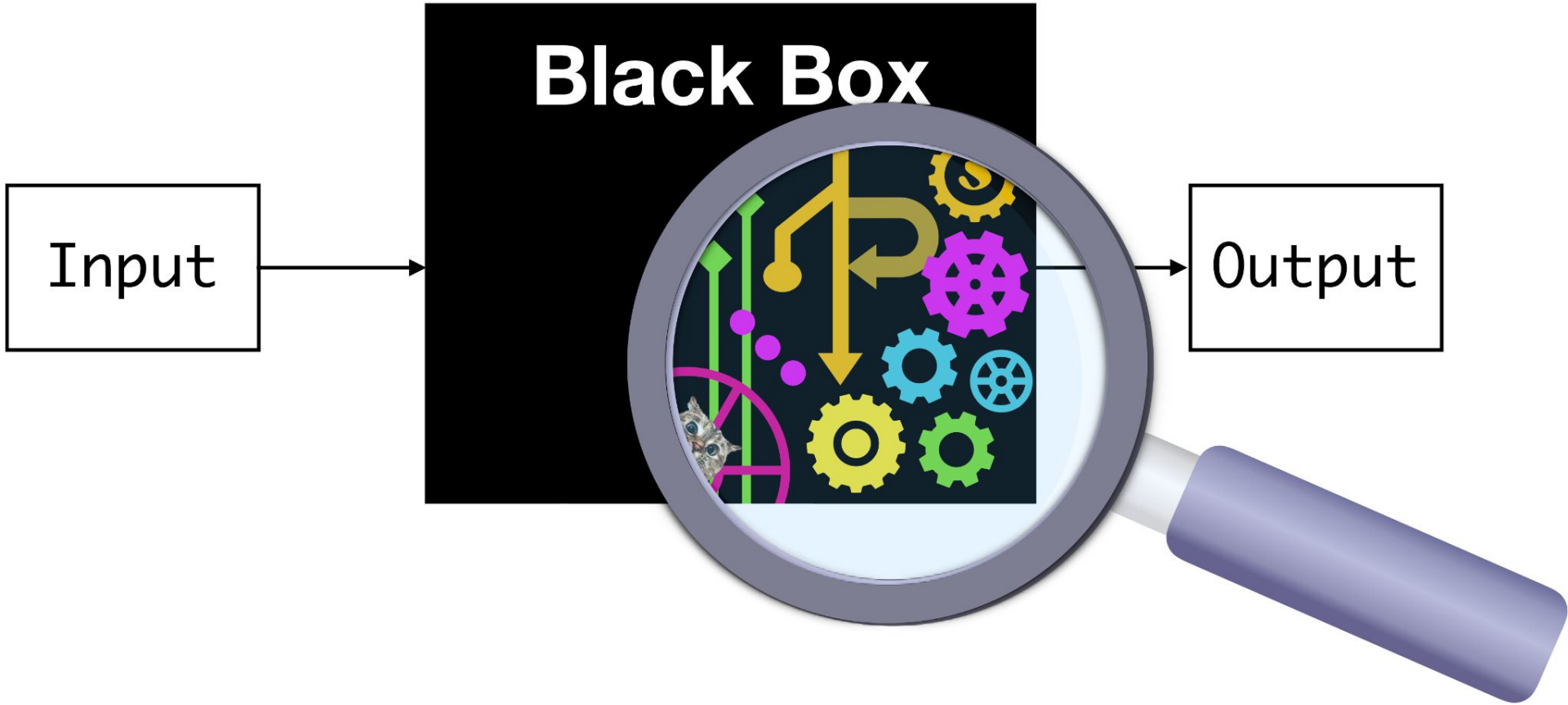


(b)

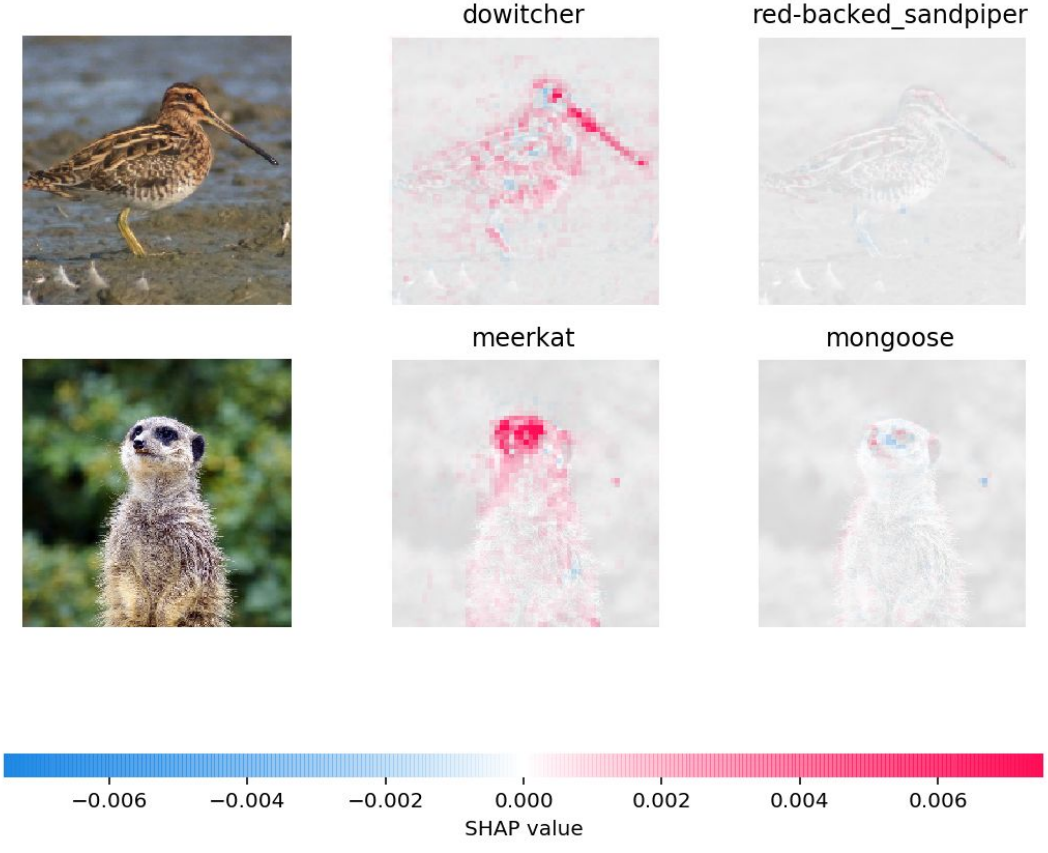


(c)

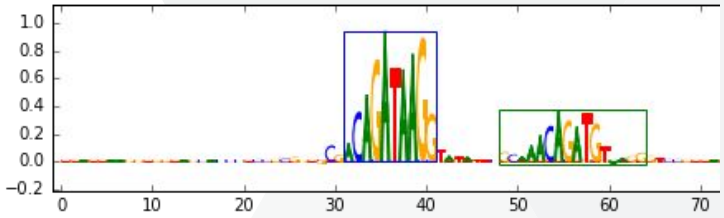
Precision-recall curves


Black Box
Input
Output

Interpreting Neural Networks

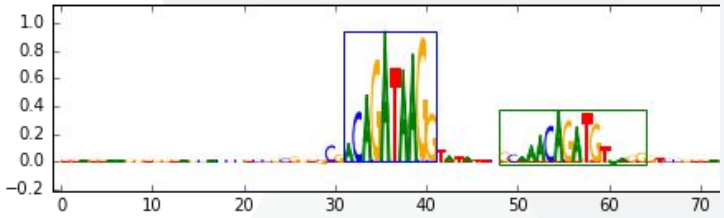


Interpreting Neural Networks

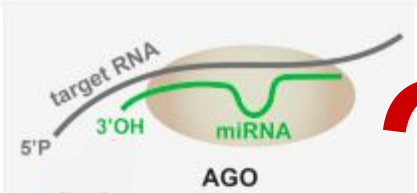


GATA motif TAL1 motif

Interpreting Neural Networks



GATA motif TAL1 motif



?

