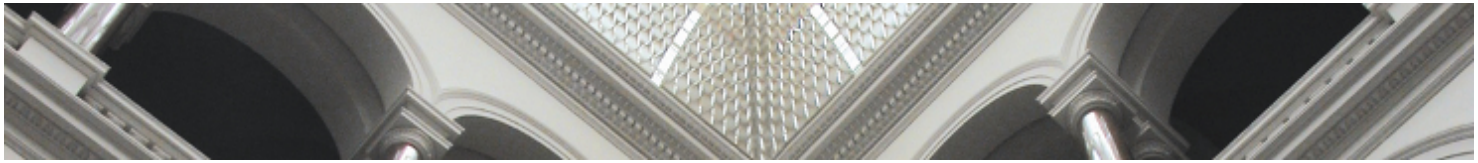
Technische Universität Berlin

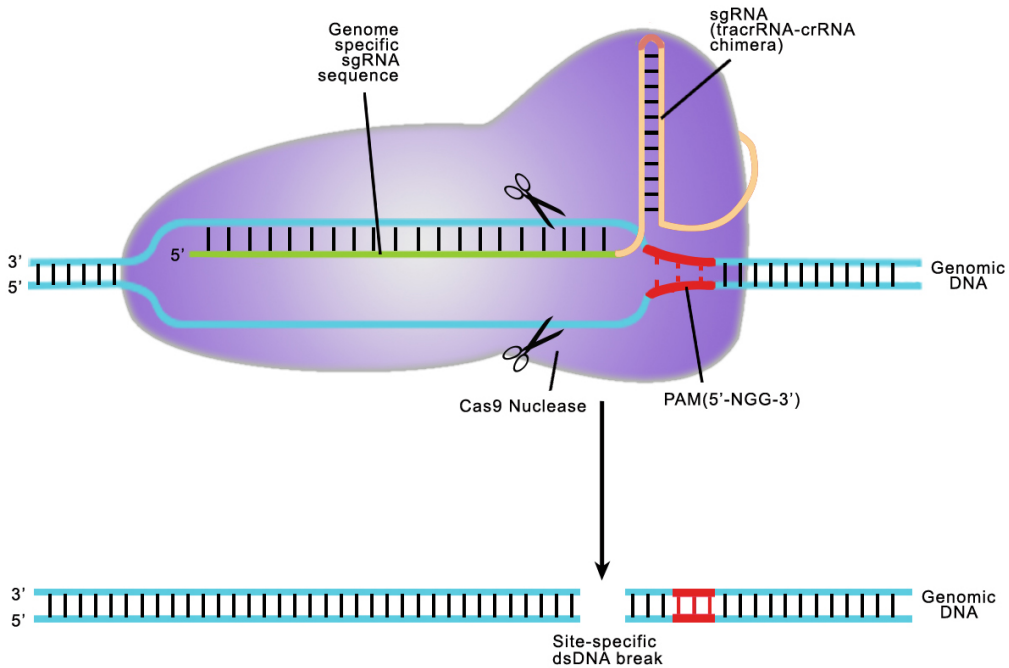
PAVOOC - Prediction and visualization of on- and off-targets for CRISPR

Moritz Schäfer | Technische Universität Berlin & Bayer Pharma | Master thesis

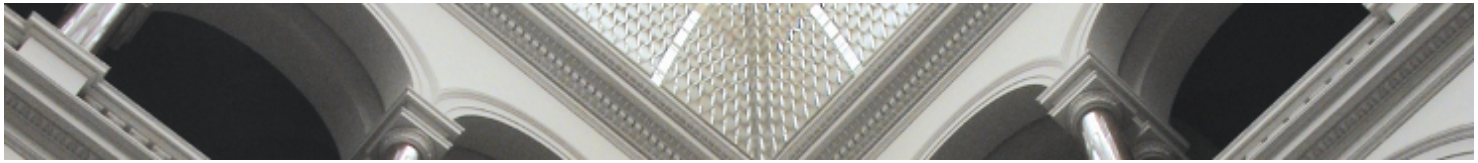


Technische Universität Berlin

Background



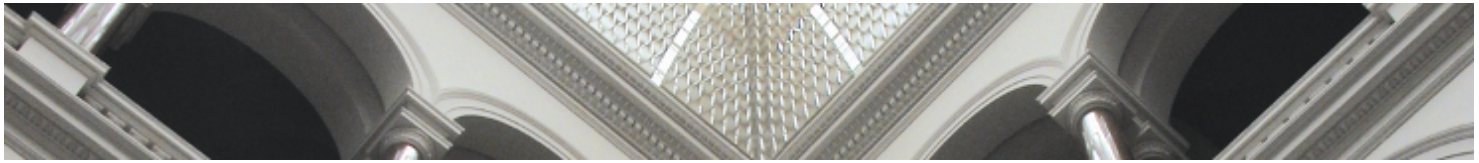
- Knockout experiments used in drug target validation
- Sequence (partially) determines efficacy



