



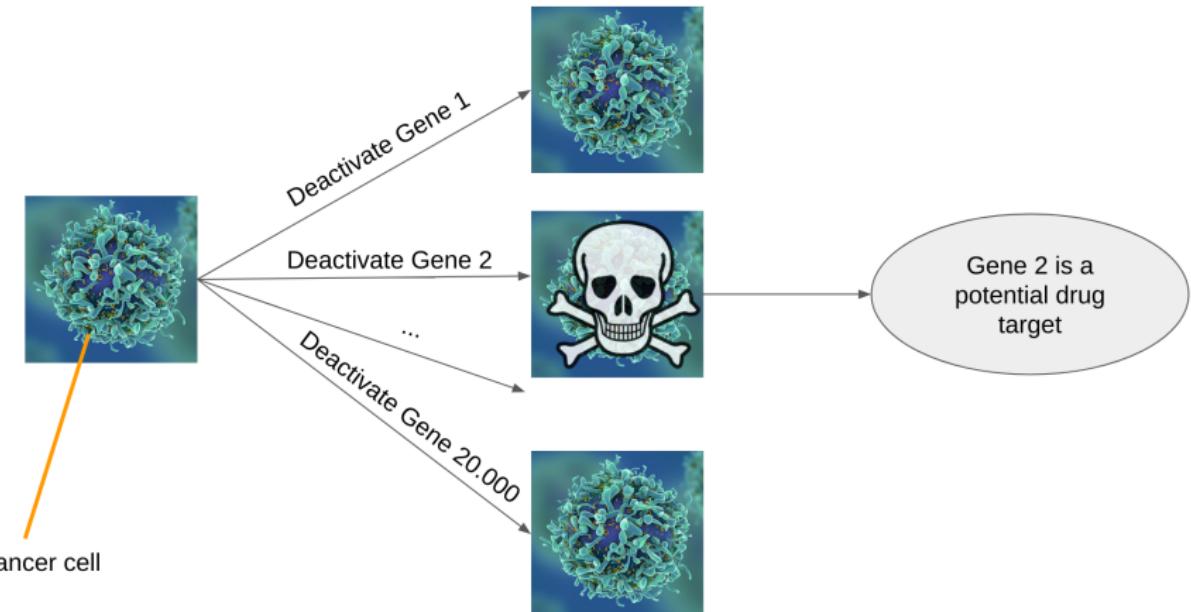
## PAVOOC - An AI integrated web-app for CRISPR target recommendation

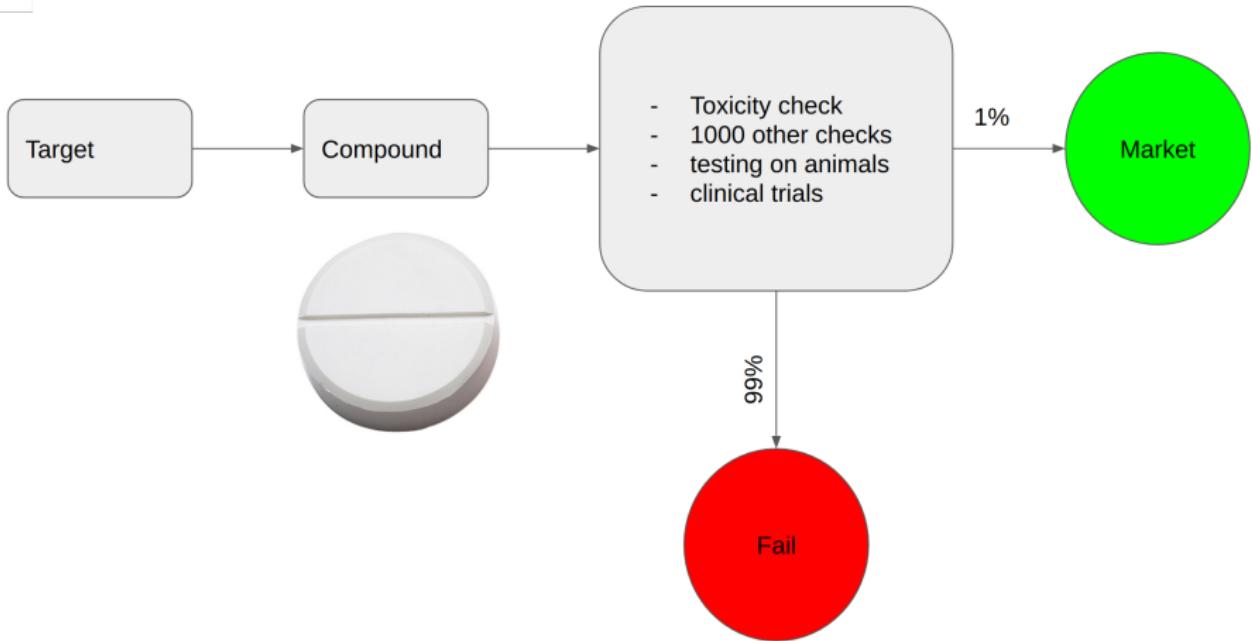
Moritz Schäfer | Technische Universität Berlin & Bayer Pharma | Prediction and visualization of on- and off-targets for CRISPR

