Explainability for Deep Neural Networks

Katarína Grešová

L-Università ta' Malta

BioGeMT

MALTAomics Summer School workshop

Katarína Grešová

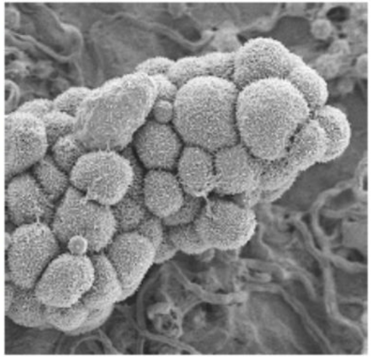
https://katarinagresova.github.io

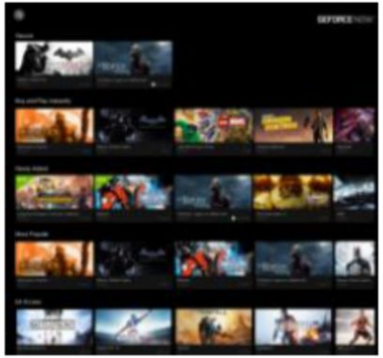
Outline

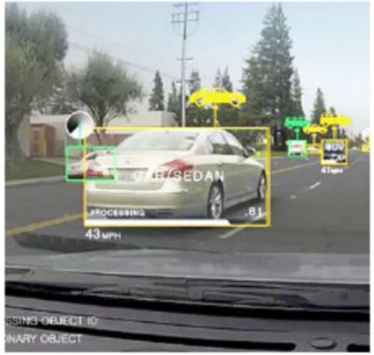
14:00 – 14:10	Introduction to explainability for deep neural networks
14:10 – 14:40	Hands on: Practical overview of explainability methods for genomic sequence data
14:40 – 15:30	Hands on: Practical overview of explainability methods for image data
15:30 – 16:00	Coffee break
16:00 – 16:30	Use case: miRNA target prediction
16:30 – 17:00	Hands on: Using DeepExperiment to interpret and visualize miRNA targeting

Introduction to explainability for deep neural networks

DEEP LEARNING EVERYWHERE







INTERNET & CLOUD

Image Classification
Speech Recognition
Language Translation
Language Processing
Sentiment Analysis
Recommendation

MEDICINE & BIOLOGY

Cancer Cell Detection
Diabetic Grading
Drug Discovery

MEDIA & ENTERTAINMENT

Video Captioning
Video Search
Real Time Translation

SECURITY & DEFENSE

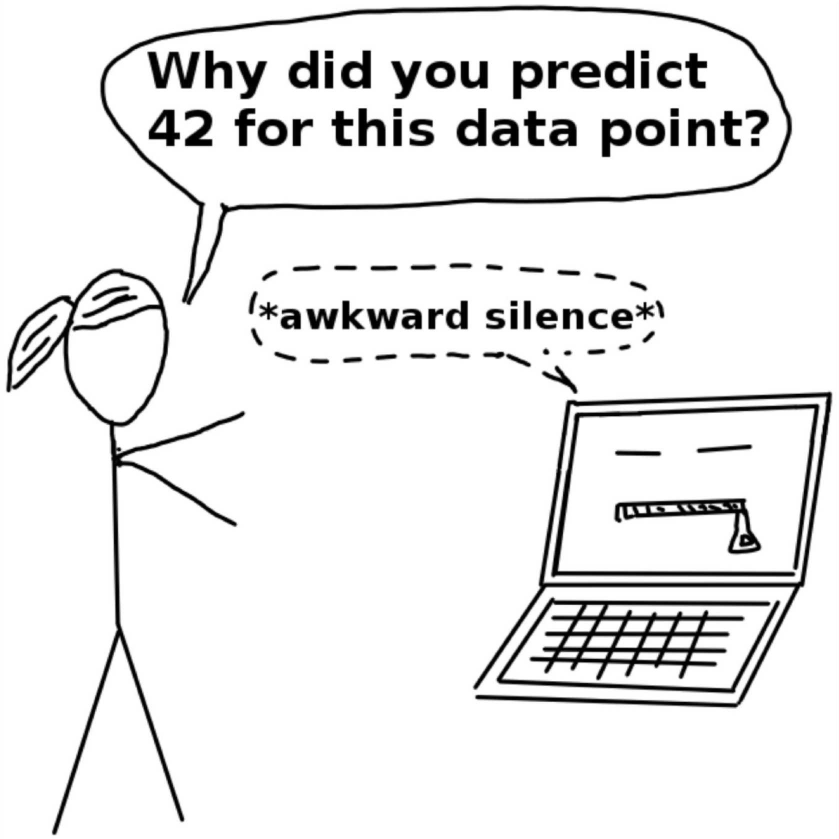
Face Detection
Video Surveillance
Satellite Imagery

AUTONOMOUS MACHINES

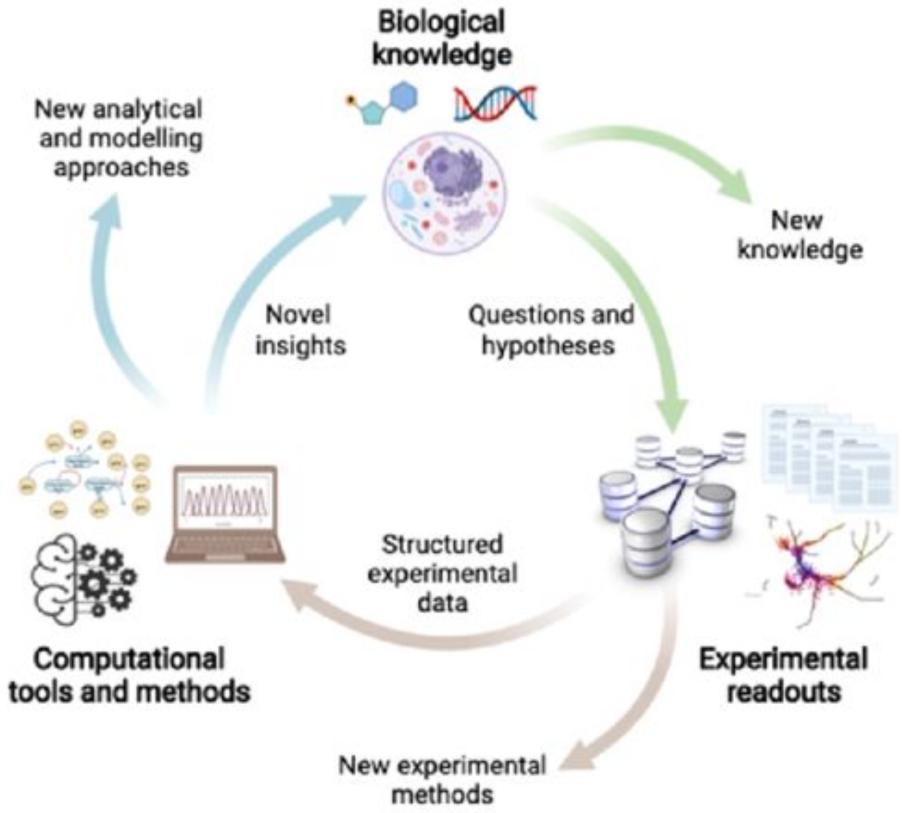
Pedestrian Detection
Lane Tracking
Recognize Traffic Sign

When do we need model interpretation?

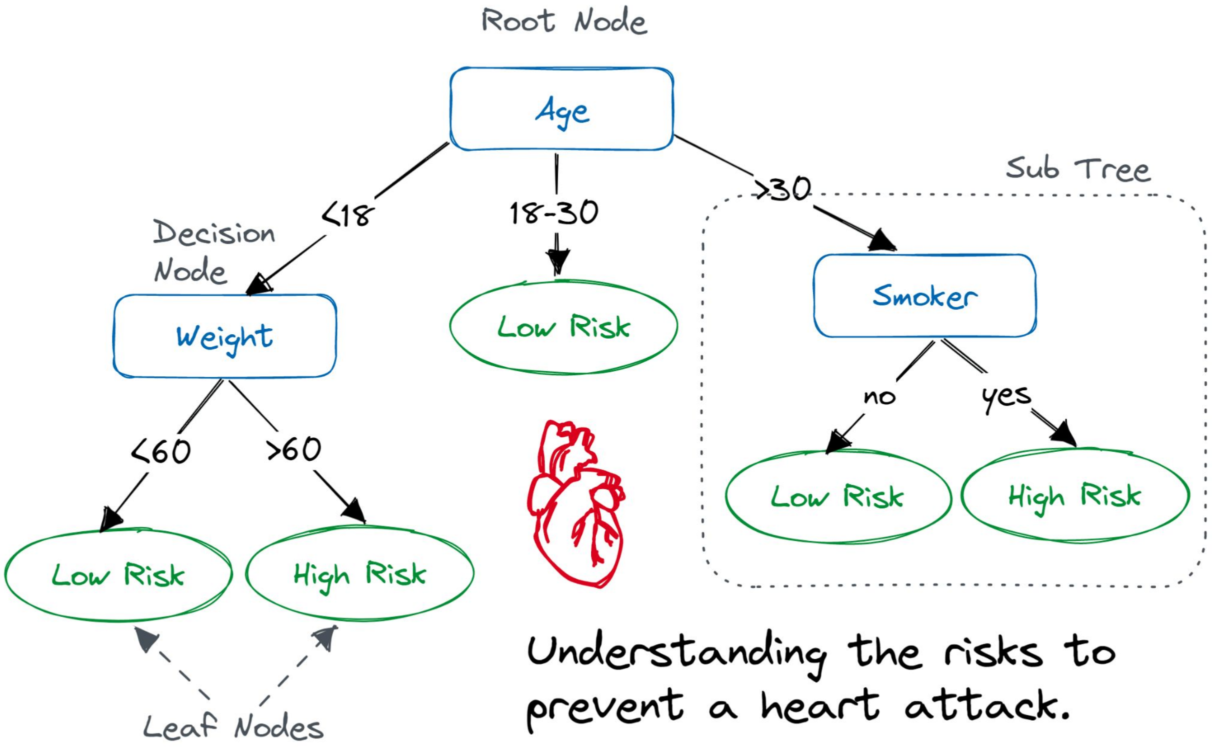
- High-stakes decision making settings
  - Impact on human lives/health/finances
  - Less studied problems, models not extensively validated
- Accuracy alone is no longer enough
  - Train/test data might not be representative of data encountered in practice
- Auxiliary criteria are also crucial
  - Nondiscrimination
  - Right to explanation

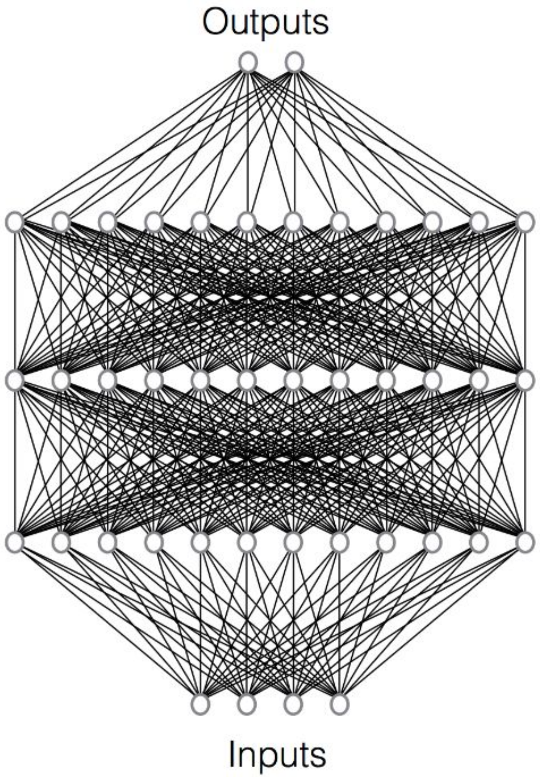


Model interpretation as scientific method?