Technische Universität Berlin

Problem

– Guide prediction scores still vary in performance

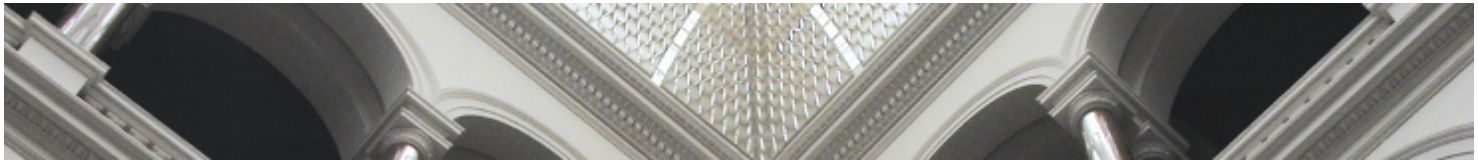


Technische Universität Berlin

Problem

- Guide prediction scores still vary in performance
- A lot of manual labor for guide selection

19	POLR2L	28,5	28	-3,00752118182 82
20	SPC24	29,5	42	-2,83403119872683
21	THAP1	31,5	27	-3,01155447143932
22	CDC123	32,5	44	-2,81699463079959
23	WDR74	37	68	2,58033320817648



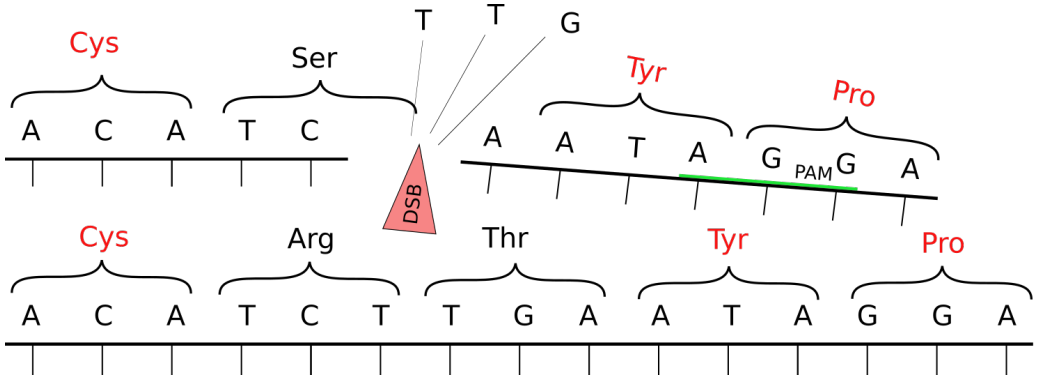
Technische Universität Berlin

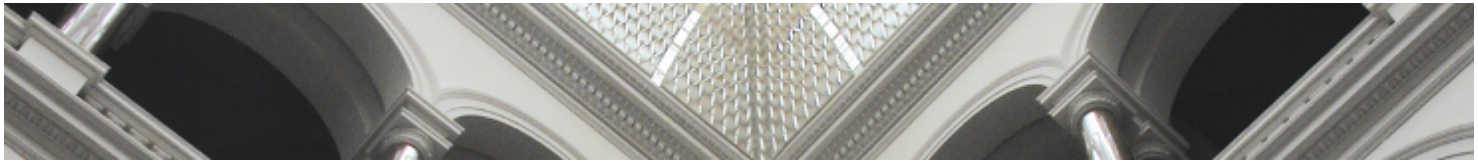
Problem

- Guide prediction scores still vary in performance
- A lot of manual labor for guide selection

19	POLR2L	28,5	28	-3,00752118182 82
20	SPC24	29,5	42	-2,83403119872683
21	THAP1	31,5	27	-3,01155447143932
22	CDC123	32,5	44	-2,81699463079959
23	WDR74	37	68	2,58033320817648

- Non-frame-shift indels cannot be circumvented





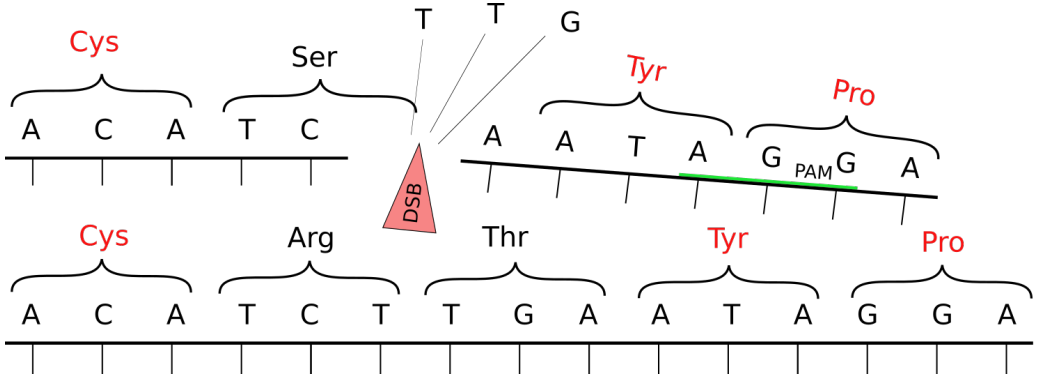
Technische Universität Berlin

Problem

- Guide prediction scores still vary in performance
- A lot of manual labor for guide selection

19	POLR2L	28,5	28	-3,00752118182 82
20	SPC24	29,5	42	-2,83403119872683
21	THAP1	31,5	27	-3,01155447143932
22	CDC123	32,5	44	-2,81699463079959
23	WDR74	37	68	2,58033320817648