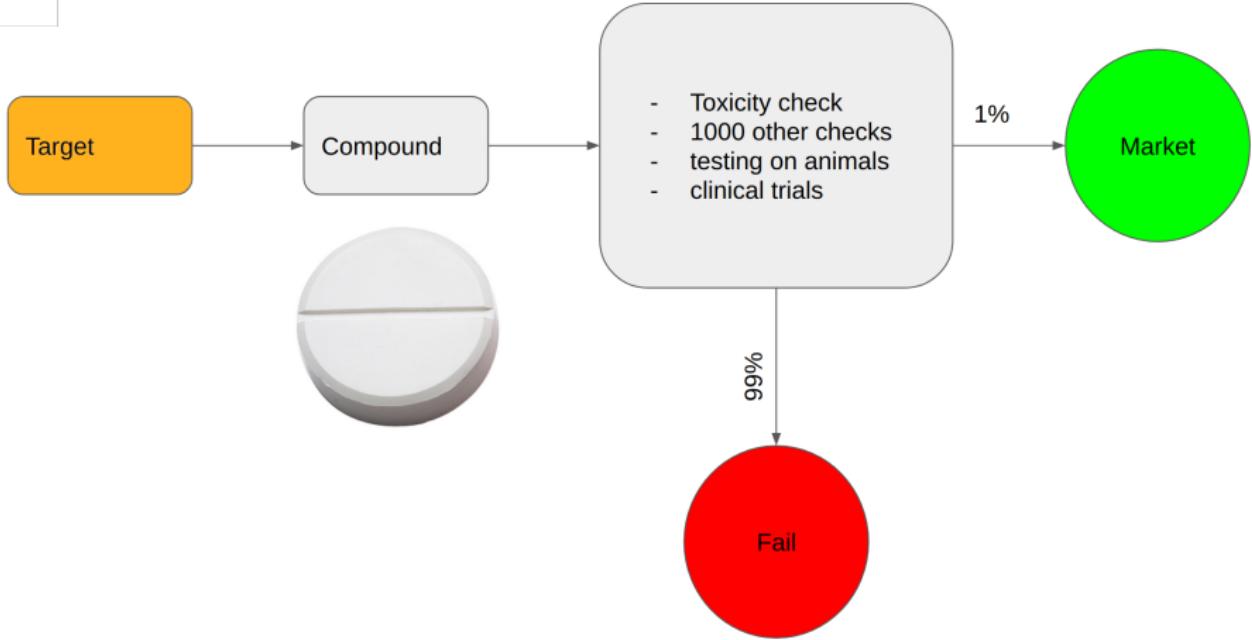




## Target discovery

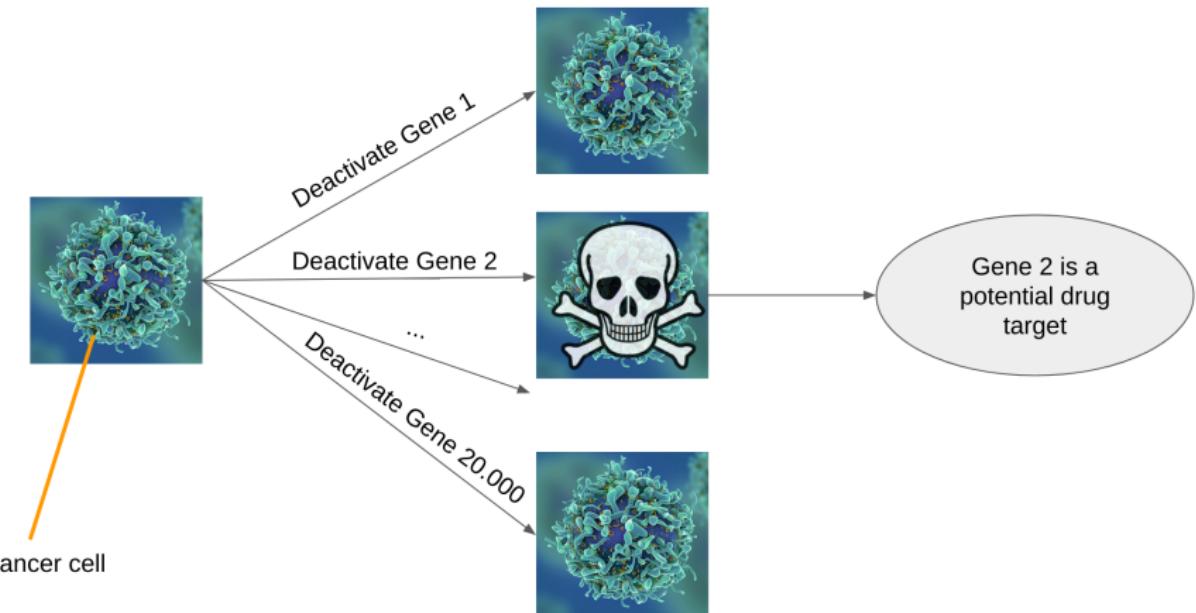








## Target discovery





