



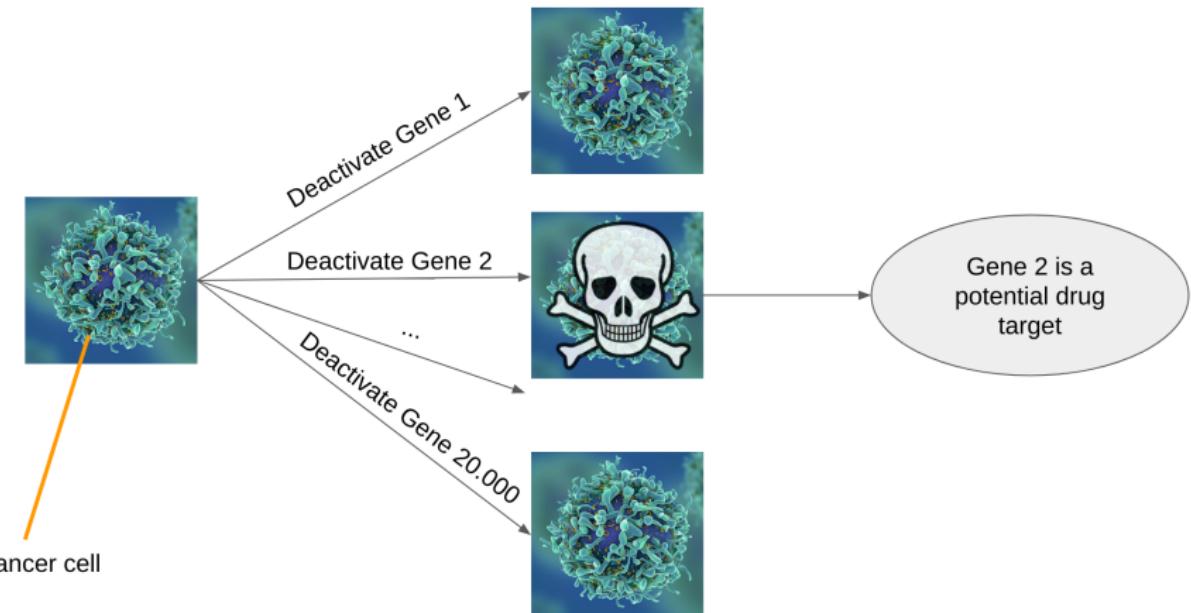
## PAVOOC - An AI integrated web-app for CRISPR target recommendation