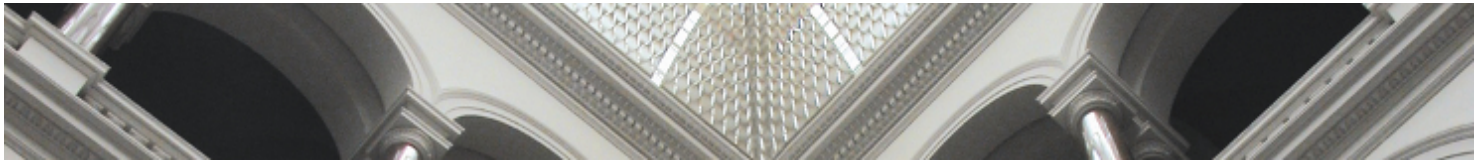
- Non-frame-shift indels cannot be circumvented



- Cancer cellline data affects certain guides

AGCATCGTAAGTGAATTACGG
PAM
- +

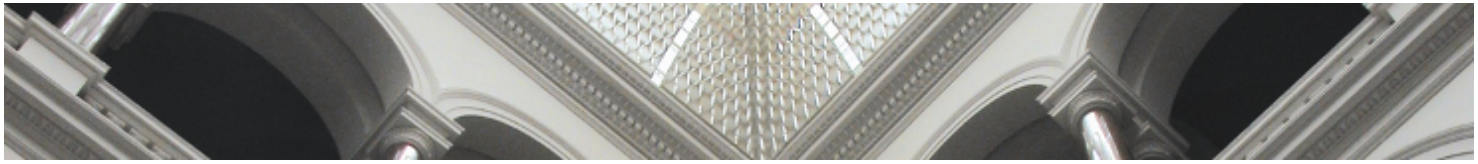
DMS153 lung cancer cellline SNP



Technische Universität Berlin

Solution

- Cutting-edge guide efficacy scoring
- All-in-one guide design tool
- Web-based
- Functional domain-aware
- Incorporate cancer cellline data



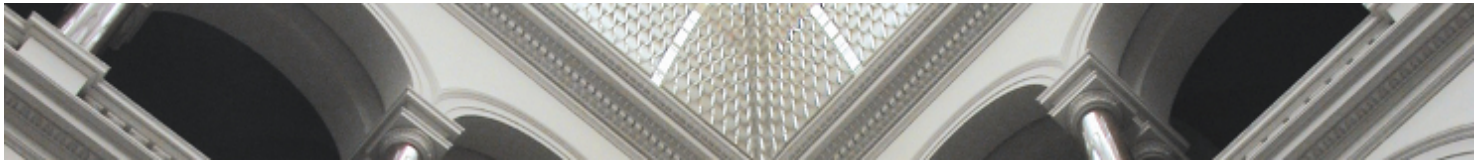
Technische Universität Berlin

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
. . .	...

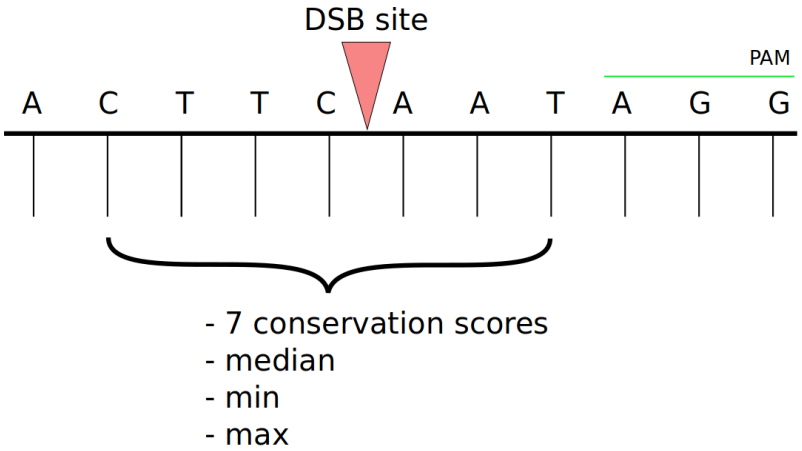
- 5310 data rows
- Efficacy relates to cell proliferation after CRISPR application

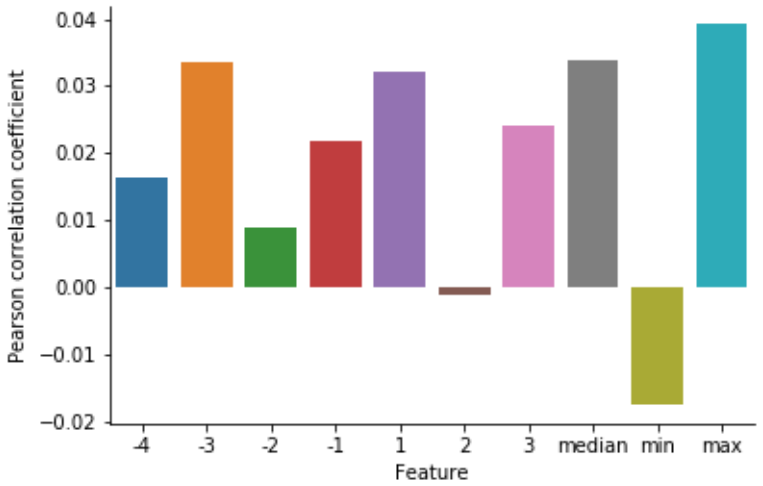
"Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9", 2016, John G. Doench et al.