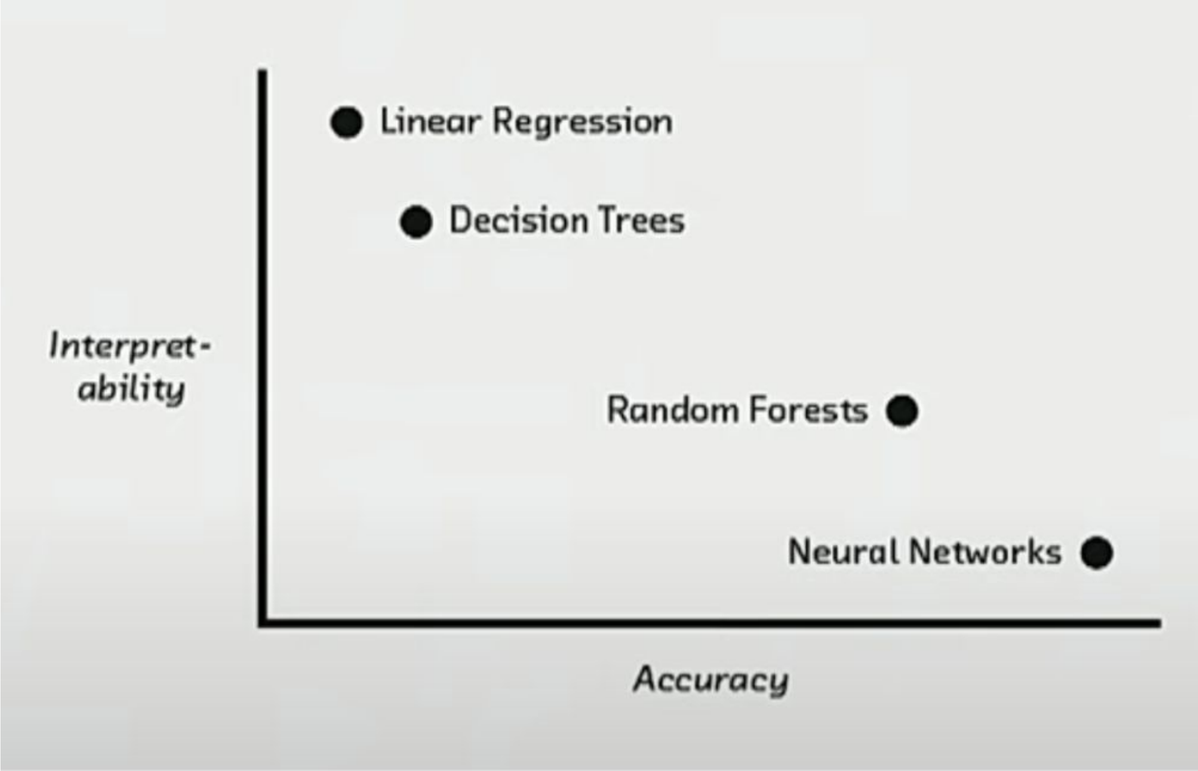


Machine Learning vs. Deep Learning

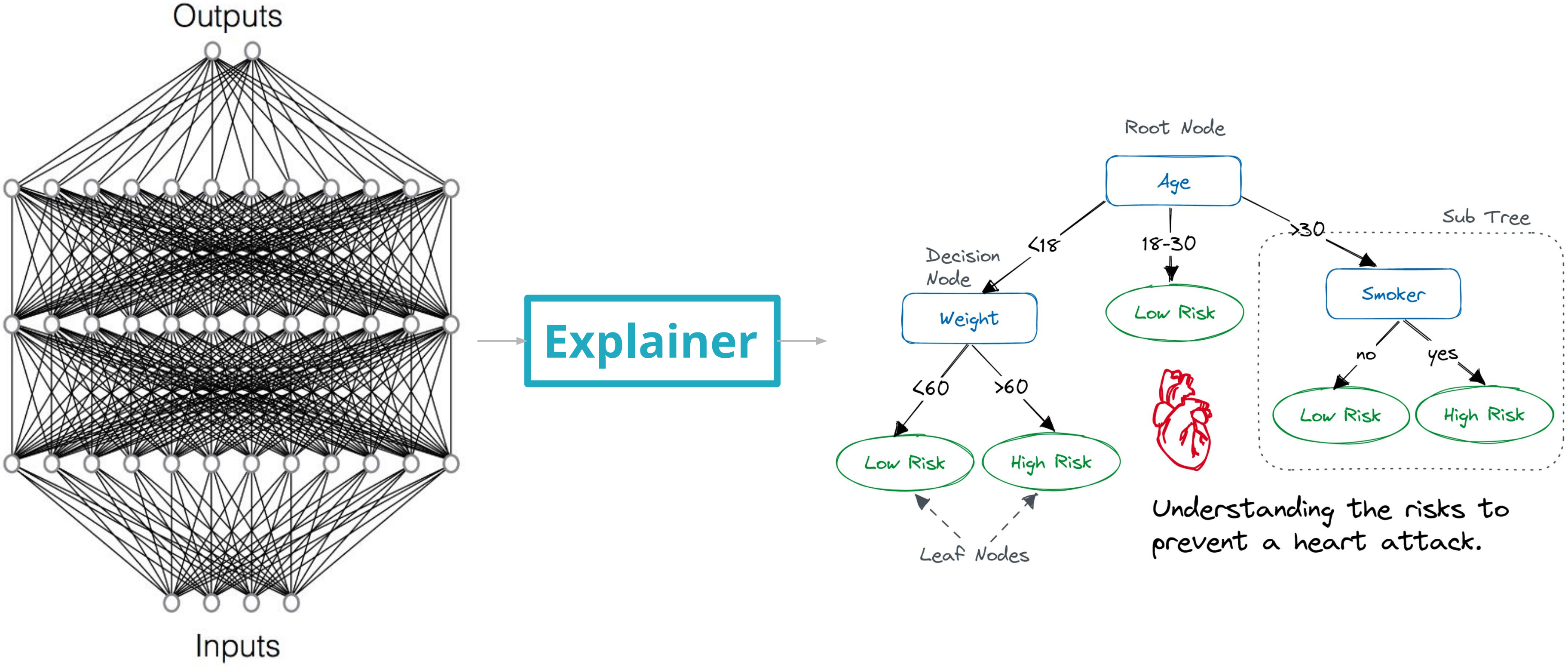




Interpretability vs Accuracy tradeoff



Post-hoc explainability



Hands on: Practical overview of explainability methods for genomic sequence data

Open the Colab notebook

Hands on: Practical overview of explainability methods for genomic sequence data (https://colab.research.google.com/drive/1Br0f8xPIBkGFIPuXZPMtDDV8GG64wkf2?usp=sharing)

Hands on: Practical overview of explainability methods for image data

Open the Colab notebook

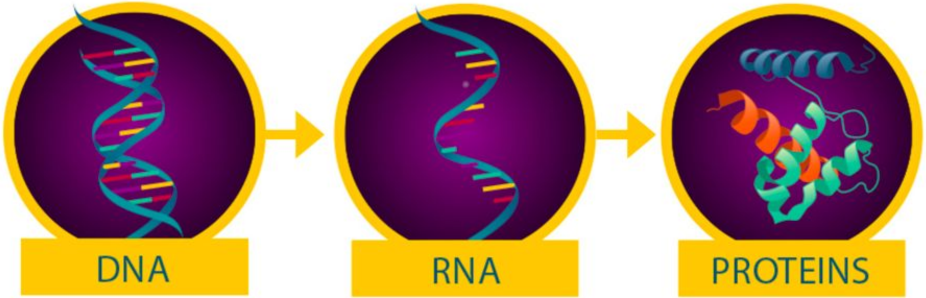
Hands on: Practical overview of explainability methods for image data
(https://colab.research.google.com/drive/1cO54Si-hoTZkrkIRS3WYxYHVm4YQUEhB?usp=sharing)

Coffee Break!

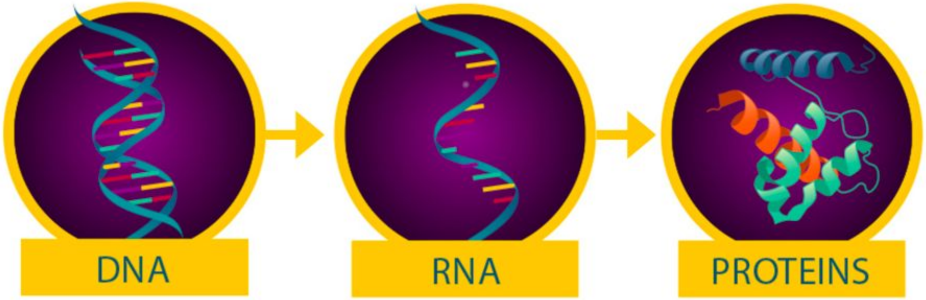
Let's start again in 30 minutes (16:00)