## How to knock out genes?

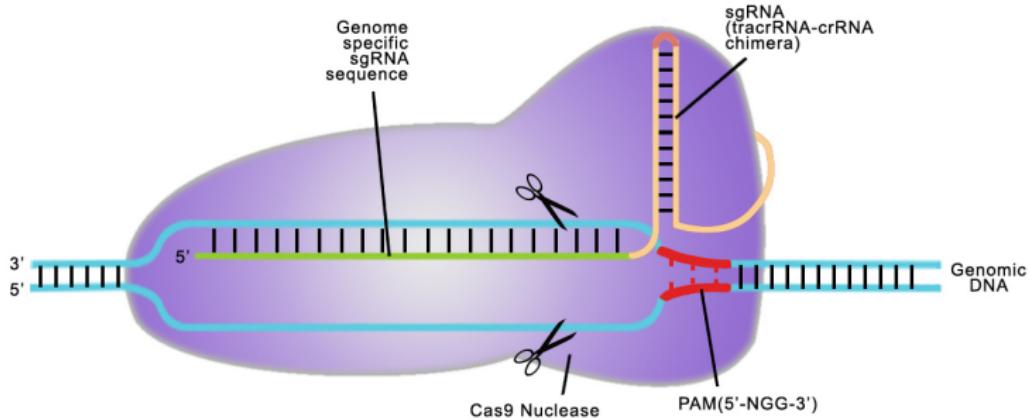


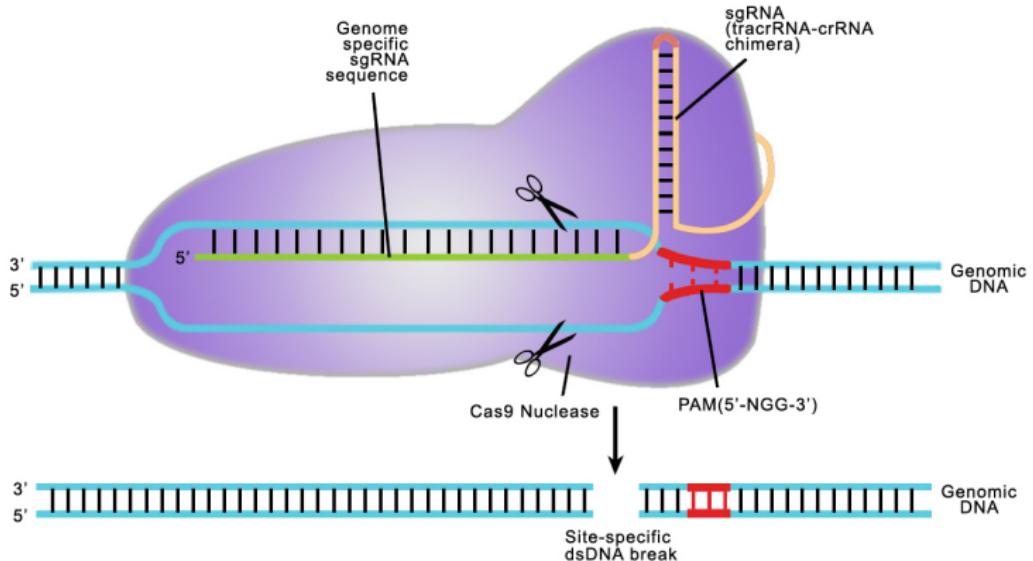
How to knock out genes?

