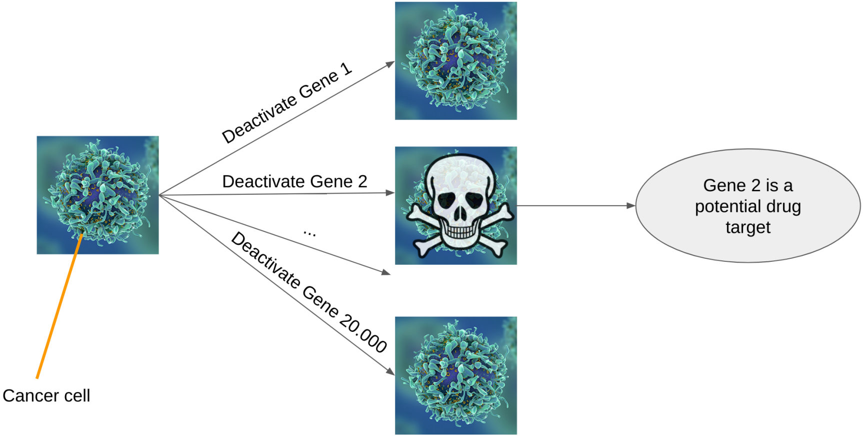
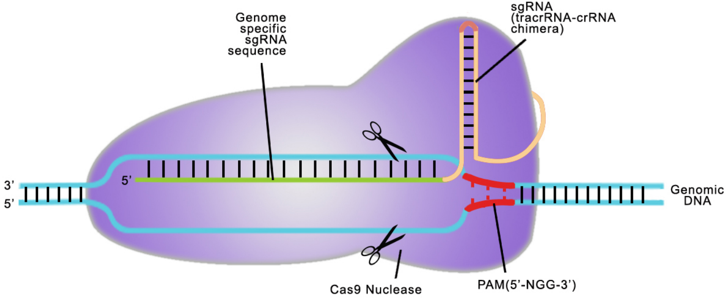
PAVOOC - An AI integrated web-app for CRISPR target recommendation

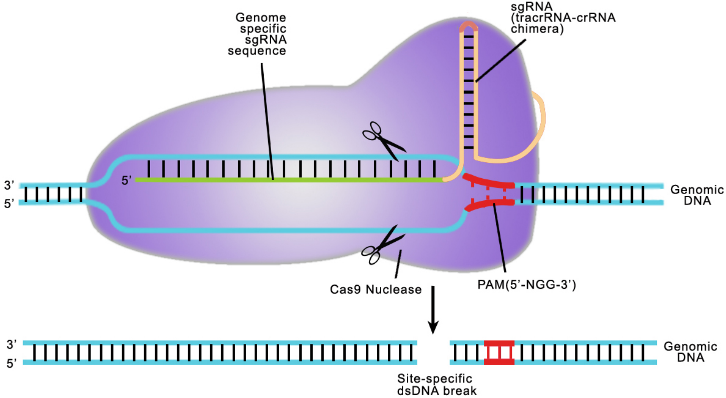
Moritz Schäfer | Technische Universität Berlin & Bayer Pharma | Prediction and visualization of on- and off-targets for CRISPR

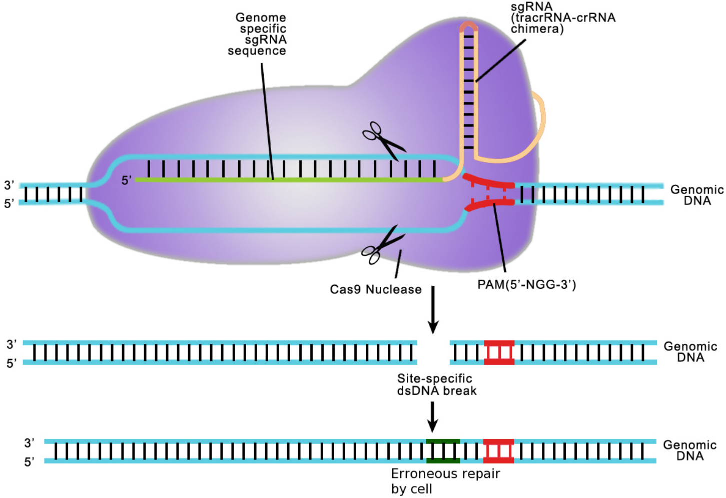

Bayer Schering Pharma

Target discovery



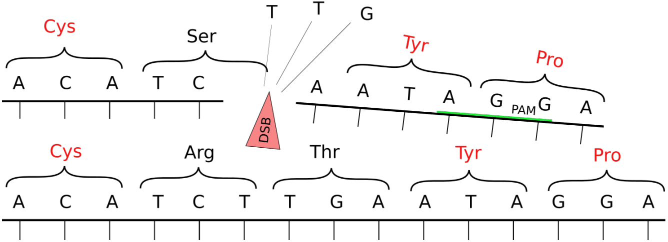

Genome
specific
sgRNA
sequence
sgRNA
(tracrRNA-crRNA
chimera)
3'
5'
5'
Genomic
DNA
Cas9 Nuclease
PAM(5'-NGG-3')