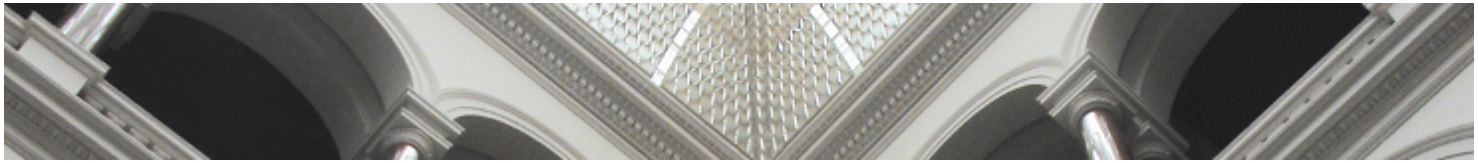
Technische Universität Berlin

Conservation features



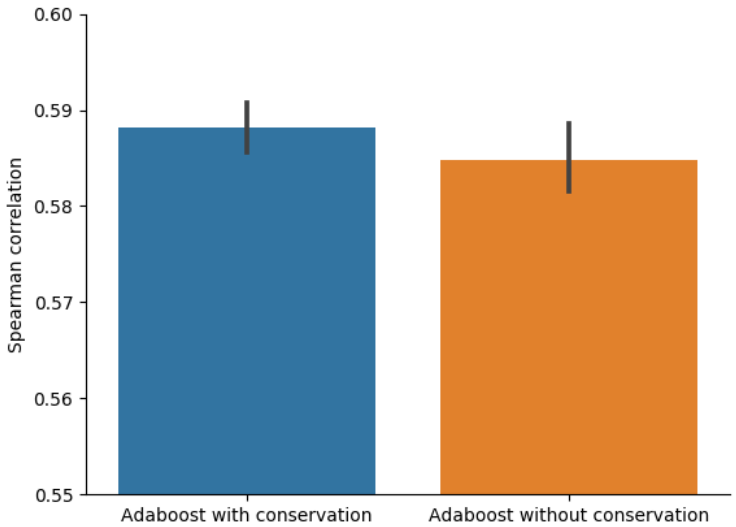


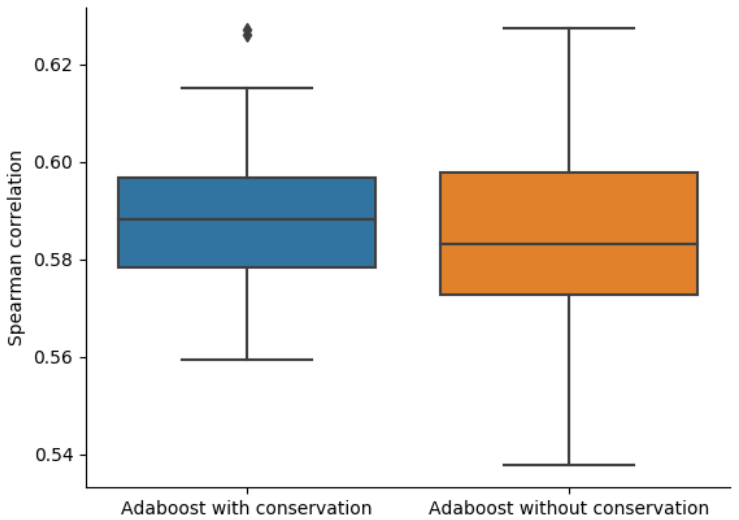
$p_{max} = 0.0043$

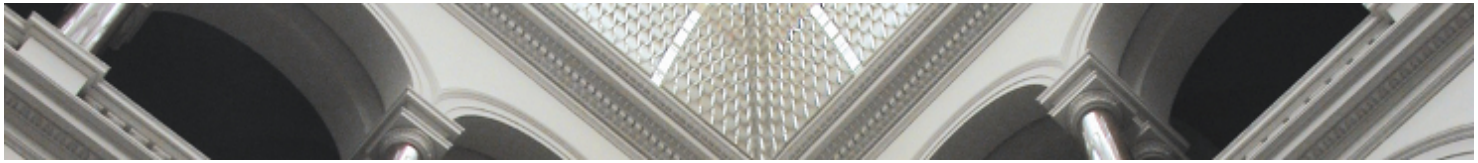


Conservation feature results