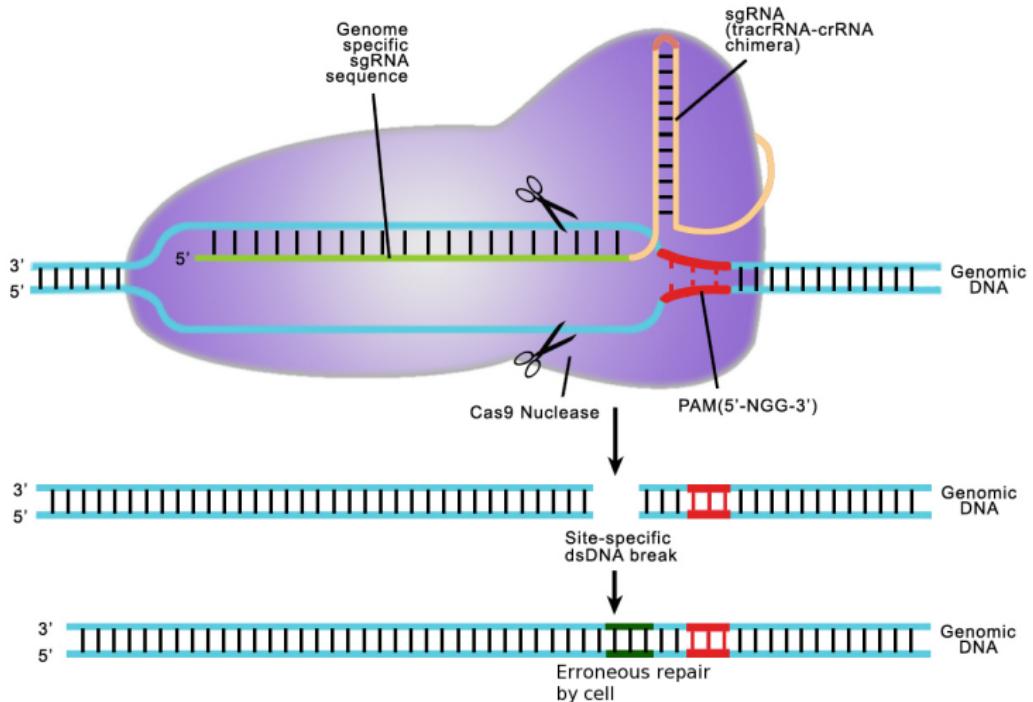
# CRISPR



Scale 1:1,000,000







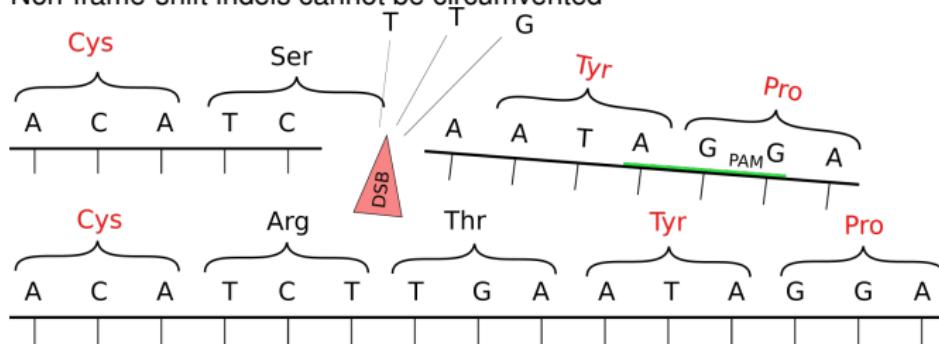


## Problems

- Guide performance varies significantly

## Problems

- Guide performance varies significantly
- Non-frame-shift indels cannot be circumvented