How to interpret interaction between sequences

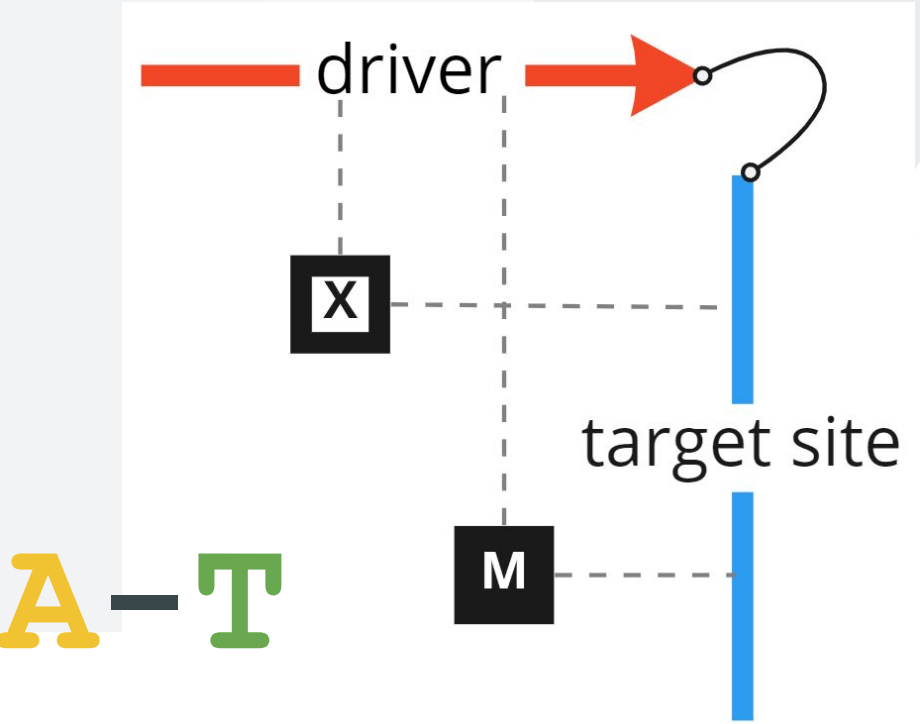
TACGTCAGTTCATGAAGCTA

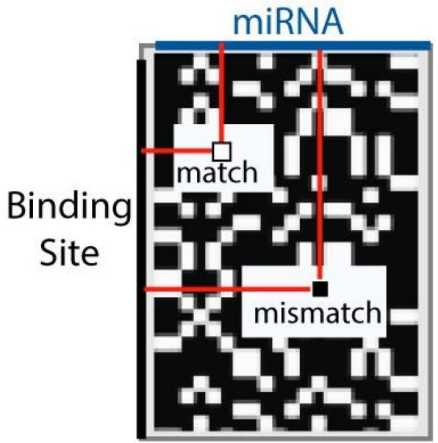
(driver ~20nt)

✖

AGTTCTAGTTCGTCCGTCAGTGTCAGTTCATGAGCACCAGTCACGTTCGTCTA

(target ~50nt)

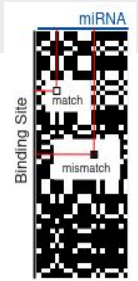


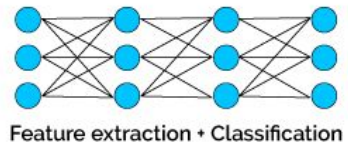


miRBind model - interpretation

miRNA:
TGAGGTAGTAGGTTGTATAG

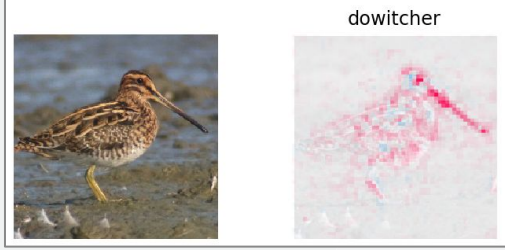
Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT





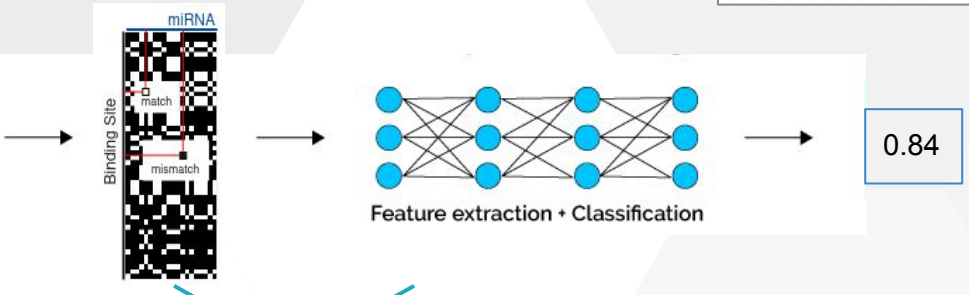
0.84

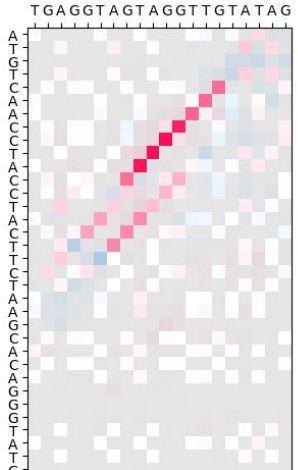
miRBind model - interpretation



miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT





Visualization

miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

Predicted alignment:

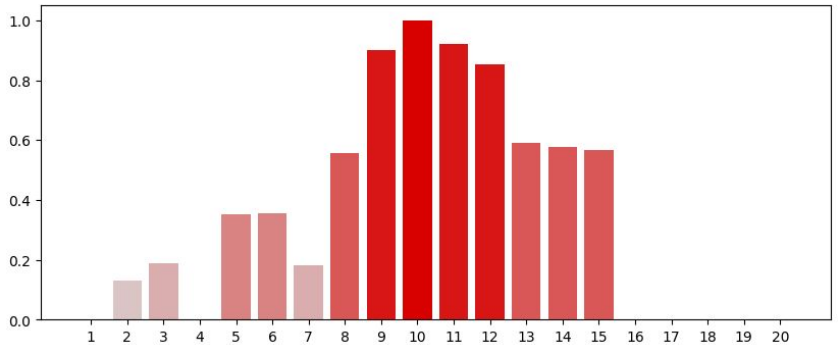
Visualization

miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

Predicted alignment:

miRNA position importance:



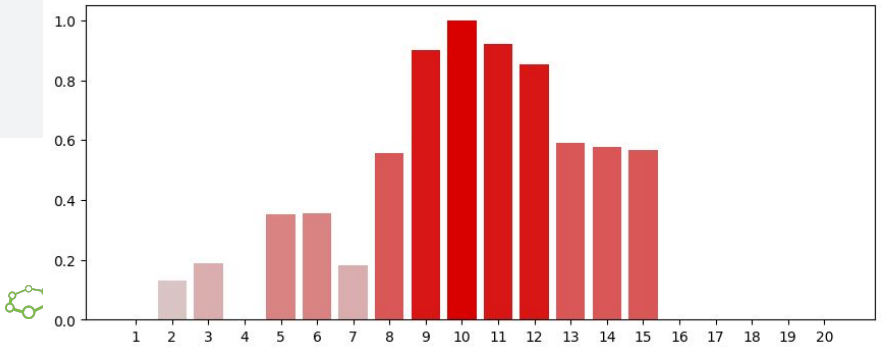
Visualization

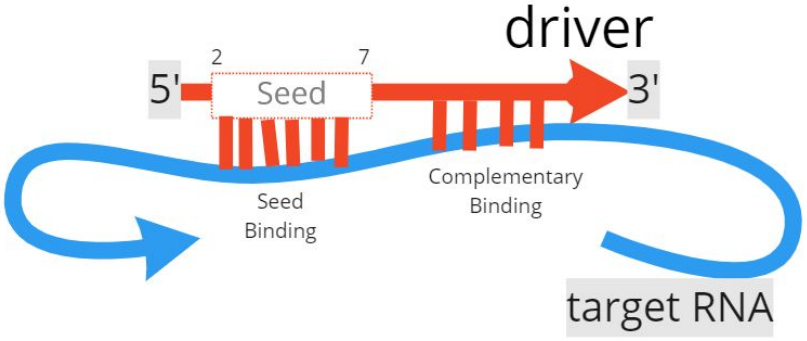
miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