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





## Target discovery



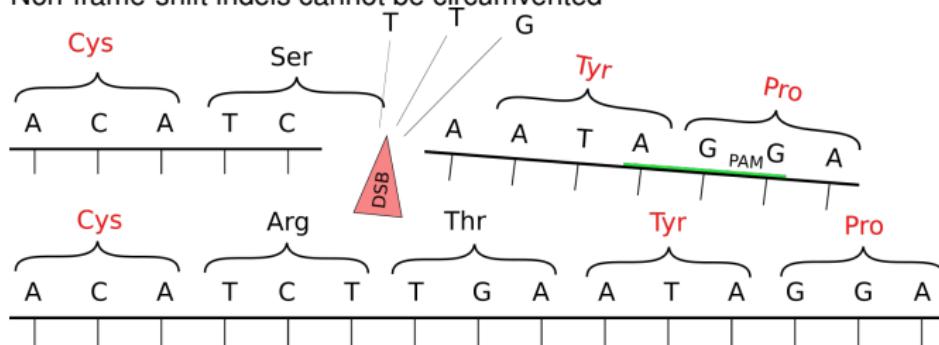


## CRISPR problems

- Guide performance varies significantly

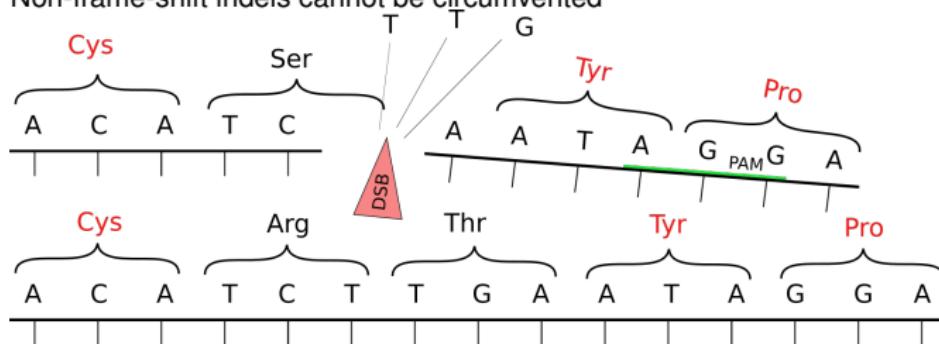
## CRISPR problems

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



## CRISPR problems

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



- Cancer mutations may affect binding site

AGCATCGTAAGTGAATTACGG  
 - + PAM

DMS153 lung cancer cellline SNP



## Solution

- Cutting-edge guide efficacy scoring



## Solution

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



## Solution

- Cutting-edge guide efficacy scoring
- Web based guide design tool
- Guide filtering for (cancer induced) single nucleotide variations



## Live Demo

<https://pavooc.me>

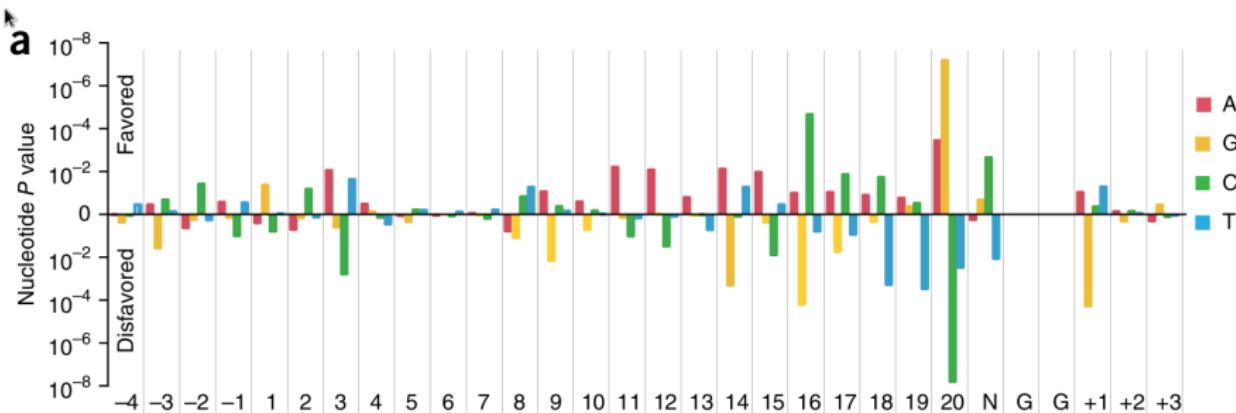
## Guide efficacy prediction – Dataset

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.



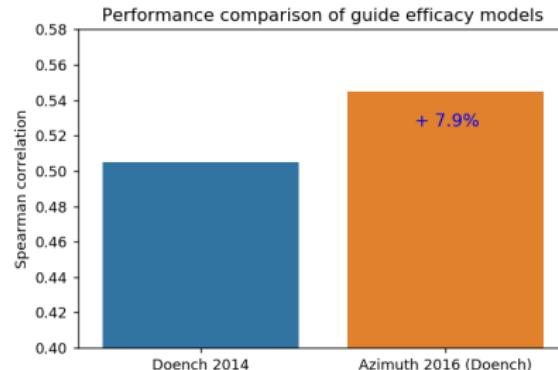
## Guide efficacy prediction – 2016 (Azimuth)

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.



