CEITEC

Central European Institute of Technology

BRNO | CZECH REPUBLIC

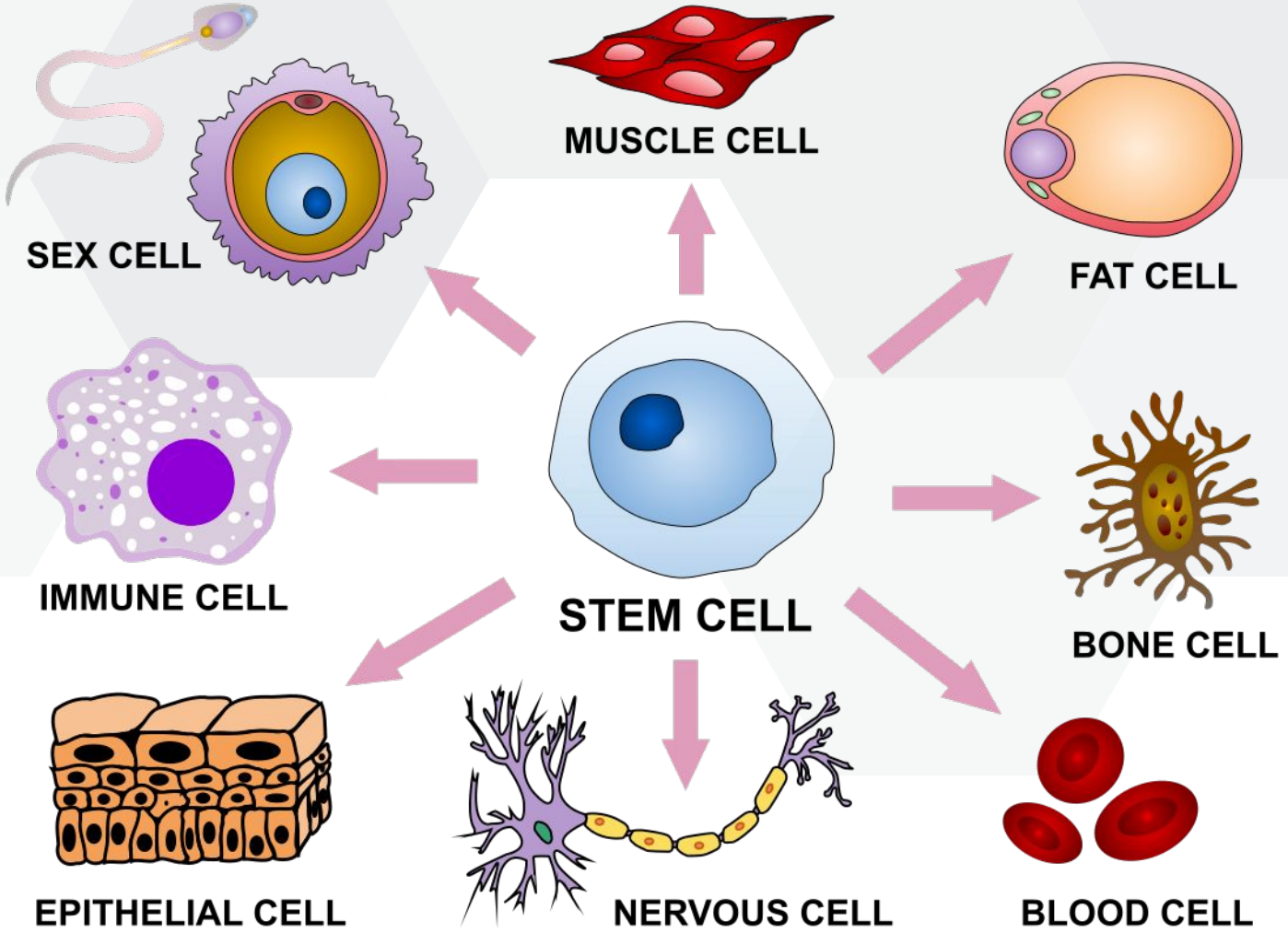
MUNI

# Understanding miRNA binding behaviour through Deep Learning

David Čechák, Katarina Grešová

CEITEC, Masaryk University, Brno, Czech republic

# Cells


SEX CELL
MUSCLE CELL
FAT CELL
IMMUNE CELL
STEM CELL
BONE CELL
EPITHELIAL CELL
NERVOUS CELL
BLOOD CELL


A–T

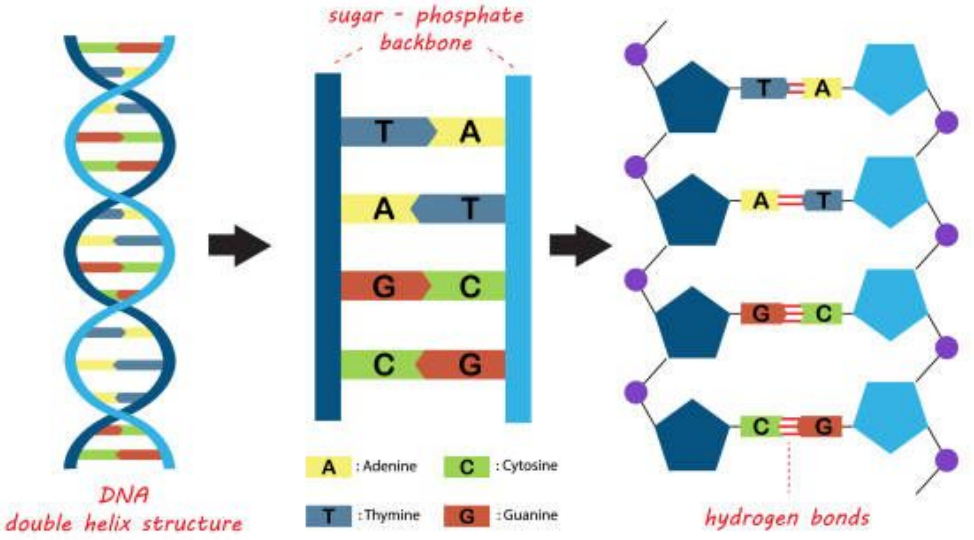
G–C


A–T

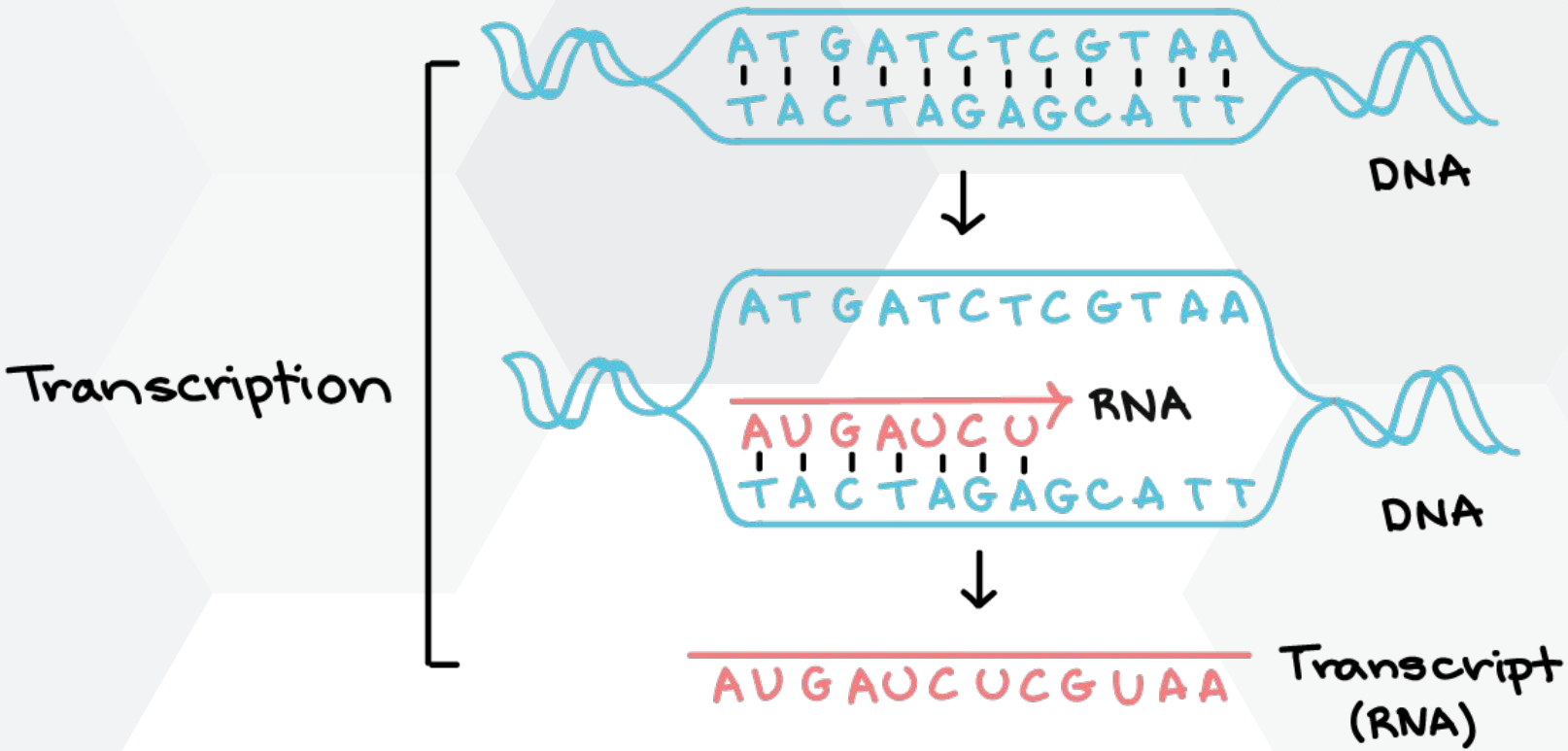
G–C

BIOLOGY

# DNA Structure




CEITEC

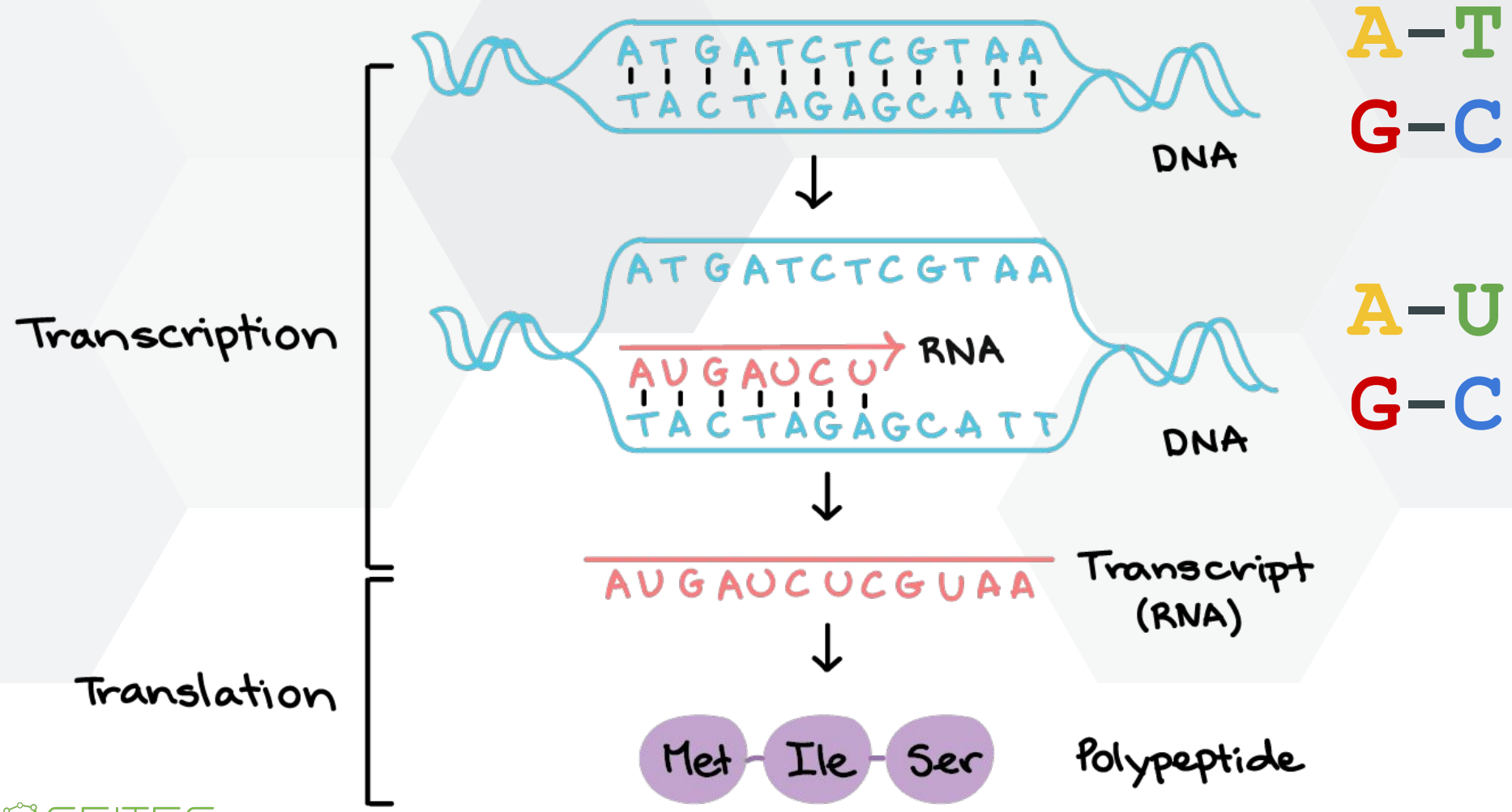

Transcription
ATGATCTCGTAA
TACTAGAGCATT
DNA
ATGATCTCGTAA
AUGAUCU
RNA
TACTAGAGCATT
DNA
AUGAUCUCGUAA
Transcript
(RNA)

A–T

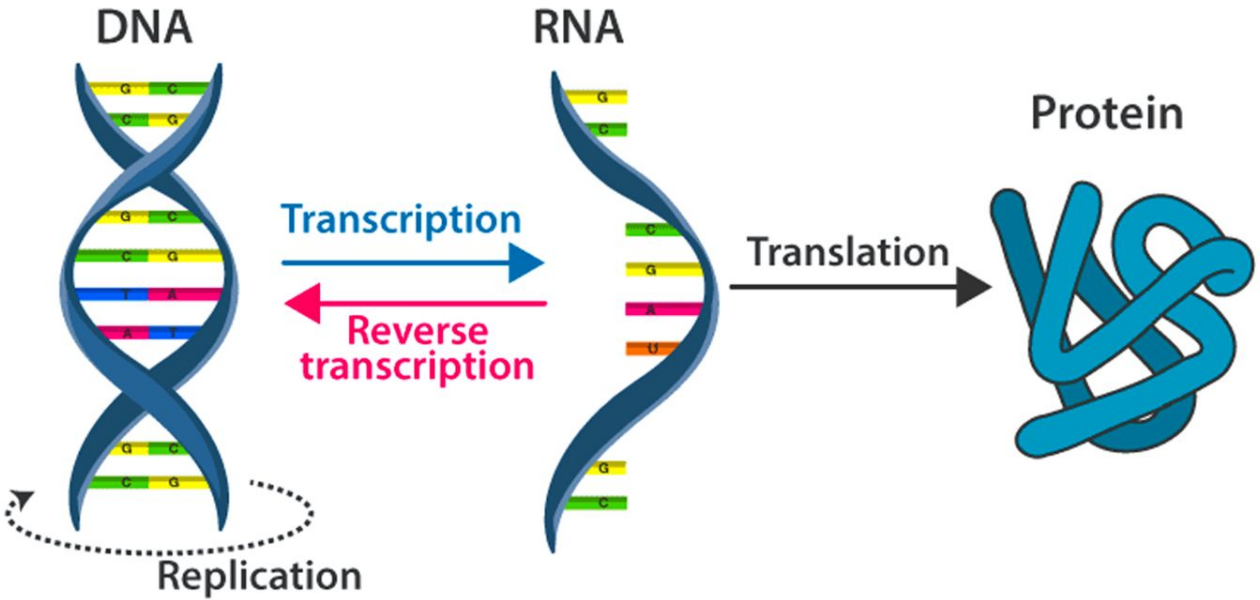
G–C

A–U

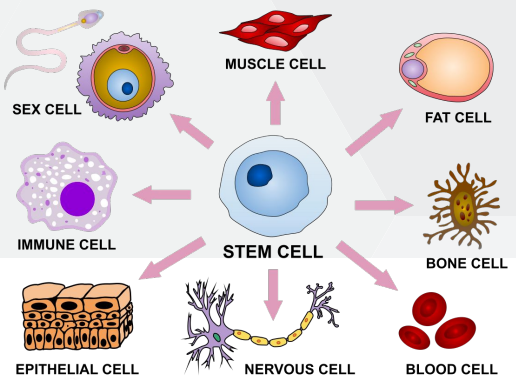
G–C

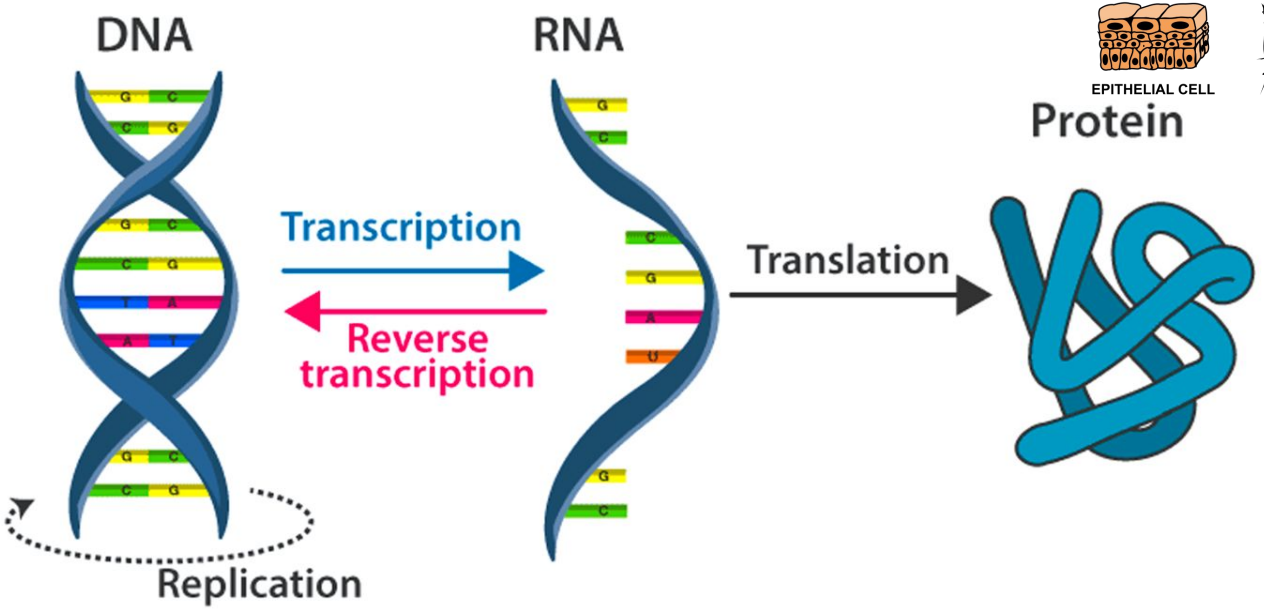

A–T
G–C
A–U
G–C
Transcription
Translation
ATGATCTCGTAA
TACTAGAGCATT
DNA
ATGATCTCGTAA
AUGAUCU
RNA
TACTAGAGCATT
DNA
AUGAUCUCGUAA
Transcript
(RNA)
Met
Ile
Ser
Polypeptide