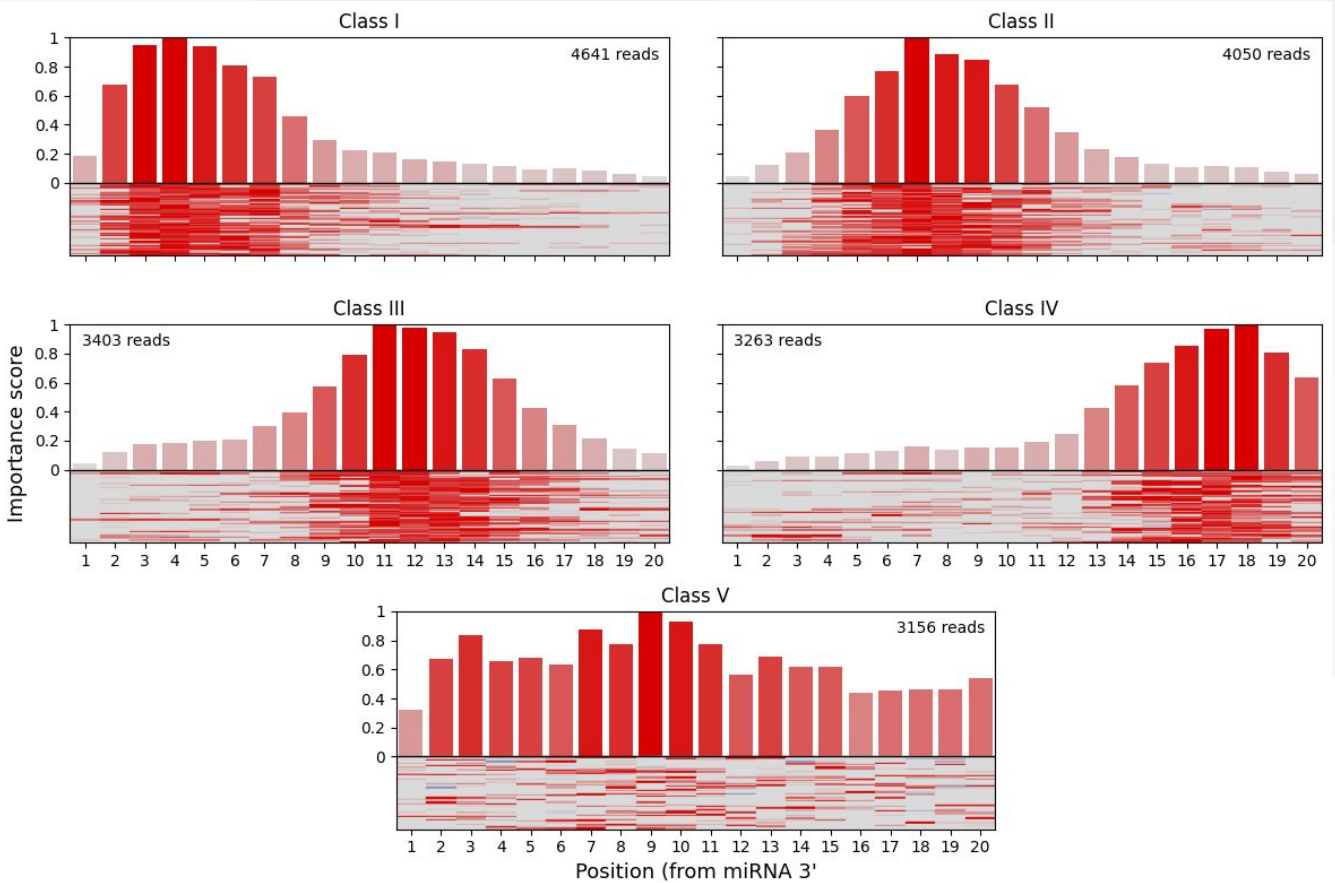
Predicted alignment:

miRNA position importance:

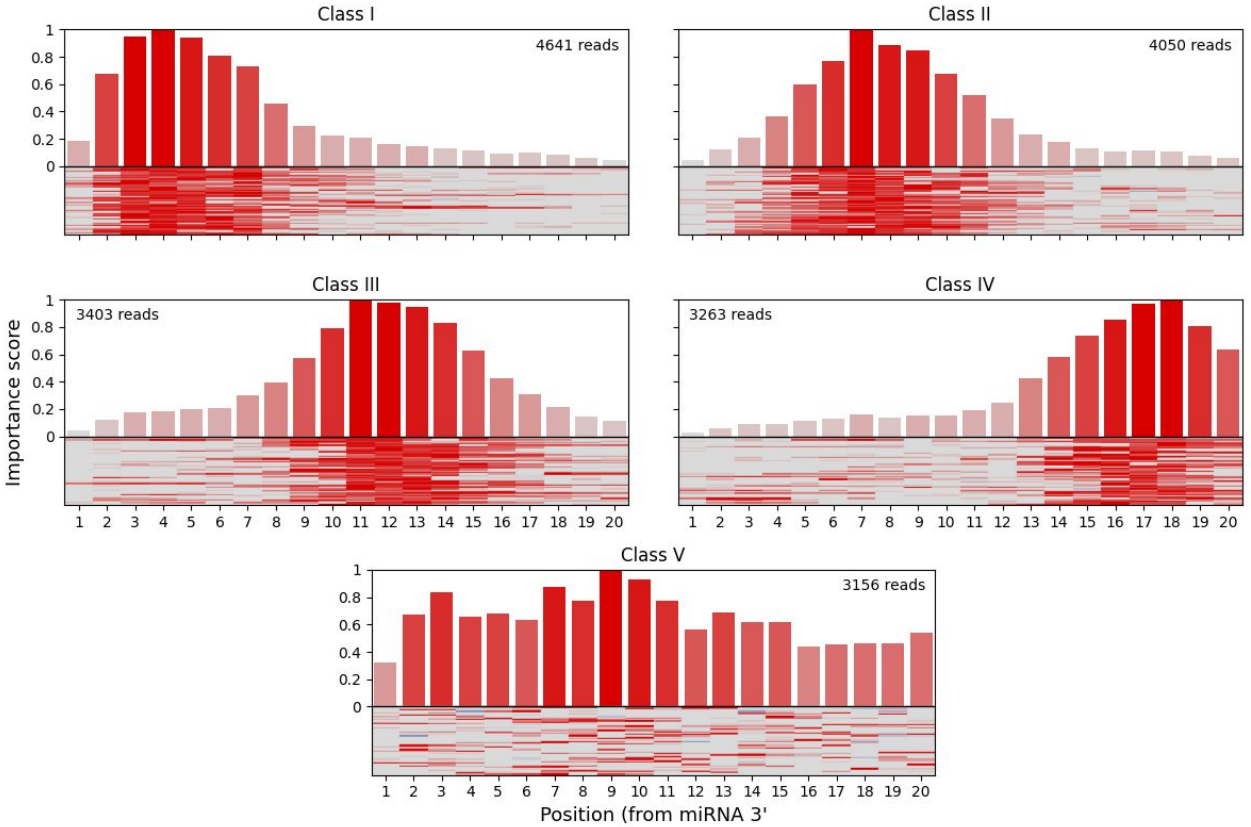


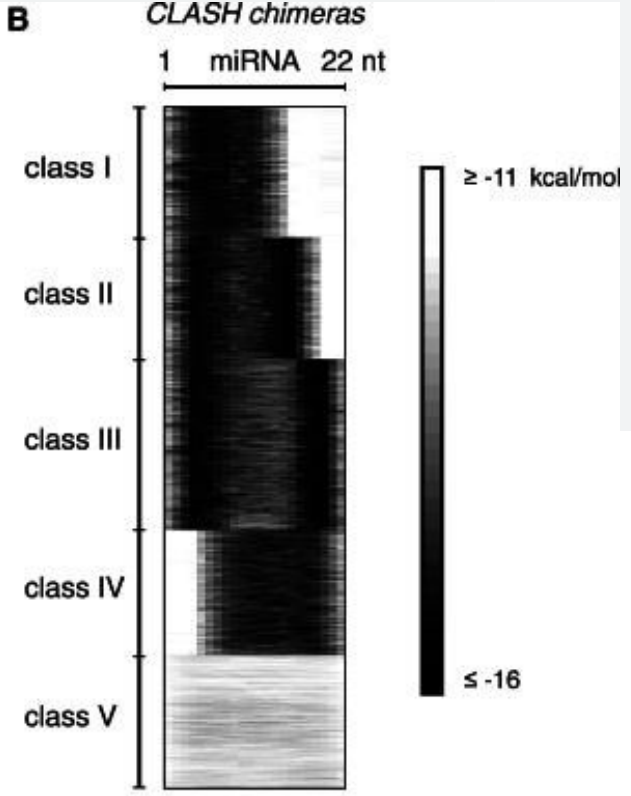


Classes of interaction



Classes of interaction





Helwak et al., 2013

Mutagenesis experiment

Open access, freely available online PLoS BIOLOGY

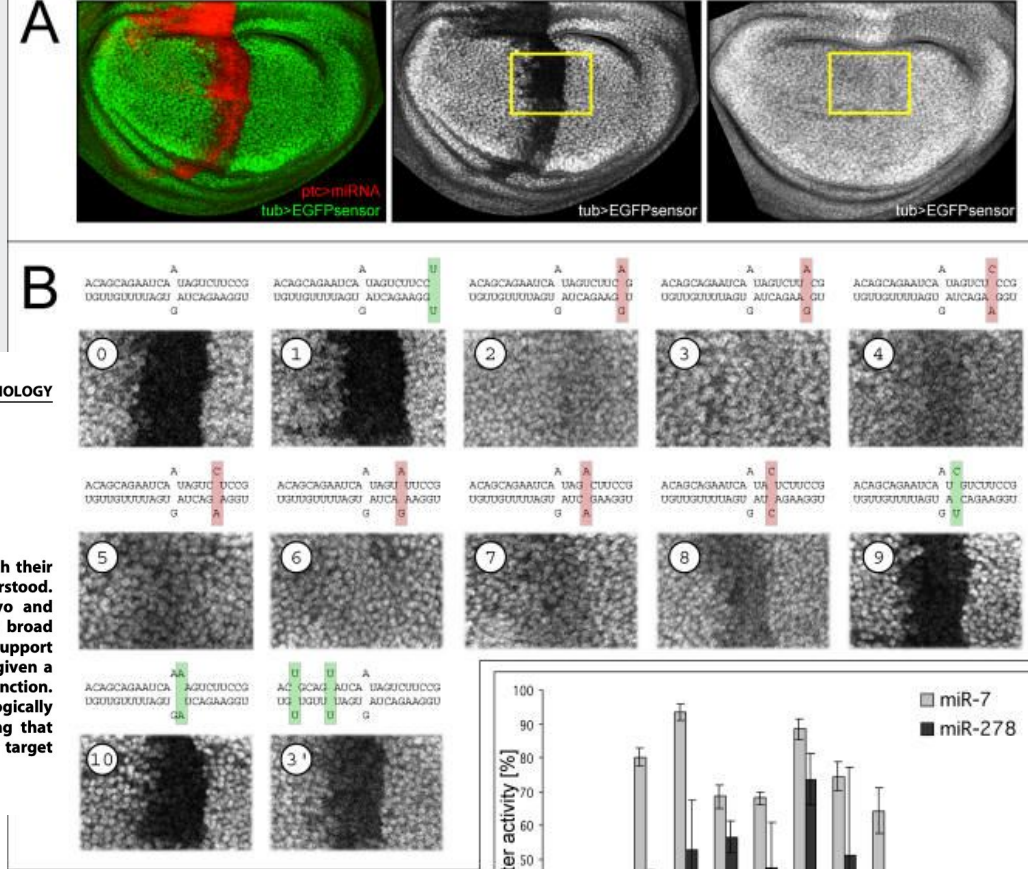
Principles of MicroRNA–Target Recognition

Julius Brennecke, Alexander Stark, Robert B. Russell, Stephen M. Cohen*

European Molecular Biology Laboratory, Heidelberg, Germany

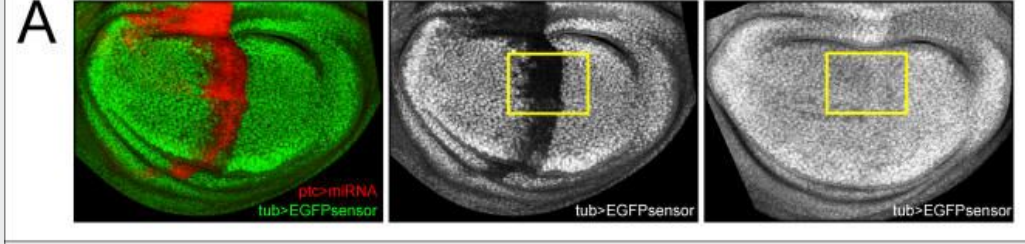
MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene expression in plants and animals. Although their biological importance has become clear, how they recognize and regulate target genes remains less well understood. Here, we systematically evaluate the minimal requirements for functional miRNA–target duplexes in vivo and distinguish classes of target sites with different functional properties. Target sites can be grouped into two broad categories. 5′ dominant sites have sufficient complementarity to the miRNA 5′ end to function with little or no support from pairing to the miRNA 3′ end. Indeed, sites with 3′ pairing below the random noise level are functional given a strong 5′ end. In contrast, 3′ compensatory sites have insufficient 5′ pairing and require strong 3′ pairing for function. We present examples and genome-wide statistical support to show that both classes of sites are used in biologically relevant genes. We provide evidence that an average miRNA has approximately 100 target sites, indicating that miRNAs regulate a large fraction of protein-coding genes and that miRNA 3′ ends are key determinants of target specificity within miRNA families.

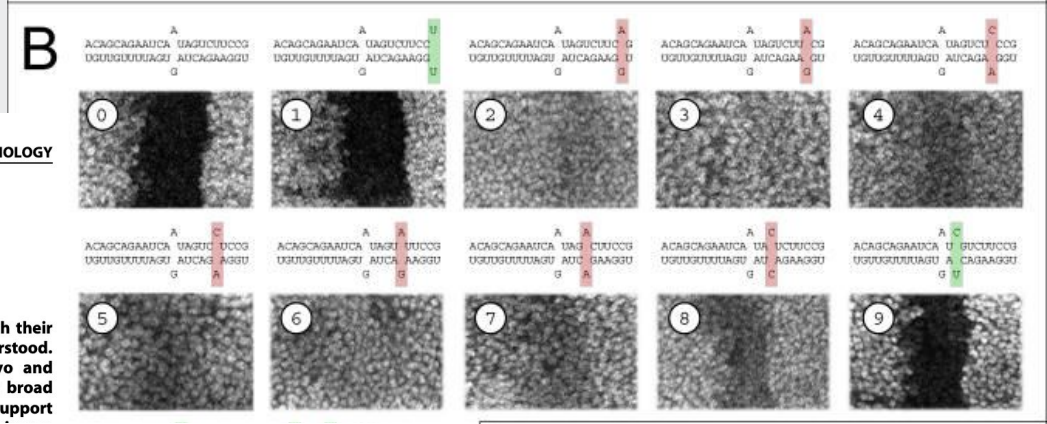
Citation: Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA–target recognition. PLoS Biol 3(3): e85.