## Guide efficacy prediction – Dataset

Guide	Measured efficacy
GTTAGGGGTCCGTACTCAGCAAGG	0.86
ACACTGCCGAGCGATGAGGATGG	0.42
AAGGTGAAGGAGGATGCGGCGGG	0.53
GAAAAGATAAGGTCACTGACCCGG	0.12
GCAAGTCACTGAGTGCAGAACGG	0.73
GCATTGGTAAGCGCACAGGAAGG	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

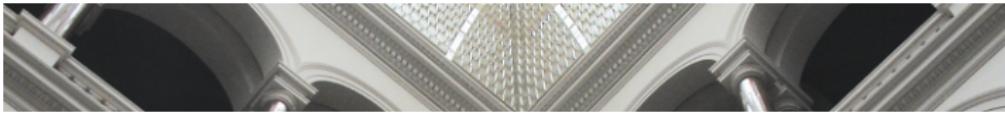
"Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9", 2016, John G. Doench et al.



## 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

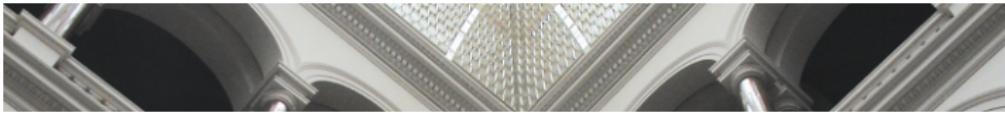
AGCT

CGCC

AGAA

CGCA

AGCT



## Convolutional neural networks

Well performing guides:

AGCA

AGGA

TGCA

AGCA

ACCA

AGCT

CGCC

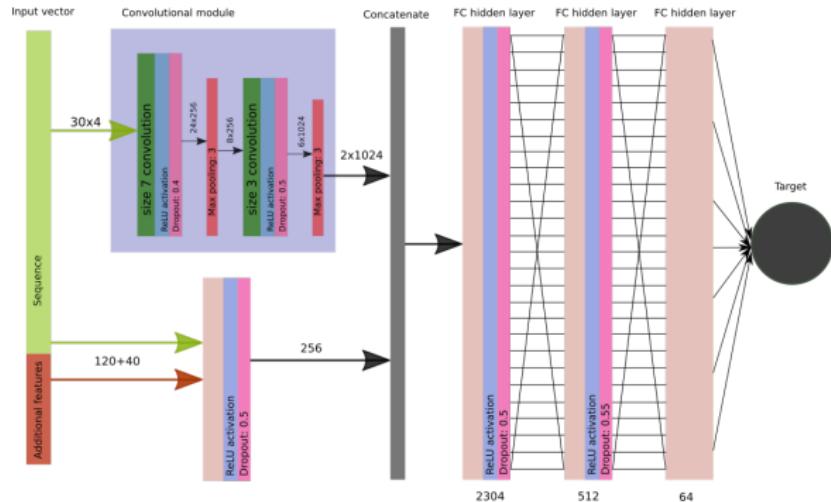
AGAA

CGCA

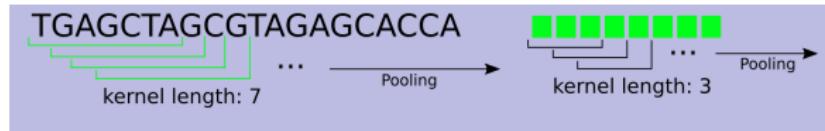
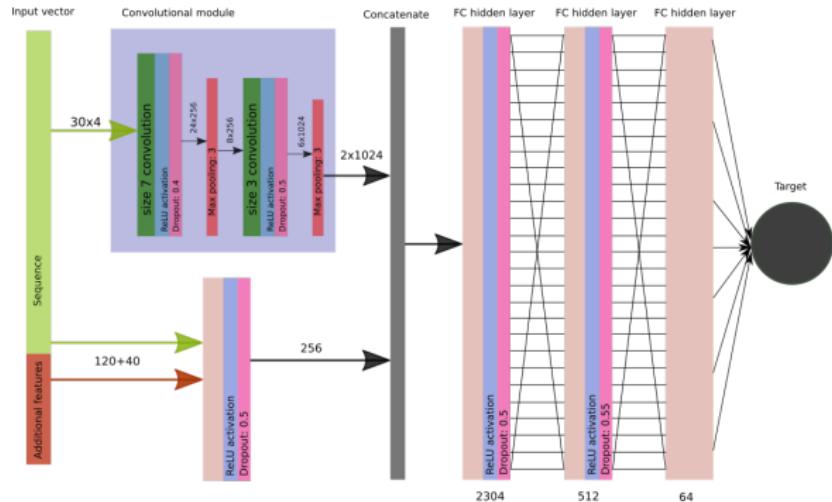
AGCT