Comparison of 100 runs on adaboost

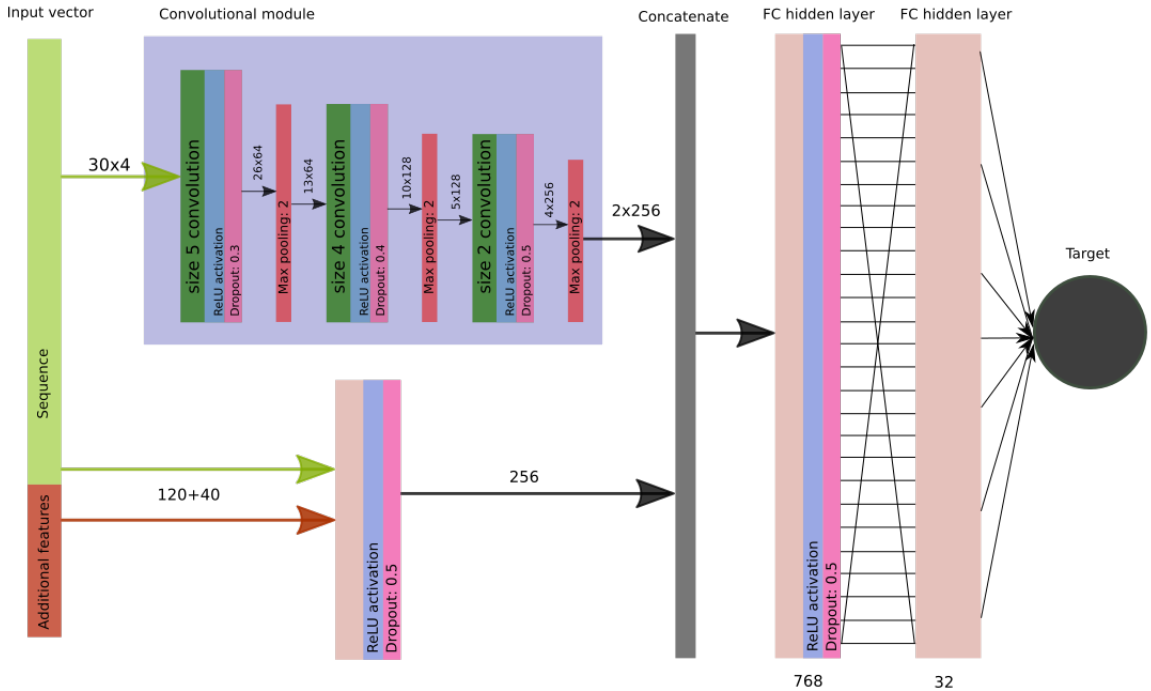


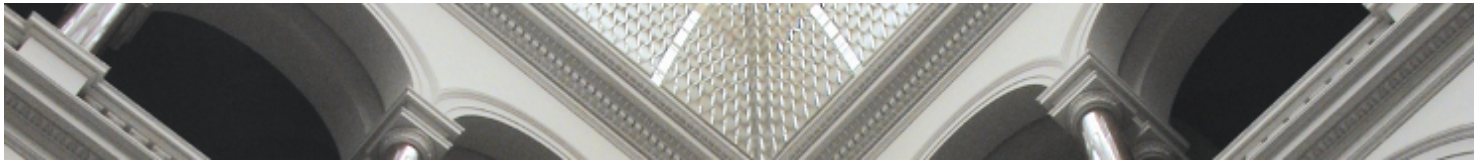




Deep Learning architecture

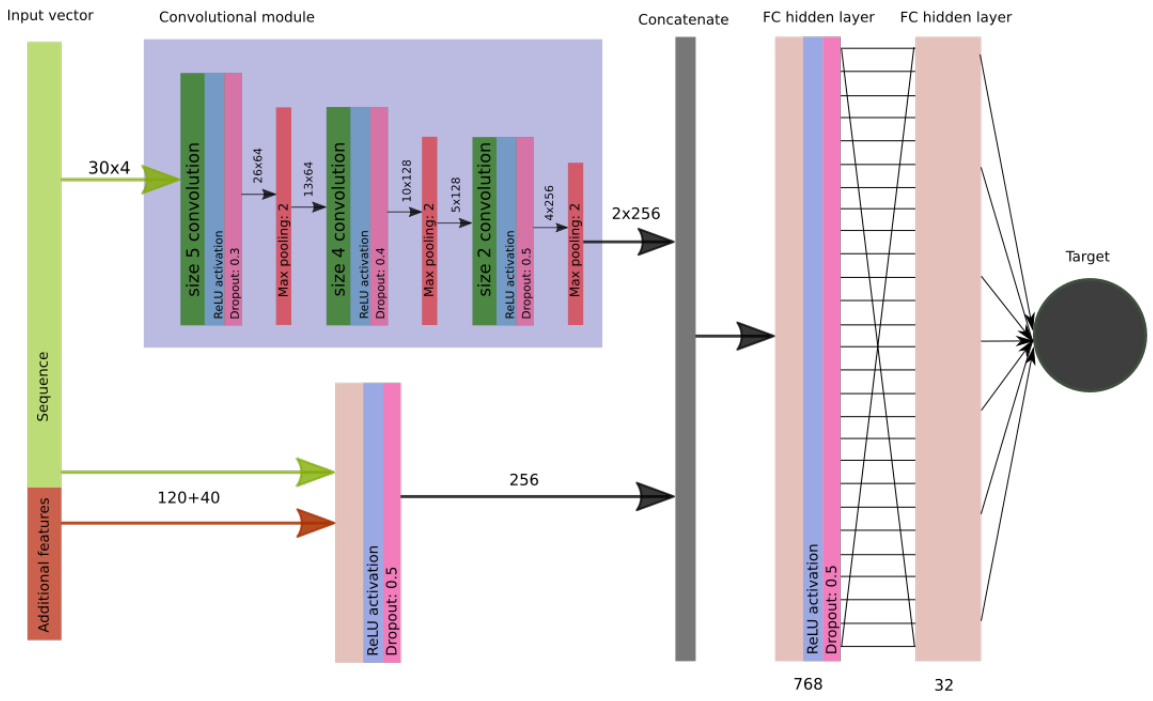
"Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning", 2015, B. Alipanahi et al.