CEITEC

Mutagenesis experiment





Open access, freely available online PLOS BIOLOGY

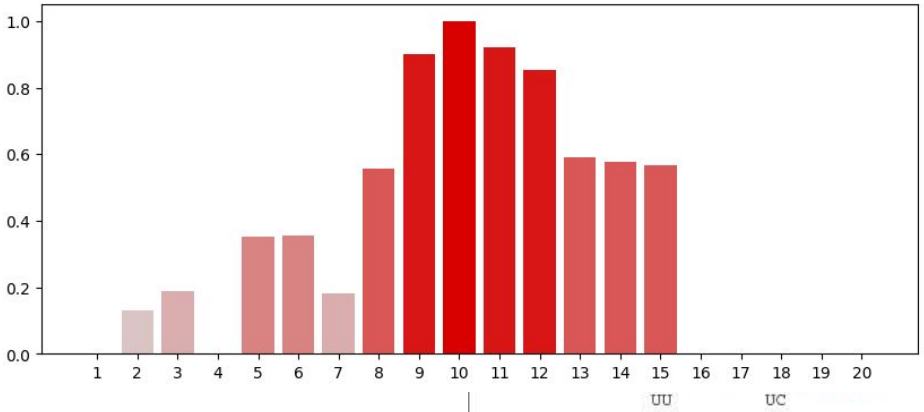
Principles of MicroRNA–Target Recognition

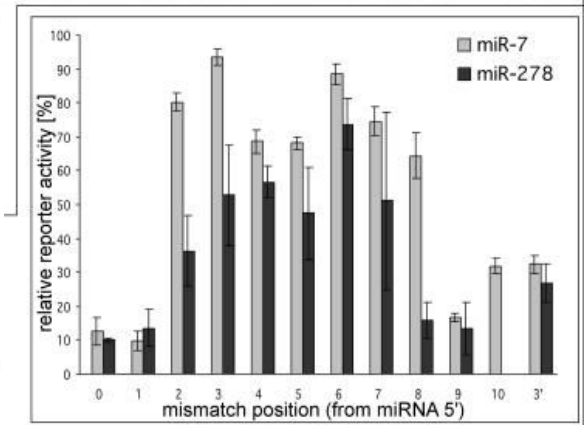
Julius Brennecke, Alexander Stark, Robert B. Russell, Stephen M. Cohen*

European Molecular Biology Laboratory, Heidelberg, Germany

MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene expression in plants and animals. Although their biological importance has become clear, how they recognize and regulate target genes remains less well understood. Here, we systematically evaluate the minimal requirements for functional miRNA–target duplexes in vivo and distinguish classes of target sites with different functional properties. Target sites can be grouped into two broad categories. 5′ dominant sites have sufficient complementarity to the miRNA 5′ end to function with little or no support from pairing to the miRNA 3′ end. Indeed, sites wi… strong 5′ end. In contrast, 3′ compensatory sites ha… We present examples and genome-wide statistical… relevant genes. We provide evidence that an ave… miRNAs regulate a large fraction of protein-codin… specificity within miRNA families.

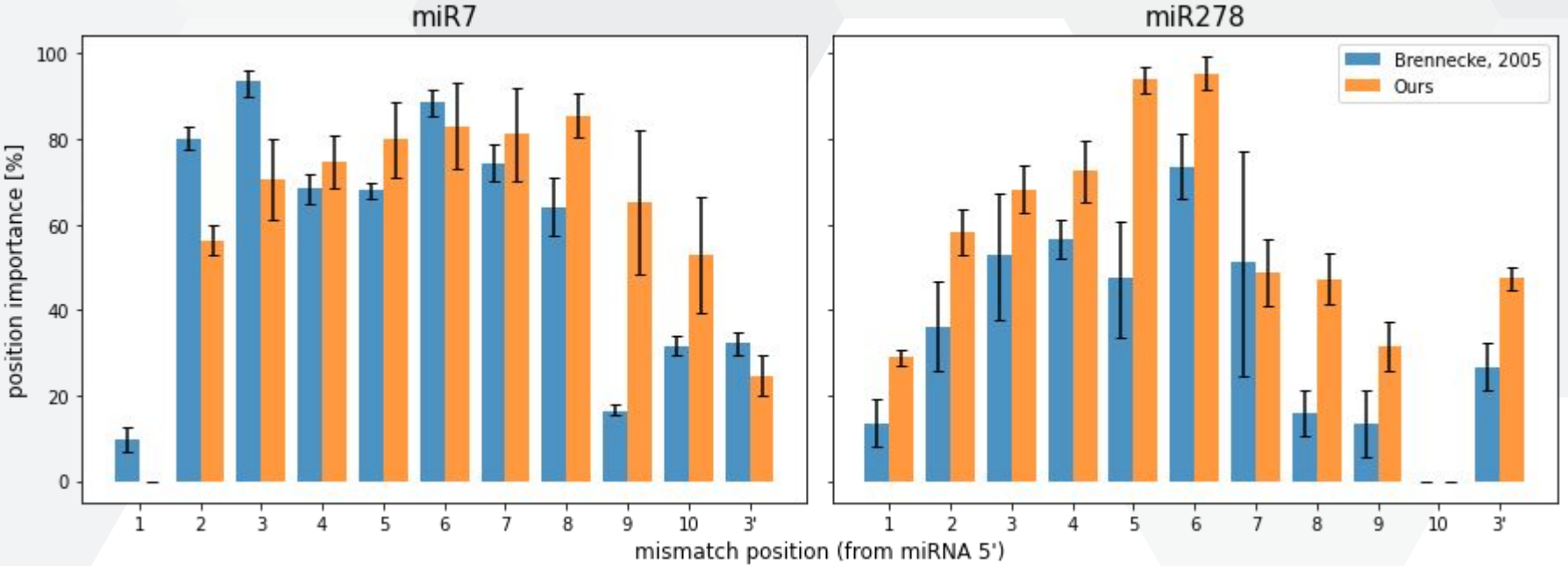
Citation: Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of r…