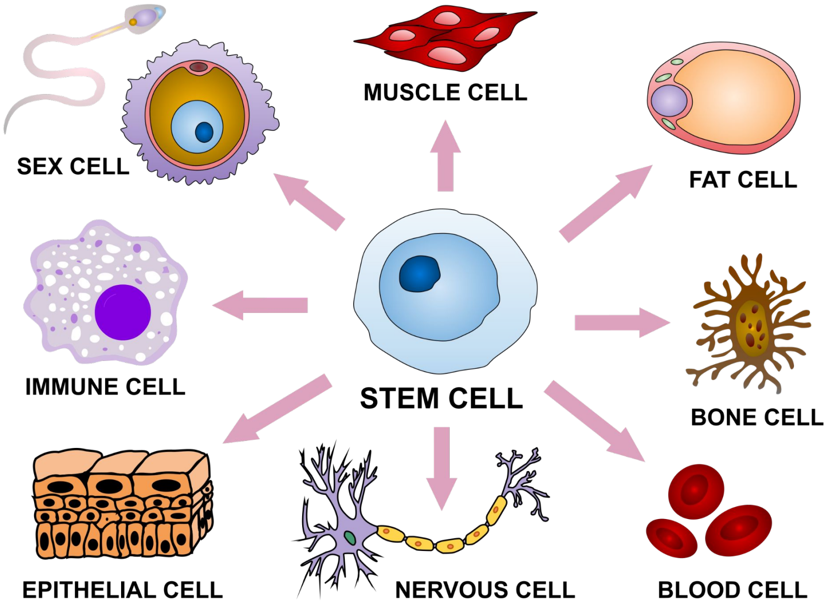
Use case: miRNA target prediction

Biological meaning

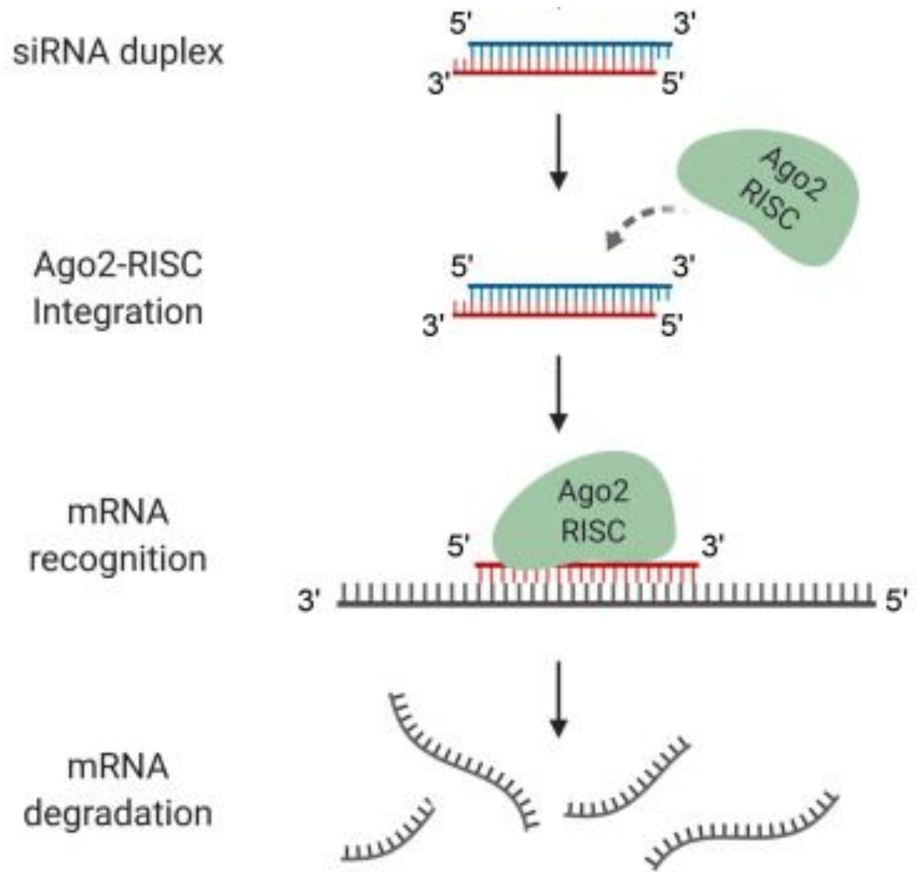


Biological meaning

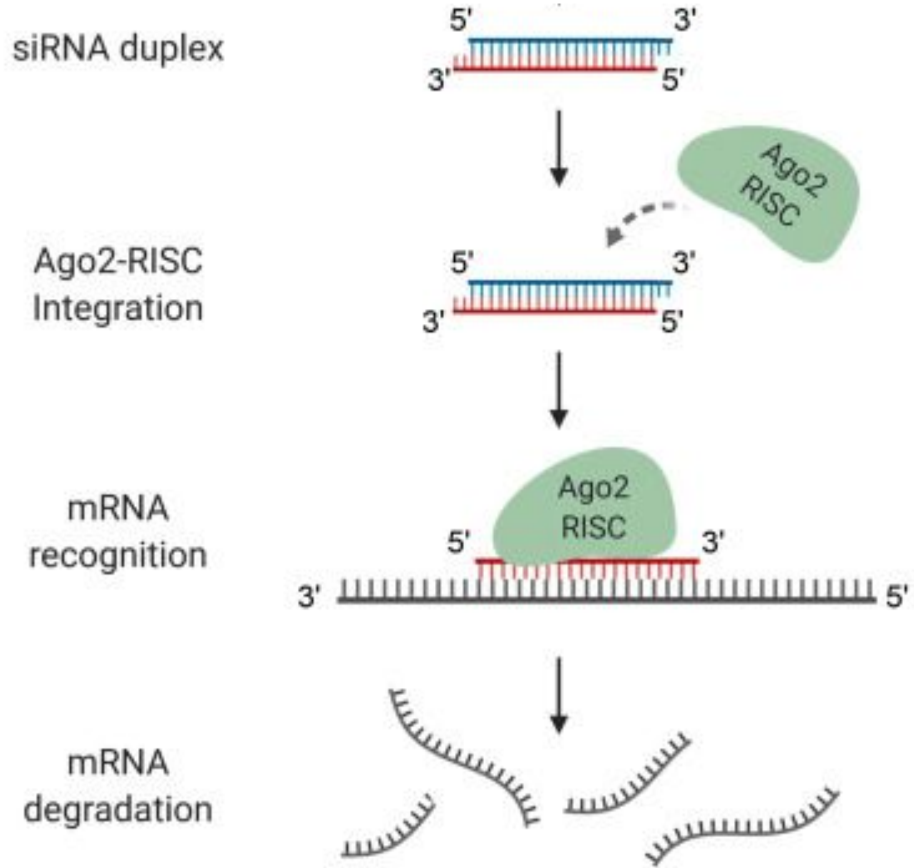


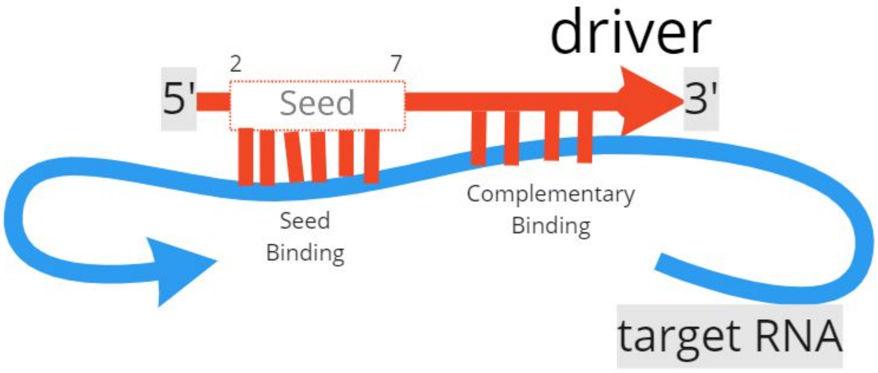


RISC (RNA-induced silencing complex)

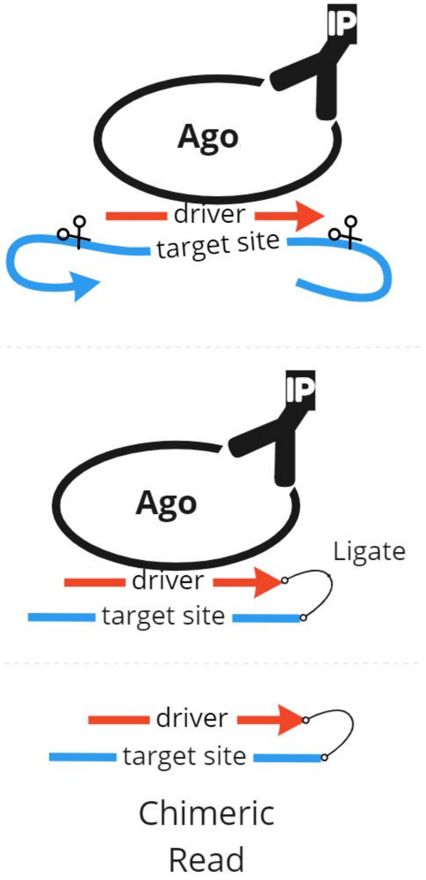


Seed Binding + More?

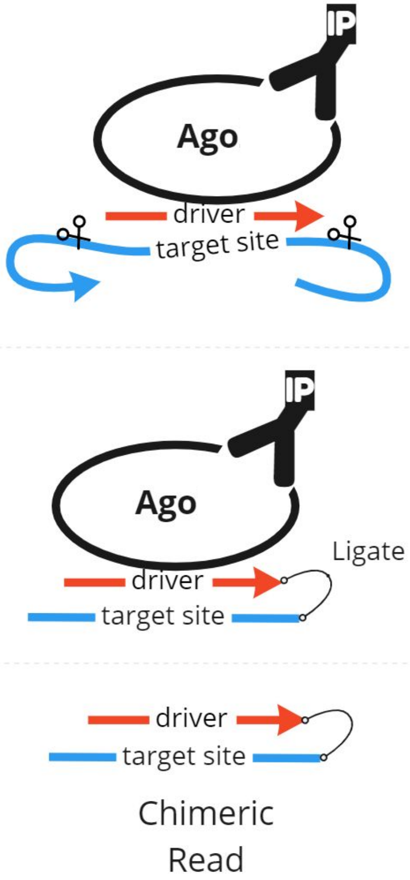




Biological experiment - CLASH



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

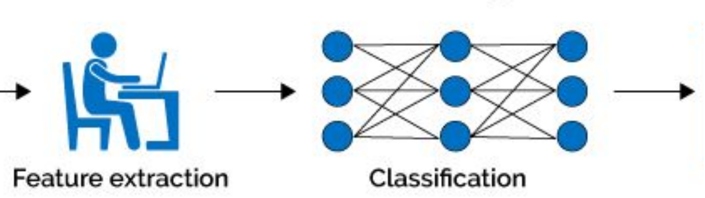
Helwak et al., 2013 CLASH dataset - 30 785 miRNA:target site pairs