# Central Dogma of Molecular Biology

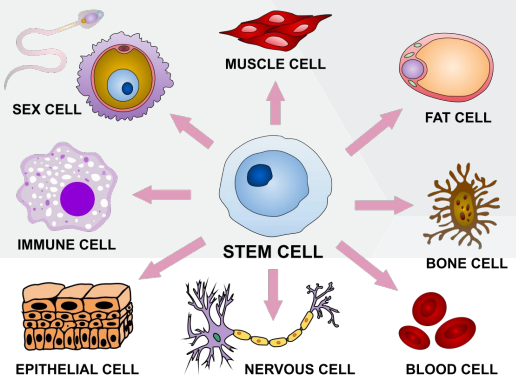


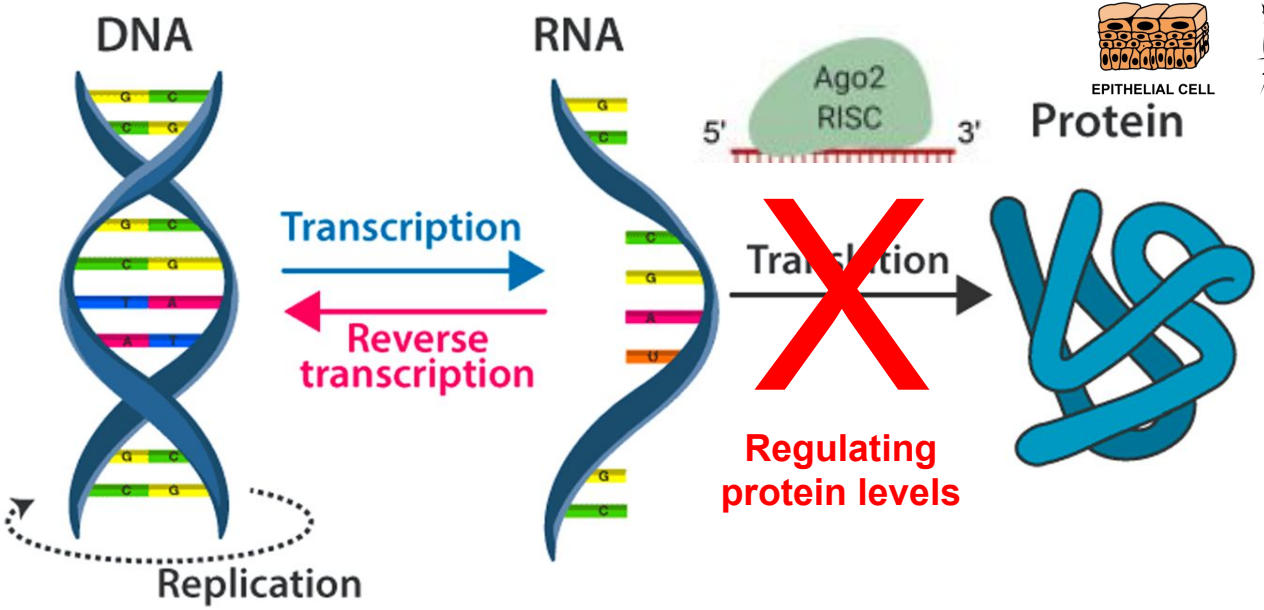
# Central Dogma of Molecular Biology



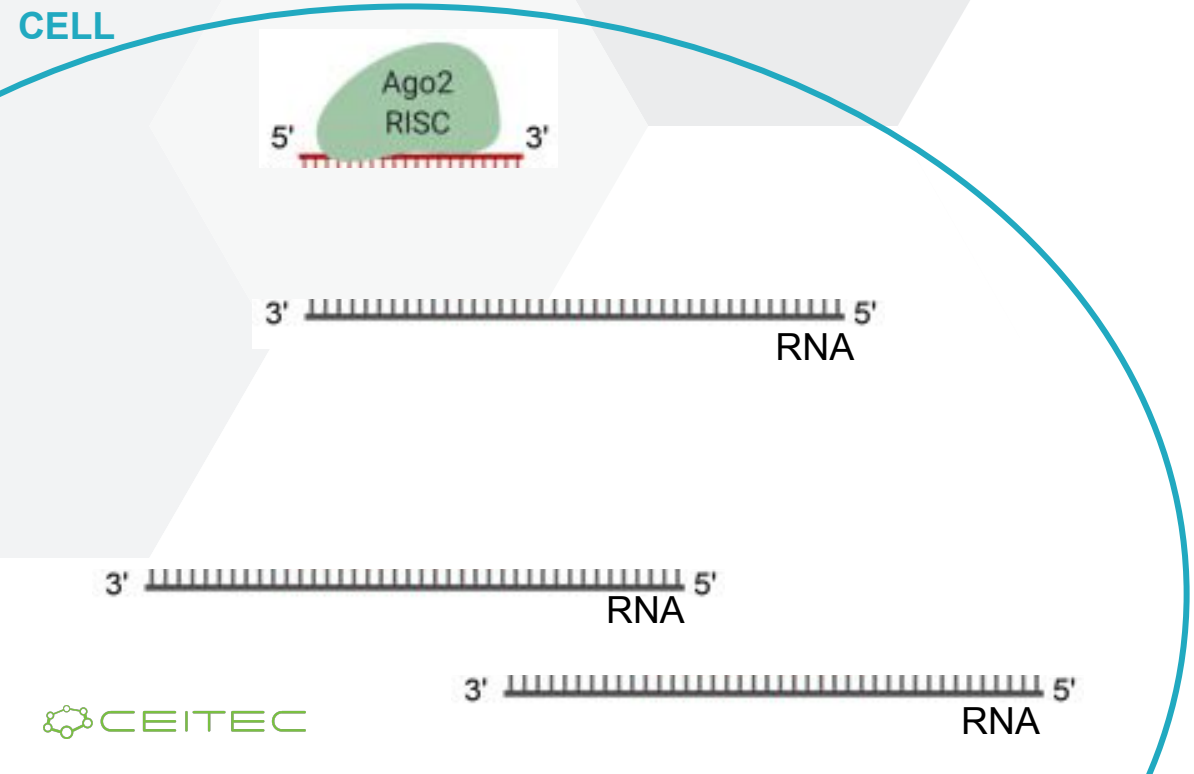



# Central Dogma of Molecular Biology



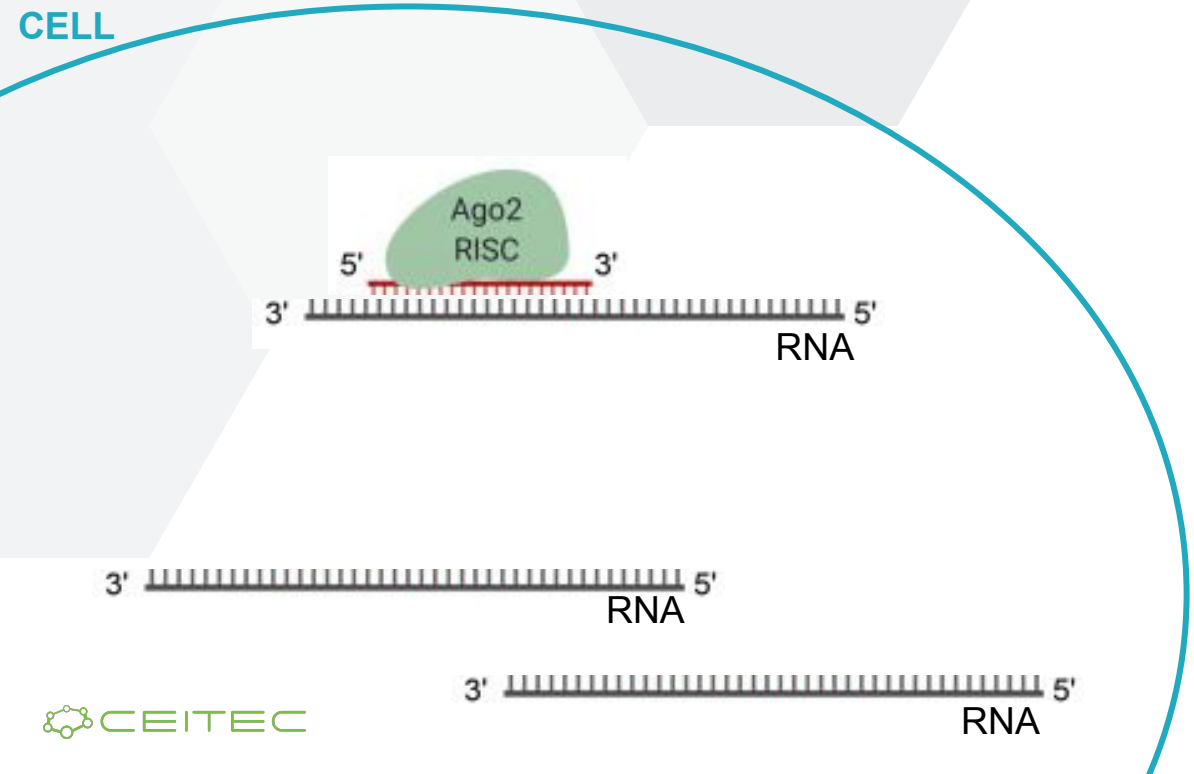



# RISC (RNA-induced silencing complex)




CEITEC

# RISC (RNA-induced silencing complex)

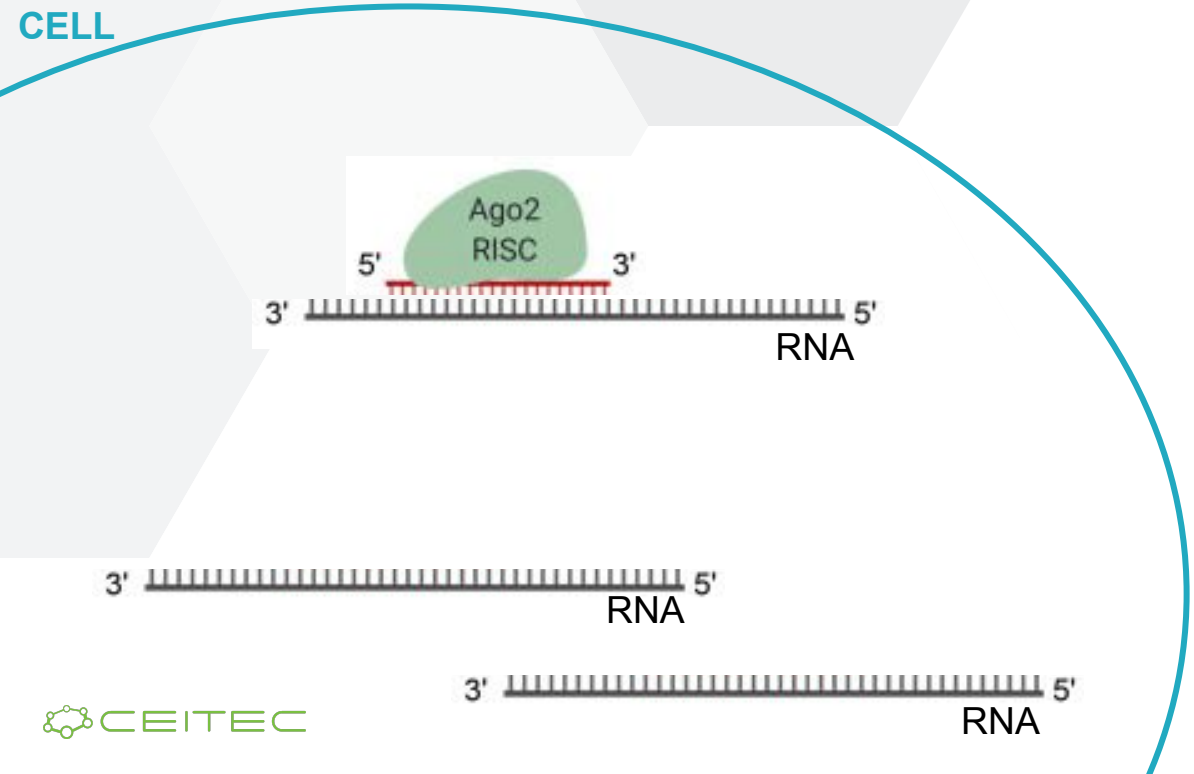



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)

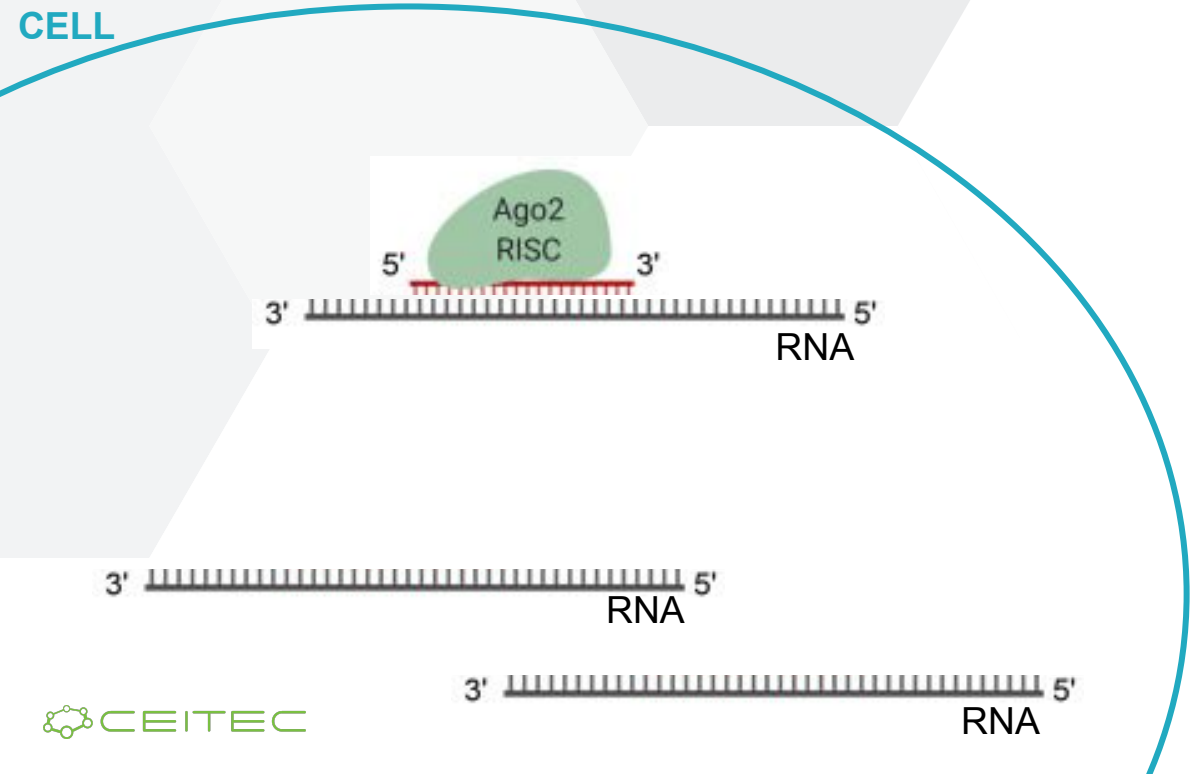



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)

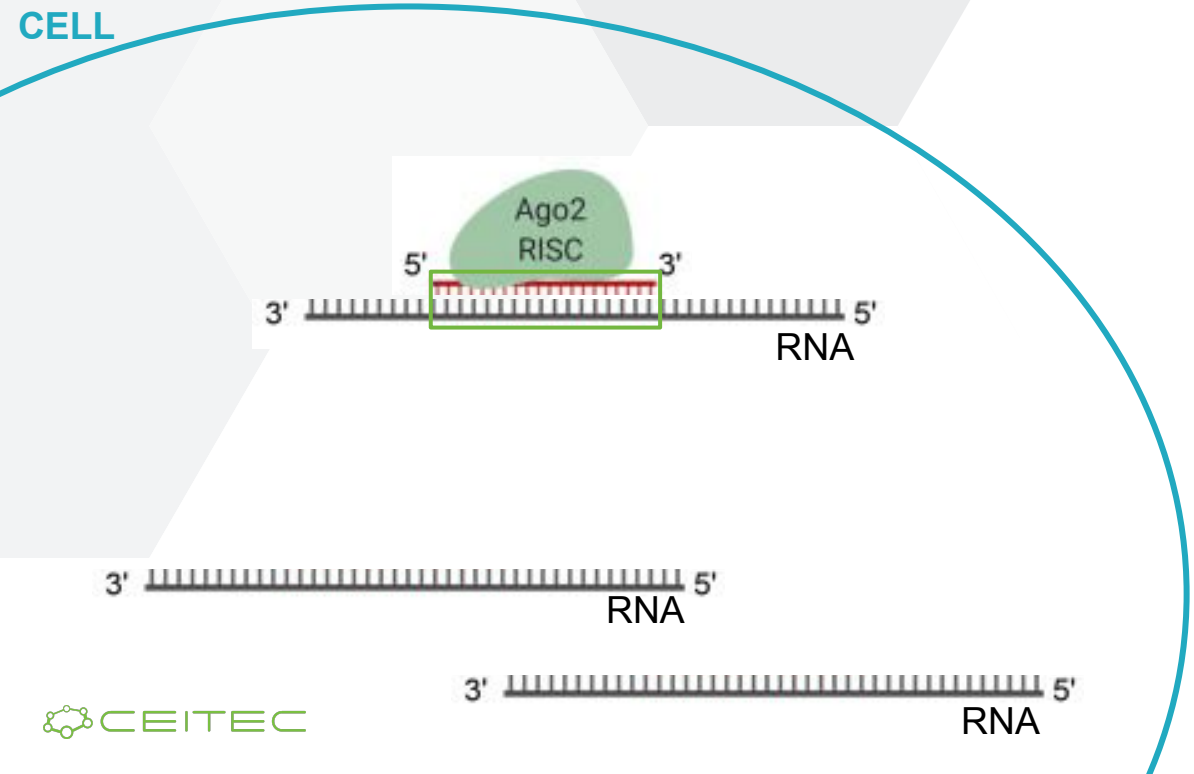



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)

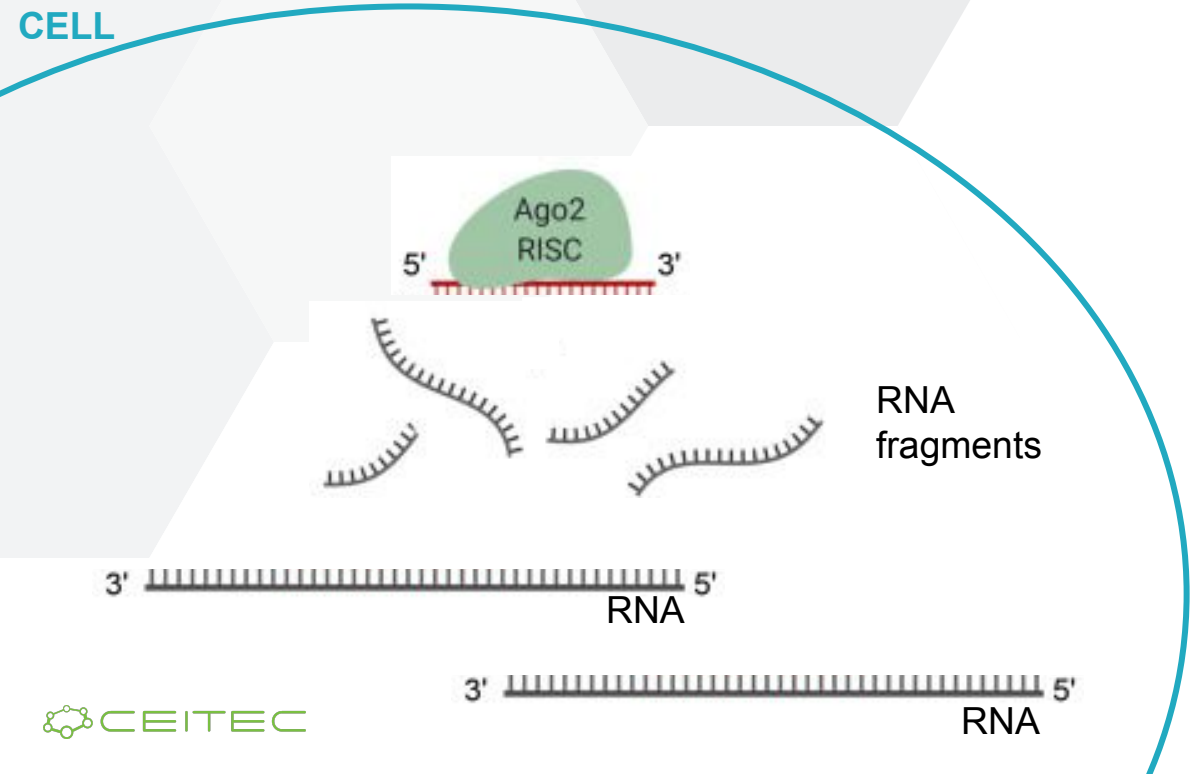



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)

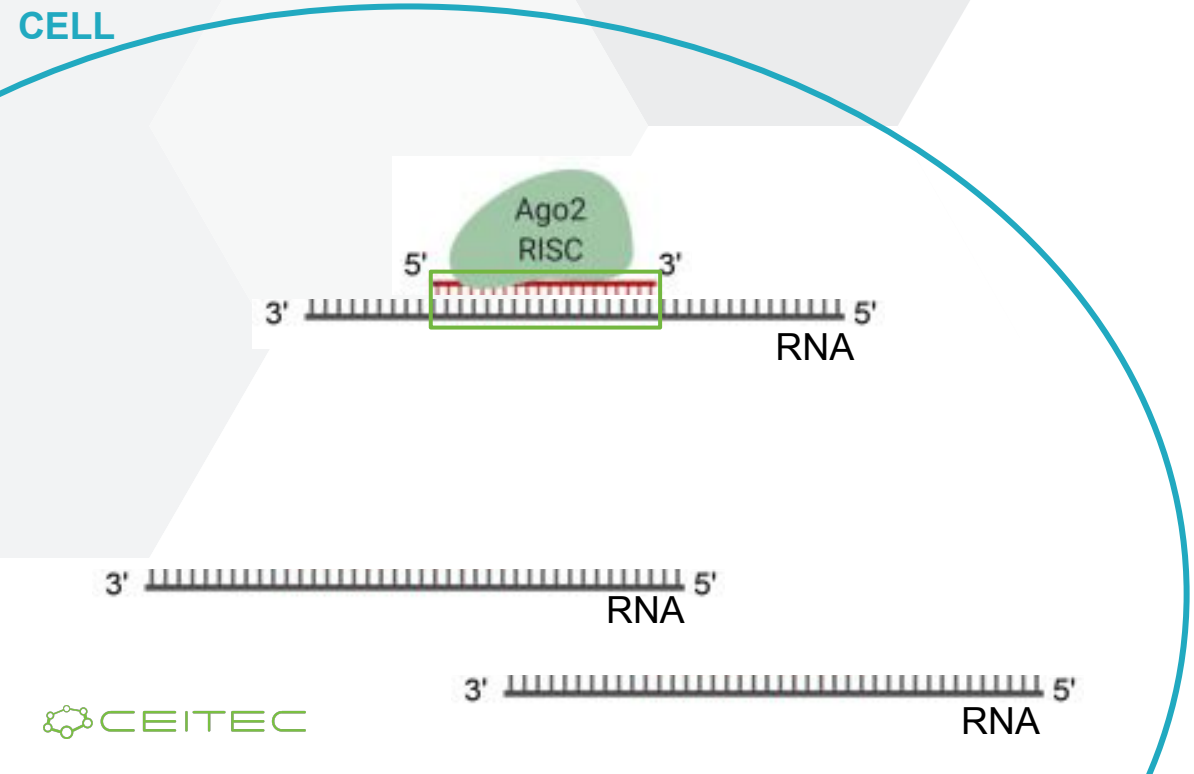



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)

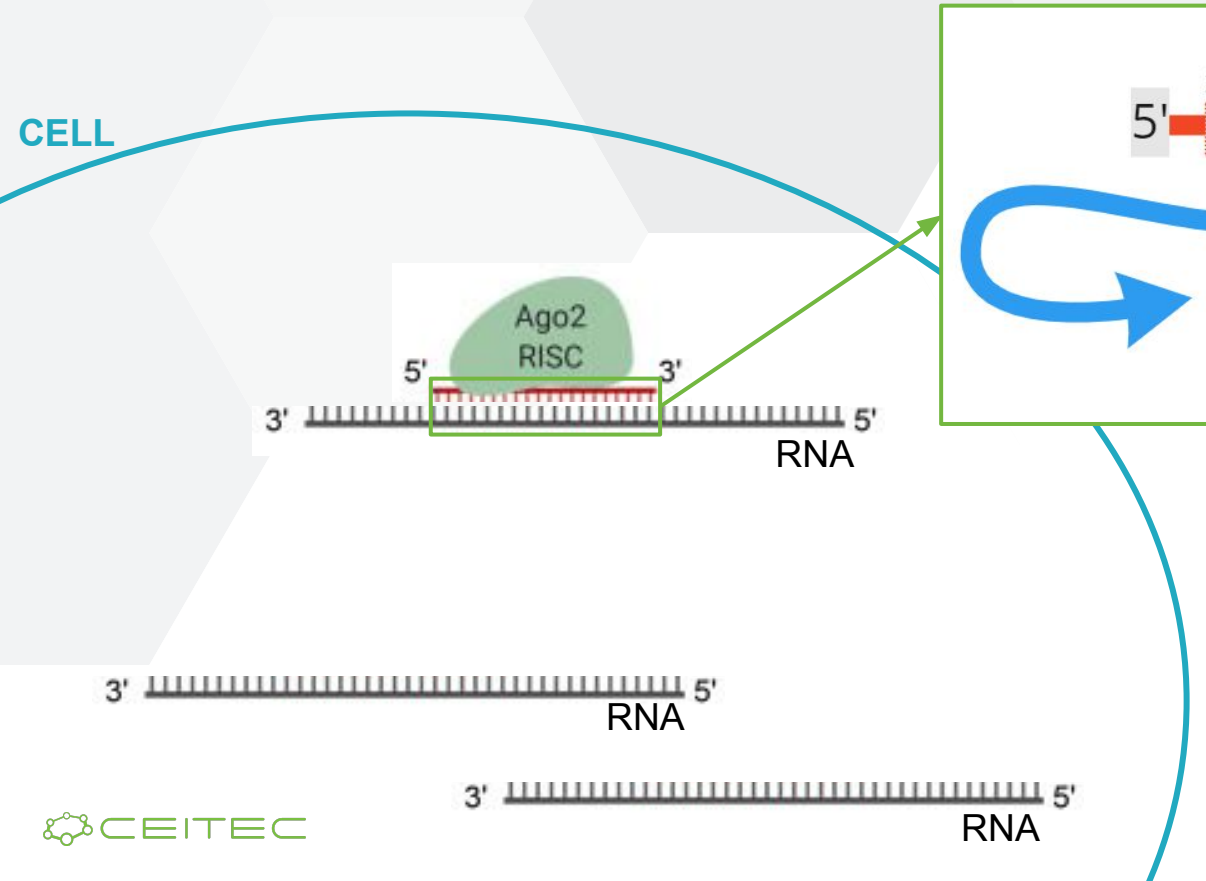


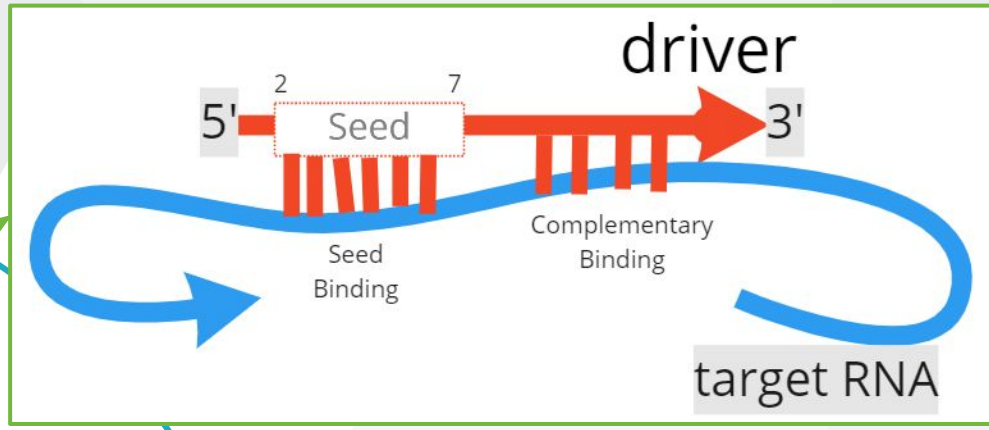

A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)