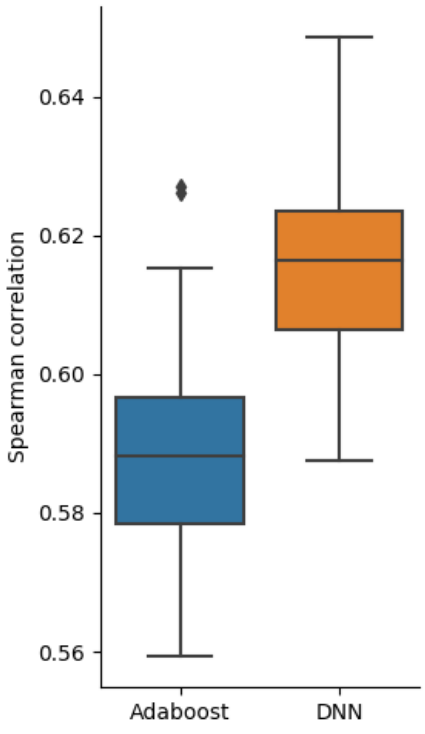




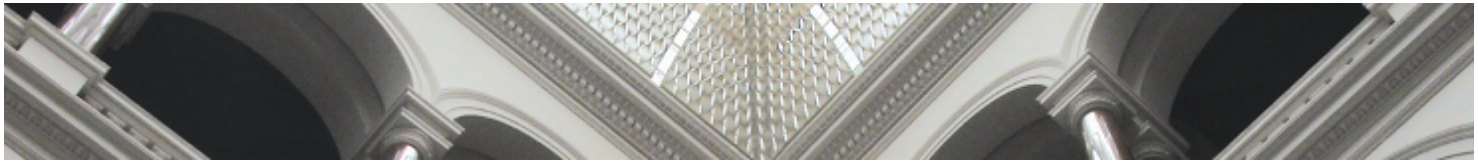
Deep Learning architecture

"Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning", 2015, B. Alipanahi et al.