Computational model

Statistical model

miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT



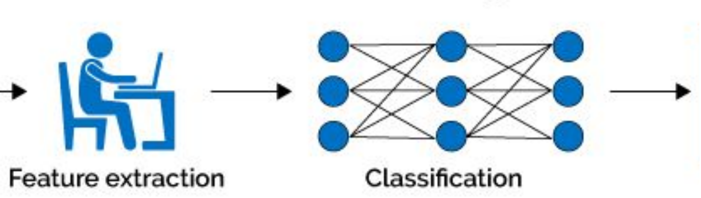
0.84

Computational model

Statistical model

miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT

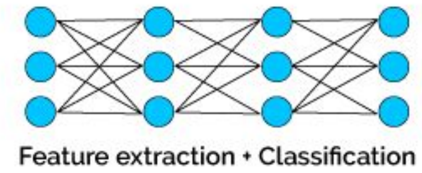


0.84

Deep neural network

miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT



0.84

miRBind model

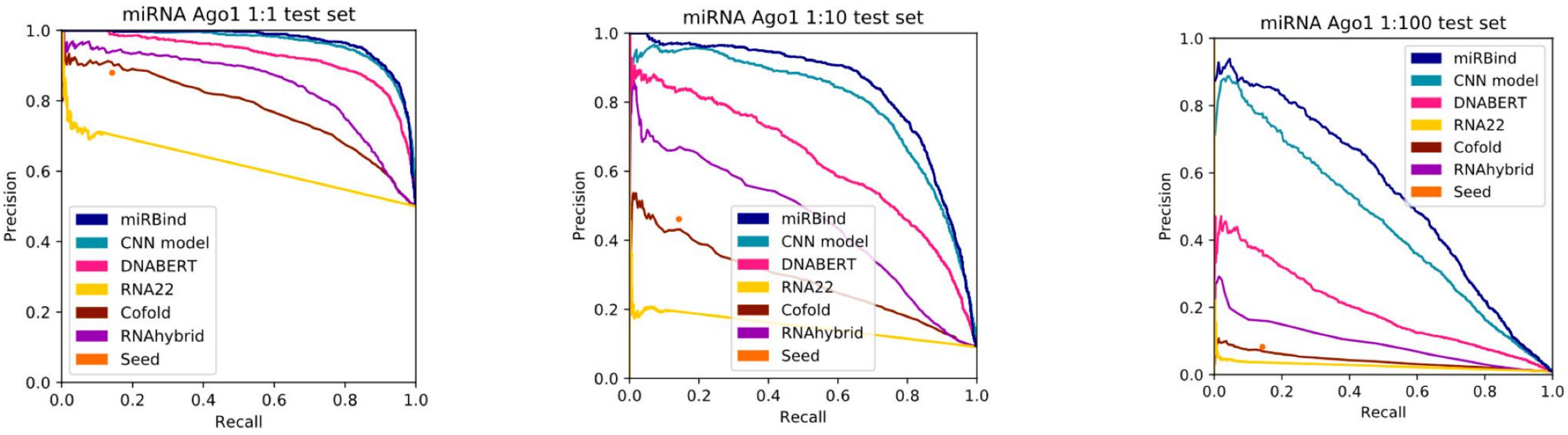
genes

MDPI

Article

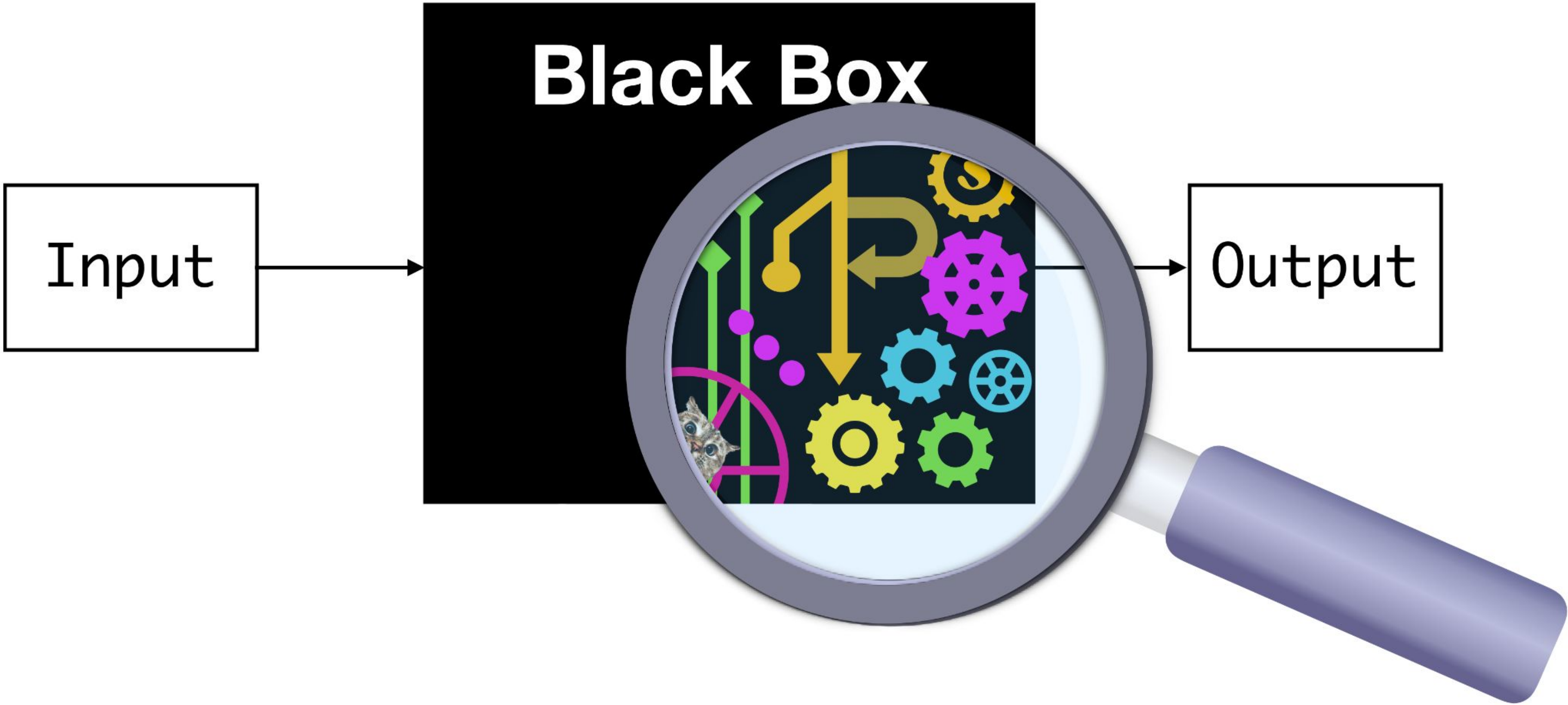
miRBind: A Deep Learning Method for miRNA Binding Classification

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

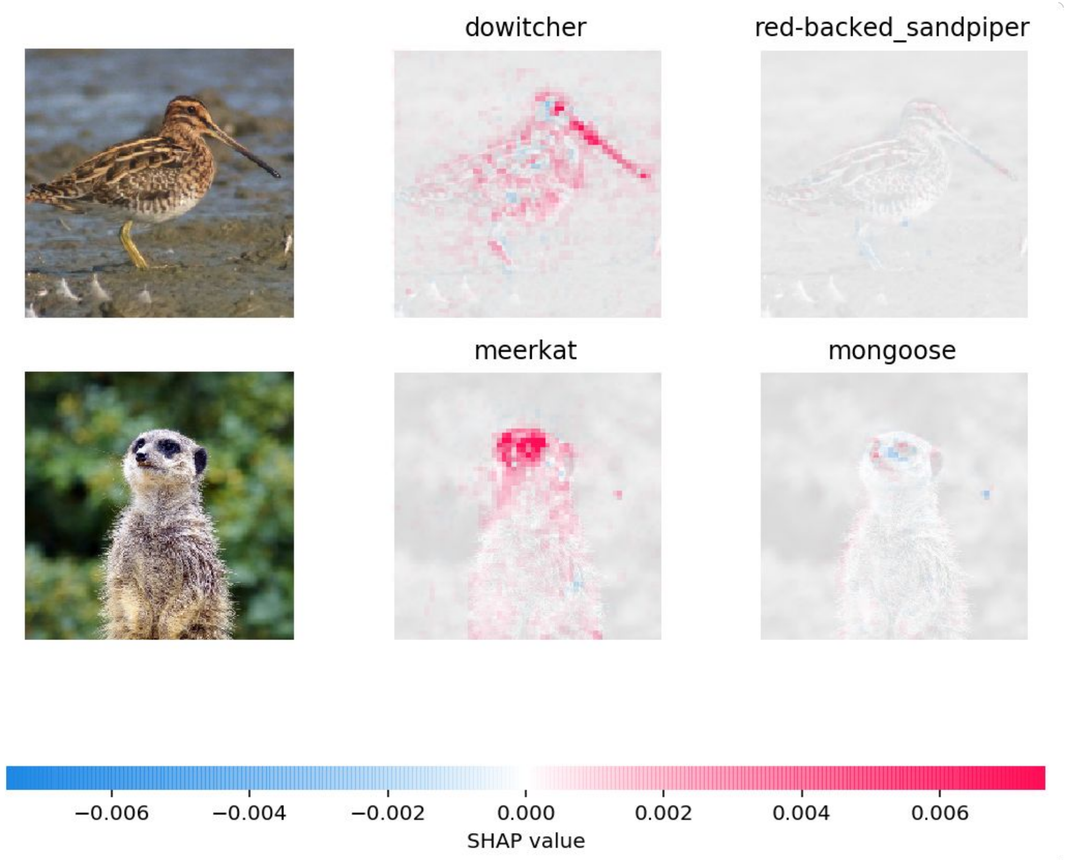


(a) (b) (c)