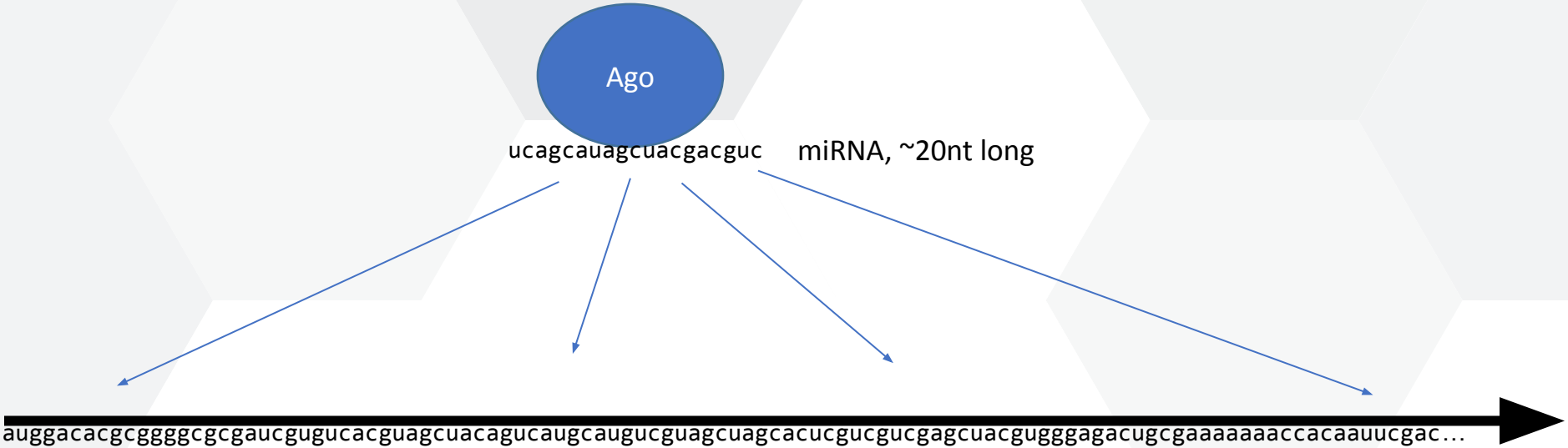
CEITEC

Verification

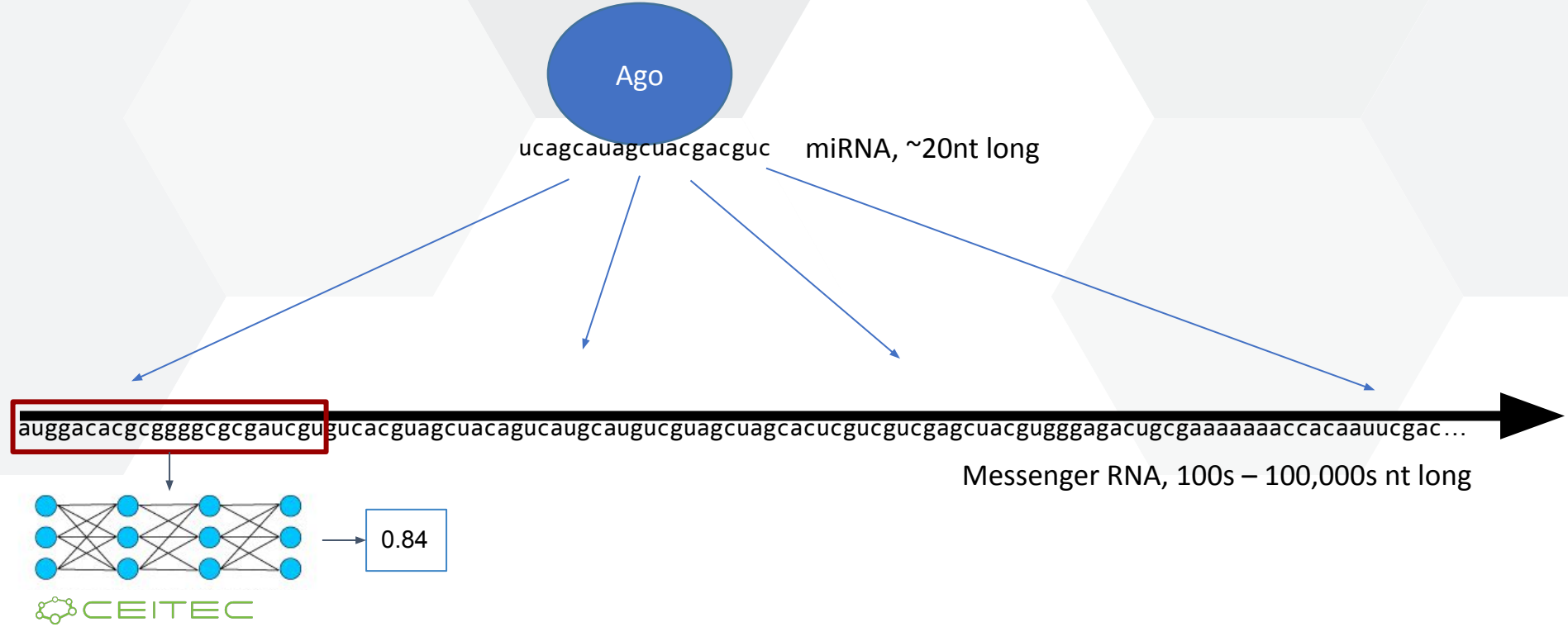


miRNA	miR-7	miR-278
correlation	0.59	0.85

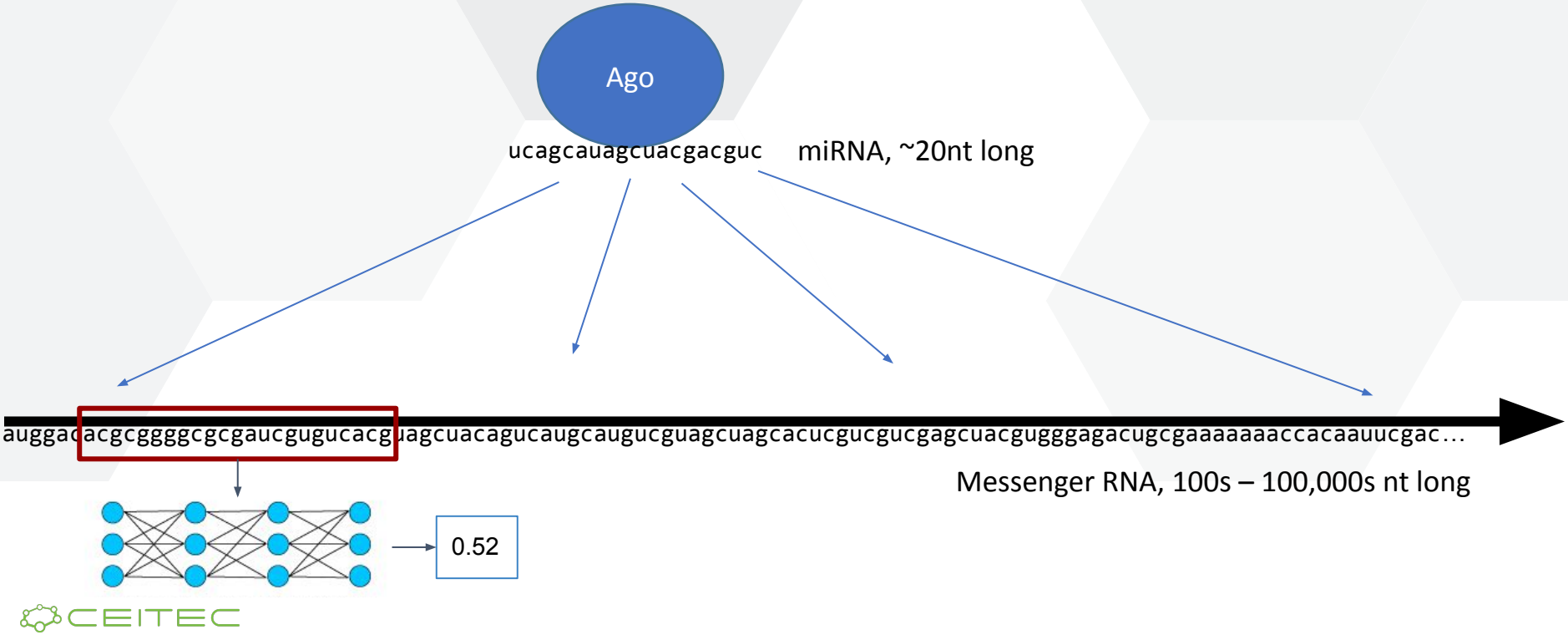
Scanning


Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long

Scanning