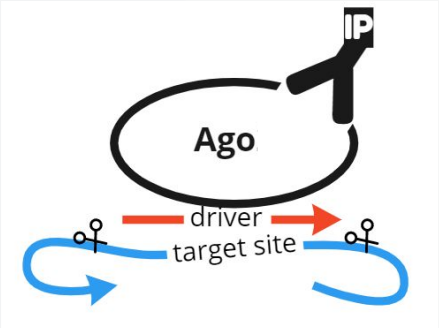




A–U

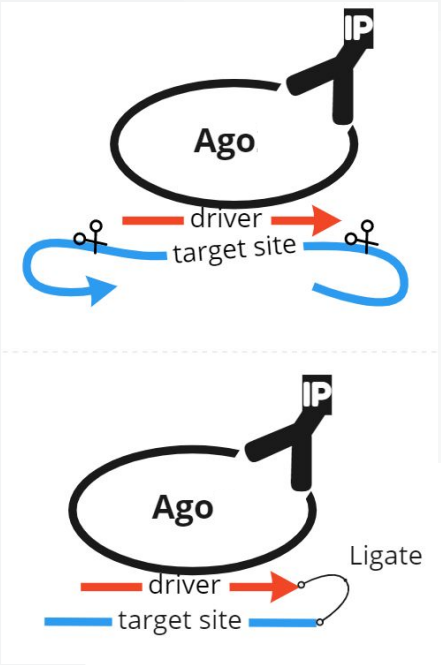
G–C

CEITEC

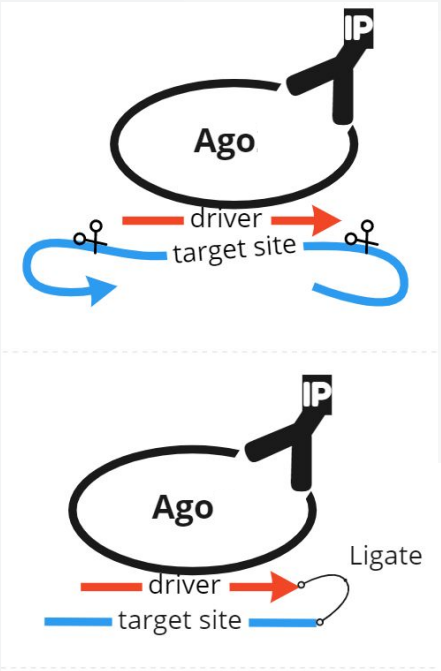
# Biological experiment - CLASH



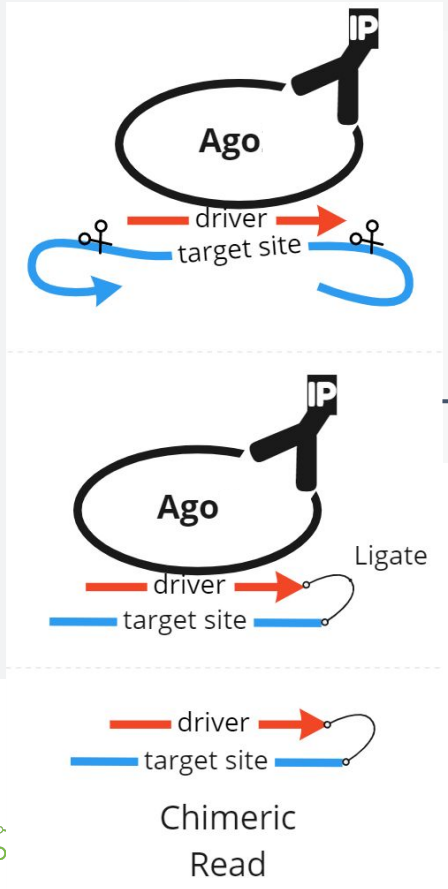
# Biological experiment - CLASH



# Biological experiment - CLASH



# Biological experiment - CLASH - positive and negative samples creation




| 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

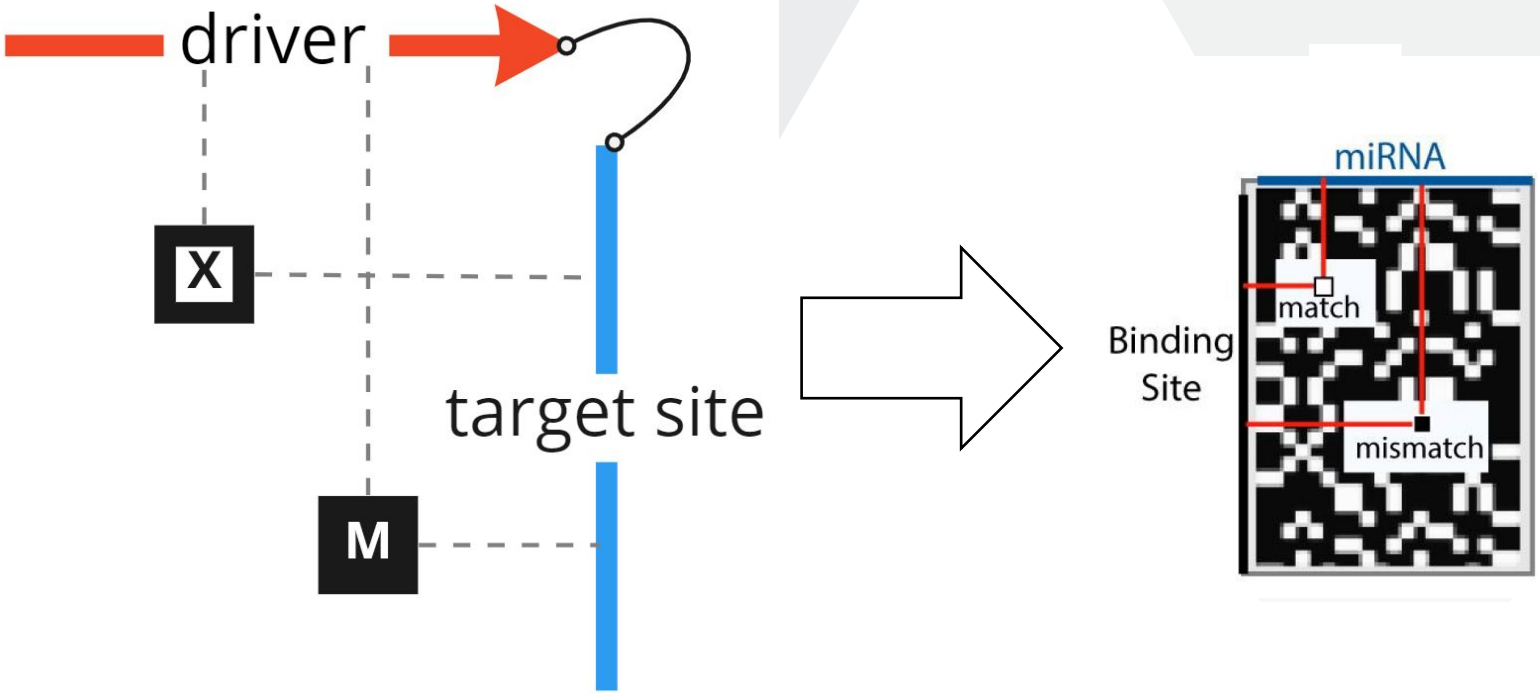
TACGTCAGTTCATGAAGCT
A

(driver ~20nt)

+

AGTTCTAGTTCGTCCGTCAGTGTCAG
TTCATGAGCACCAGTCACGTTCGTCTA

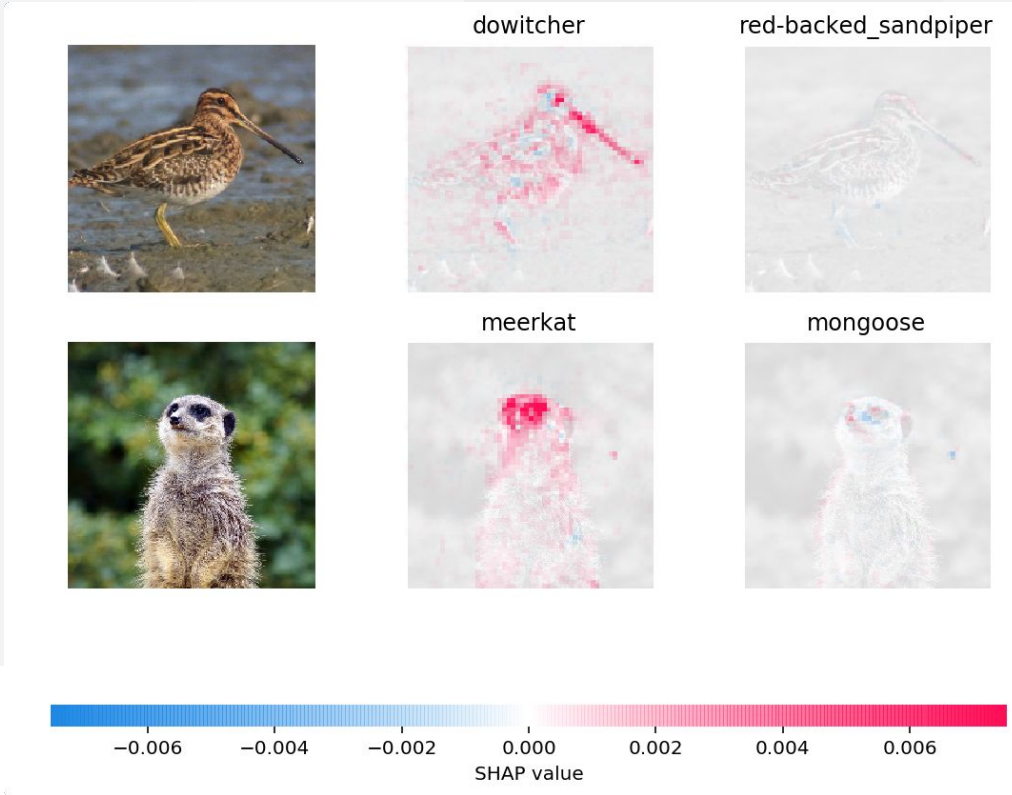
(target ~50nt)



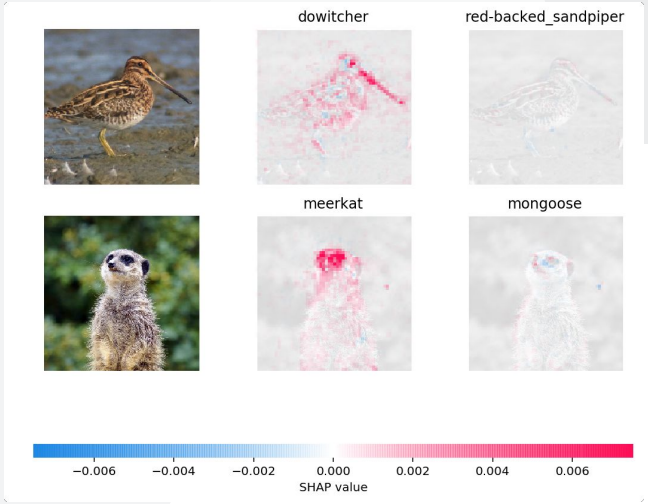

A–T

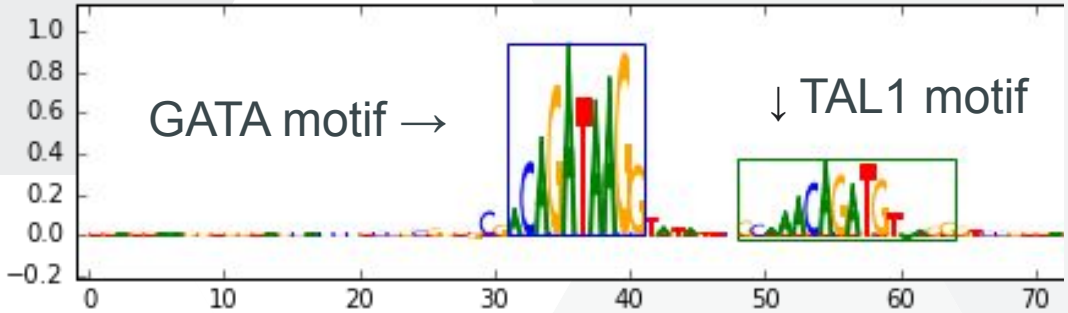
G–C

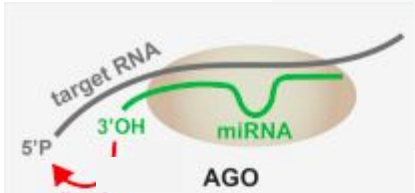
# Model interpretation - SHAP values



# Model interpretation - SHAP values




How to visualize interaction between sequences

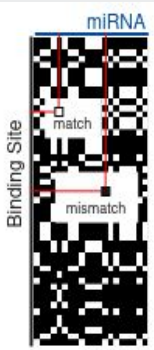
?

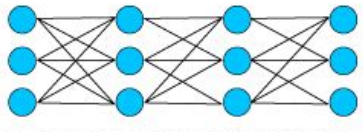
CEITEC

# miRBind model - interpretation

miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT



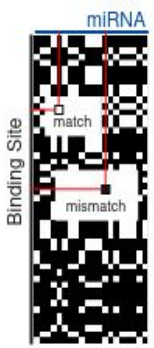
CNN

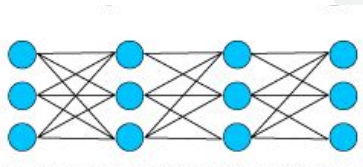
Binding site

# miRBind model - interpretation

miRNA:
TGAGGTAGTAGGTTGTATAG

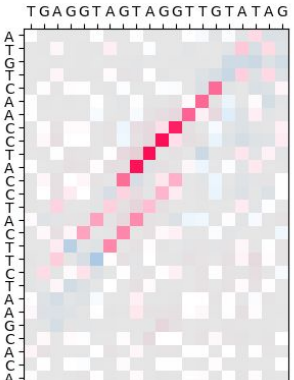
Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT