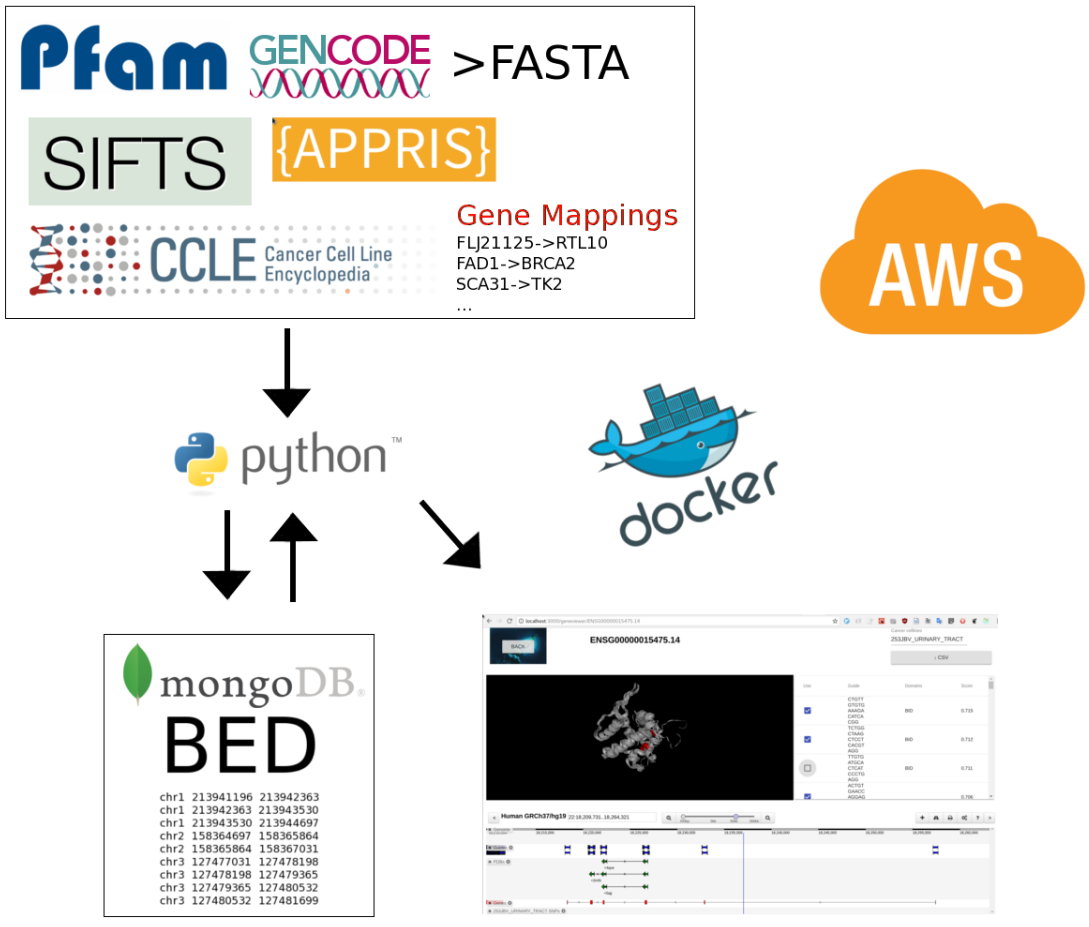


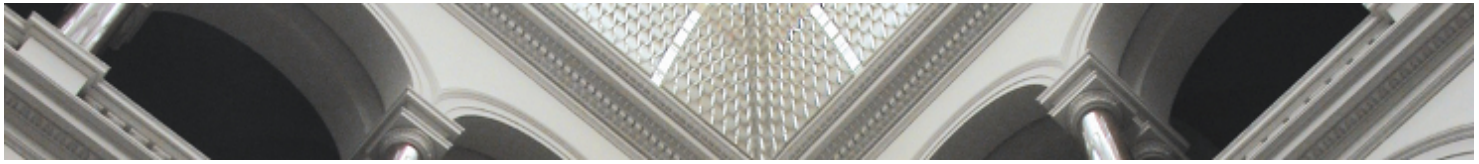
Conclusion: Deep Learning can improve guide efficacy prediction