Genome specific sgRNA sequence
sgRNA (tracrRNA-crRNA chimera)
3'
5'
5'
Genomic DNA
Cas9 Nuclease
PAM(5'-NGG-3')
3'
5'
Genomic DNA
Site-specific dsDNA break


Genome specific sgRNA sequence
sgRNA (tracrRNA-crRNA chimera)
3'
5'
5'
Genomic DNA
Cas9 Nuclease
PAM(5'-NGG-3')
3'
5'
Genomic DNA
Site-specific dsDNA break
3'
5'
Genomic DNA
Erroneous repair by cell

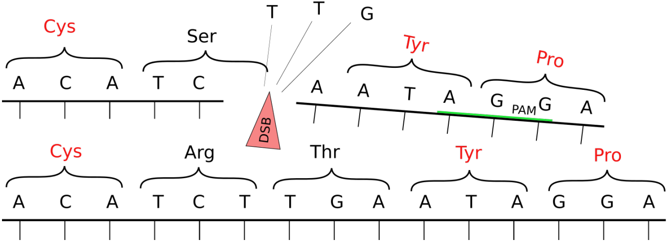
Problems

– Non-frame-shift indels cannot be circumvented



Problems

– Non-frame-shift indels cannot be circumvented



– Guide performance varies significantly

Solution

- Cutting-edge guide efficacy scoring

Solution

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

Guide efficacy prediction – Dataset

Guide	Measured efficacy
GTAGGGGTCCGTACTCAGCAAGG	0.86
ACACTGCCGAGCGATGAGGATGG	0.42
AAGGTGAAGGAGGATGCGGCGGG	0.53
GAAAAGATAGGTCACTGACCCGG	0.12
GCAAGTCACTGAGTGCAGAACGG	0.73
GCATTGGTAAGCGCACAGGAAGG	0.70
AAGACTGGCGCATGGTCCACTGG	0.57
...	...

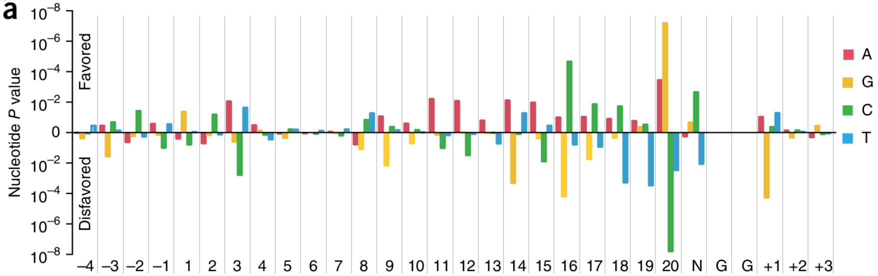
– 1,837 data rows from 2014

– 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.

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

ACTATCTATCGTACGATTGA

Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features

ACTATCTATCGTACGATTGA

ACTATCTATCGTACGACAAG

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

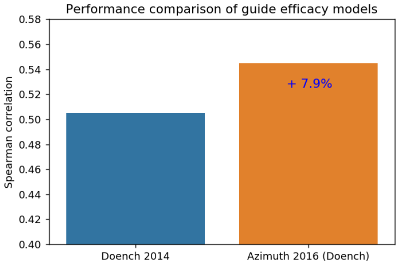
Guide efficacy prediction – 2016 (Azimuth)

Pairwise nucleotide features

ACTATCTATCGTACGATTGA

ACTATCTATCGTACGACAAG

"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

AGCGTAGAGCGCGCTGCGCC

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

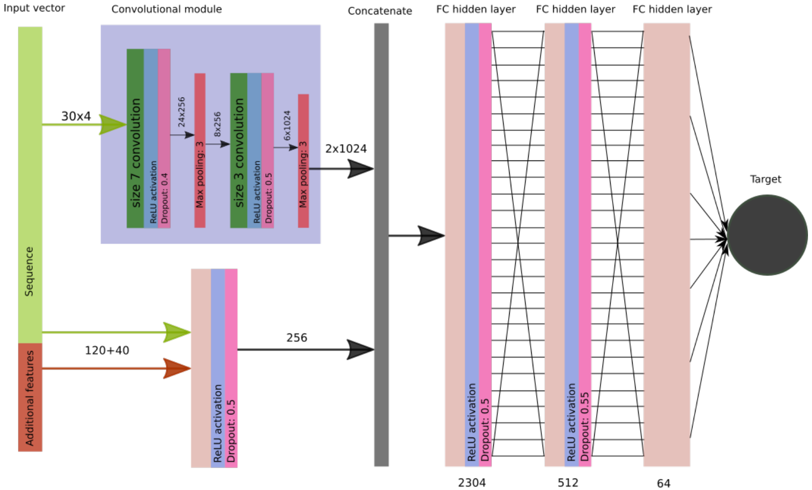
AGAA

CGCA

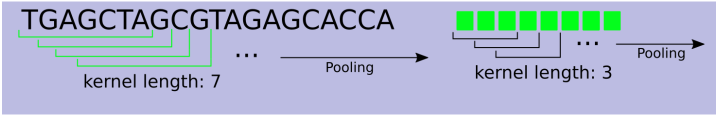
AGCT

Learned filter: [A G C A]