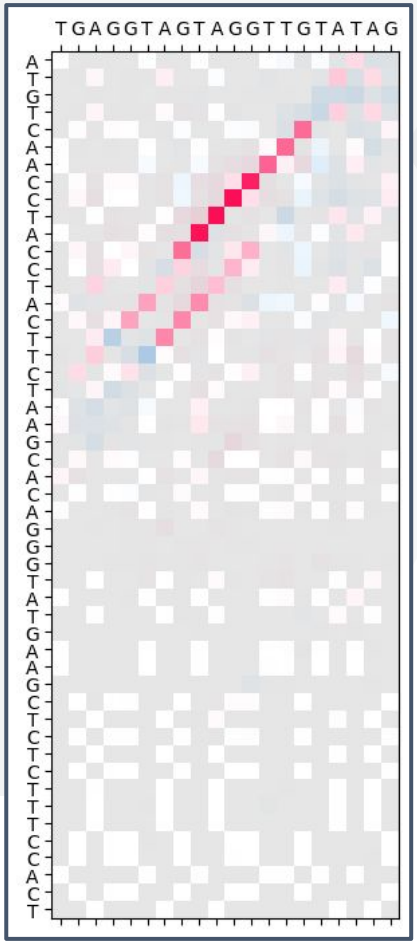



CNN

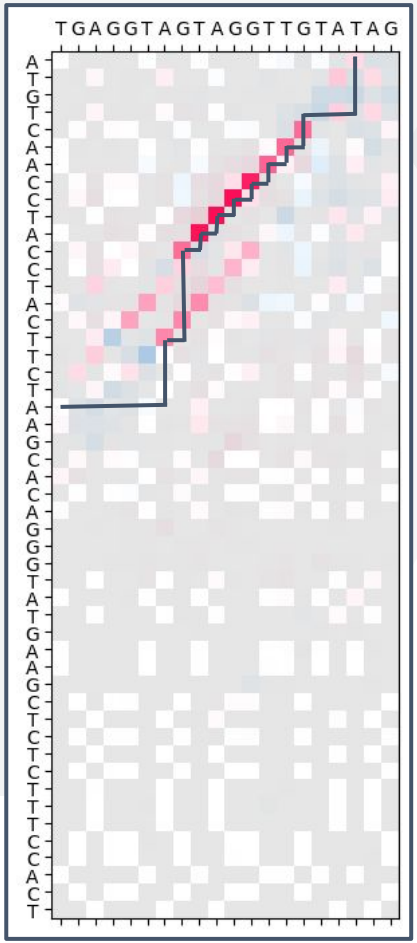
Binding site



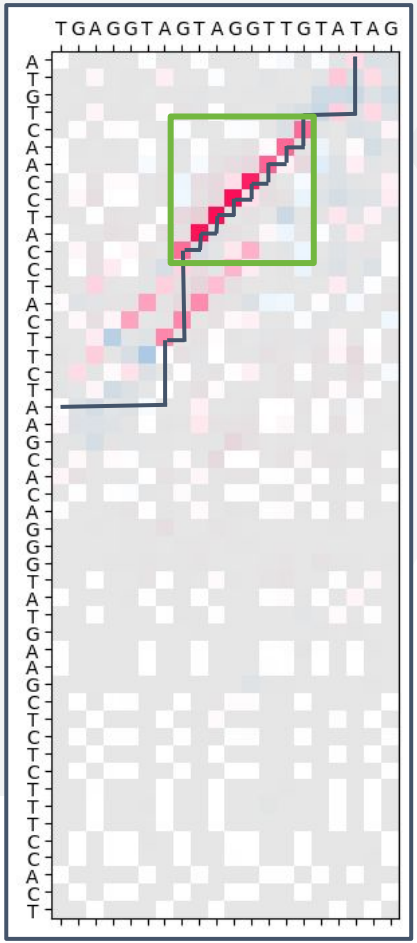

CEITEC


TGAGGTAGTAGGTTGTATAG
ATGTCAAACCCTACCCTTACACTTTCCTAAAGCACACAGGGGTATATGAAGCCTCTCTTTTCCCACT

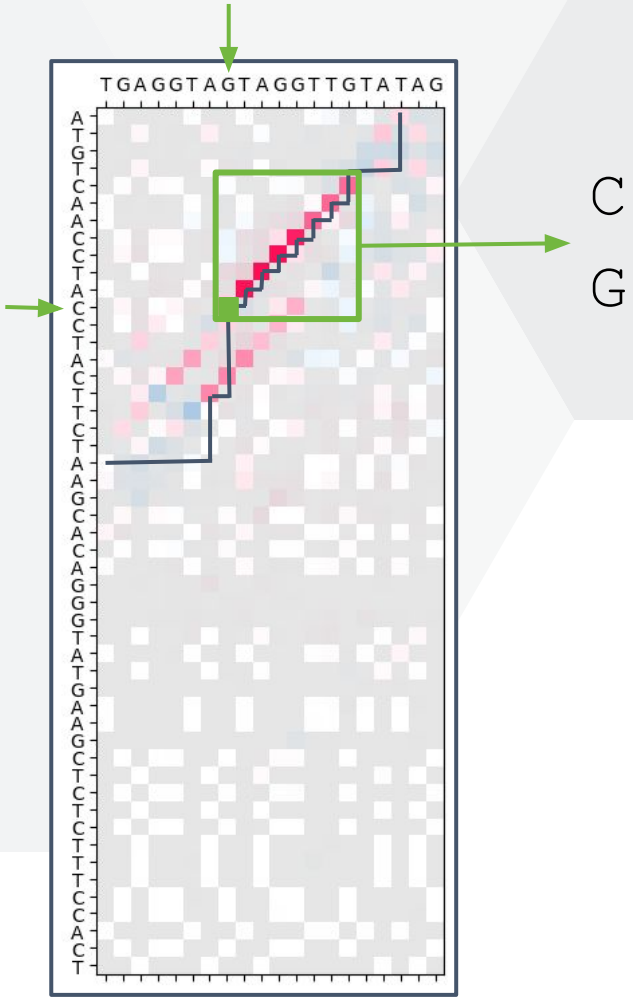
CEITEC


TGAGGTAGTAGGTTGTATAG
ATGTCAACCCTACCCTACTTCTAAAGCACACAGGGTATGAAGCCTCTCTTTCCCACT

CEITEC

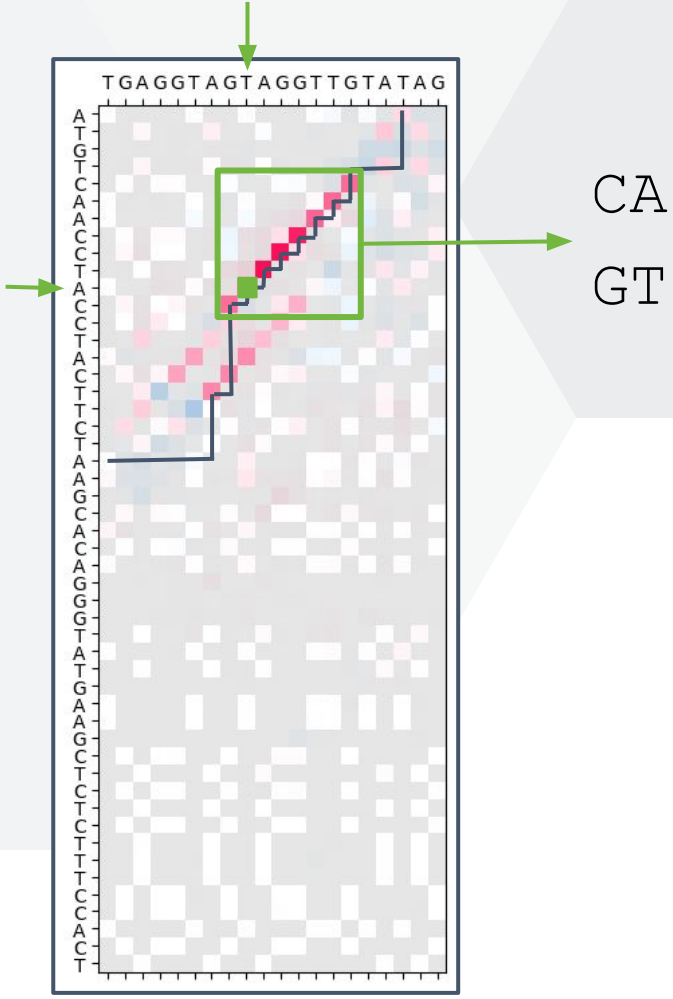

TGAGGTAGTAGGTTGTATAG
ATGTCAACCCTACCCTACTTCTAAAGCACACAGGGTATGAAGCCTCTCTTTCCACT

CEITEC


TGAGGTAGTAGGTTGTATAG
ATGTCAACACCCTACCCTACTTCTAAAGCACACAGGGTATGAAGCCTCTCTTTCCACT

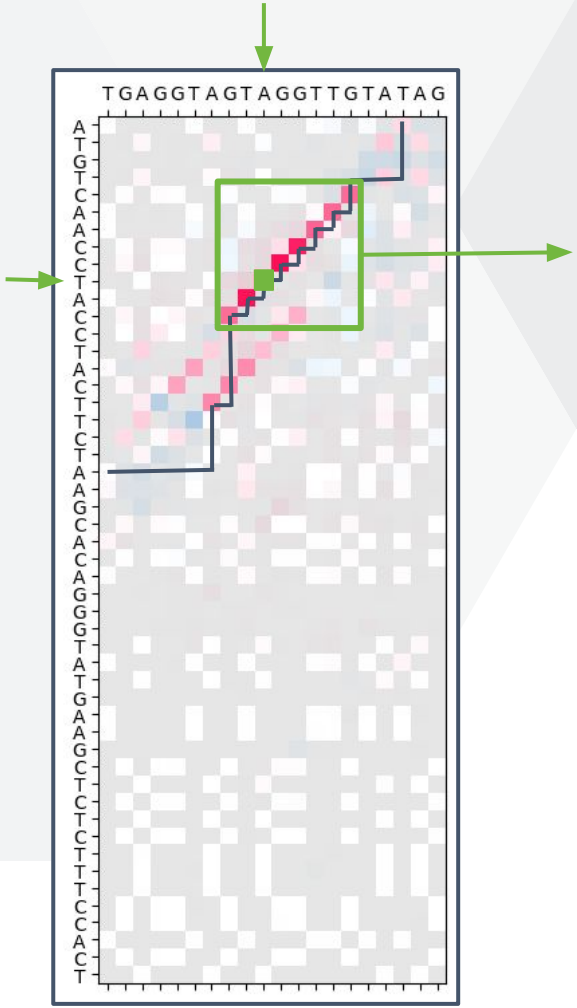
C

G


TGAGGTAGTAGGTTGTATAG
ATGTCAACACCCTACCCTACTTCTAAAGCACACAGGGGTATGAAGCCTCTCTTTTCCACT

CA

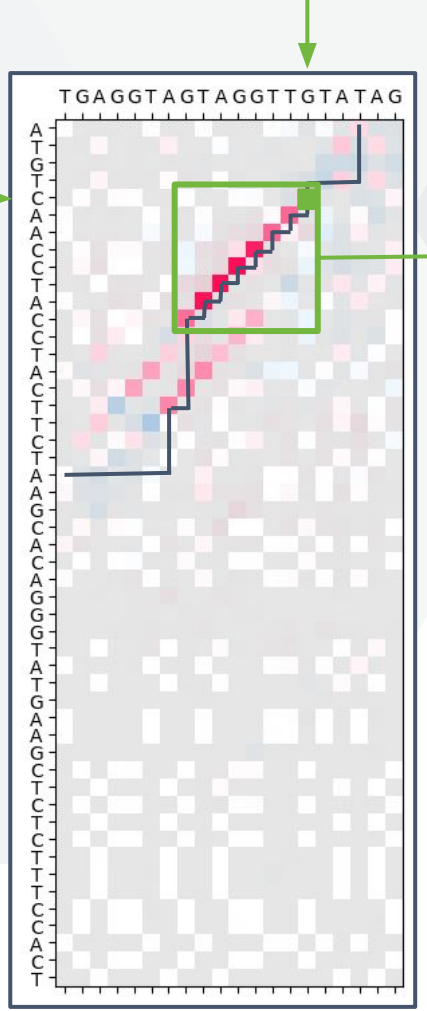
GT




CAT

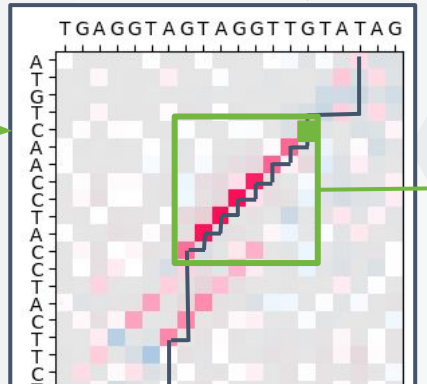
GTA

CEITEC




CATCCAAC

GTAGGTTG

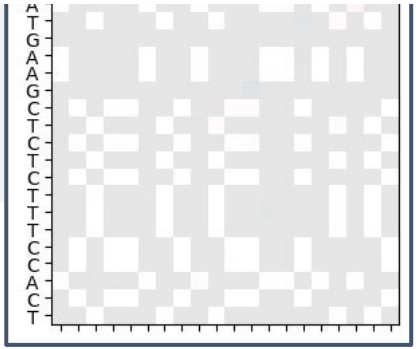



CATCCAAC

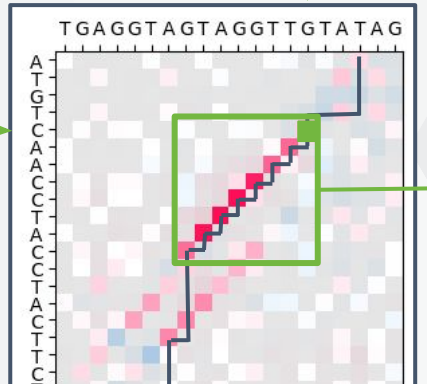
GTAGGTTG

TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-

-----------------------TGAGGTA-GTAGGTTGTATAG



CEITEC

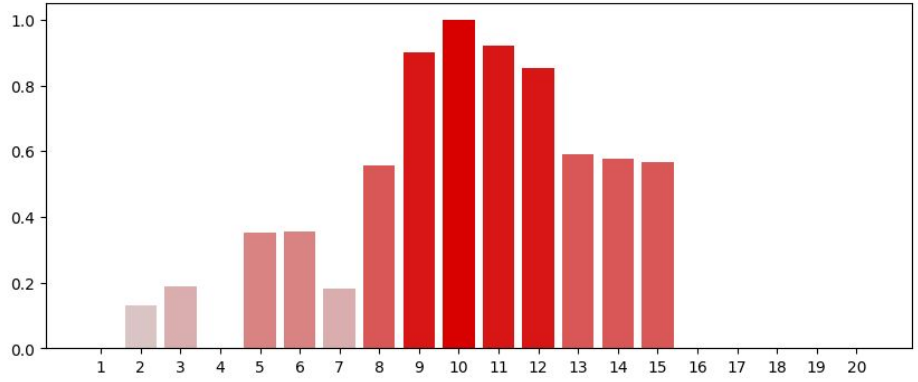


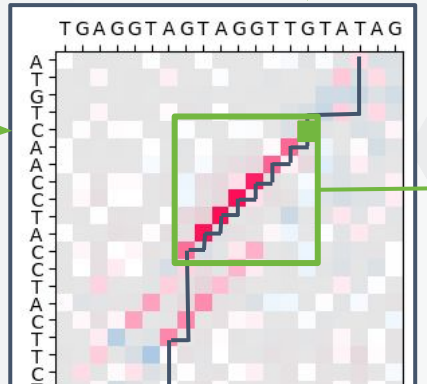

CATCCAAC

GTAGGTTG

TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-

TGAGGTA-GTAGGTTGTATAG



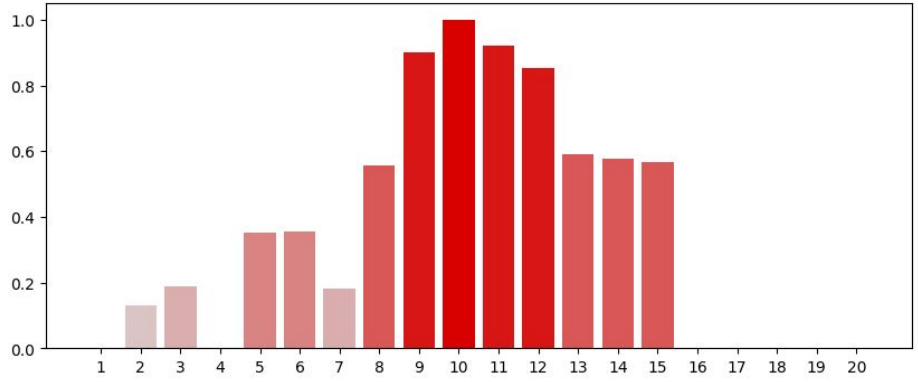


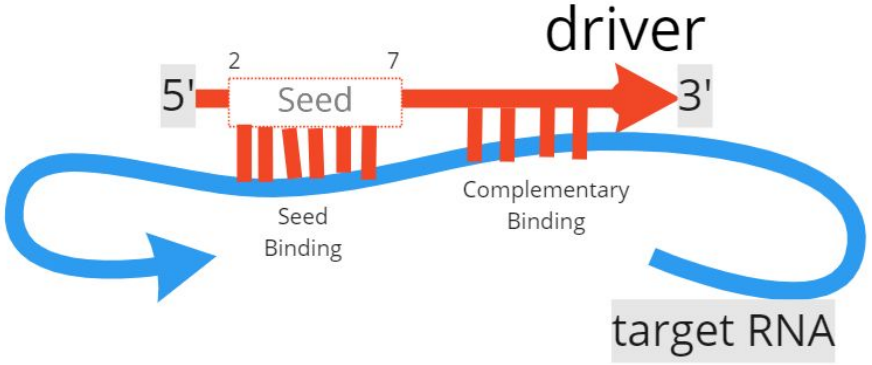

CATCCAAC

GTAGGTTG

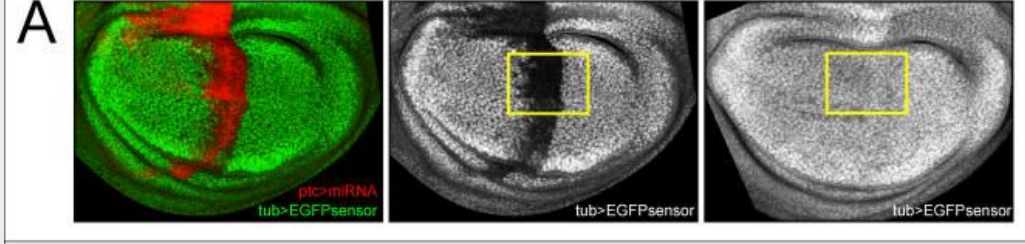
TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-

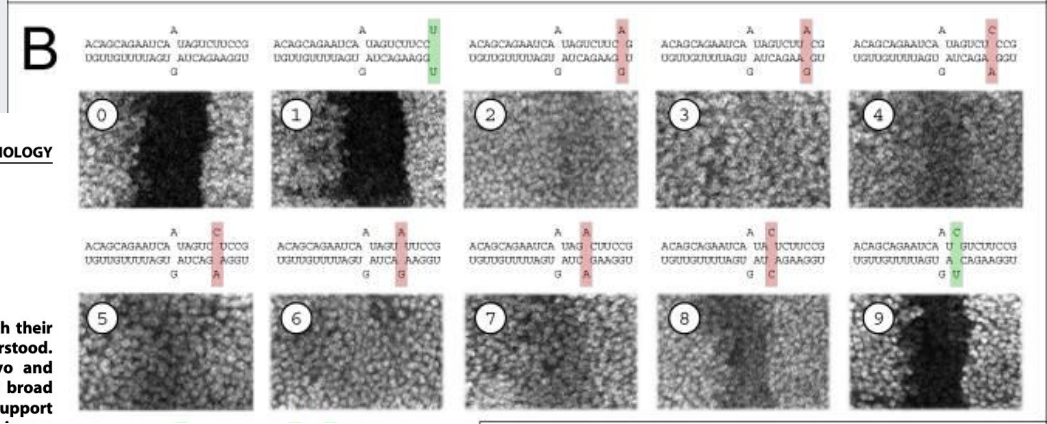
TGAGGTA-GTAGGTTGTATAG



# Mutagenesis experiment




*Open access, freely available online* PLOS BIOLOGY

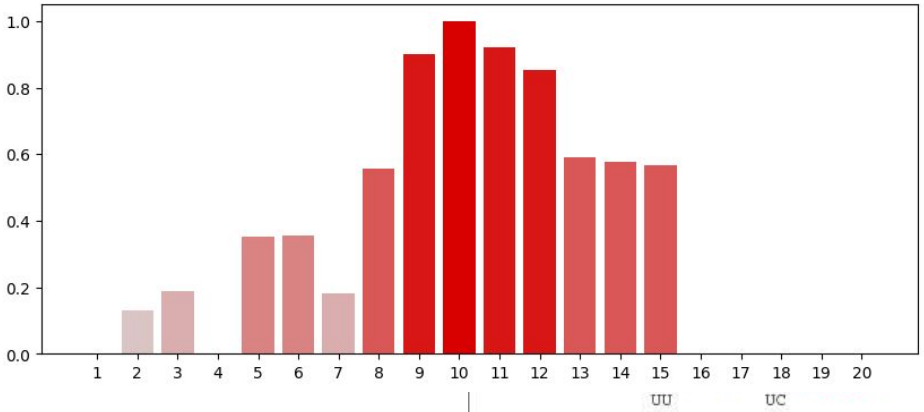
## Principles of MicroRNA–Target Recognition

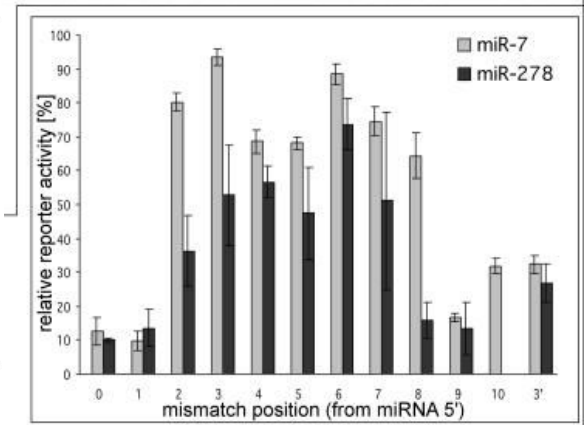
**Julius Brennecke, Alexander Stark, Robert B. Russell, Stephen M. Cohen***

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.**

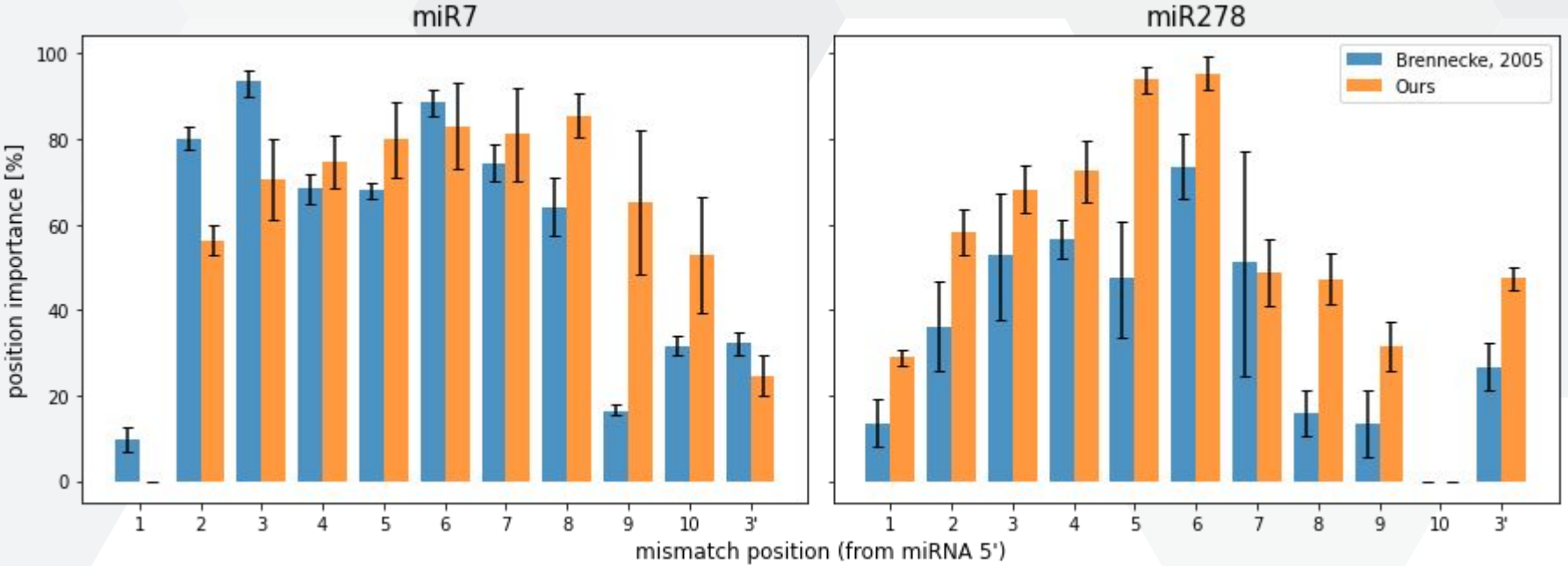
Citation: Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of r…




CEITEC

# Verification - correlation with in vitro experiment




| miRNA | miR-7 | miR-278 |
|---|---|---|
| correlation | 0.59 | 0.85 |

CEITEC

# Functional MicroRNA Targeting

Simple miRNA-mRNA binding model

How will the amount of the products (proteins) of a gene change if a certain miRNA is introduced into the environment in larger quantities?

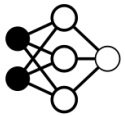
# Task overview


ucagcauagcuacgacguc miRNA, ~20nt long

**Search for binding.**
**If binds → suppress the mRNA.**

auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…

Messenger RNA, 100s – 100,000s nt long

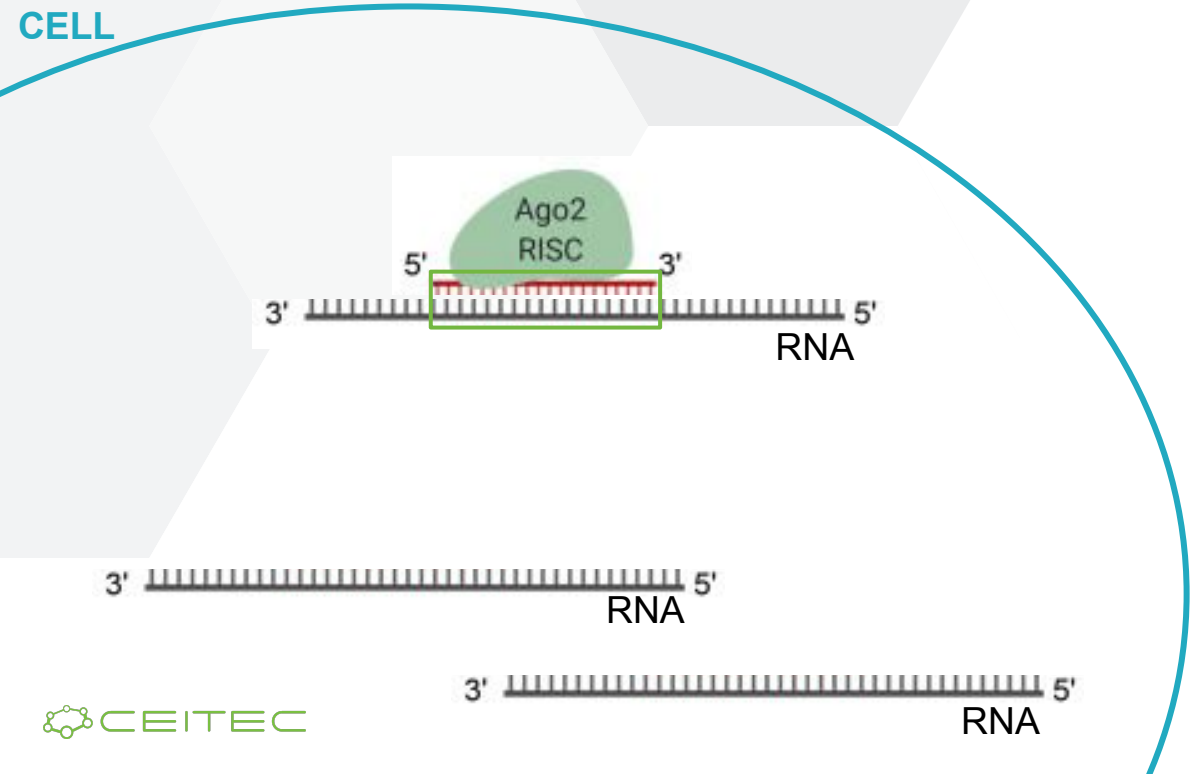


regression

How much less protein products will we get?
(in comparison to a normal cell); Approximate range <0, -2>