## Solution

- Cutting-edge guide efficacy scoring



## Solution

- Cutting-edge guide efficacy scoring
- Web based guide design tool



## Live Demo

<http://pavooc.me>

## Guide efficacy prediction – Dataset

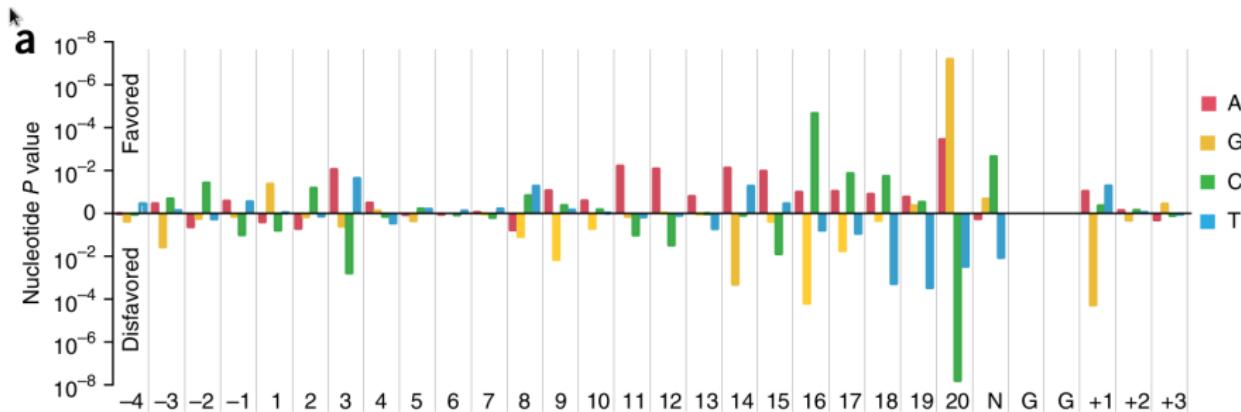
Guide	Measured efficacy
GTTAGGGTCCGTACTCAGCAAGG	0.86
ACACTGCCGAGCGATGAGGATGG	0.42
AAGGTGAAGGAGGATGCGGCGGG	0.53
GAAAAGATAAGTCAGTGACCCGG	0.12
GCAAGTCACTGAGTGCAGAACGG	0.73
GCATTGTAAGCGCACAGGAAGG	0.70
AAGACTGGCGCATGGTCCACTGG	0.57
...	...

- 1,837 data rows from 2014
- 3,473 data rows from 2016
- Efficacy relates to cell proliferation after CRISPR application

"Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9", 2016, John G. Doench et al.



## Guide efficacy prediction – 2014



"Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation", 2014, John G. Doench et al.



## Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features



## Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features

ACTATCTATCGTACGA**TT**GA



## Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features

ACTATCTATCGTACGA**TT**GA

ACTATCTATCGTACGAC**AAG**

"Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9", 2016, John G. Doench et al.



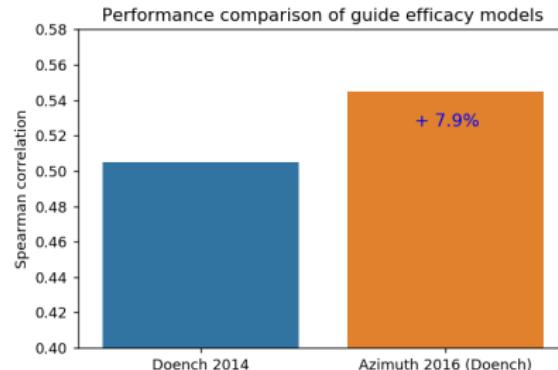
## Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features

ACTATCTATCGTACGA**TT**GA

ACTATCTATCGTACGAC**AAG**

"Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9", 2016, John G. Doench et al.





## Convolutional neural networks

Aim: Finding spatial patterns



## Convolutional neural networks

Aim: Finding spatial patterns



10x10 Filter





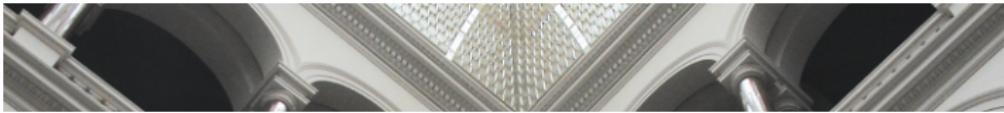
## Convolutional neural networks

[A C A A]