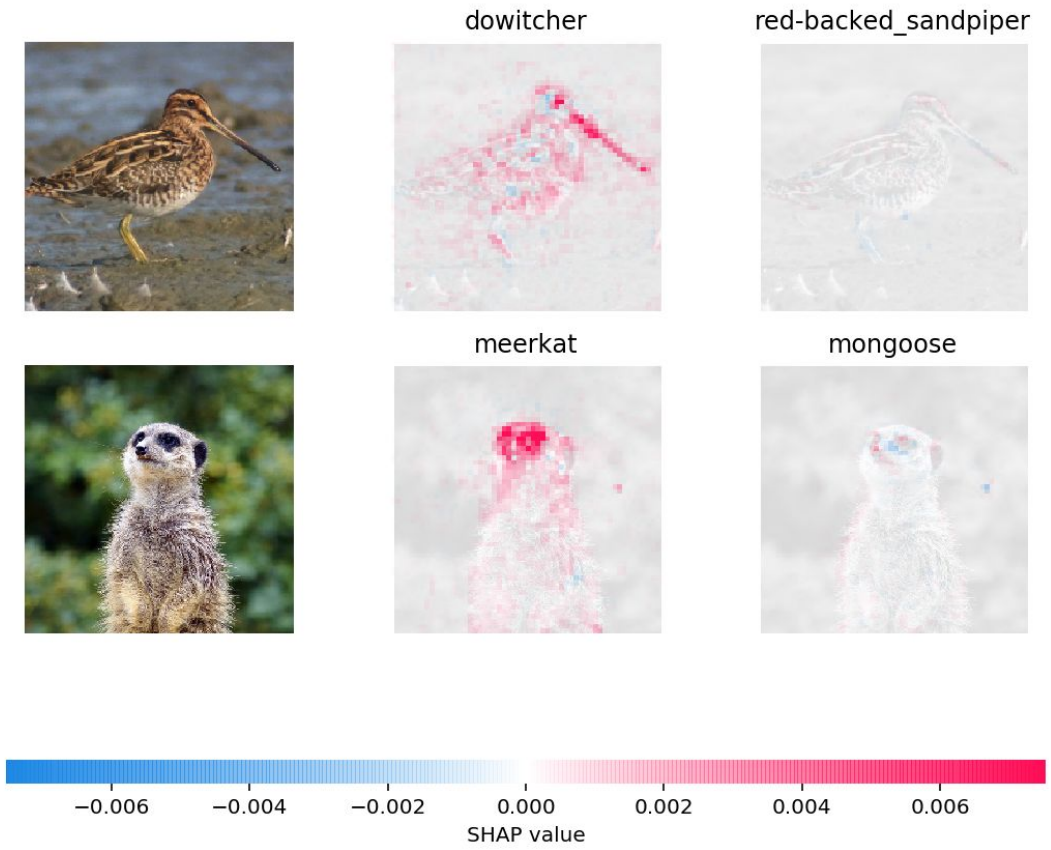
Precision-recall curves

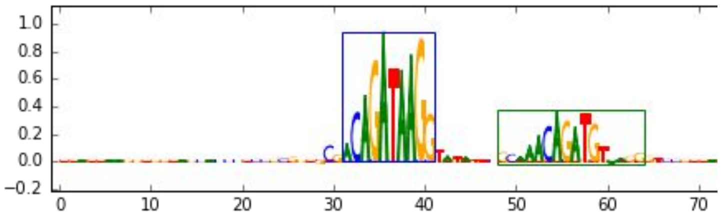

Black Box
Input
Output

Interpreting Neural Networks



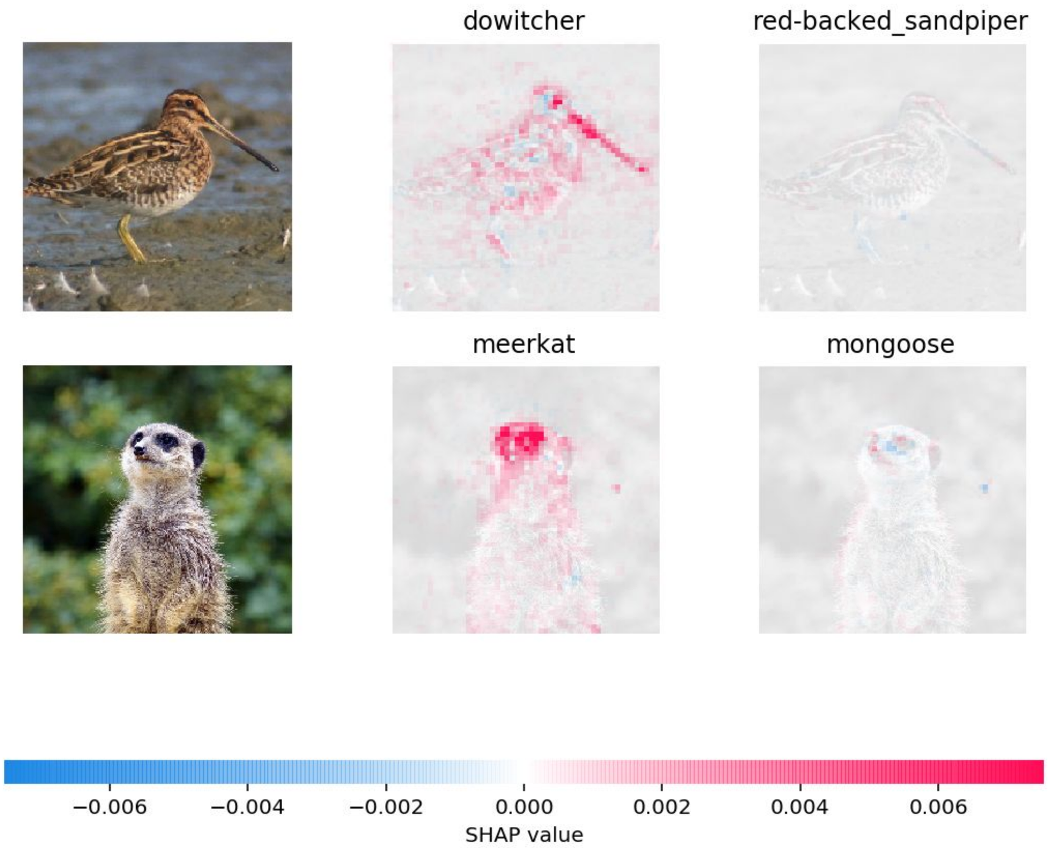
Interpreting Neural Networks

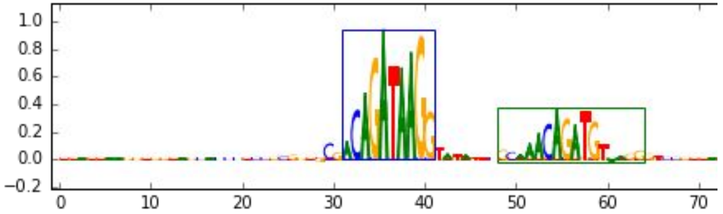




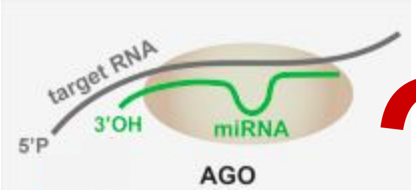
GATA motif TAL1 motif

Interpreting Neural Networks





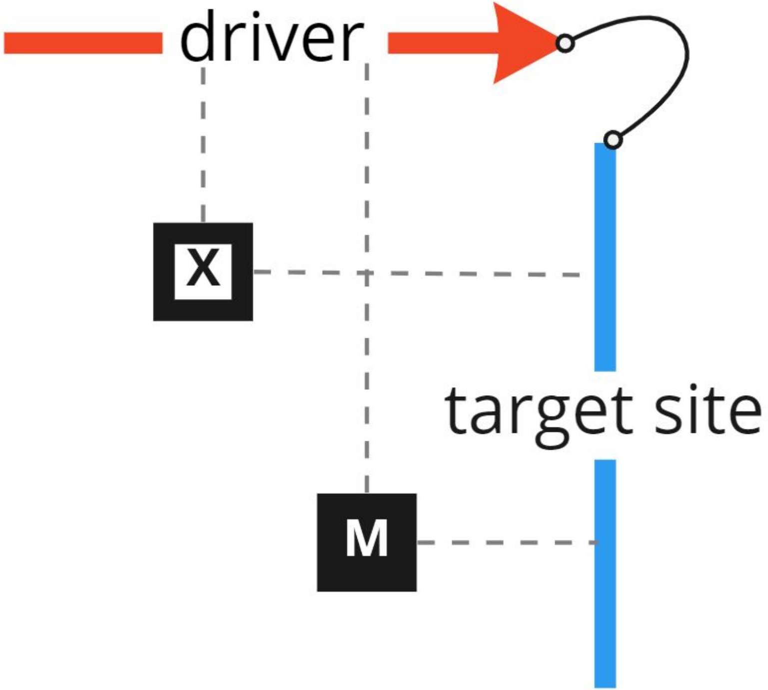
GATA motif TAL1 motif

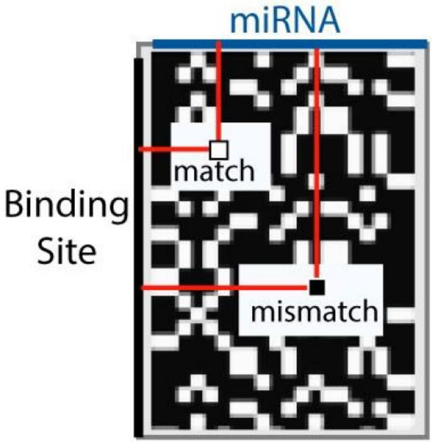


How to interpret interaction between sequences ?

TACGTCAGTTCATGAAGCT
A
(driver ~20nt)

AGTTCTAGTTCGTCCGTCAGTGTCAG
TTCATGAGCACCAGTCACGTTCGTCTA
(target ~50nt)

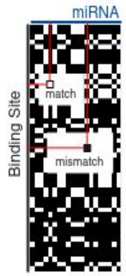


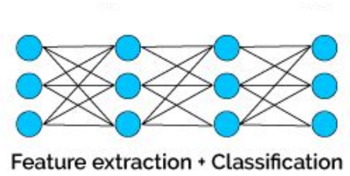


miRBind model - interpretation

miRNA:
TGAGGTAGTAGGTTGTATAG

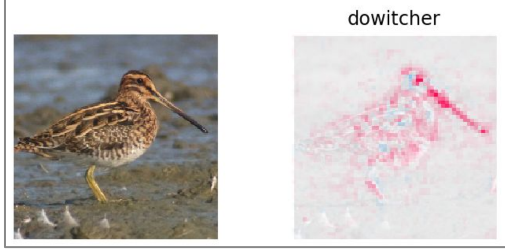
Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT





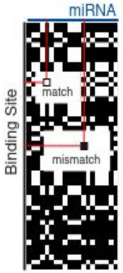
0.84

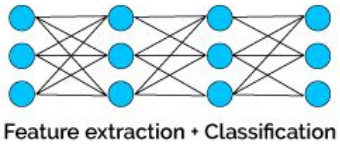
miRBind model - interpretation



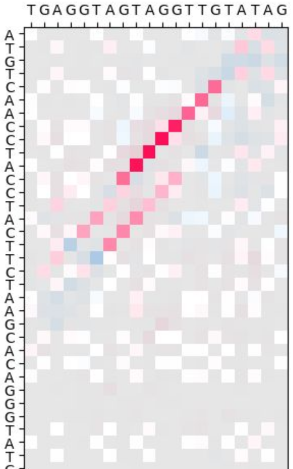
miRNA:
TGAGGTAGTAGGTTGTATAG

Binding site:
ATGTCAACCTACCTACTTCTAAGCA
CAGGGTATGAAGCTCTCTTTCCACT





0.84



Visualization

miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

Predicted alignment:

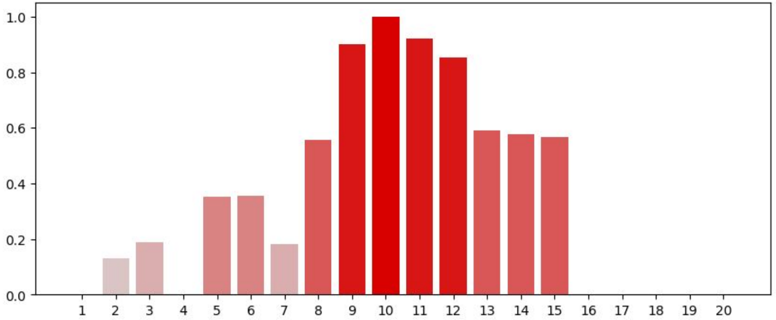
Visualization

miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

Predicted alignment:

miRNA position importance:



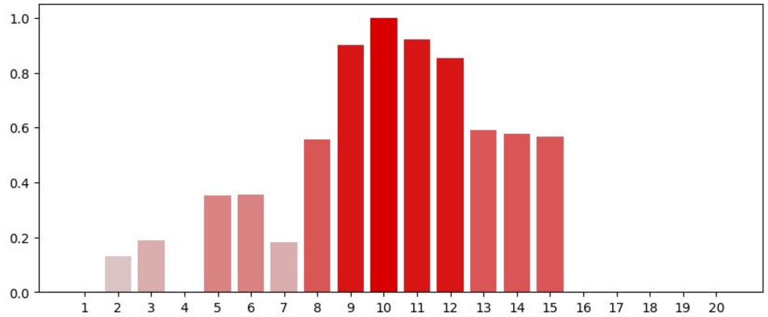
Visualization

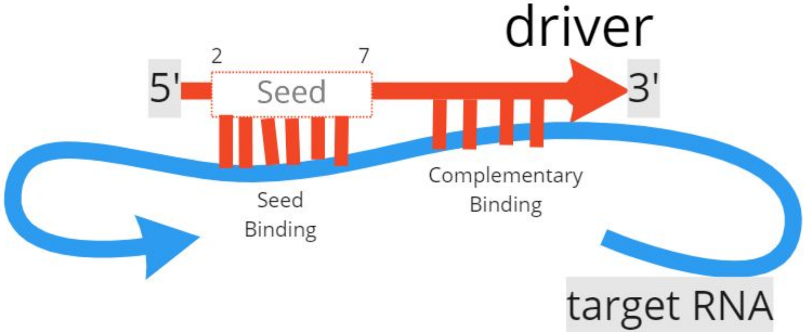
miRNA: TGAGGTAGTAGGTTGTATAG

Binding site: ATGTCAACCTACCTACTTCTAAGCACAGGGTATGAAGCTCTCTTTCCACT

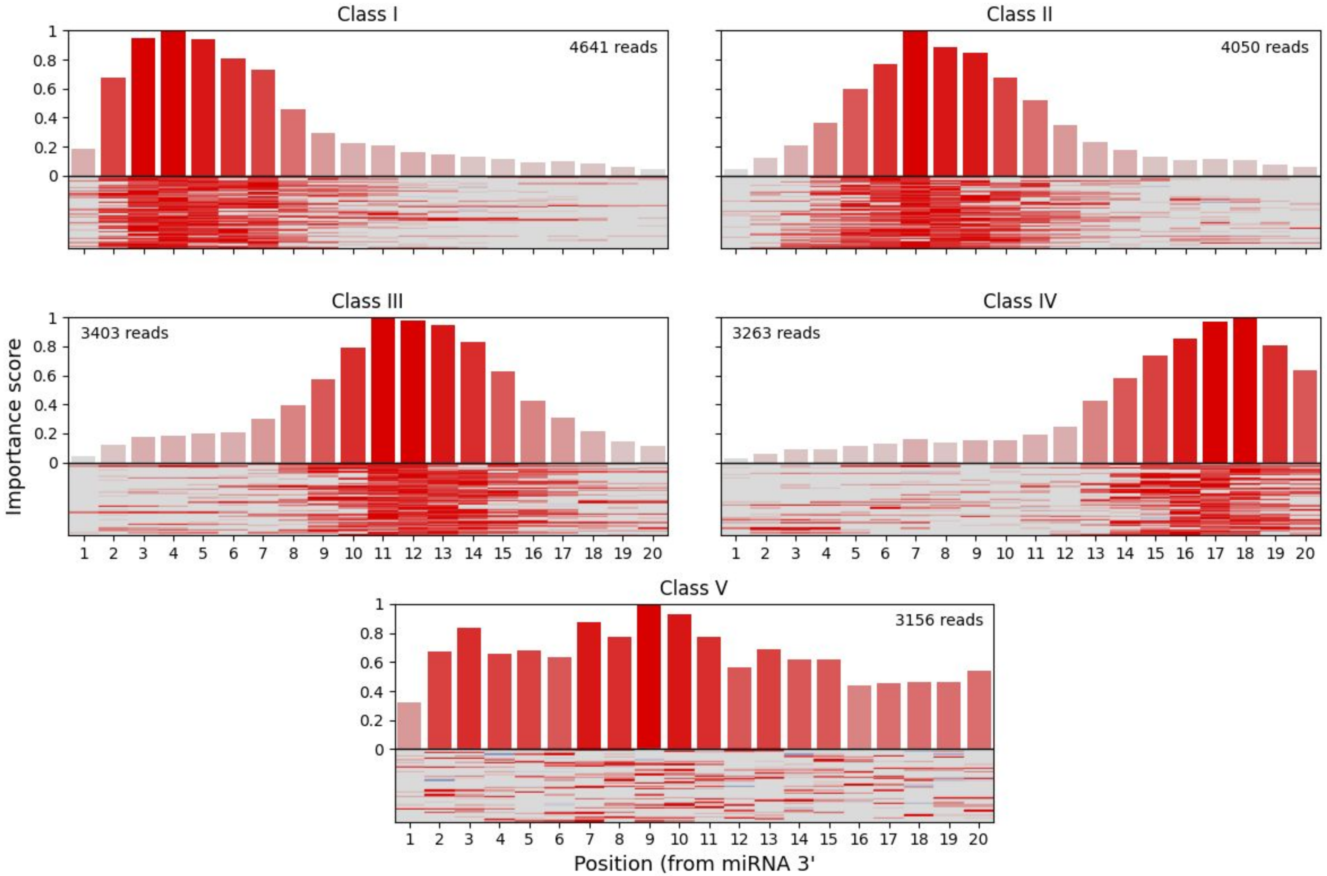
Predicted alignment:

miRNA position importance:

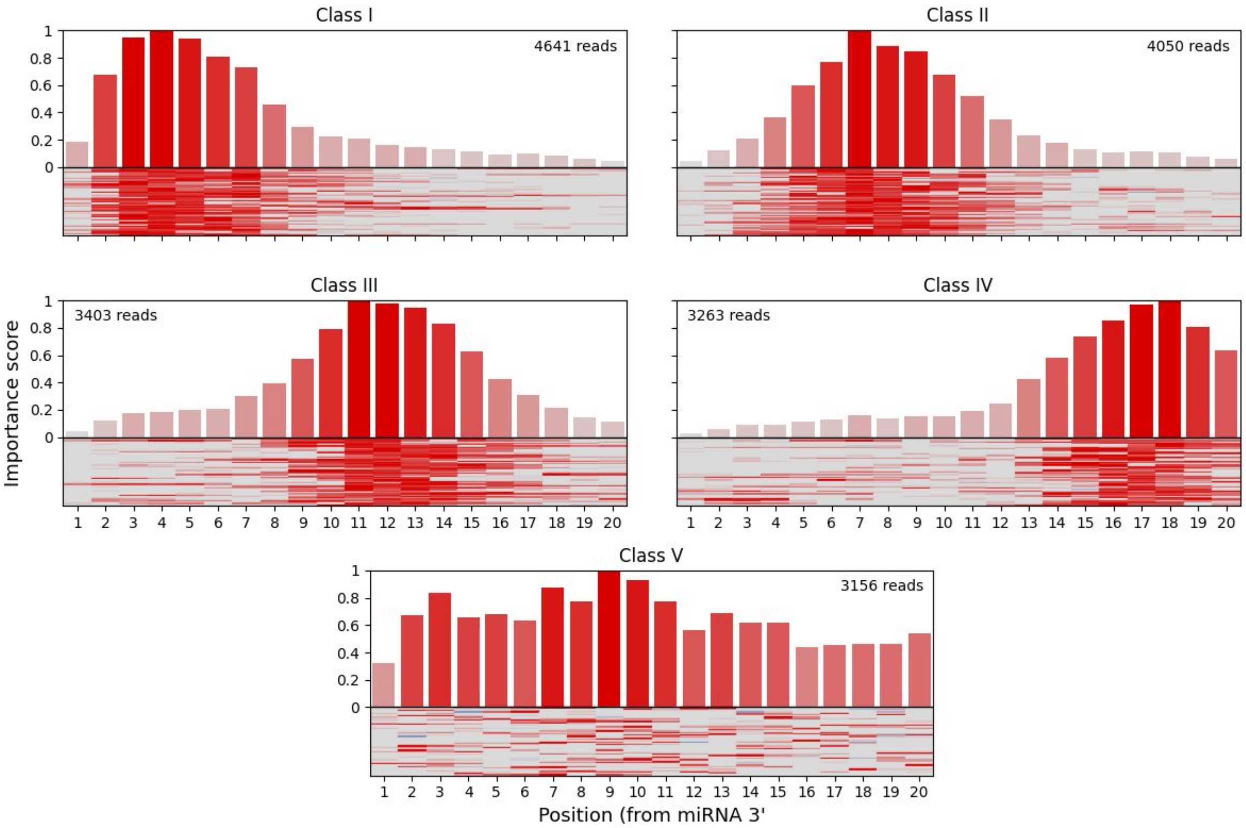


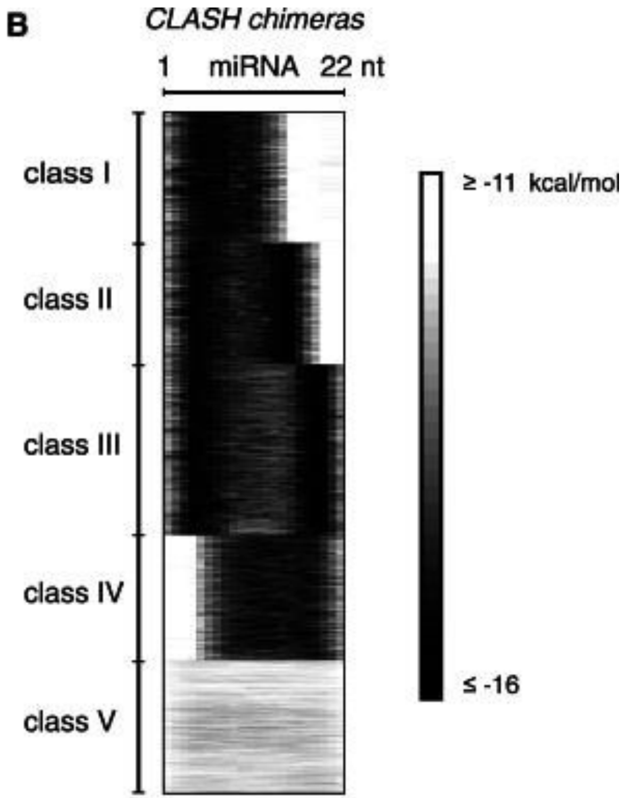


Classes of interaction



Classes of interaction





Helwak et al., 2013

Mutagenesis experiment

Open access, freely available online PLoS BIOLOGY

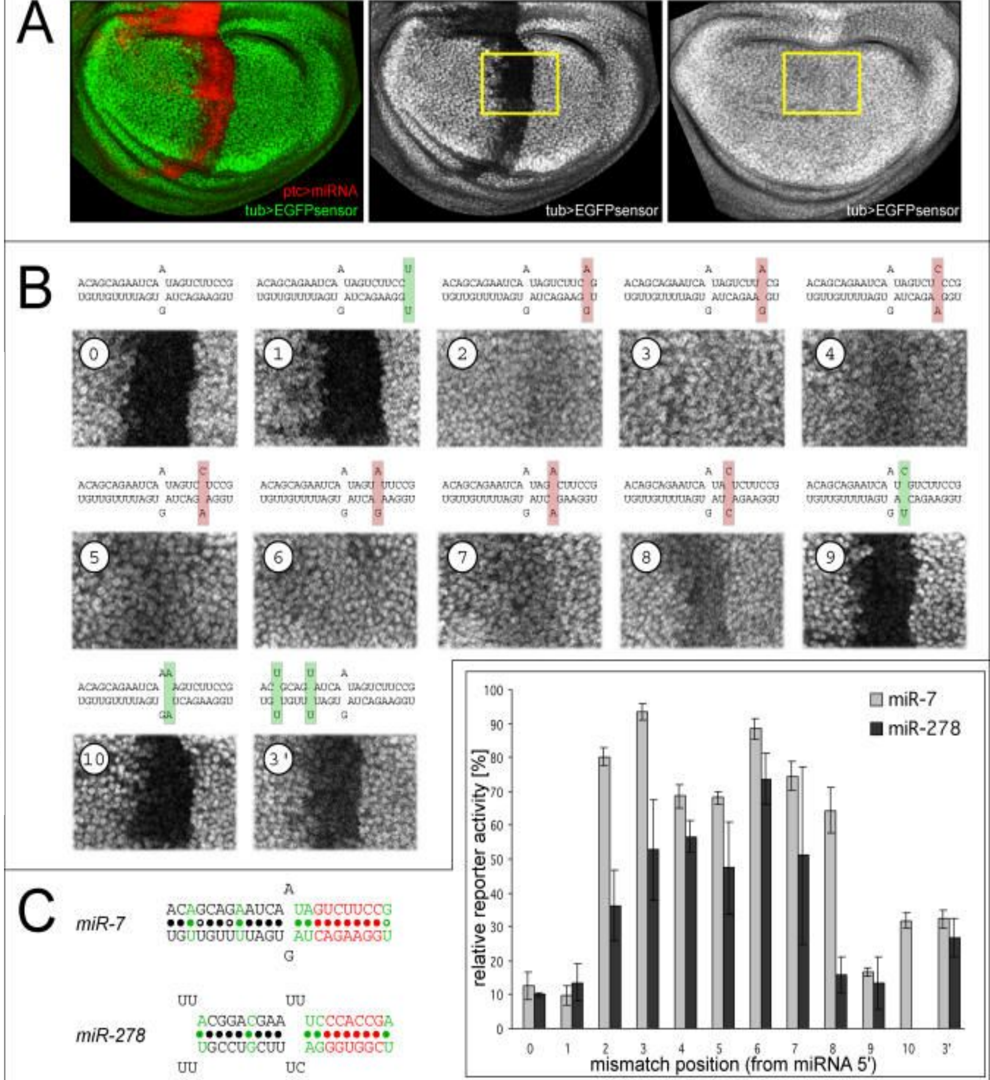
Principles of MicroRNA–Target Recognition

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.