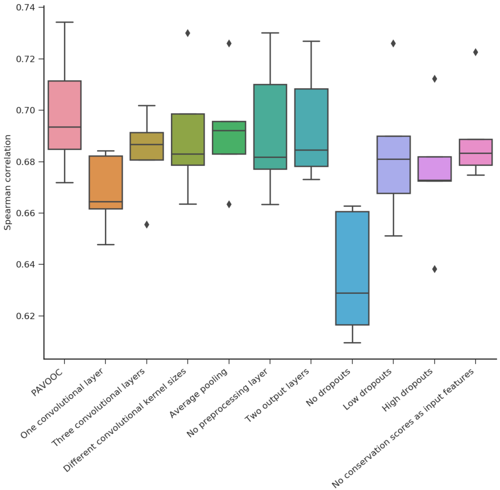
Model architecture



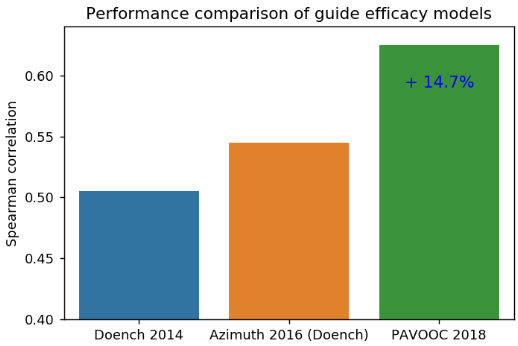
Model architecture – Convolution module



Model optimization



Model architecture



Technische Universität Berlin

Conclusion and Takeaways

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

Future Work & Discussion

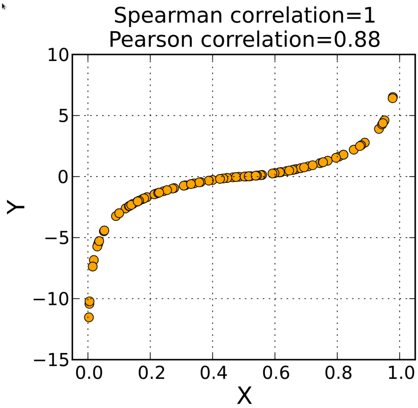
- Additional input features (chromatin accessibility)
- Evaluate model on different datasets
- Support additional species

Live Demo

http://pavooc.me

End

Spearman correlation



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:

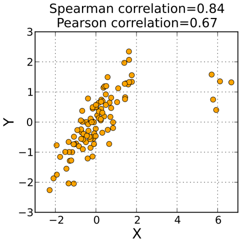
- Ranking mimicks the task of guide selection

Spearman correlation

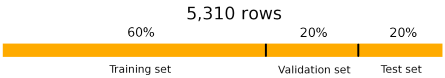
– Test set labels: [0.2, 0.3, 0.5, 0.9]
– Test set predictions: [0.4, 0.6, 0.7, 0.8]
– Spearman correlation: 1.0

Advantages:

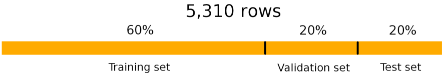
– Ranking mimicks the task of guide selection
– Robust against outliers

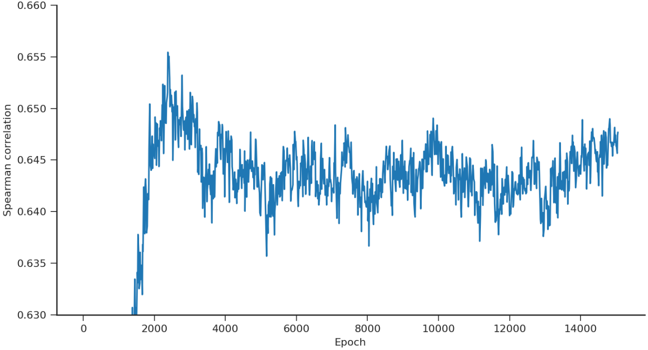


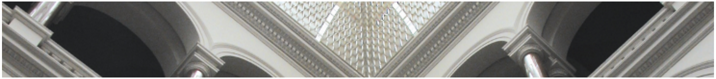
Performance evaluation



Performance evaluation







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