# Convolutional neural networks

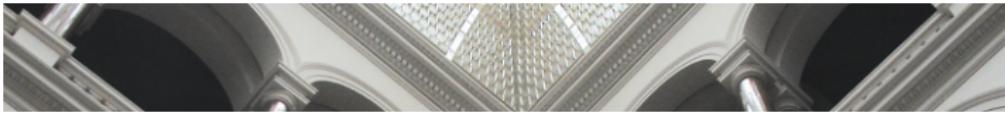
[A C A A]



## Convolutional neural networks

Well performing guides:

GTAGGGGTCCGTACTCAGCA  
CAGGGTCCGTACTCAGAGGA  
CTAGCGTAGAGCGCACTGCA  
ACTGAGCTAGCGTAGAAGCA  
TGAGCTAGCGTAGAGCACCA  
ACTGAGCTAGCGTAGTAGCT  
AGCGTAGAGCGCGCTGCC  
GAGCGCACTGAGCTAGAGAA  
ATAGAGCGCCTGAGCTCGCA  
CGTAGAGCGCACTGAGAGCT



## Convolutional neural networks

Well performing guides:

AGCA

AGGA

TGCA

AGCA

ACCA

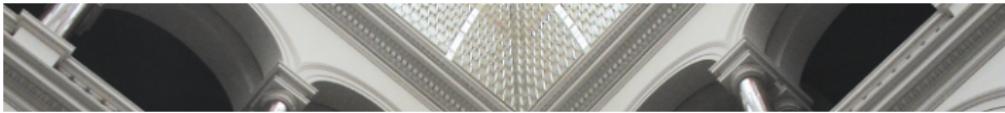
AGCT

CGCC

AGAA

CGCA

AGCT



## Convolutional neural networks

Well performing guides:

AGCA

AGGA

TGCA

AGCA

ACCA

AGCT

CGCC

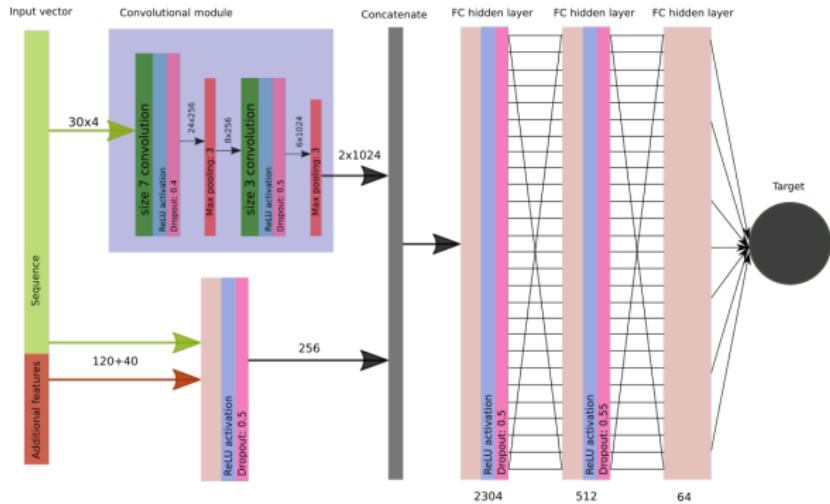
AGAA

CGCA

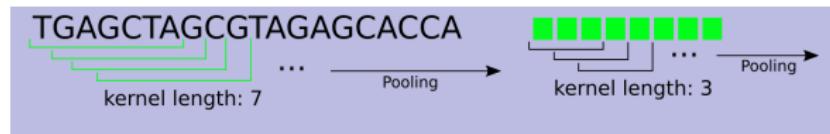
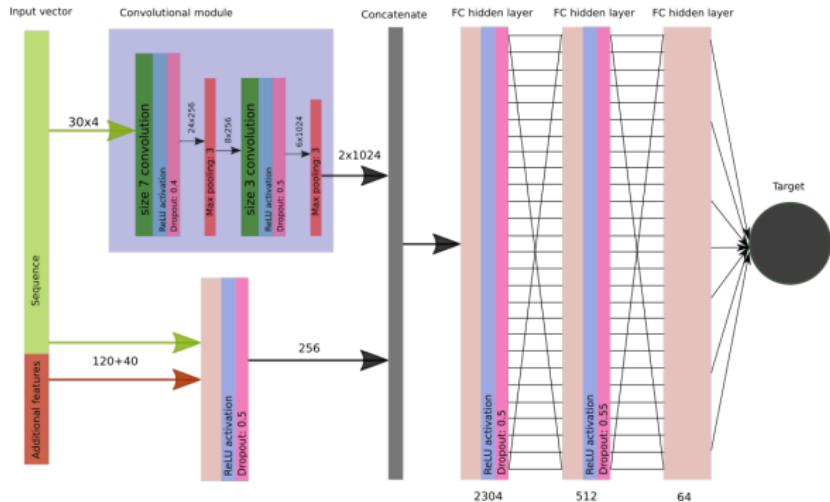
AGCT