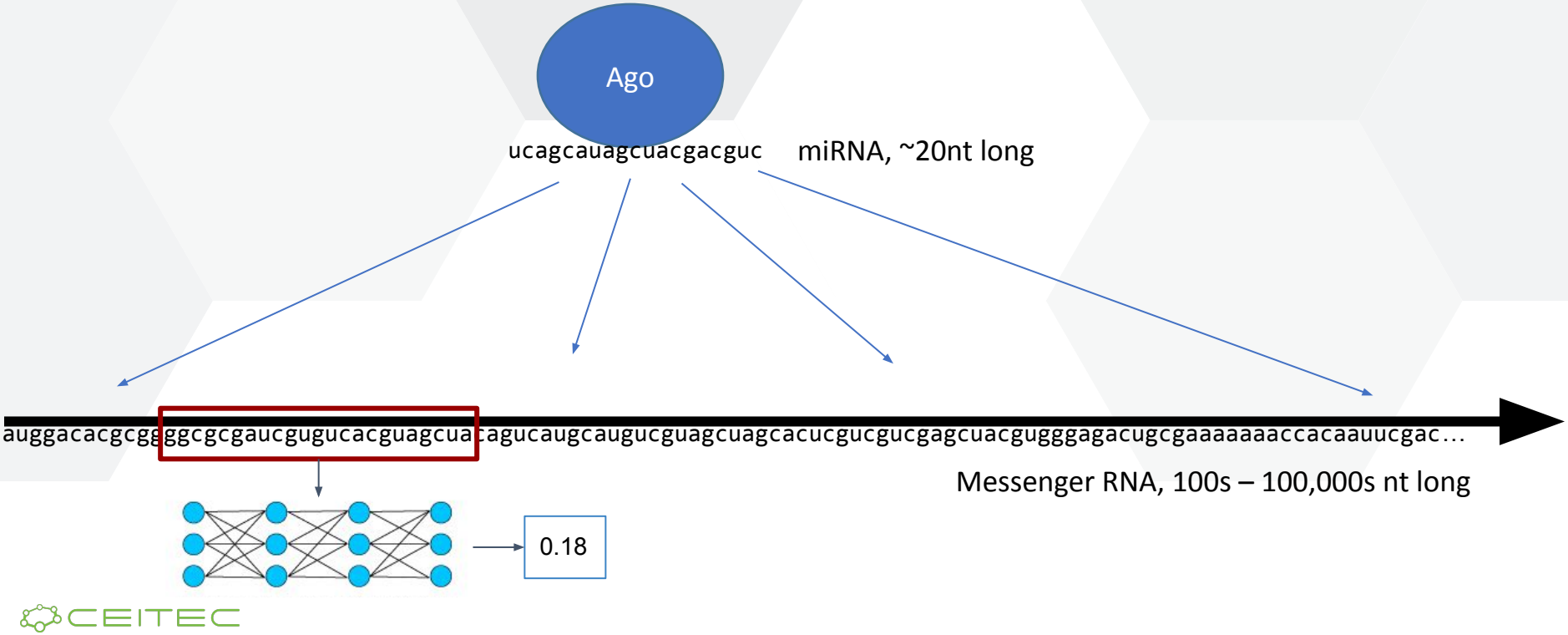

Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long
0.84
CEITEC

Scanning



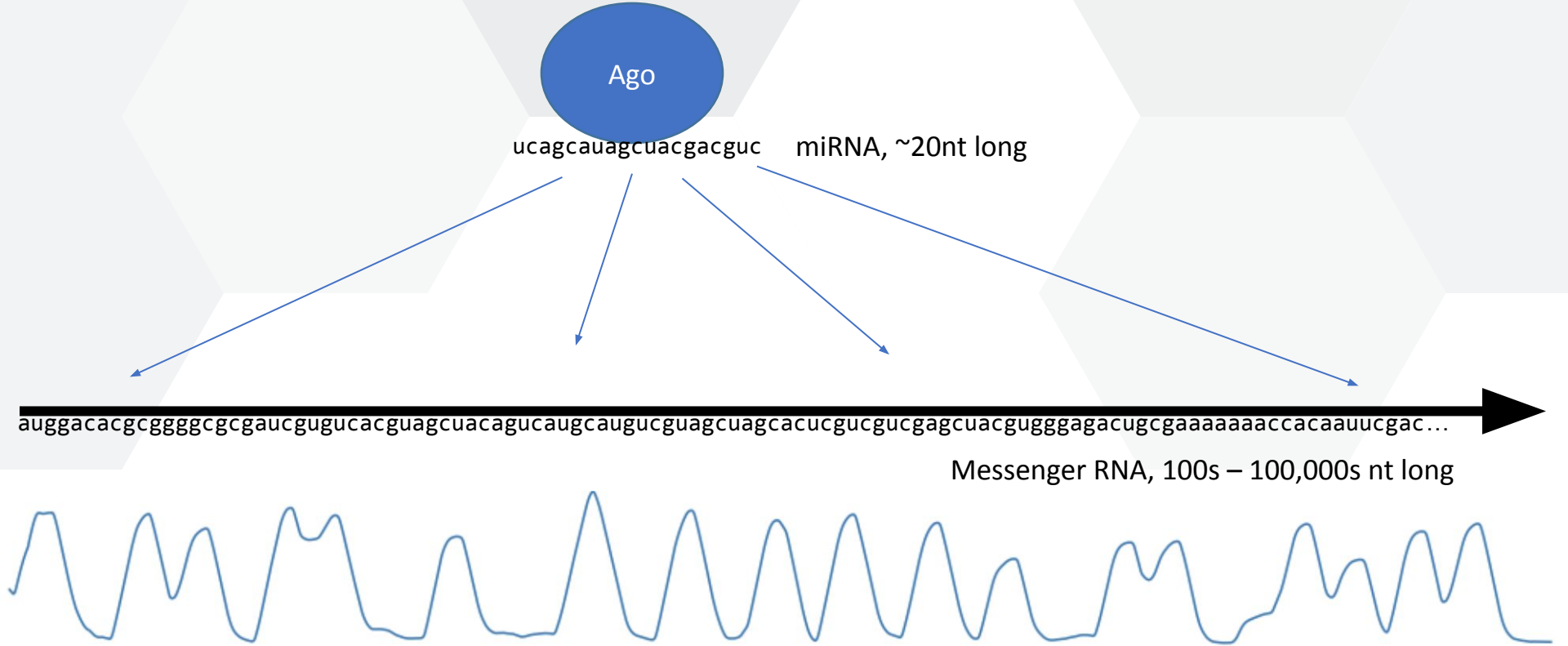
CEITEC

Scanning

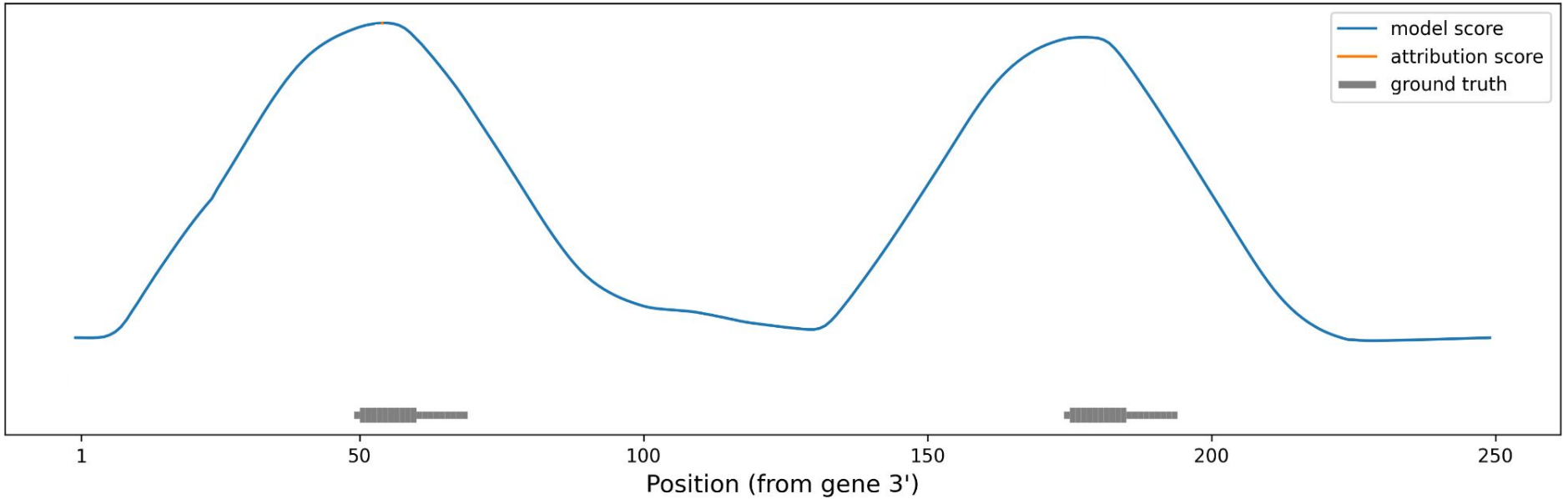


CEITEC

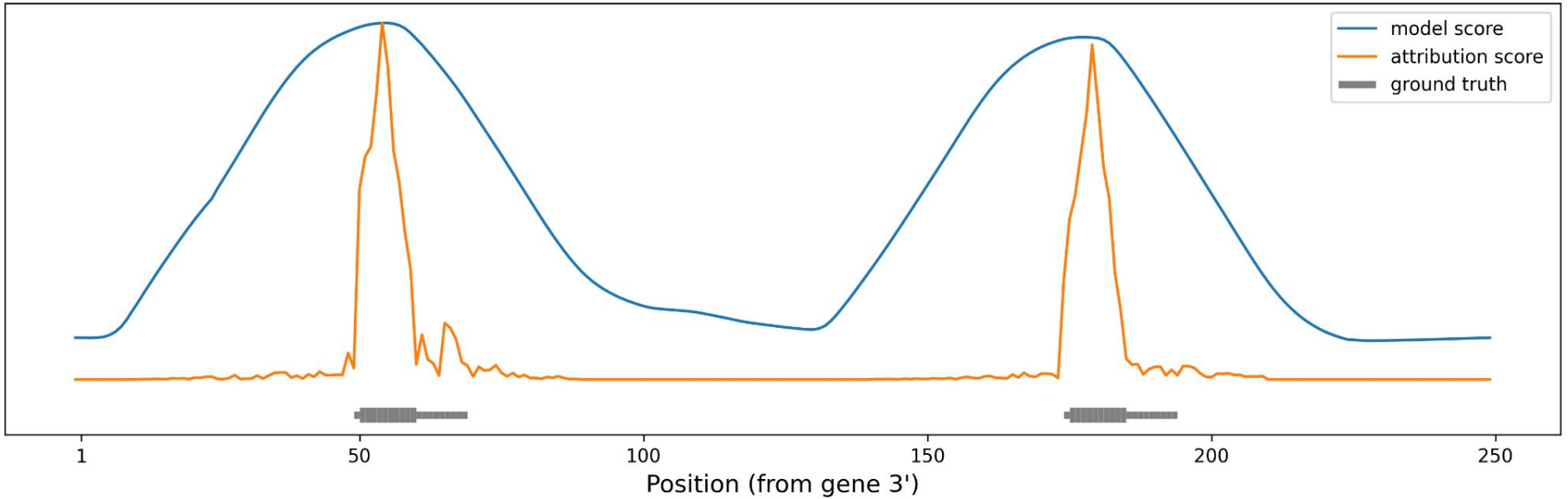
Scanning


Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long

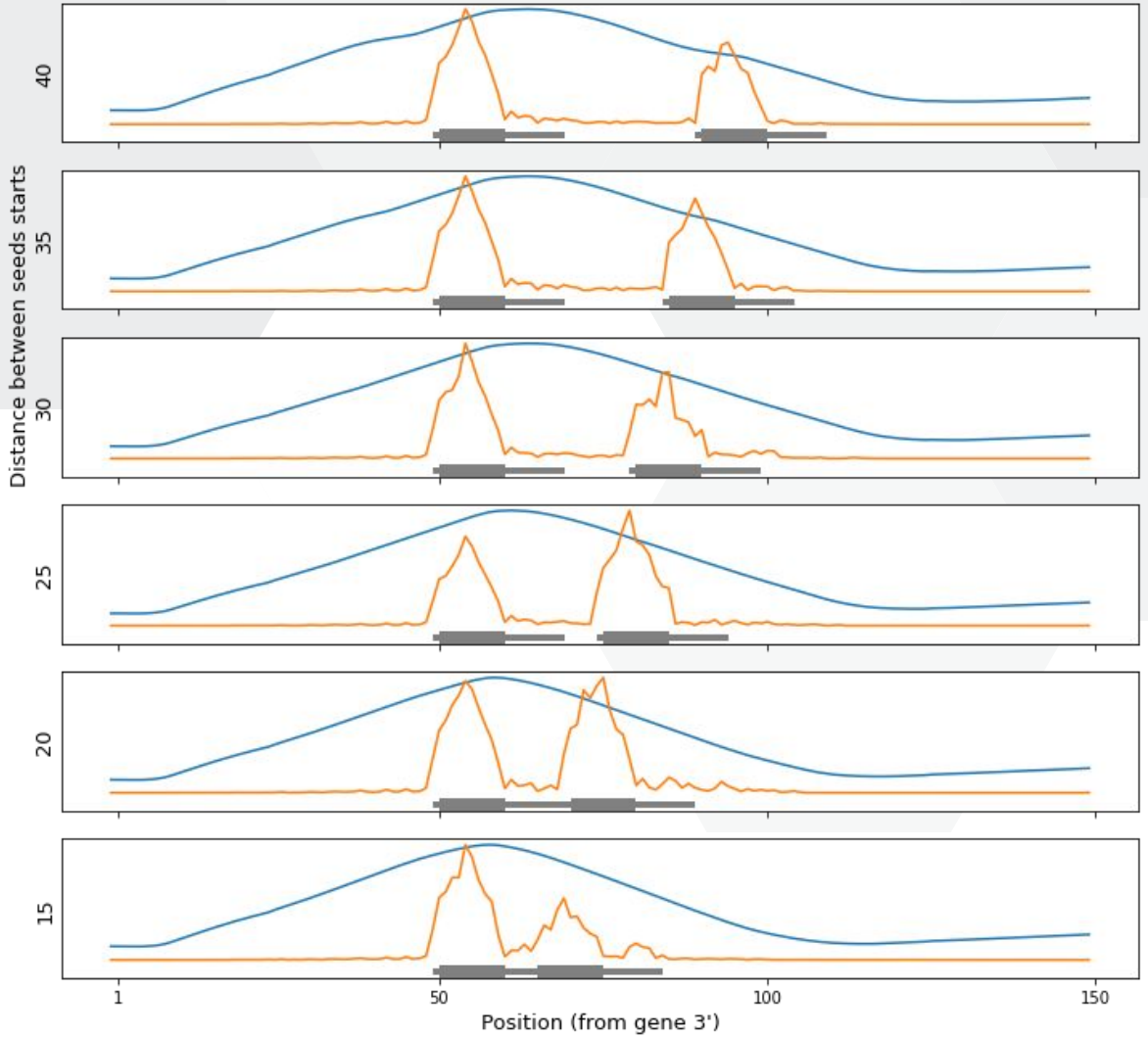
Narrowing the peaks



Narrowing the peaks

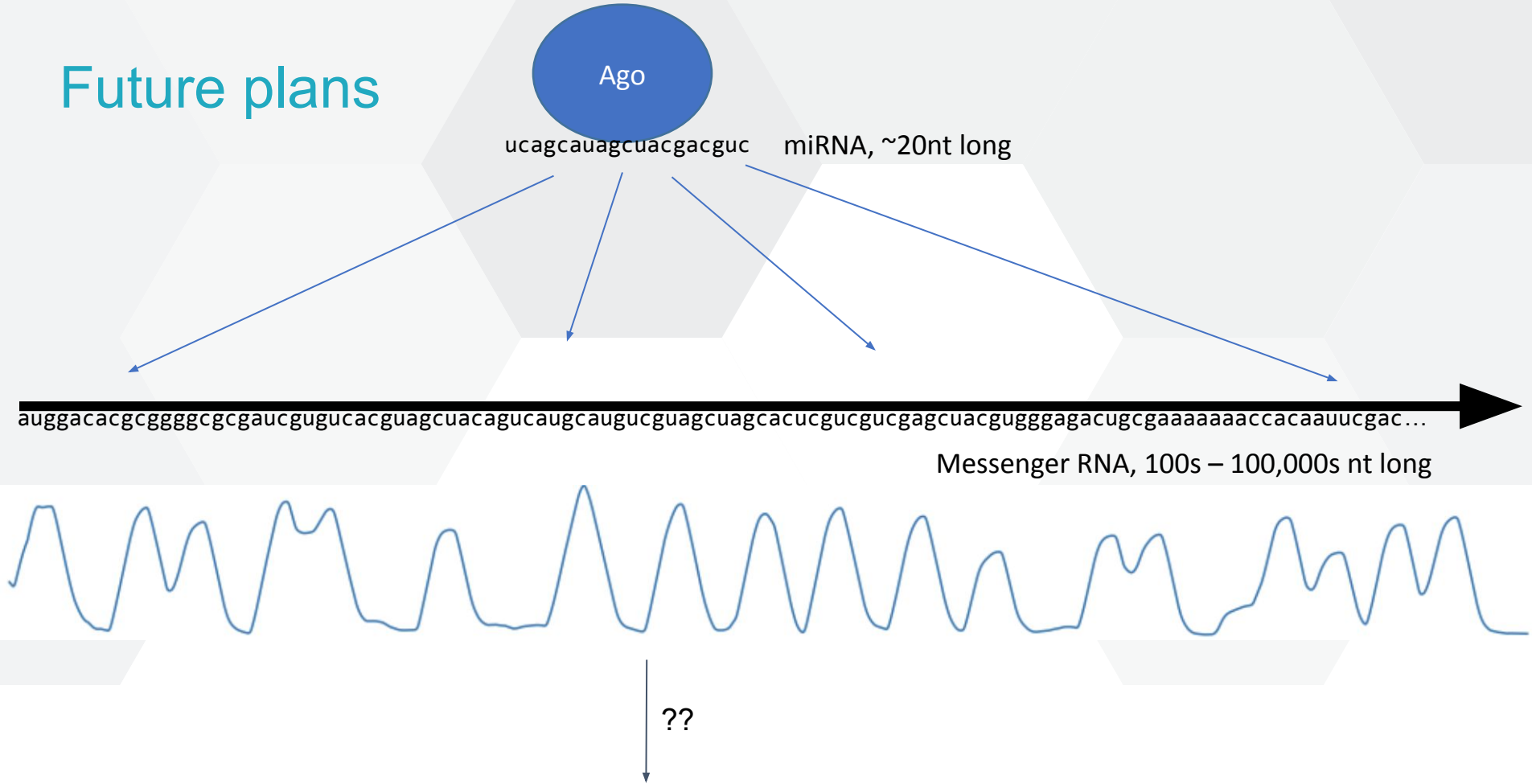


Close by peaks



CEITEC

Future plans



CEITEC

Thank you for your Attention!

514001@mail.muni.cz