Mutagenesis experiment

Open access, freely available online PLoS BIOLOGY

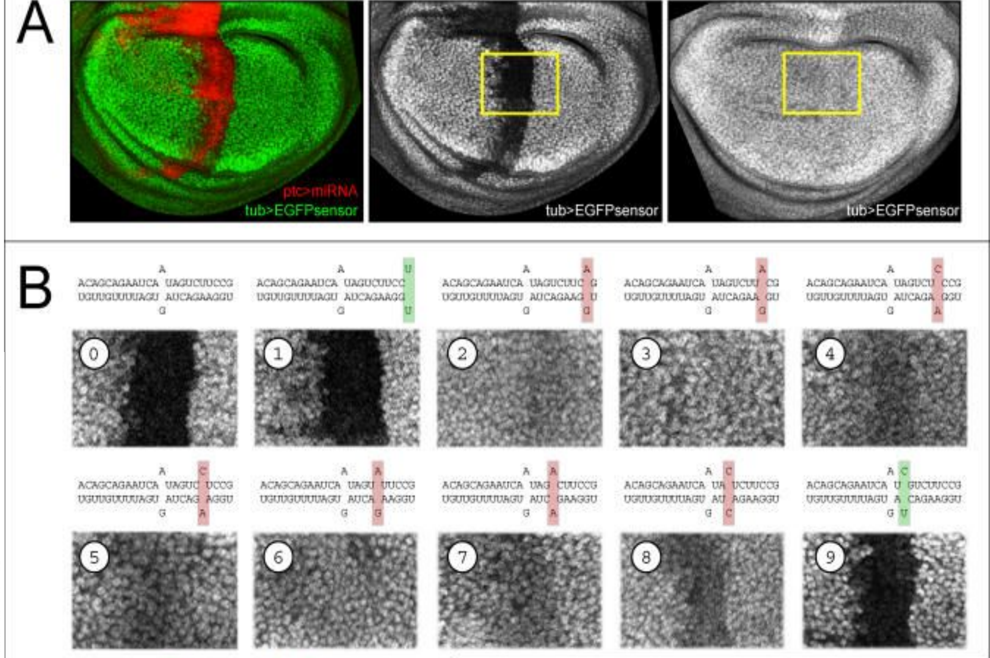
Principles of MicroRNA–Target Recognition

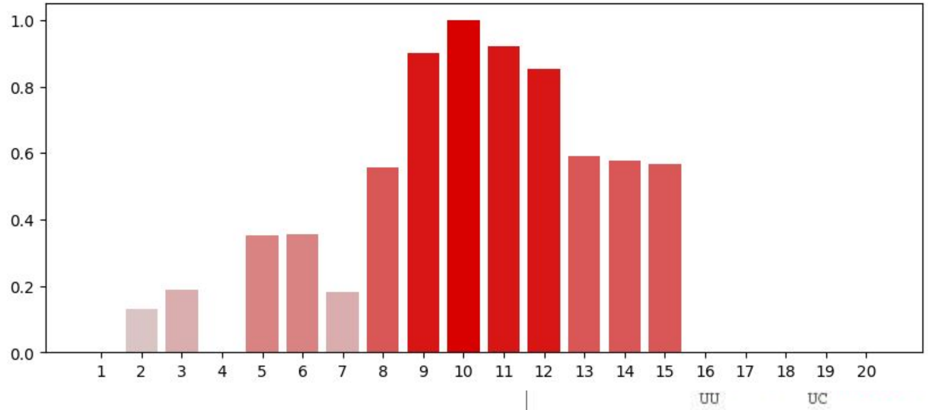
Julius Brennecke[©], Alexander Stark[©], Robert B. Russell, Stephen M. Cohen[*]

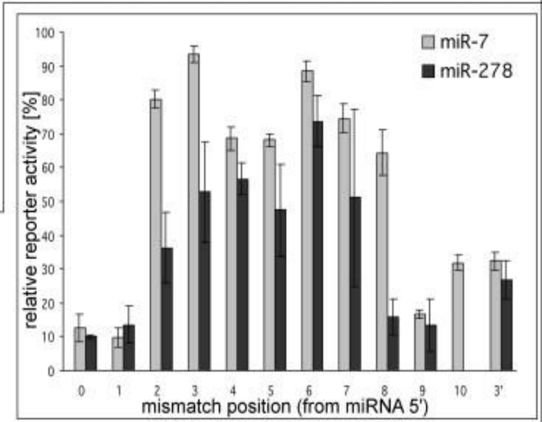
European Molecular Biology Laboratory, Heidelberg, Germany

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.

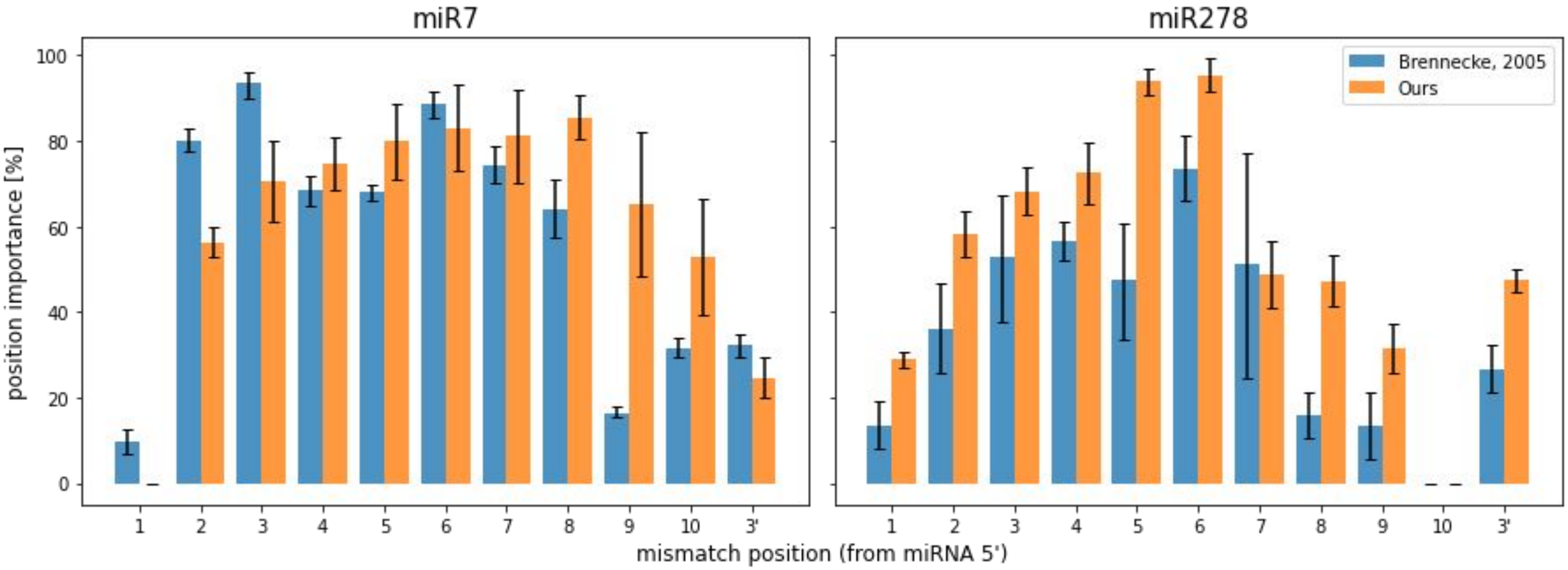
Citation: Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of r