Learned filter: [A G C A]

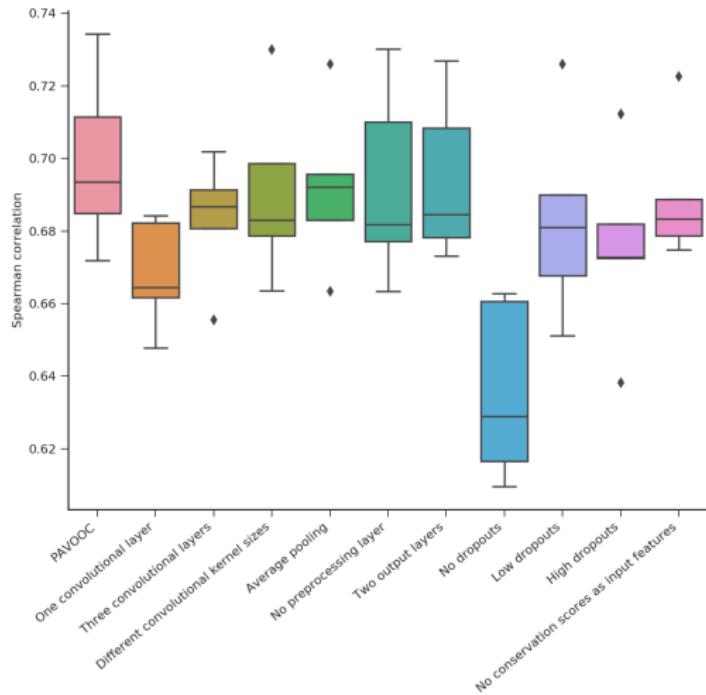
## Model architecture



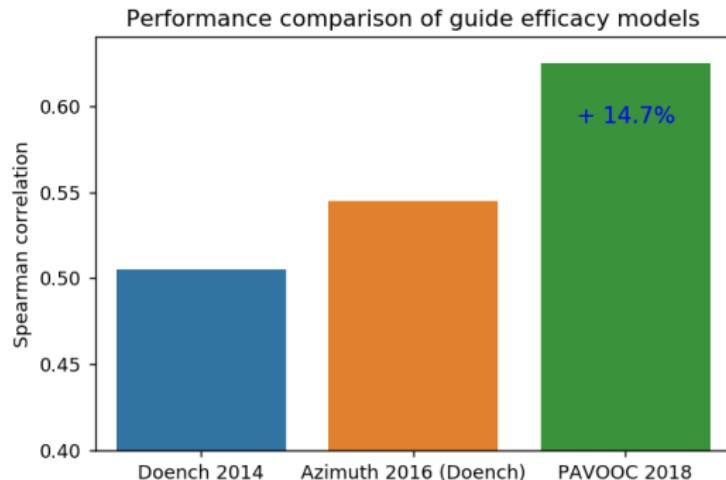
## Model architecture



## Model optimization

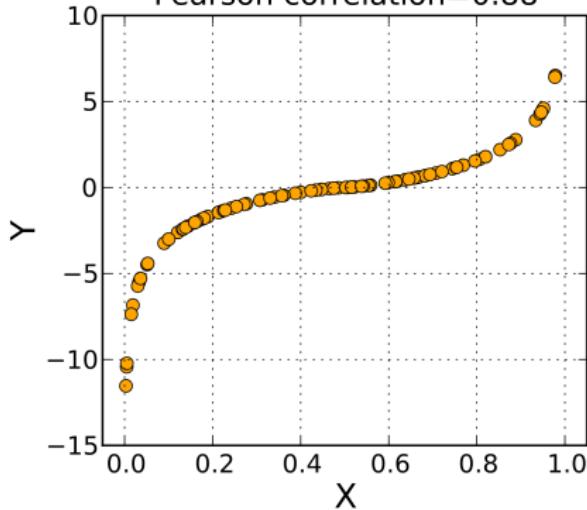


## Model architecture



## Spearman correlation

Spearman correlation=1  
Pearson correlation=0.88





## Spearman correlation

- Test set labels: [0.2, 0.3, 0.5, 0.9]
- Test set predictions: [0.4, 0.6, 0.7, 0.8]
- Spearman correlation: 1.0