CEITEC

# RISC (RNA-induced silencing complex)

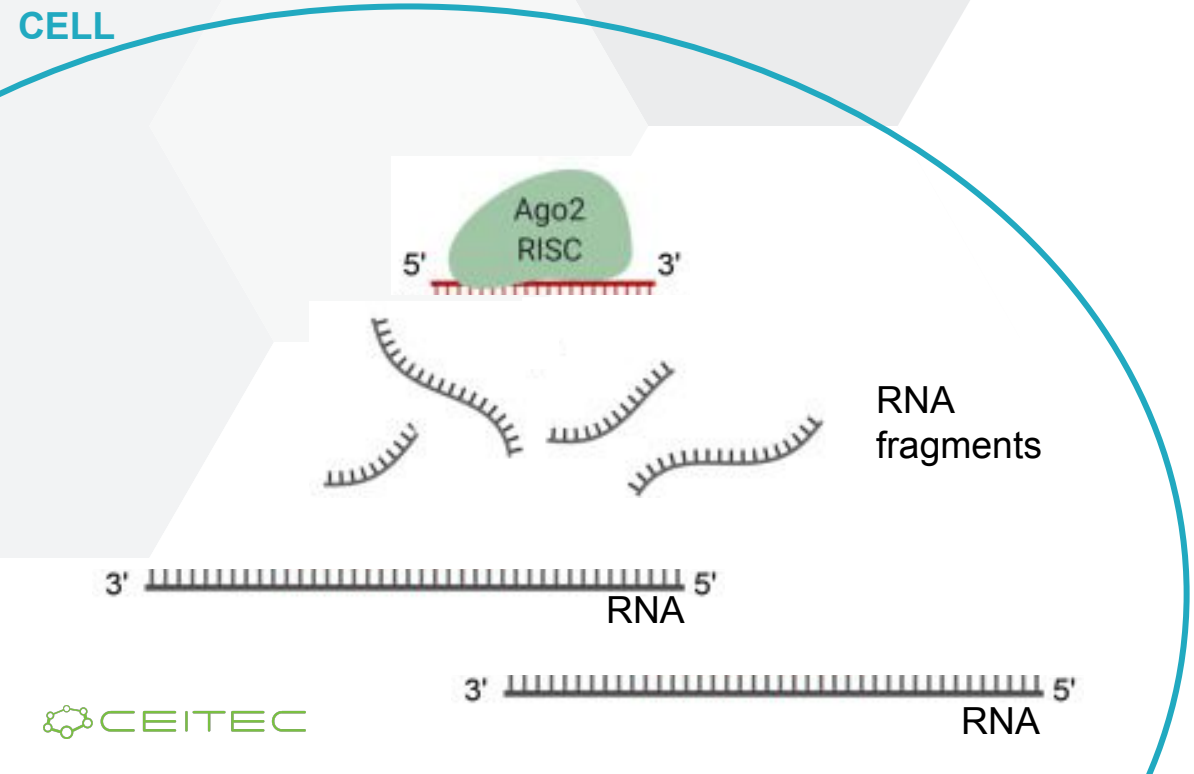



A–U

G–C

CEITEC

# RISC (RNA-induced silencing complex)



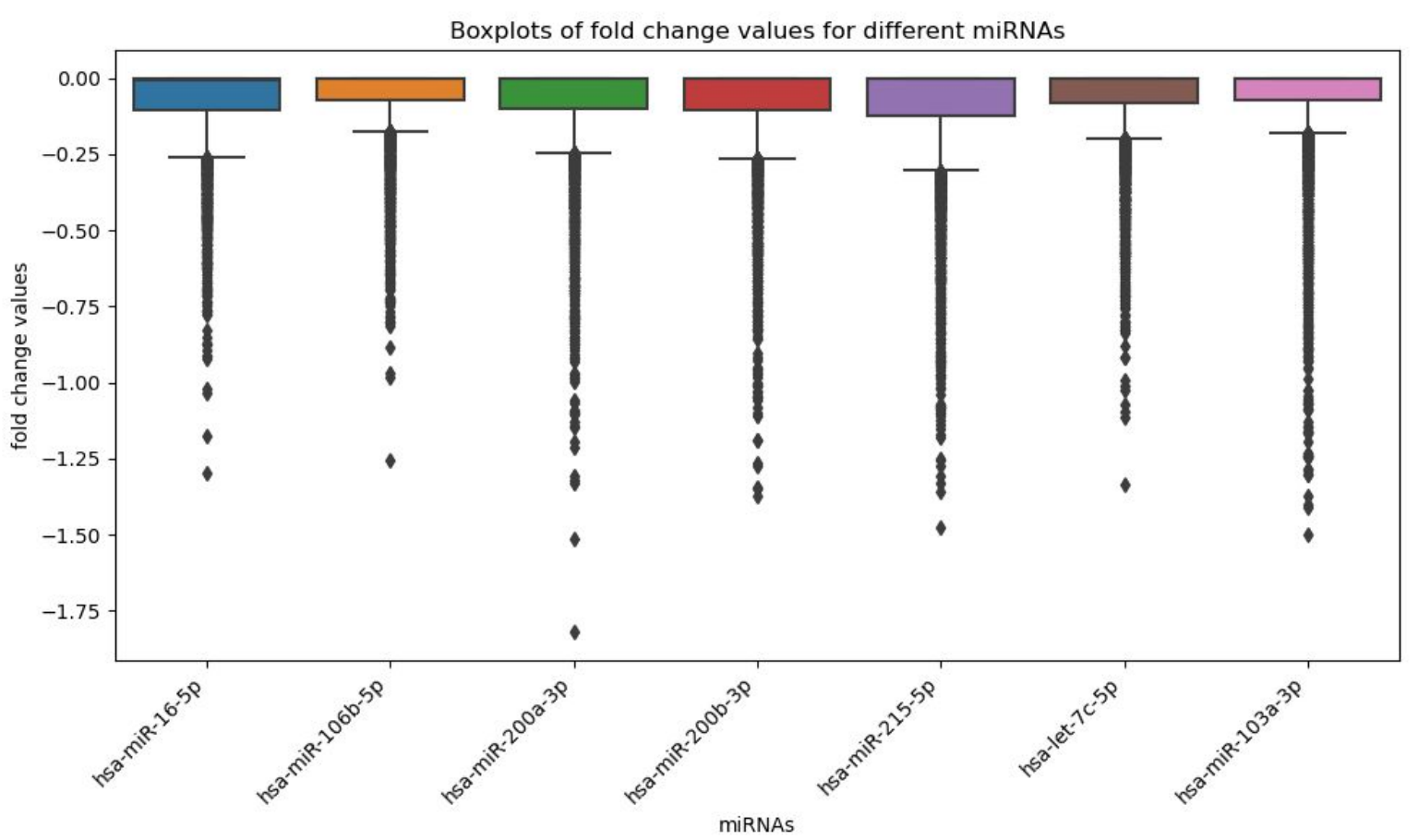

A–U

G–C

CEITEC

# Dataset - labels

-




CEITEC

# Dataset - labels

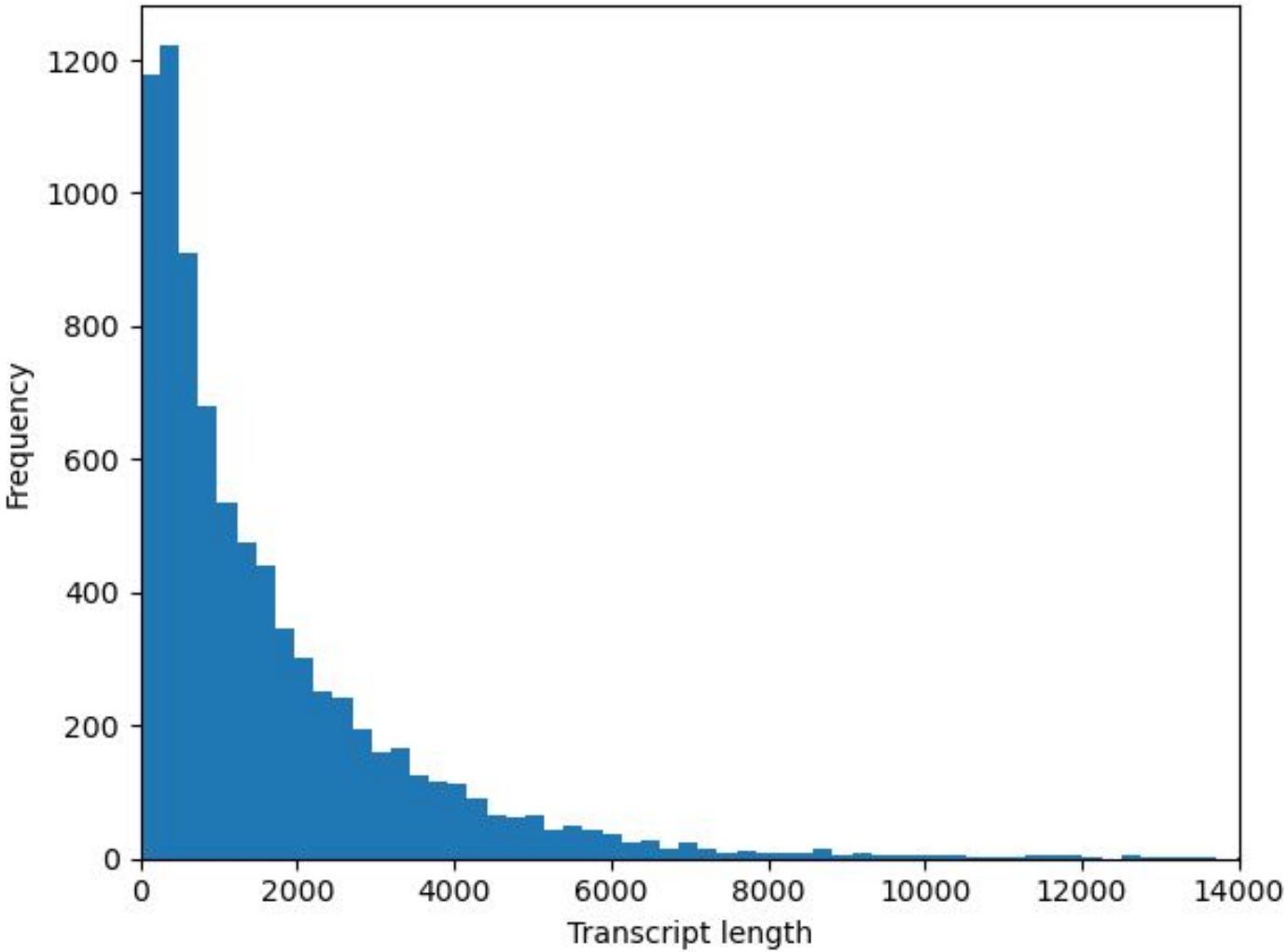
| microRNA | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| hsa-miR-16-5p | 7915 | -0.072 | 0.122 | -1.297 | -0.104 | -0.004 | 0 | 0 |
| hsa-miR-106b-5p | 7902 | -0.056 | 0.109 | -1.255 | -0.07 | 0 | 0 | 0 |
| hsa-miR-200a-3p | 7934 | -0.075 | 0.144 | -1.82 | -0.098 | 0 | 0 | 0 |
| hsa-miR-200b-3p | 7966 | -0.077 | 0.139 | -1.372 | -0.105 | 0 | 0 | 0 |
| hsa-miR-215-5p | 7976 | -0.089 | 0.164 | -1.477 | -0.122 | 0 | 0 | 0 |
| hsa-let-7c-5p | 8002 | -0.063 | 0.119 | -1.334 | -0.079 | 0 | 0 | 0 |
| hsa-miR-103a-3p | 7489 | -0.069 | 0.152 | -1.498 | -0.072 | 0 | 0 | 0 |
| average | 7883.429 | -0.07158 | 0.135495 | -1.43614 | -0.09286 | -0.00057 | 0 | 0 |

f.e. hsa-let-7c-5p train:4046 test:2050

-- because removing transcripts without signal

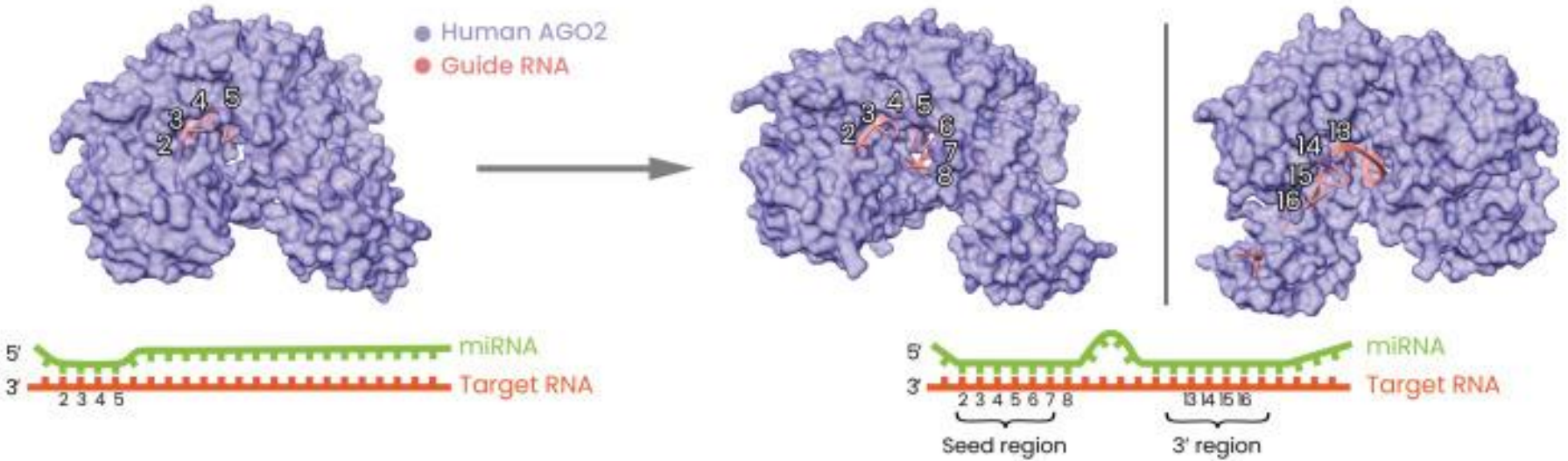
# Dataset - inputs

Lengths of transcript sequences



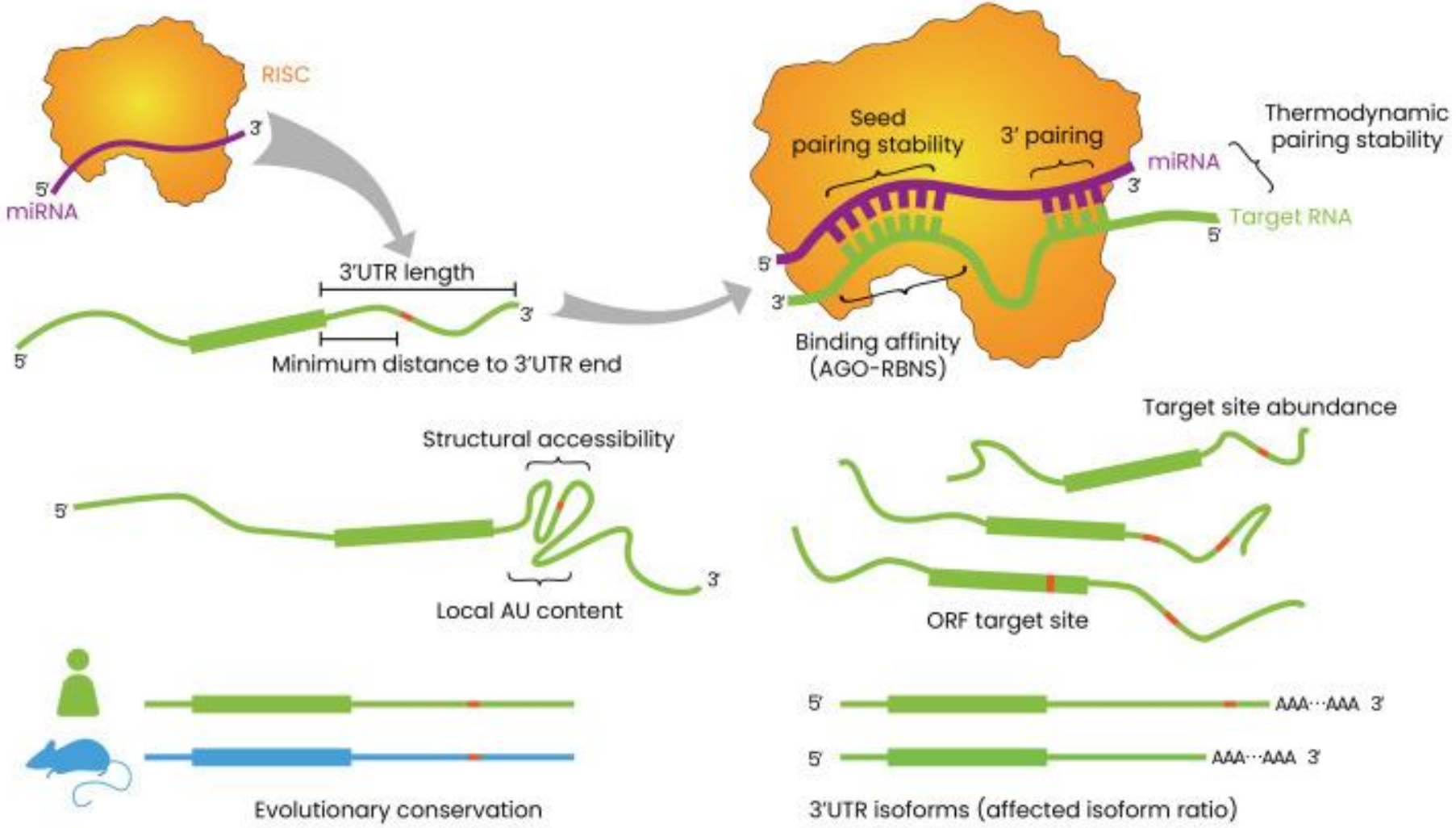

CEITEC

# State-of-the-art so far - manual feature extraction




| Canonical sites | | 1 2 3 4 5 6 7 8 9 |
|---|---|---|
| | 8-mer | AOOOOOOON |
| | 7-mer m8 | BOOOOOOON |
| | 7-mer A1 | AOOOOOOØN |
| | 6-mer | BOOOOOOØN |

| Noncanonical sites | | 1 2 3 4 5 6 7 8 9 |
|---|---|---|
| | 6-mer A1 | AOOOOOØØN |
| | offset 7-mer | BØOOOOOOO |
| | offset 6-mer | BØOOOOOON |
| | CDNST 1 | NNØOOOOOB |
| | CDNST 2 | NNOOØOOOA |
| | CDNST 3 | OOOØOØOØN |
| | CDNST 4 | NOØØØOOOA |


RISC
3'
5'
miRNA
3'UTR length
3'
5'
Minimum distance to 3'UTR end
Seed
pairing stability
3' pairing
Thermodynamic
pairing stability
miRNA
3'
Target RNA
5'
5'
3'
Binding affinity
(AGO-RBNS)
Structural accessibility
5'
3'
Local AU content
Target site abundance
ORF target site
5'
AAA···AAA 3'
5'
AAA···AAA 3'
Evolutionary conservation
3'UTR isoforms (affected isoform ratio)

## Table 1

A table of representative computational tools for miRNA target prediction and the determinants they use

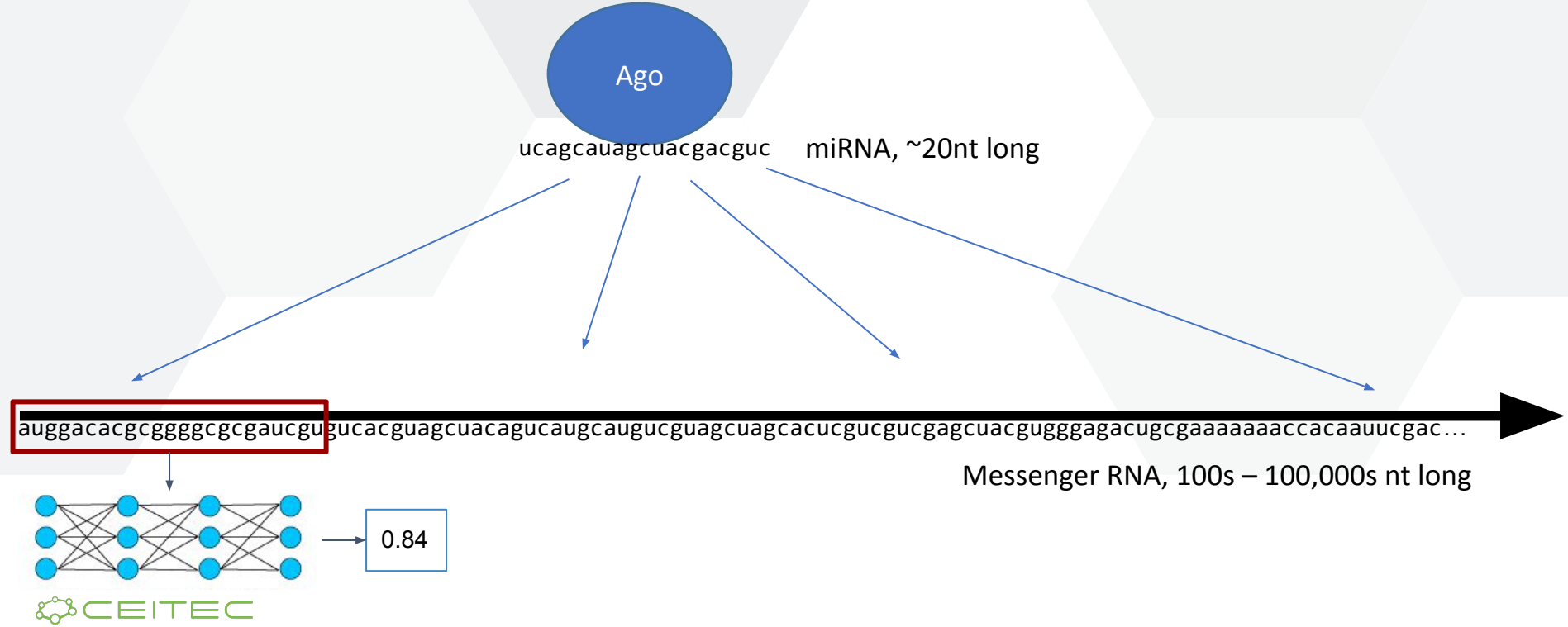
| Model | Seed | TPS | EC | SA | Dist. | AU | Len. | 3Sup. | TA | ORFS |
|---|---|---|---|---|---|---|---|---|---|---|
| TargetScan7 | 0 | SPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8m |
| miRanda-mirSVR | 0 | X | 0 | 0 | 0 | 0 | 0 | 0 | X | X |
| DIANA-microT-CDS | 0 | 0 | 0 | 0 | 0 | 0 | X | X | X | 0 |
| MIRZA-G | 0 | 0 | 0 | 0 | 0 | X | X | X | X | X |
| PITA | Opt. | 0 | X | 0 | X | X | X | X | X | X |
| PicTar | 0 | 0 | 0 | X | X | X | X | X | X | X |
| RNAhybrid | Opt. | 0 | X | X | X | X | X | X | X | X |
| MicroTar | 0 | 0 | X | X | X | X | X | X | X | X |

Open in a separate window

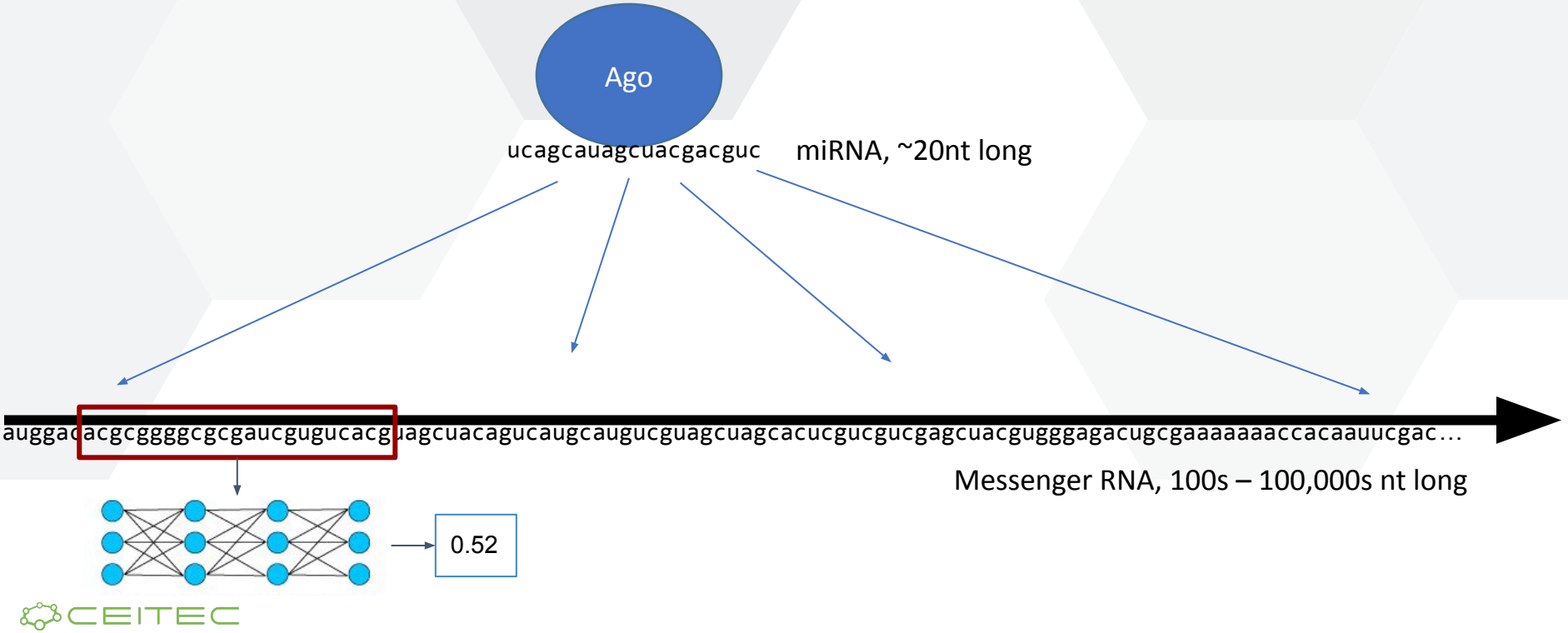
Seed, seed match or site type; TPS, thermodynamic pairing stability; EC, evolutionary conservation; SA, structural accessibility; Dist., distance to 3′UTR ends or relative position of the target sites in the 3′UTR; AU, AU or GC content; Len., length of transcript or UTR; 3Sup., 3′ supplementary pairing; TA, target abundance; ORFS, ORF or CDS sites; Opt., optional; SPS, seed pairing stability; 8m, number of 8-mer sites in the ORF.