Verification



miRNA	miR-7	miR-278
correlation	0.59	0.85

Scanning

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

Scanning


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

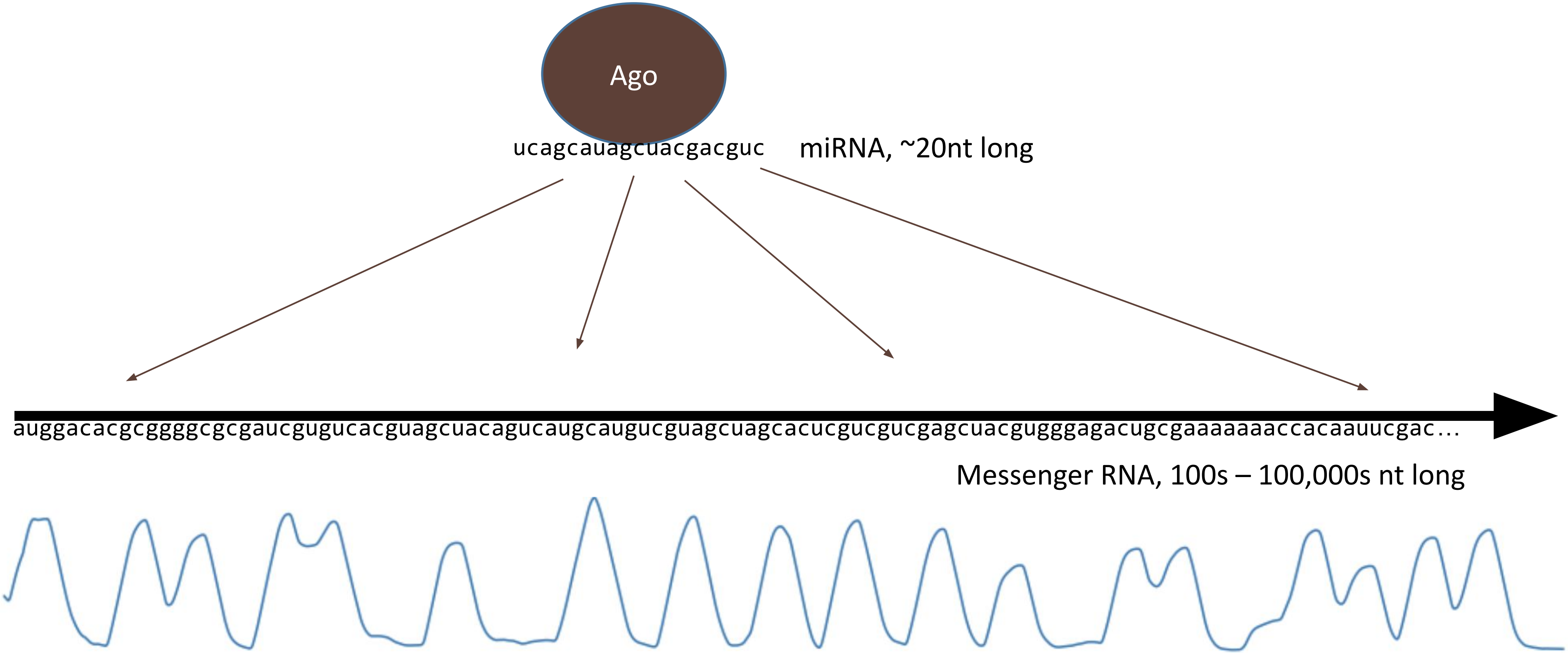
Scanning

Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long
0.52

Scanning

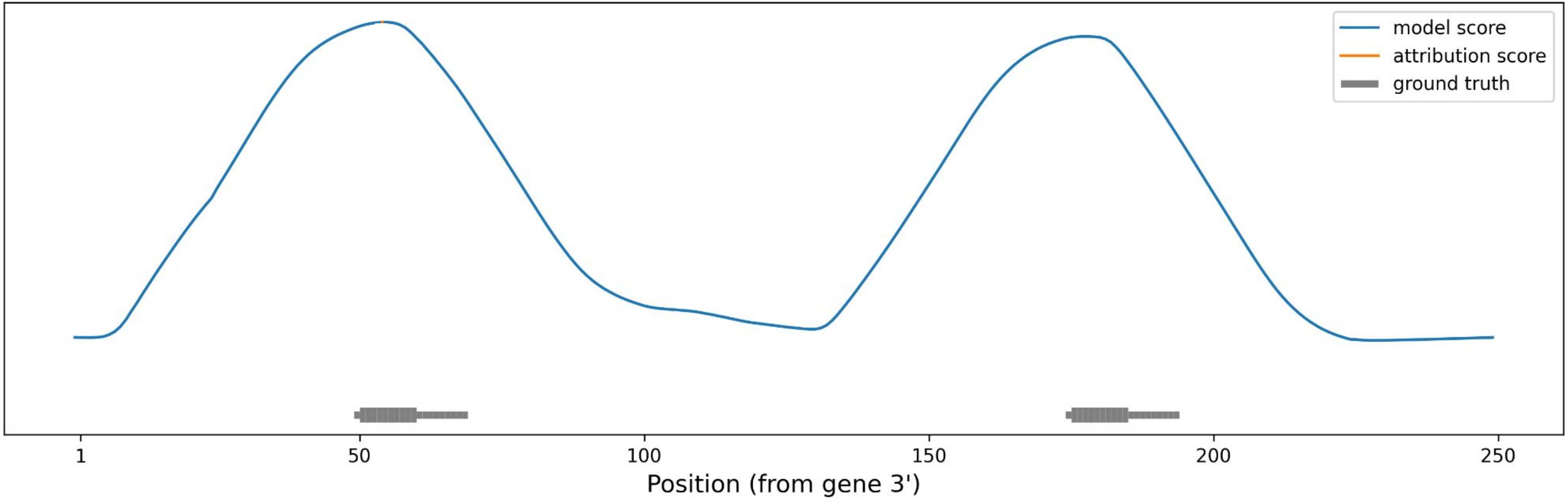

Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long
0.18

Scanning

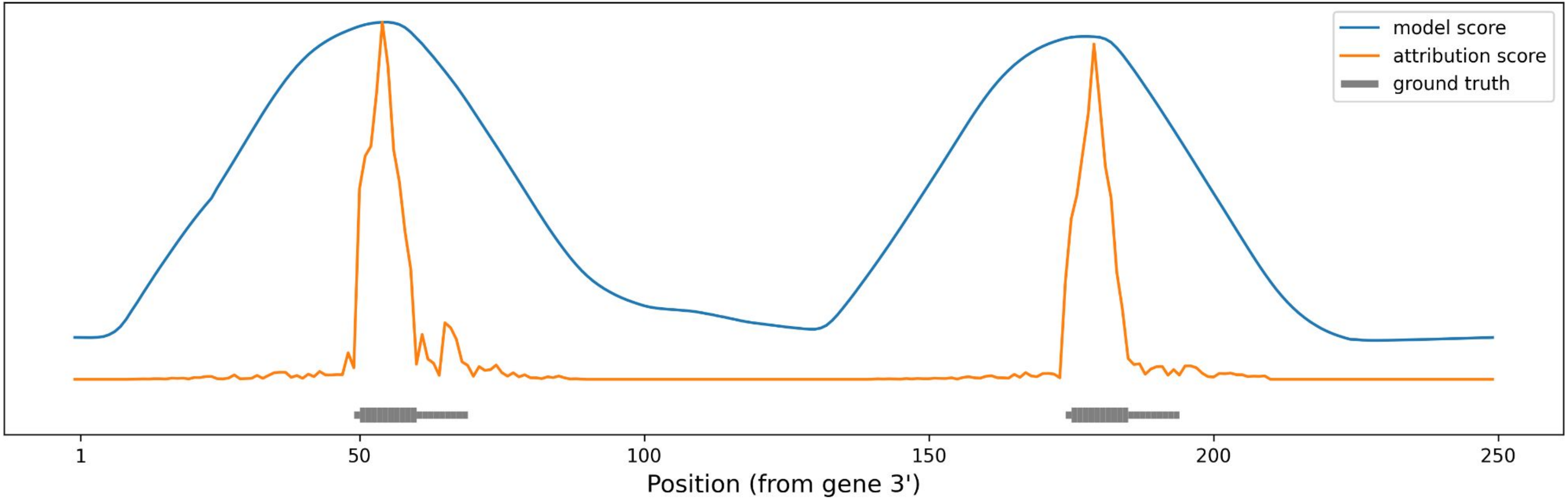


Ago
ucagcauagcuacgacguc
miRNA, ~20nt long
auggacacgcggggcgcgaucgugucacguagcuacagucaugcaugucguagcuagcacucgucgucgagcuacgugggagacugcgaaaaaaccacaauucgac…
Messenger RNA, 100s – 100,000s nt long

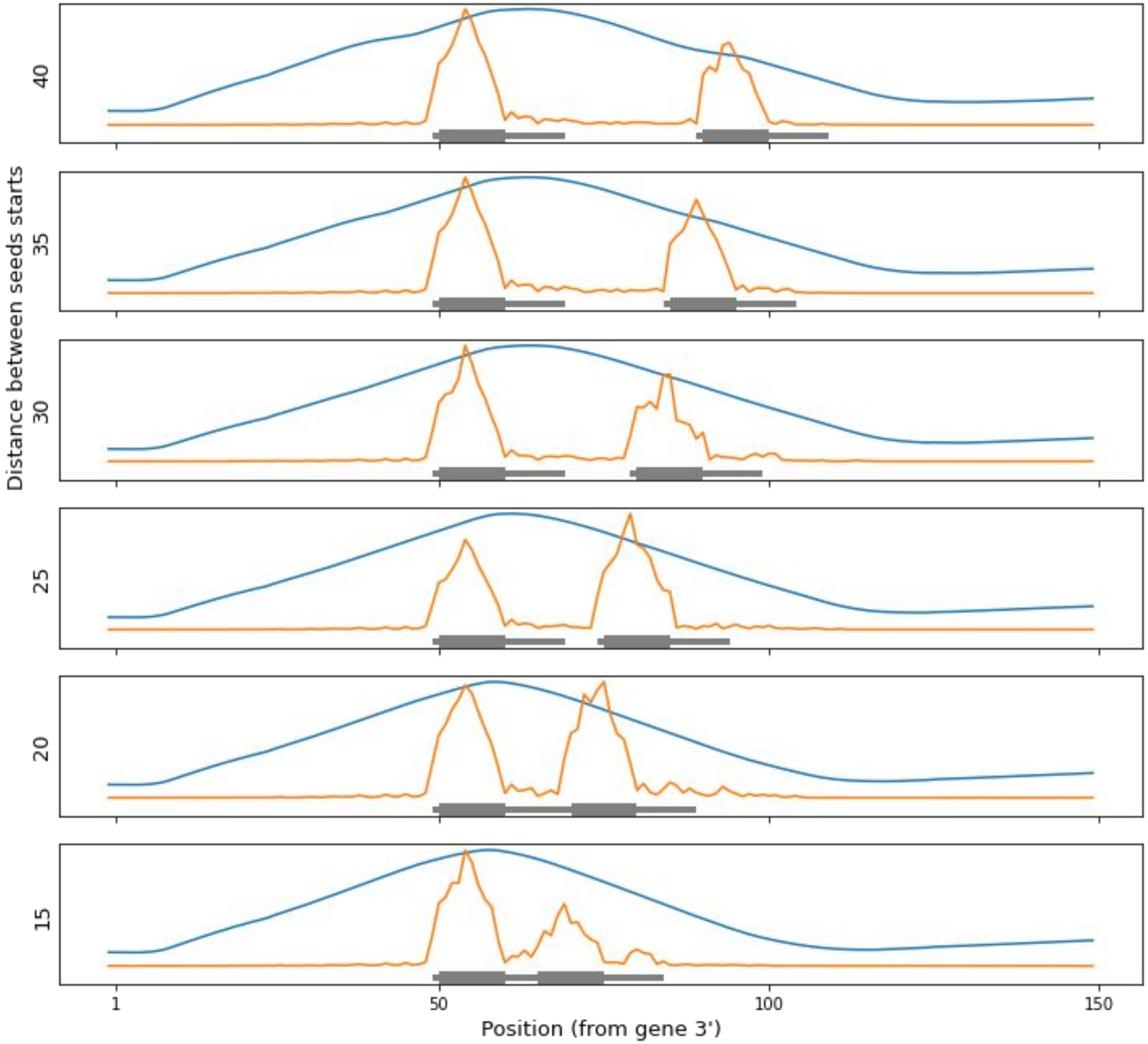
Narrowing the peaks



Narrowing the peaks



Close by peaks



Hands on: Using DeepExperiment to interpret and visualize miRNA targeting

Open the Colab notebook

[Hands on: Using DeepExperiment to interpret and visualize miRNA targeting](https://colab.research.google.com/drive/1lelArVN_BJ4P9Uex3yhB8hM3MfEPGay2?usp=sharing)
(https://colab.research.google.com/drive/1lelArVN_BJ4P9Uex3yhB8hM3MfEPGay2?usp=sharing)

Conclusions

- There are many techniques for interpreting neural networks
- They use different principles and produce different results
- Personal tip: don’t use just one, try multiple interpretation techniques

Thank you for your attention!

Deep Neural Networks are like a complex organisms and interpretation techniques help us perform experiments to better understand them.

https://katarinagresova.github.io