## Spearman correlation

- Test set labels: [0.2, 0.3, 0.5, 0.9]
- Test set predictions: [0.4, 0.6, 0.7, 0.8]
- Spearman correlation: 1.0

### Advantages:

- Ranking mimicks the task of guide selection

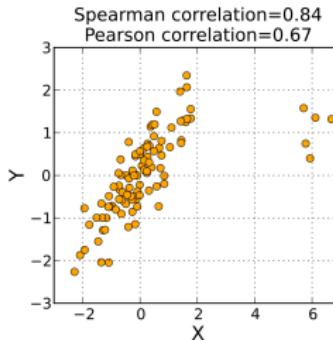


## Spearman correlation

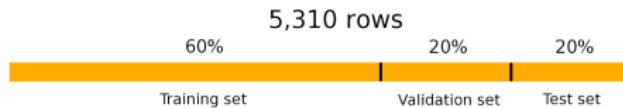
- Test set labels: [0.2, 0.3, 0.5, 0.9]
- Test set predictions: [0.4, 0.6, 0.7, 0.8]
- Spearman correlation: 1.0

### Advantages:

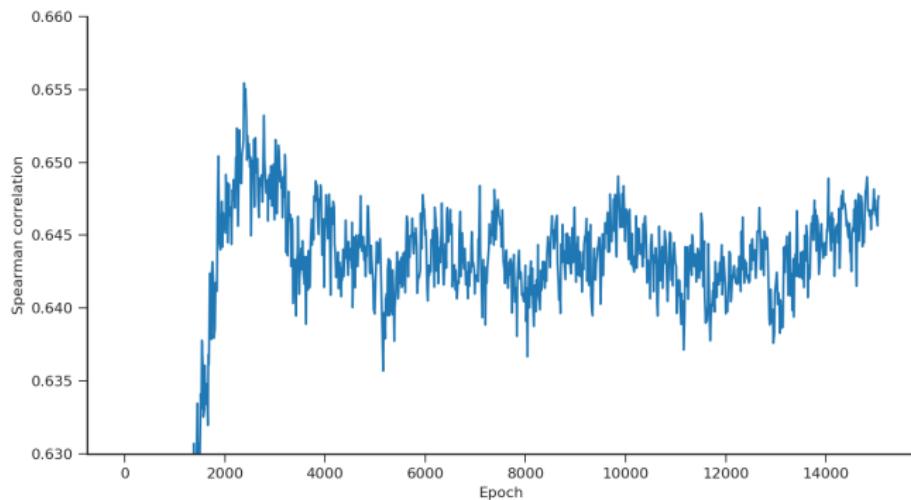
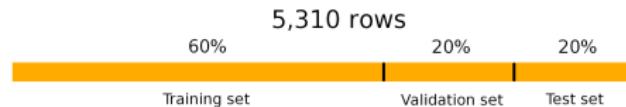
- Ranking mimicks the task of guide selection
- Robust against outliers

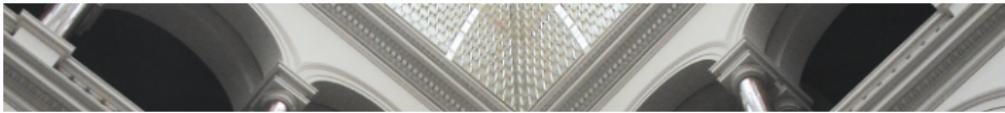


## Performance evaluation



## Performance evaluation





## Conclusion and Takeaways

- Cas9 efficacy depends on complex biological coherences
- DL improves guide efficacy prediction
- DL feasible with ~5,000 rows



## Acknowledgements

- Prof. Manfred Opper (supervisor at TU Berlin)
- Dr. Andreas Steffen (supervisor at Bayer)
- Dr. Djork-Arne Clevert (machine learning scientist at Bayer)
- Robin Winter (PhD student at Bayer)
- Bayer Pharma AG



## Future Work & Discussion

- Additional input features (chromatin accessibility)
- Evaluate model on different datasets
- Support additional species