Learned filter: [A G C A]

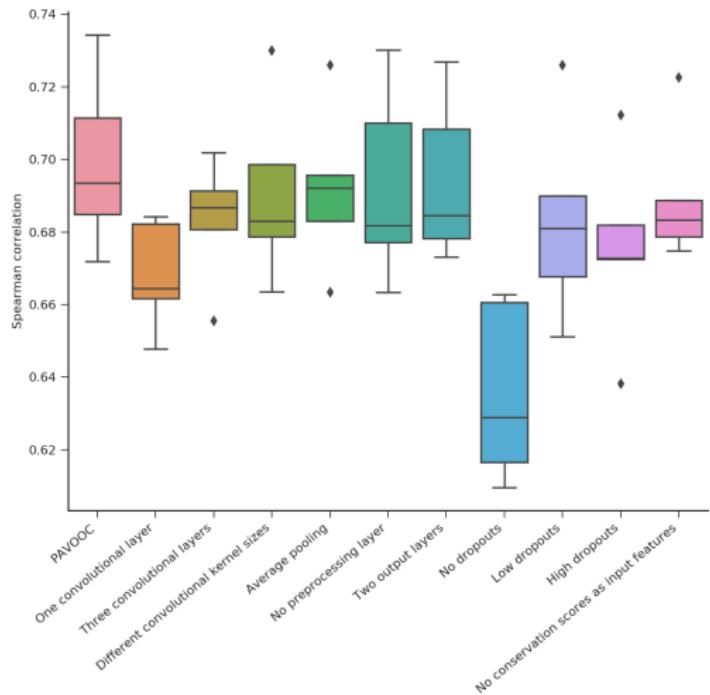
## Model architecture



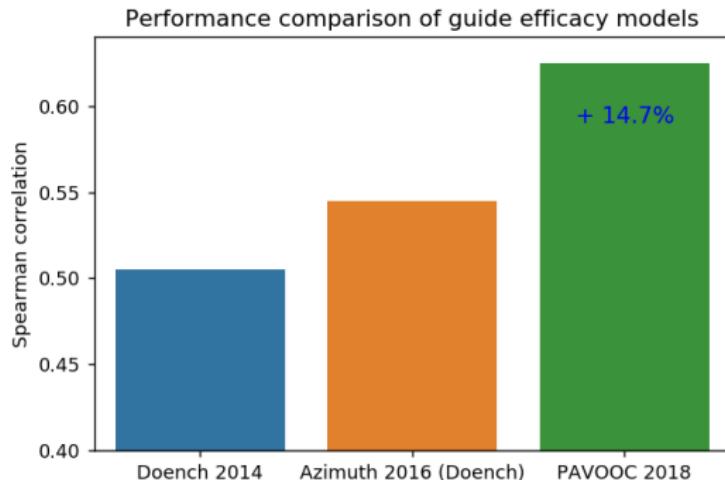
## Model architecture



## Model optimization



## Model architecture



## Conclusion and Takeaways

- PAVOOC provides amino acid level bindings
- Cas9 efficacy depends on complex biological coherences
- DL improves guide efficacy prediction
- DL feasible with ~5,000 rows

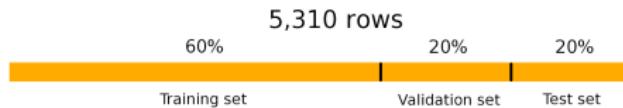
Disclaimer: My model is not employed in the website for publication reasons.



## Future Work & Discussion

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

## Performance evaluation



## Performance evaluation