# Scanning - prediction only

# Scanning - prediction only

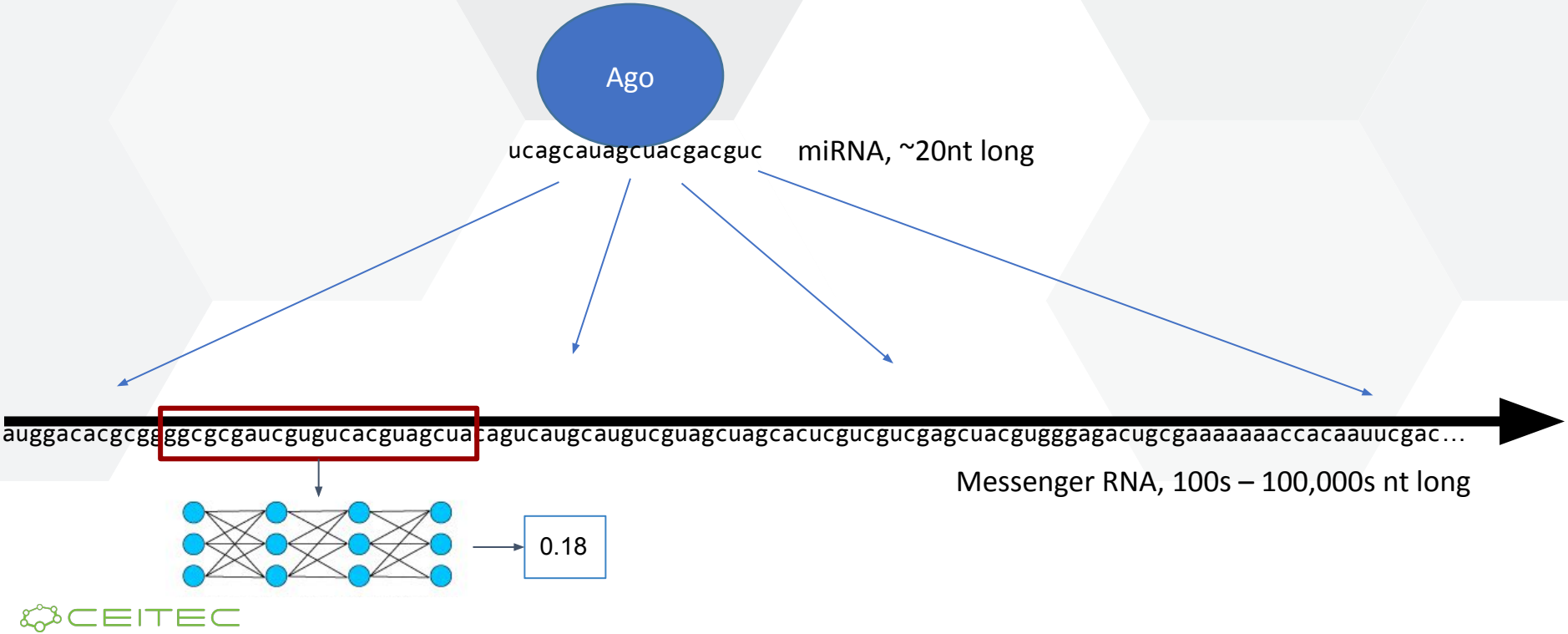


# Scanning - prediction only



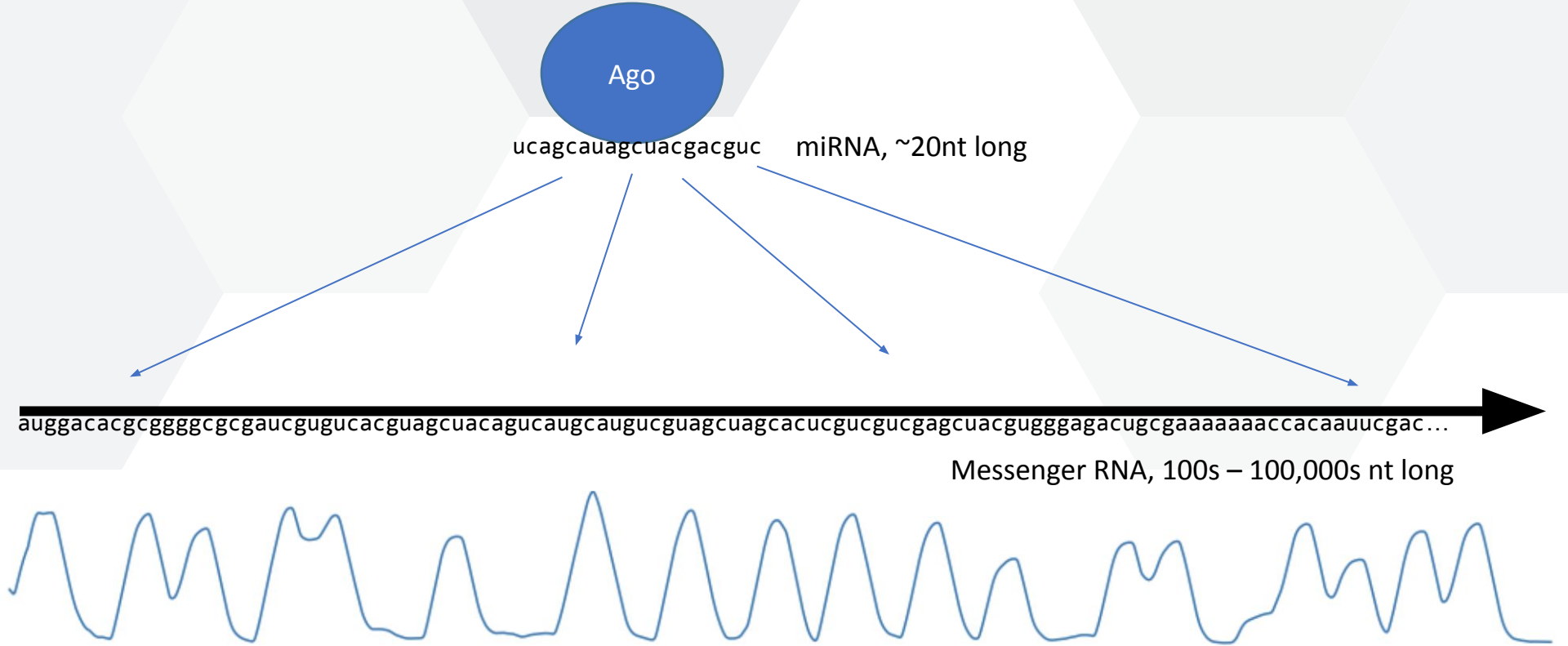

CEITEC

# Scanning - prediction only

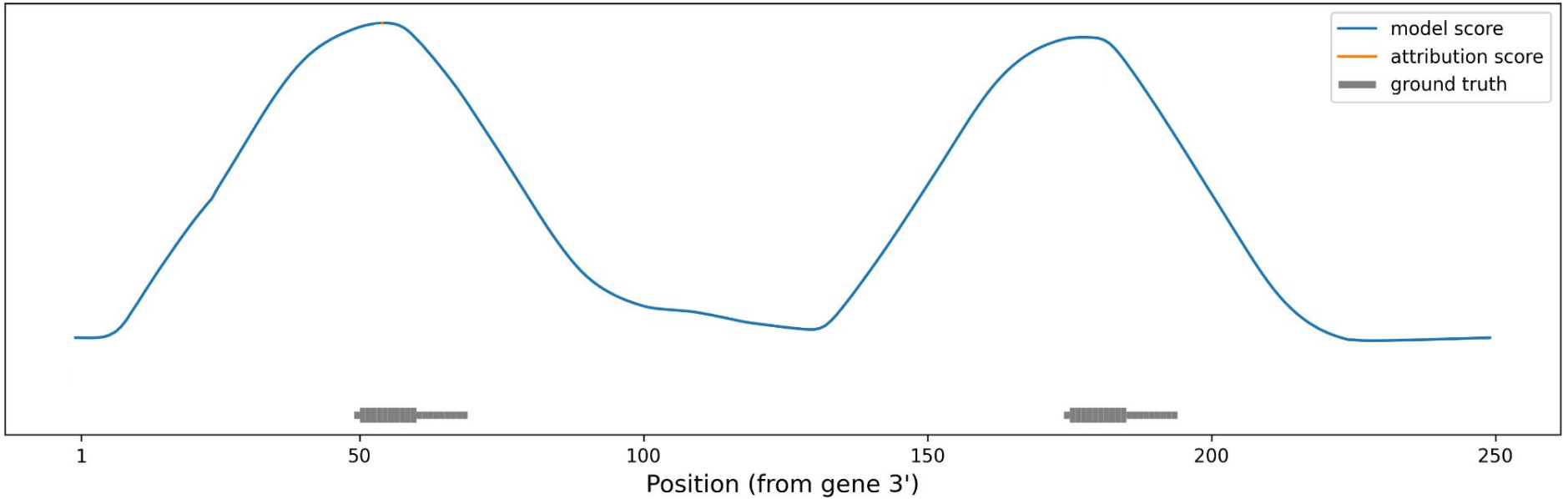



CEITEC

# Scanning - prediction only

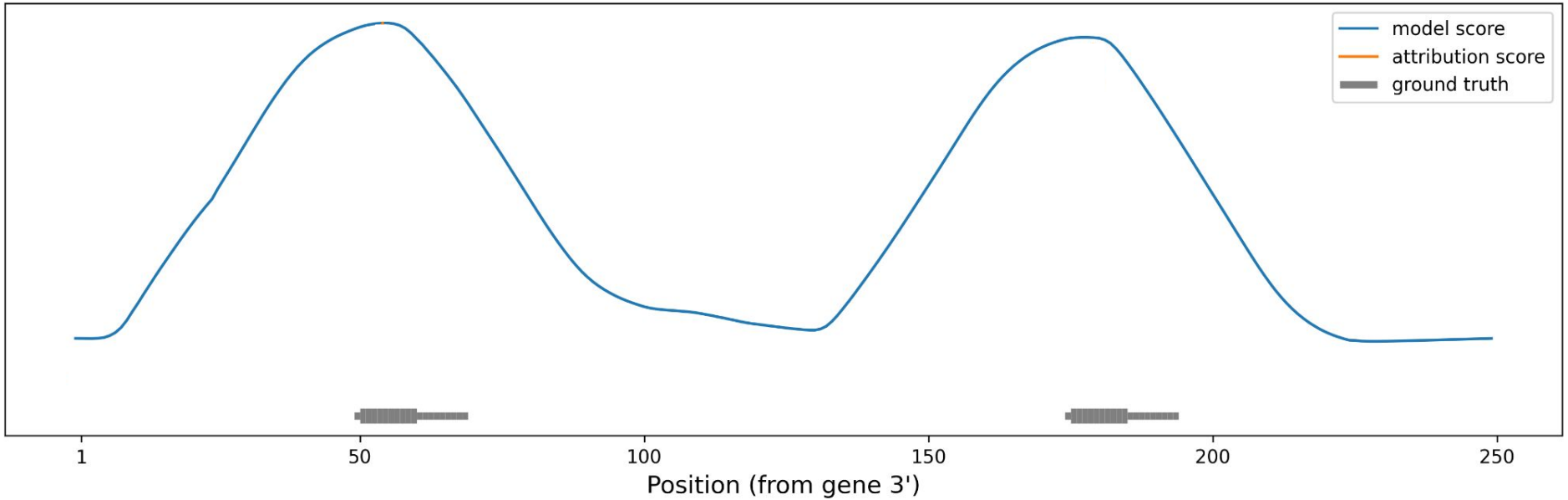


# Scanning - prediction only
# Narrowing the peaks



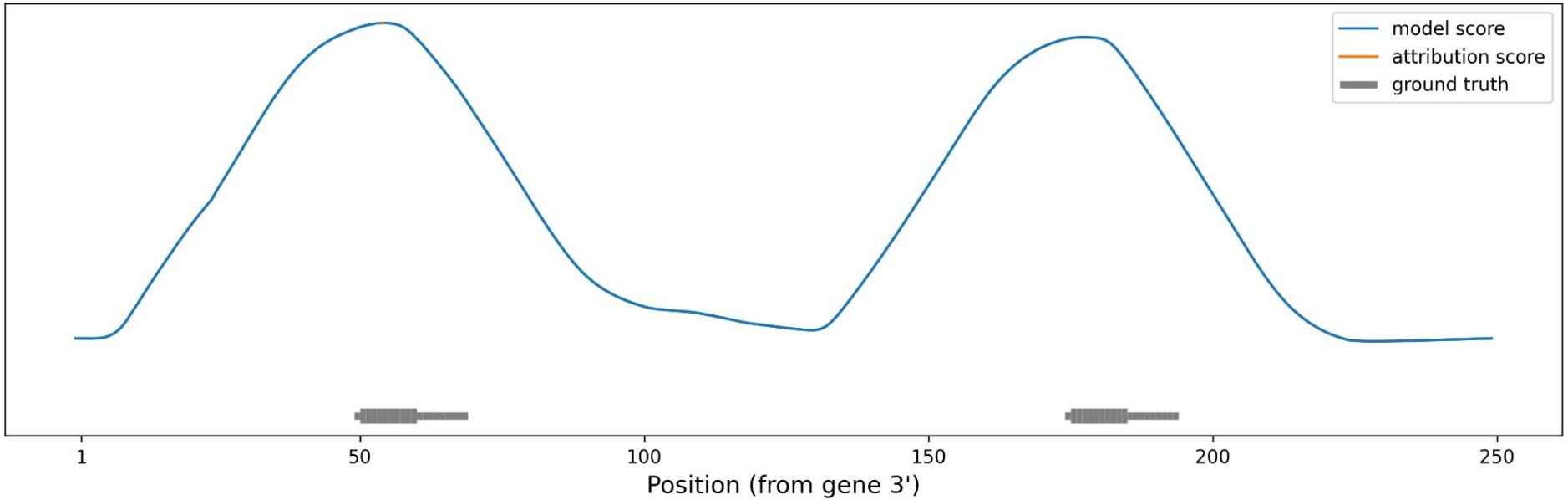

CEITEC

# Scanning - including attribution score
# Narrowing the peaks

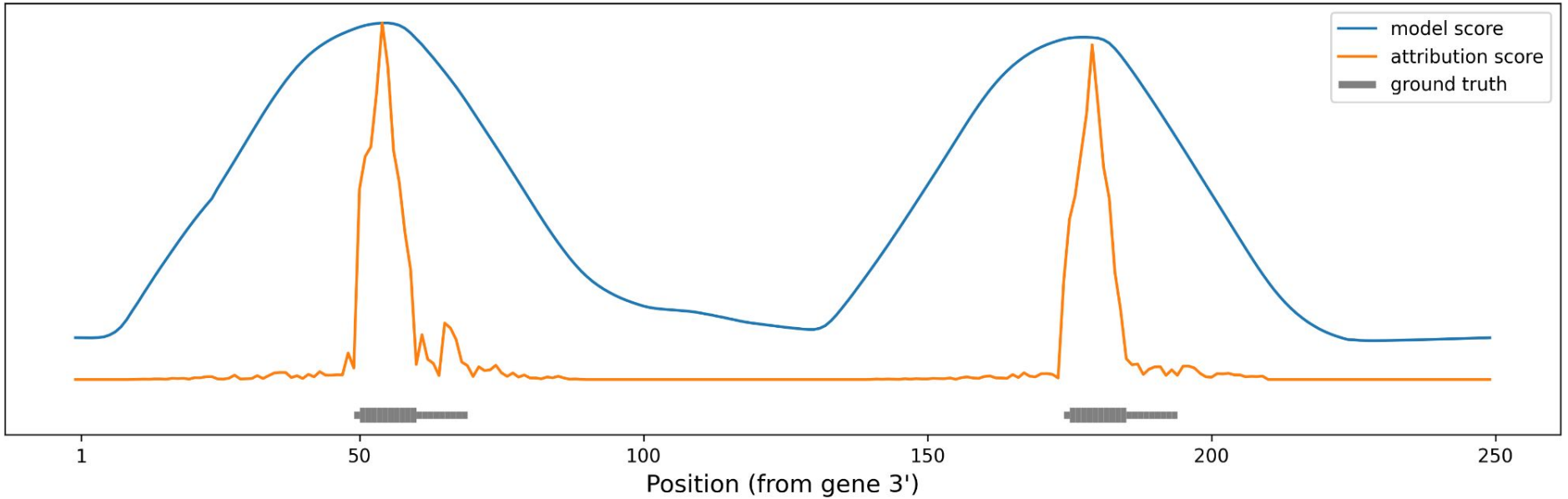

CEITEC

# Scanning - including attribution score
# Narrowing the peaks

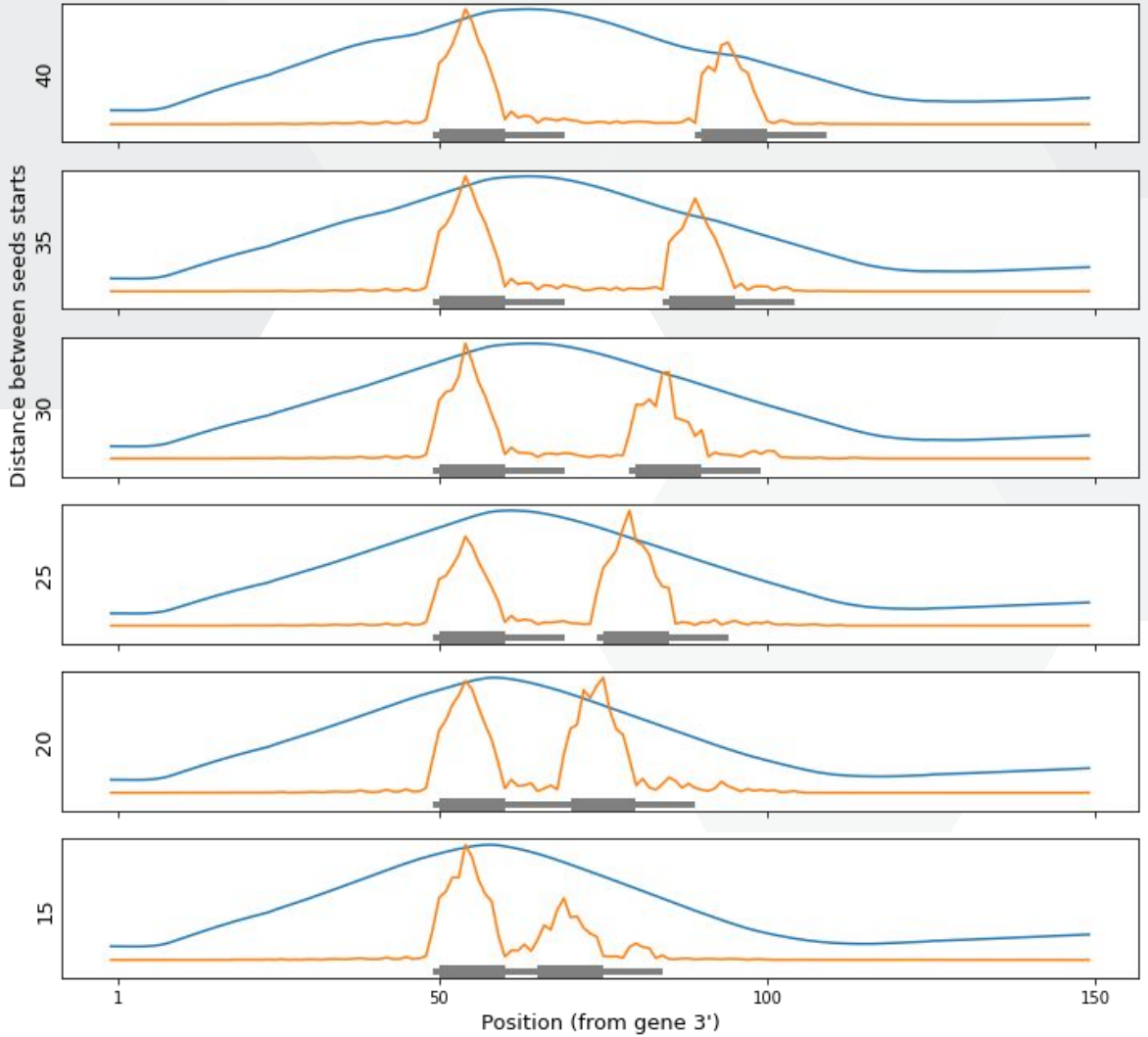

CEITEC

# Scanning - including attribution score
# Narrowing the peaks


CEITEC

# Close by peaks

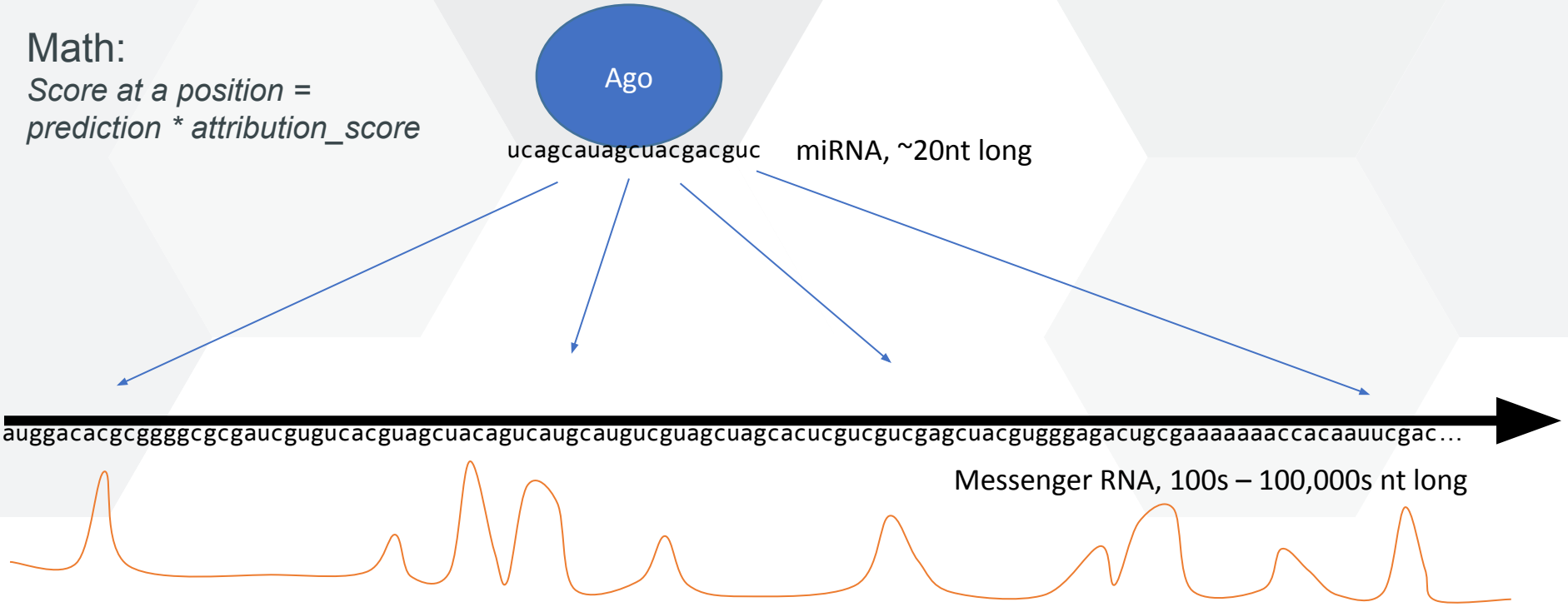



CEITEC

# Scanning - using attribution score

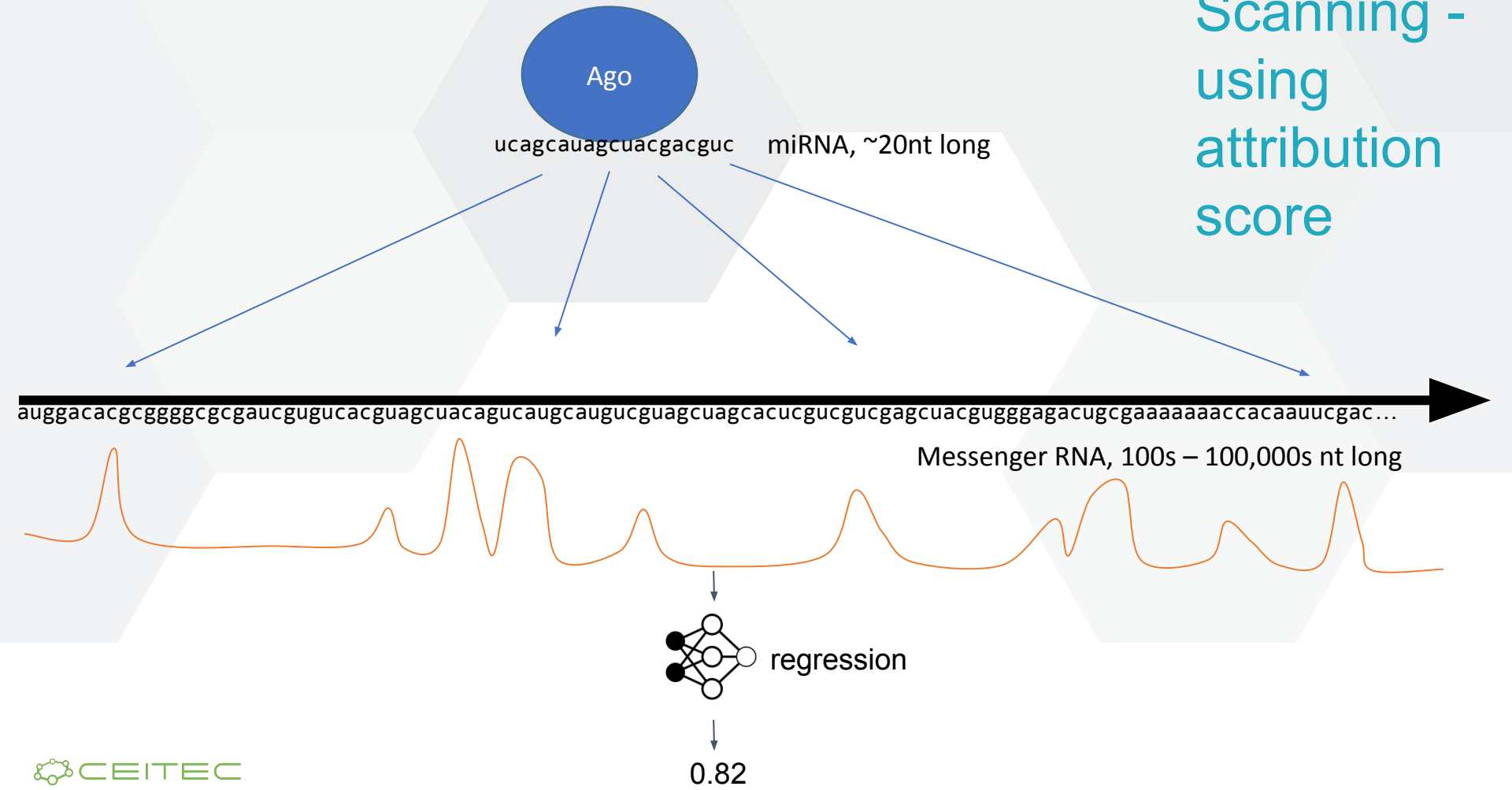
Math:

*Score at a position =*
*prediction * attribution_score*

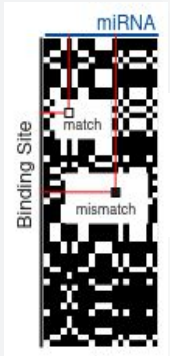



CEITEC

# Scanning - using attribution score




CEITEC




CNN

ucagcauagcuacgacguc

**miRNA, ~20nt long**

**regression model**

**prediction**

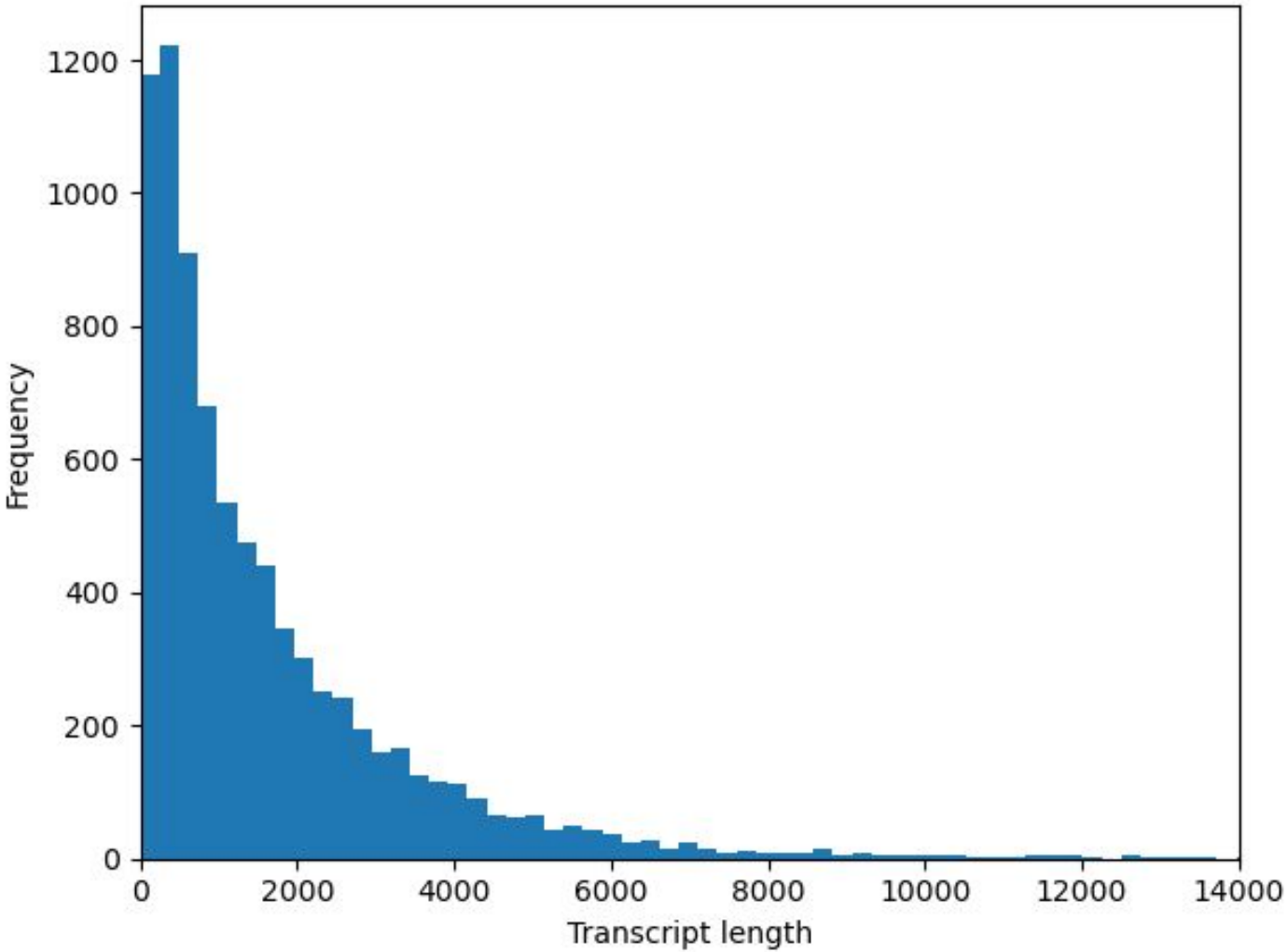
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaaccacaauucgac…

**Messenger RNA, 100s – 100,000s nt long**

# Dataset - inputs

Lengths of transcript sequences

Sequences too long



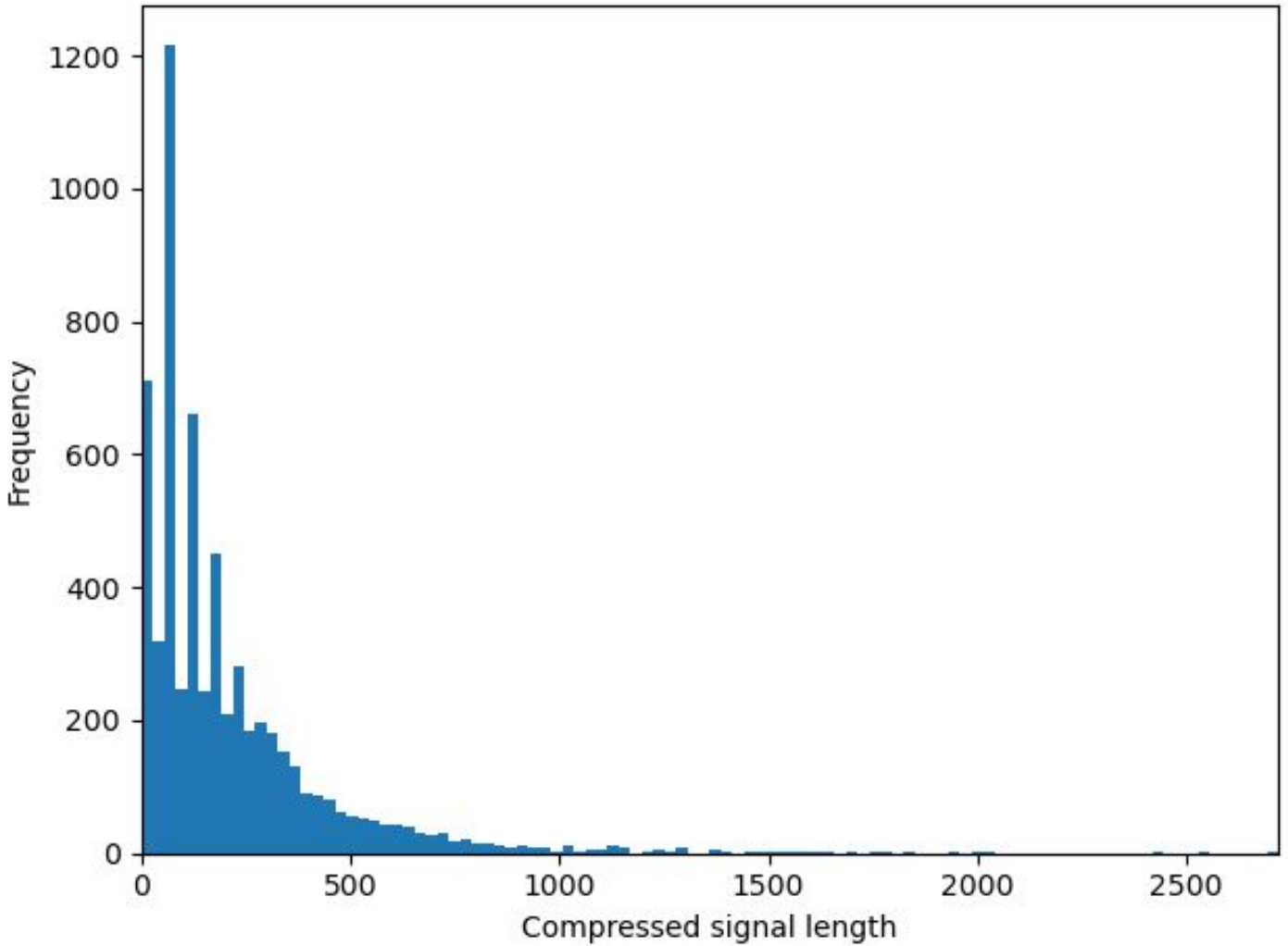

CEITEC

# Compressed inputs

Lengths of signals compressed

**Per transcript**

Longest: 2719




CEITEC

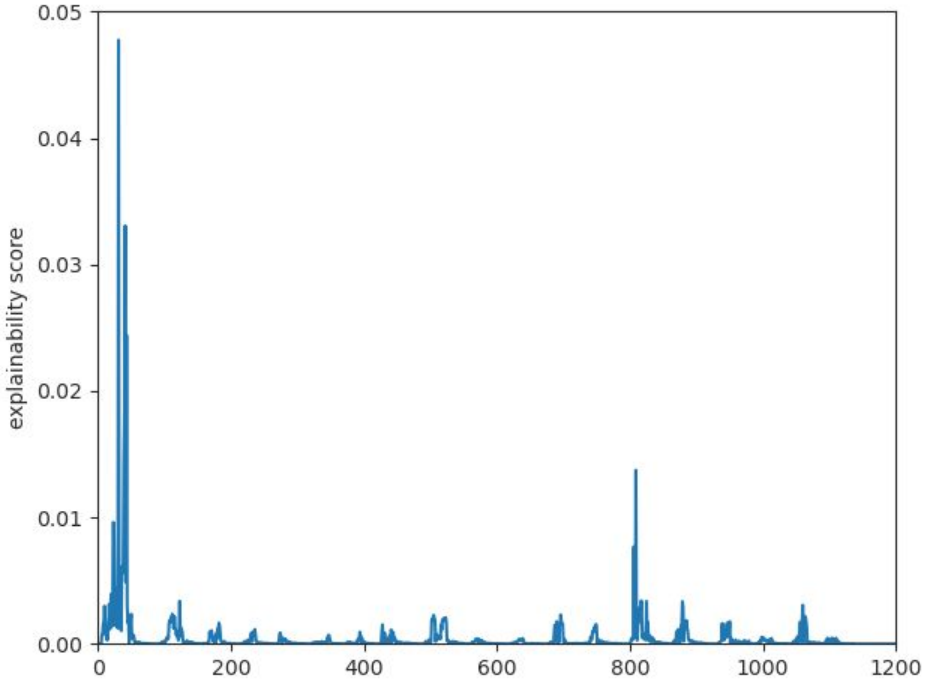
# Signals preprocessing

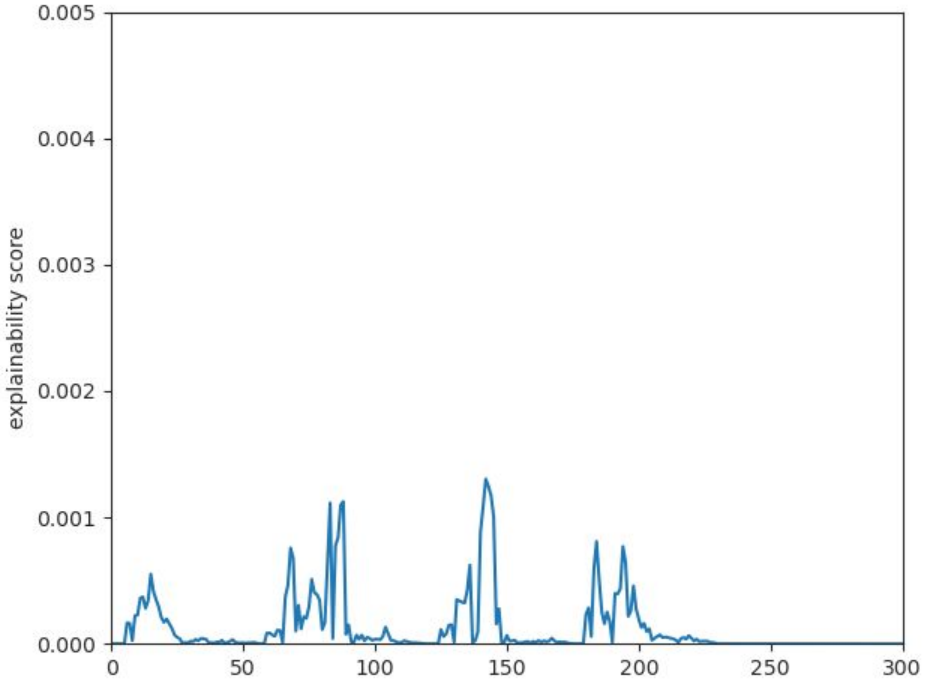
Highly sparse → compression: *(number_of_zeroes % 100) +1*

Normalization to <0.00001, 1>

0 used for padding

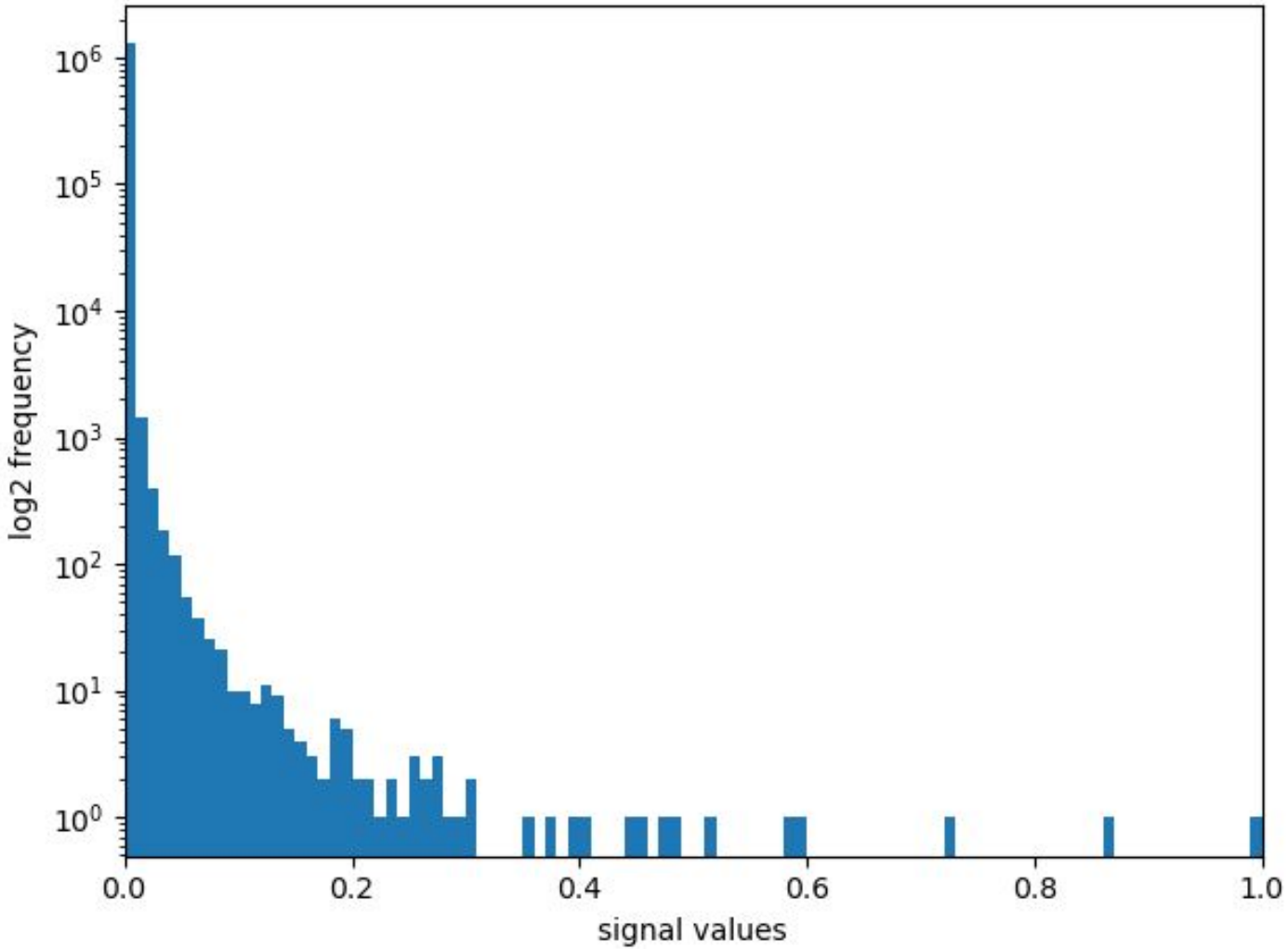
# Signal samples (compressed)



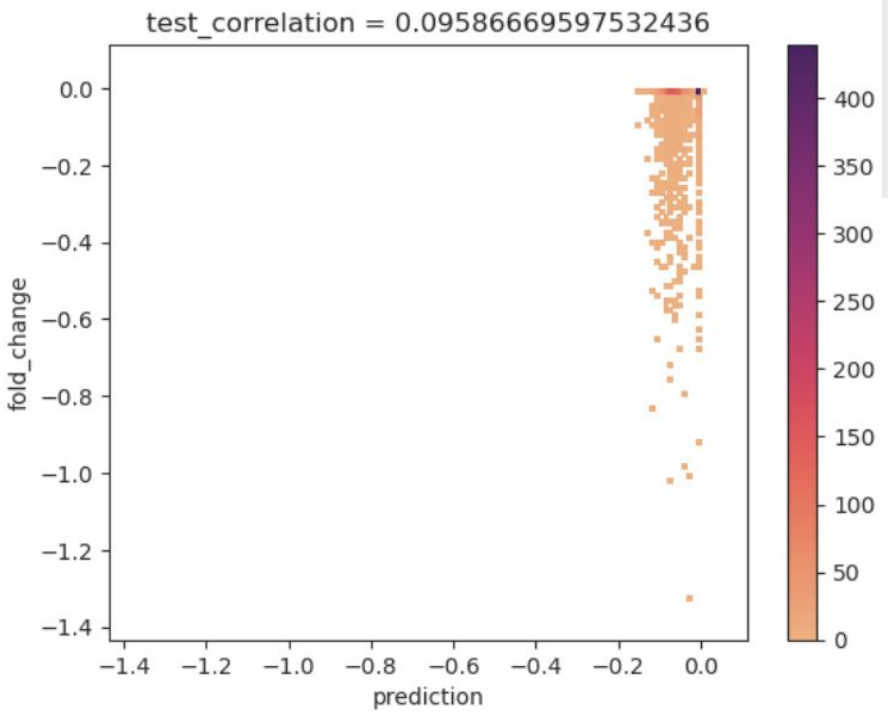



# Signal Values histogram over all samples

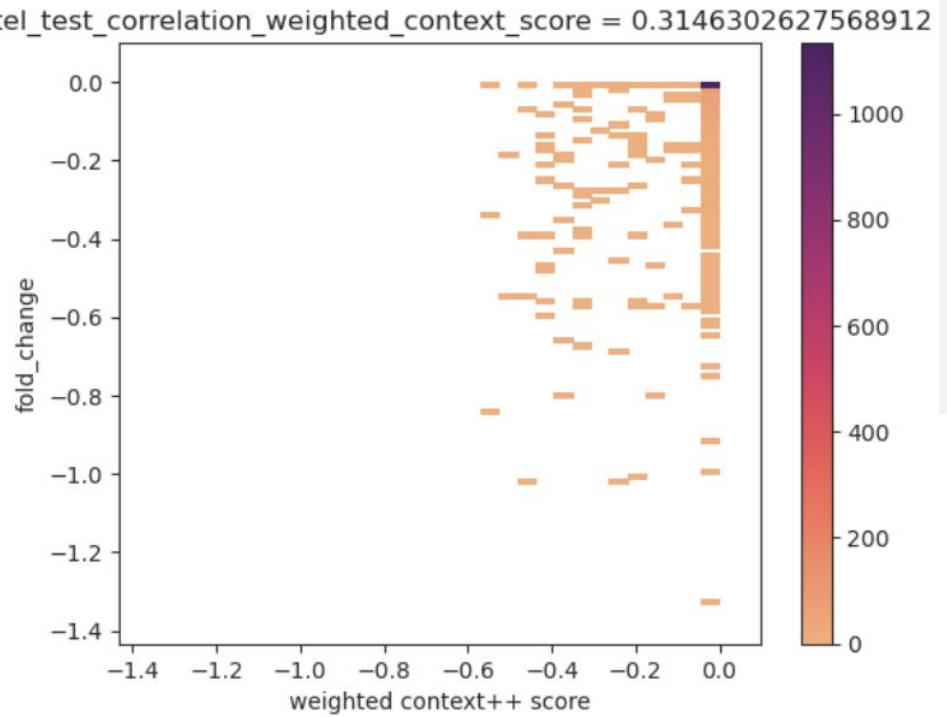
(before padding)



## CNN + RNN + pooling

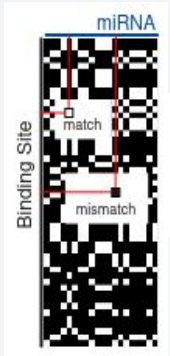


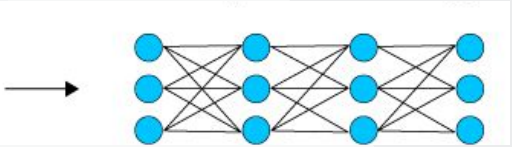

## **State-of-the-art** based on feature selection (TargetScan)



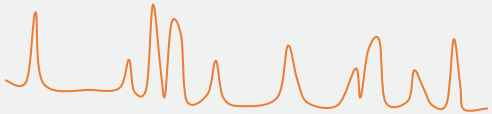

CEITEC

# Summary



CNN



regression model

prediction

ucagcauagcuacgacguc

**miRNA, ~20nt long**

auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…

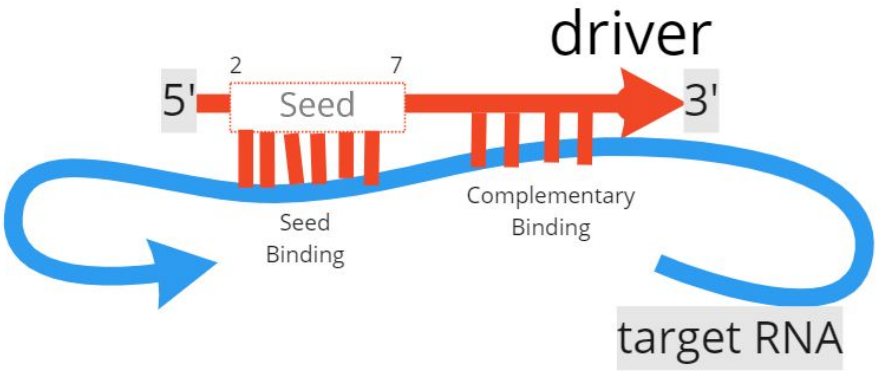
**Messenger RNA, 100s – 100,000s nt long**

# Advantages & Disadvantages of our two-part approach

1. Shields from sequence and overfitting on simple patterns like seed binding
2. Generalizes across miRNAs

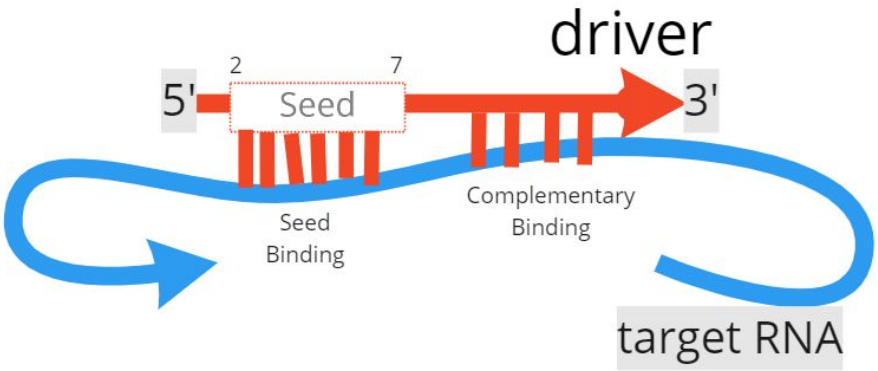
1. First model is not perfect which leads to accumulation of mistakes to the second model
2. Cannot propagate error through second model to the first model

# Summary


driver
2
7
5'
Seed
3'
Seed
Binding
Complementary
Binding
target RNA

**miRNA:**
TGAGGTAGTA
GGTTGTATAG

**Binding site:**
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT

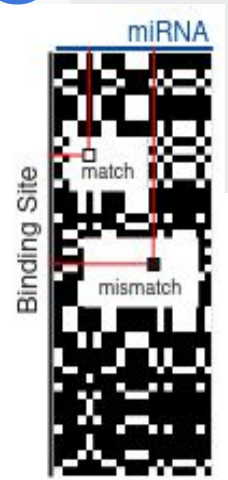


A–T

G–C

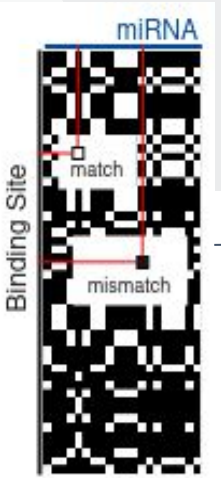
**miRNA:**
TGAGGTAGTA
GGTTGTATAG

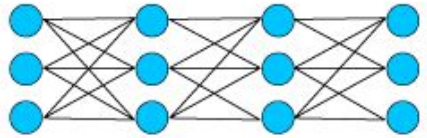
**Binding site:**
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT



**miRNA:**
TGAGGTAGTA
GGTTGTATAG

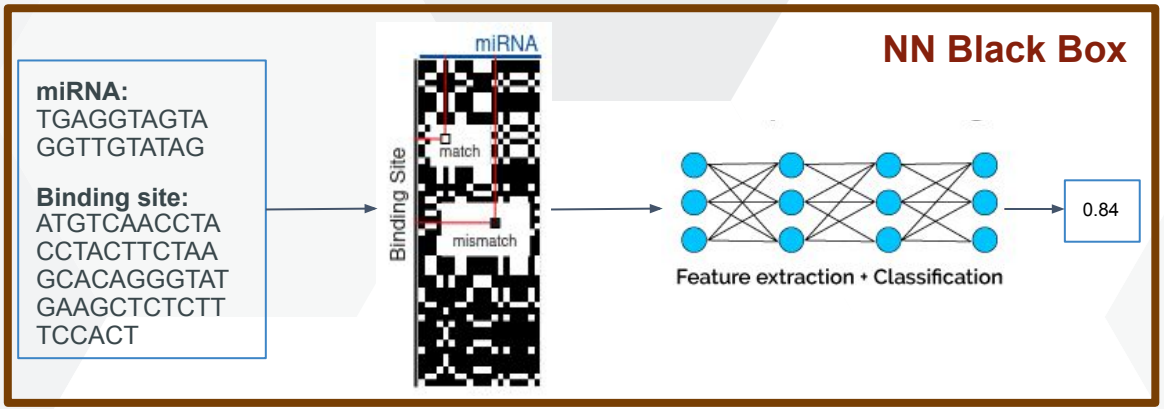
**Binding site:**
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT

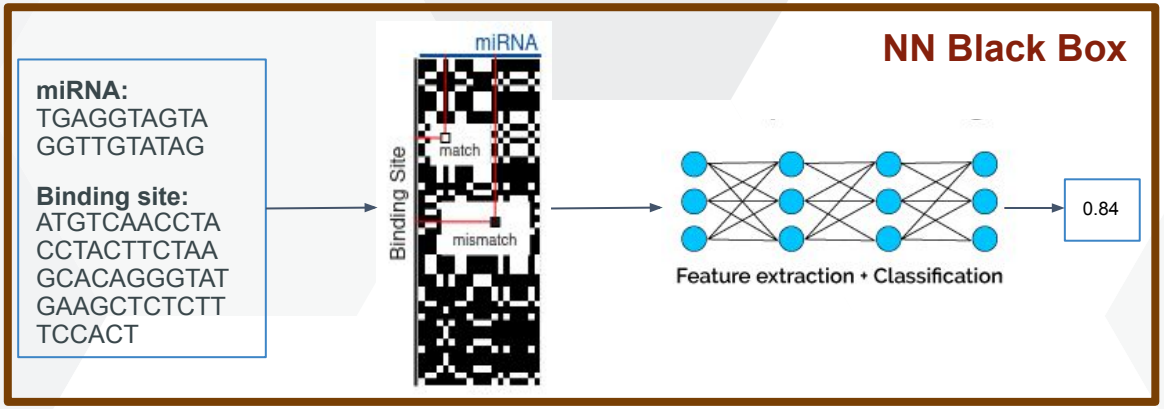


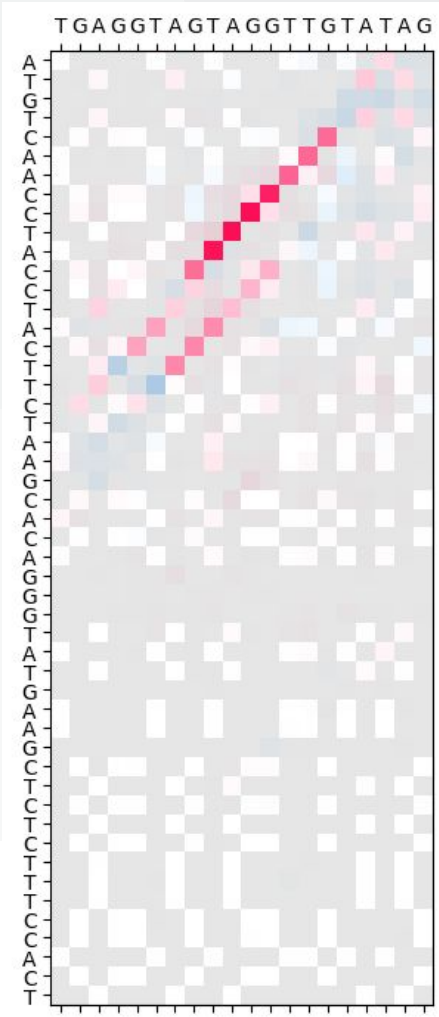



Feature extraction + Classification

0.84

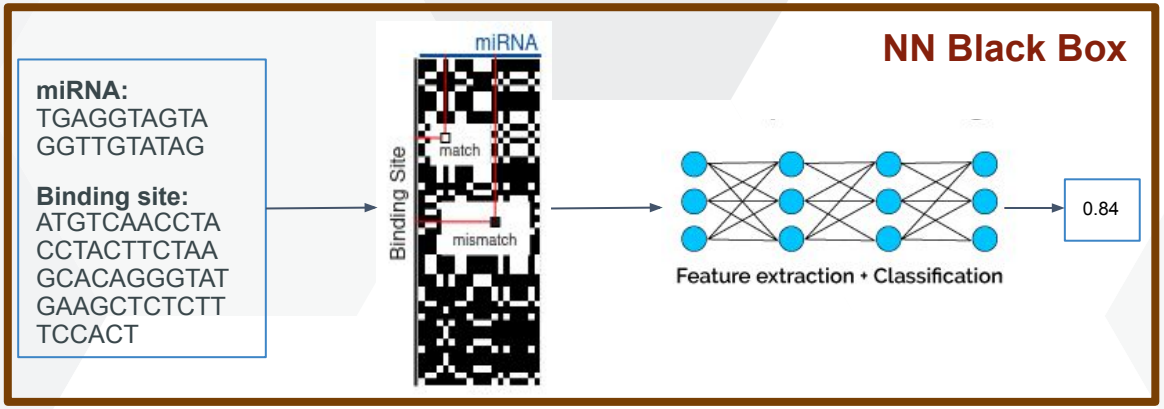

NN Black Box
miRNA:
TGAGGTAGTA
GGTTGTATAG
Binding site:
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT
miRNA
Binding Site
match
mismatch
Feature extraction + Classification
0.84

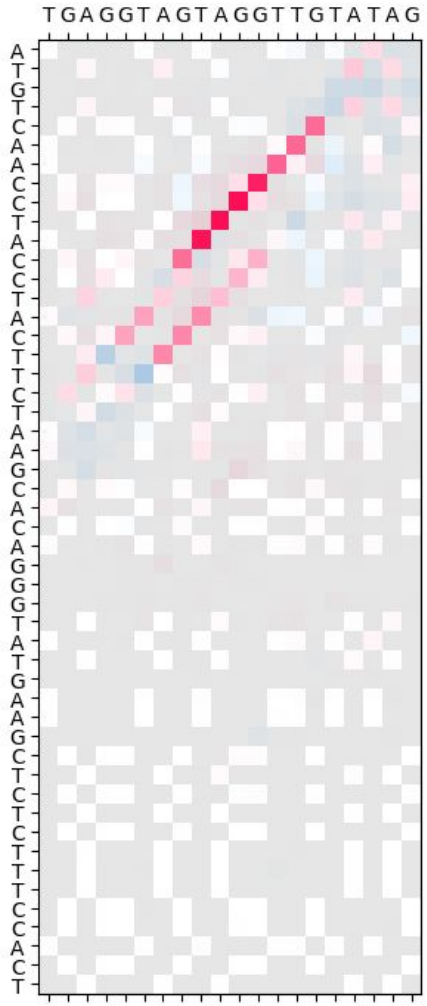

NN Black Box
miRNA:
TGAGGTAGTA
GGTTGTATAG
Binding site:
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT
miRNA
Binding Site
match
mismatch
Feature extraction + Classification
0.84


TGAGGTAGTAGGTTGTATAG
ATGTCAACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

CEITEC

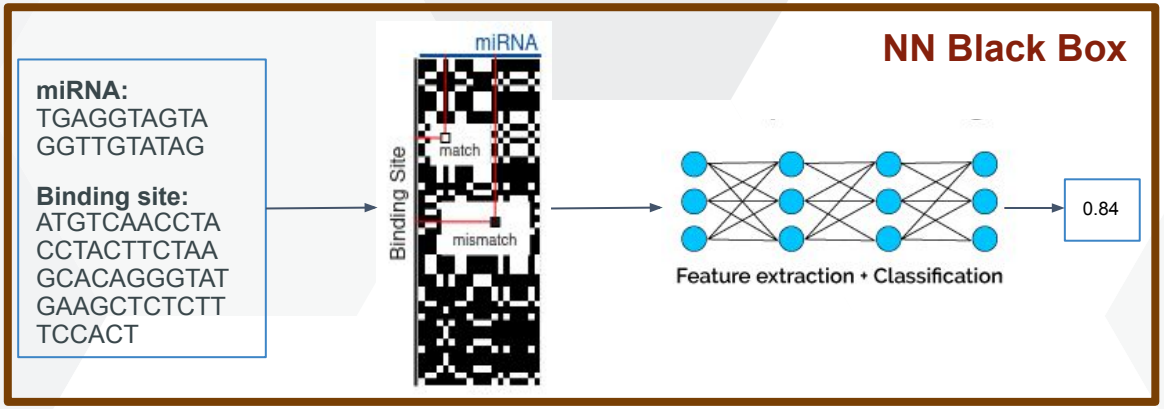

TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-
TGAGGTA-GTAGGTTGTATAG

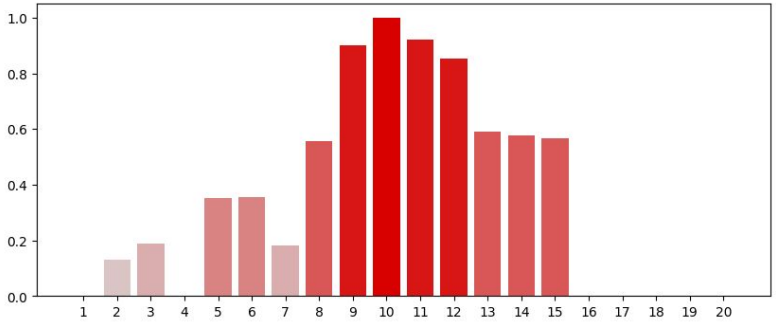

NN Black Box
miRNA:
TGAGGTAGTA
GGTTGTATAG
Binding site:
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT
miRNA
Binding Site
match
mismatch
Feature extraction + Classification
0.84

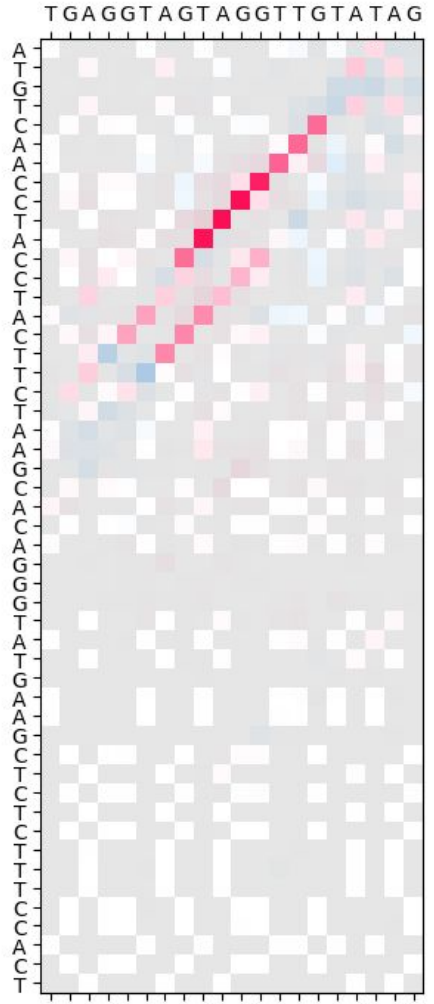

TGAGGTAGTAGGTTGTATAG
ATGTCAACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

CEITEC


TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-
TGAGGTA-GTAGGTTGTATAG

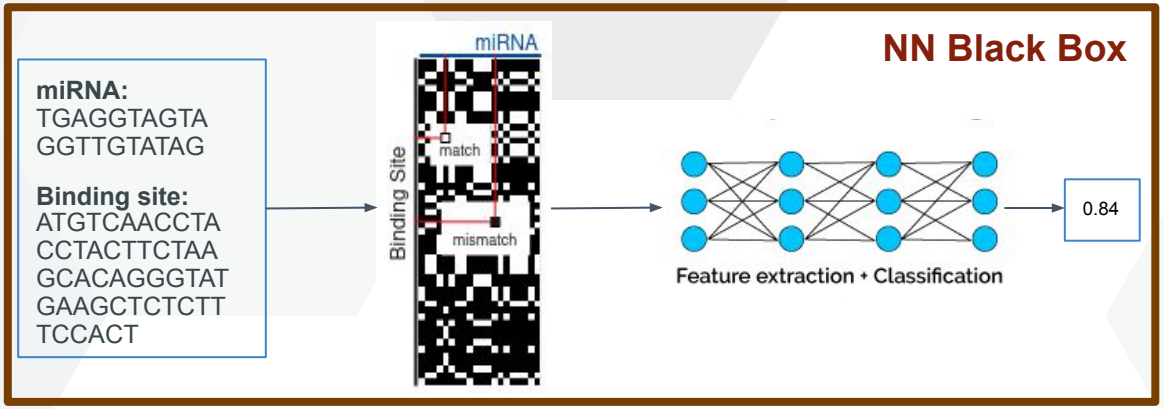

NN Black Box
miRNA:
TGAGGTAGTA
GGTTGTATAG
Binding site:
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT
miRNA
Binding Site
match
mismatch
Feature extraction + Classification
0.84

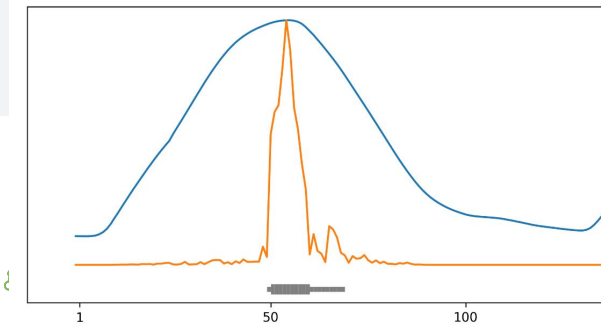

1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

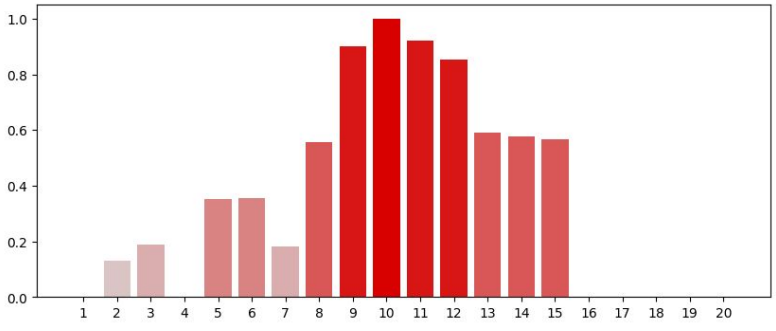

TGAGGTAGTAGGTTGTATAG
ATGTCAACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

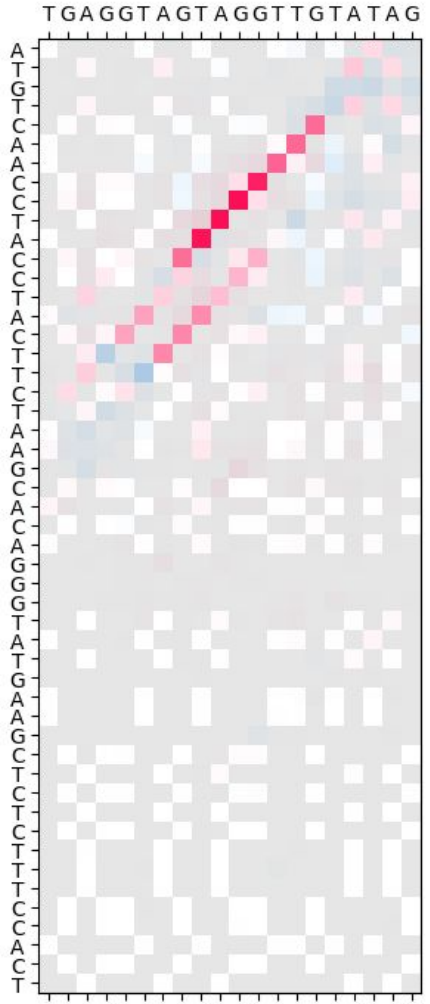
CEITEC

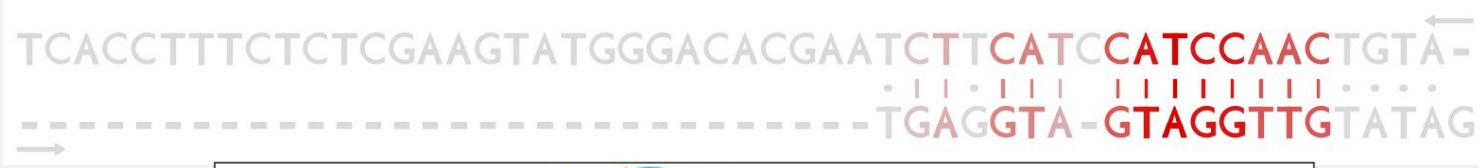

TCACCTTTCTCTCGAAGTATGGGACACGAATCTTCATCCATCCAACTGTA-
TGAGGTA-GTAGGTTGTATAG

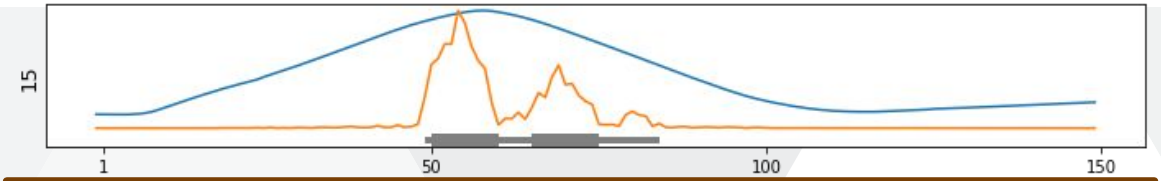

NN Black Box
miRNA:
TGAGGTAGTA
GGTTGTATAG
Binding site:
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT
miRNA
Binding Site
match
mismatch
Feature extraction + Classification
0.84


1
50
100


1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20


TGAGGTAGTAGGTTGTATAG
ATGTCAACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

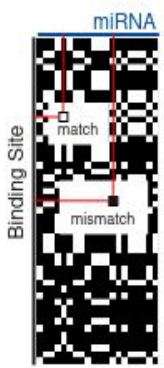


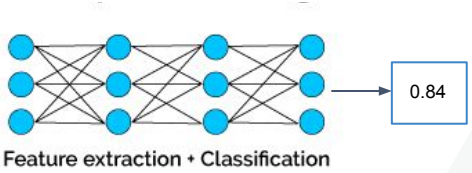




**NN Black Box**

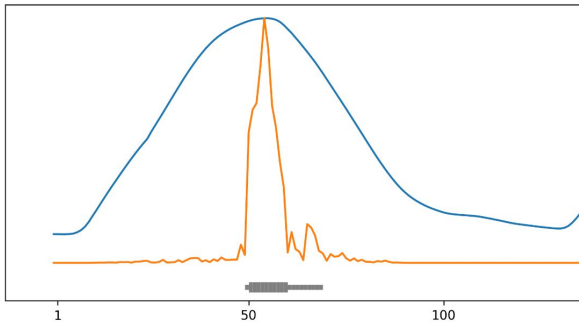
**miRNA:**
TGAGGTAGTA
GGTTGTATAG

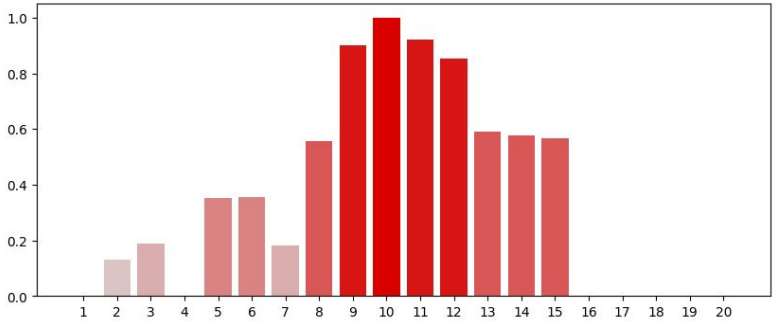
**Binding site:**
ATGTCAACCTA
CCTACTTCTAA
GCACAGGGTAT
GAAGCTCTCTT
TCCACT

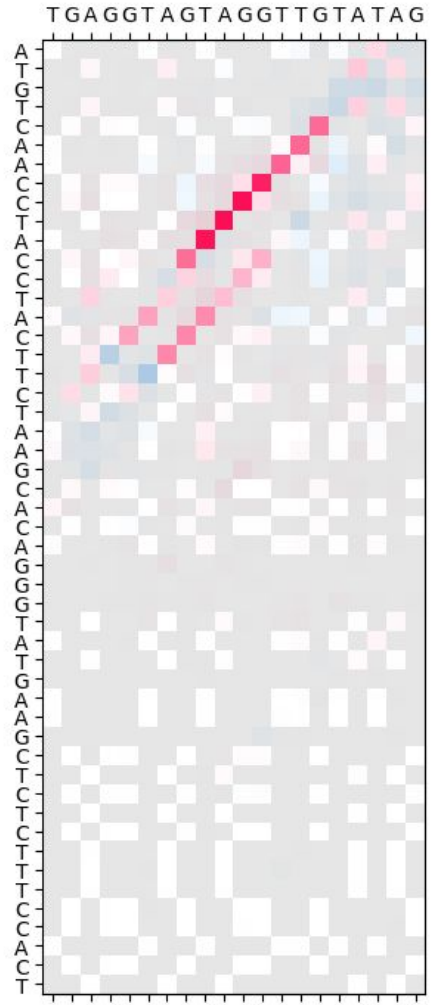


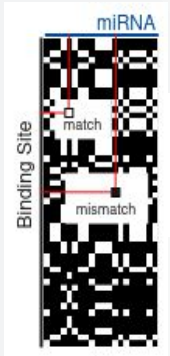

0.84

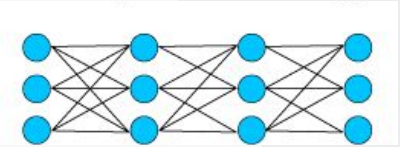


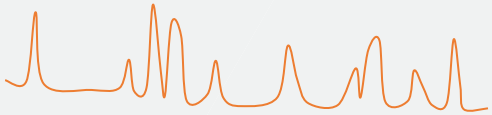



# Summary



CNN



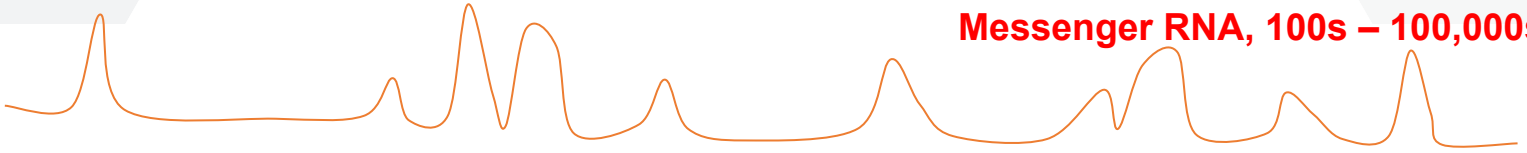
regression model

prediction

ucagcauagcuacgacguc **miRNA, ~20nt long**

auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaaccacaauucgac…

**Messenger RNA, 100s – 100,000s nt long**



# Future work

1. Include other features
    a. Genomic conservation - score / multiple sequence alignment / tree
    b. RNA Binding Proteins - binding sites
    c. Sequence?
    d. …
    e. Ablation studies

2. If two-part approach does not work
    a. Simplify regression to classification task?
    b. Skip the two-part approach and go with sequence? (in progress)
        i. HyenaDNA - pretrained single nucleotide resolution transformer for long sequences
            1. Use for embeddings
            2. Use full with regression head

# Sources

SHAP https://github.com/shap/shap

Determinants of Functional MicroRNA Targeting
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9880601/

miRBind: A Deep Learning Method for miRNA Binding Classification
https://www.mdpi.com/2073-4425/13/12/2323

Using Attribution Sequence Alignment to Interpret Deep Learning Models for miRNA Binding Site Prediction https://www.mdpi.com/2079-7737/12/3/369

# Thank you for your Attention!