Technische Universität Berlin

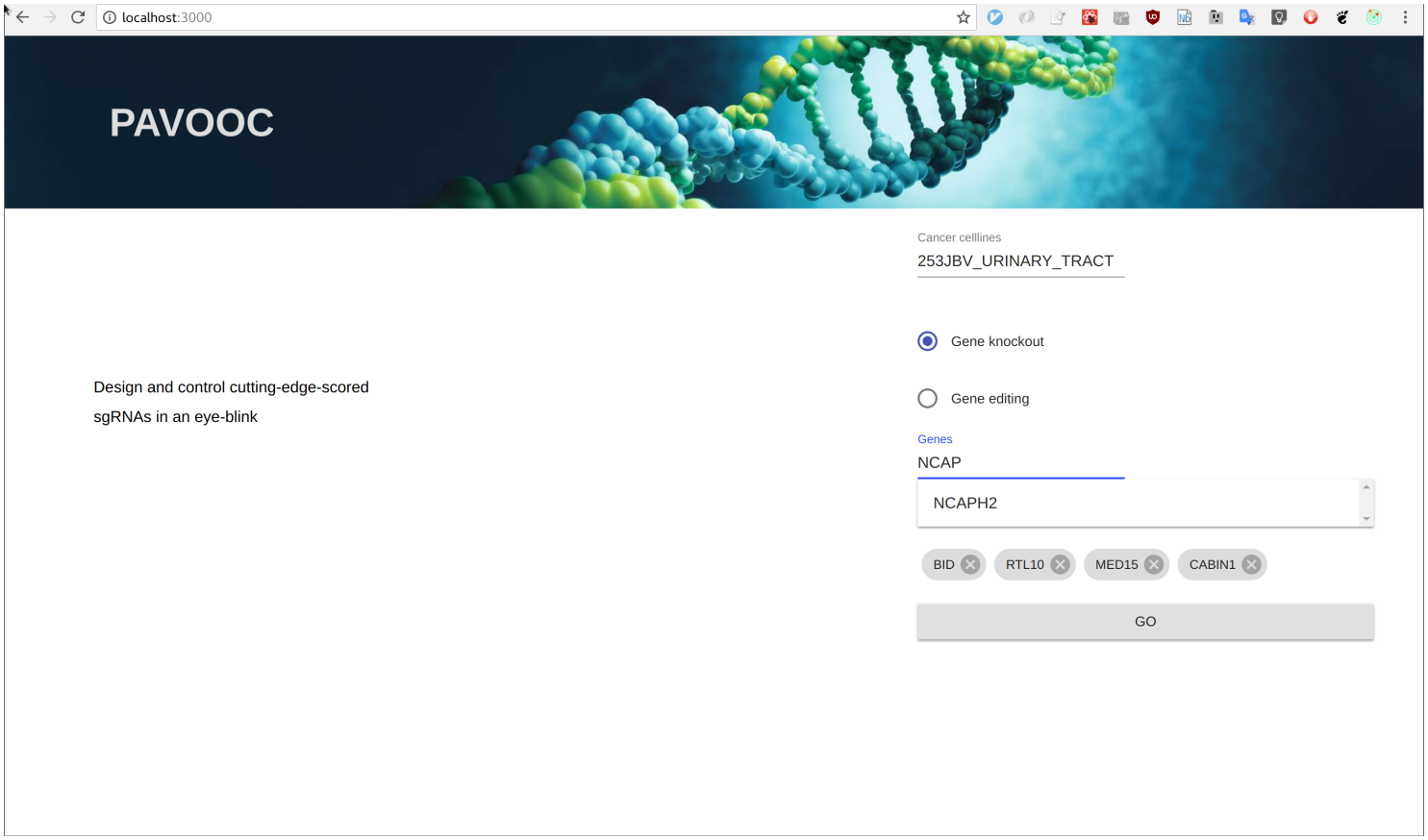
Architecture

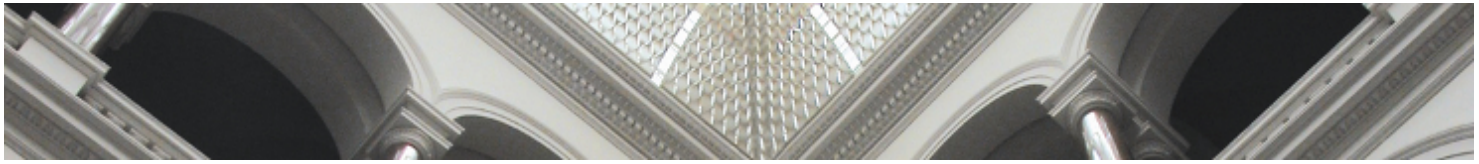




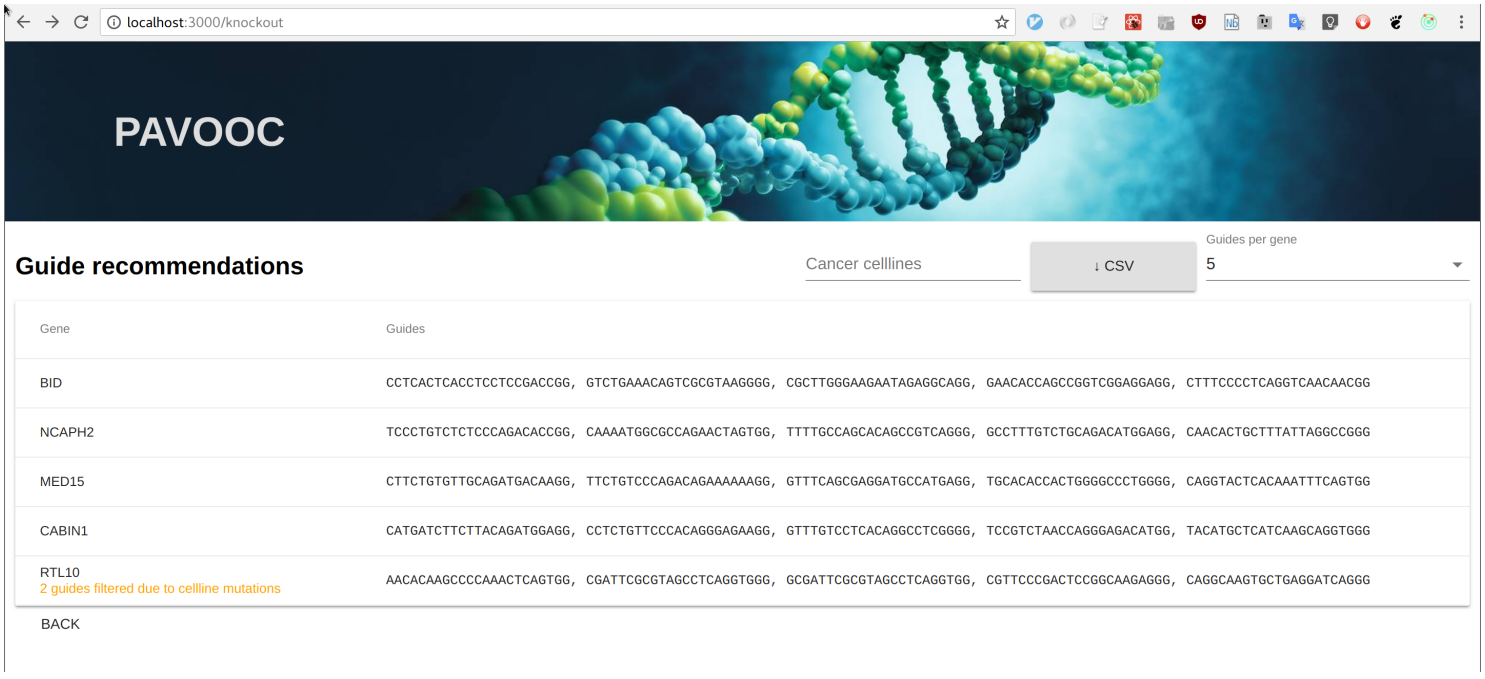
Technische Universität Berlin

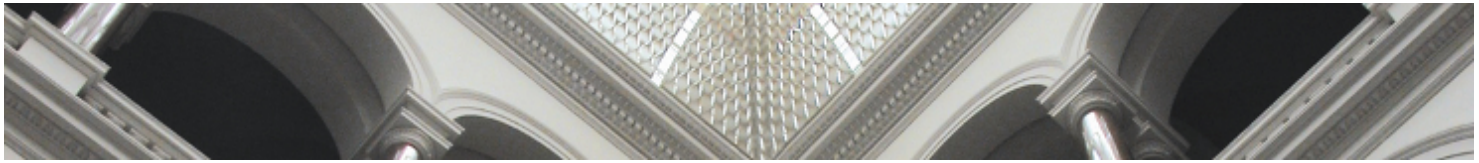
Application overview





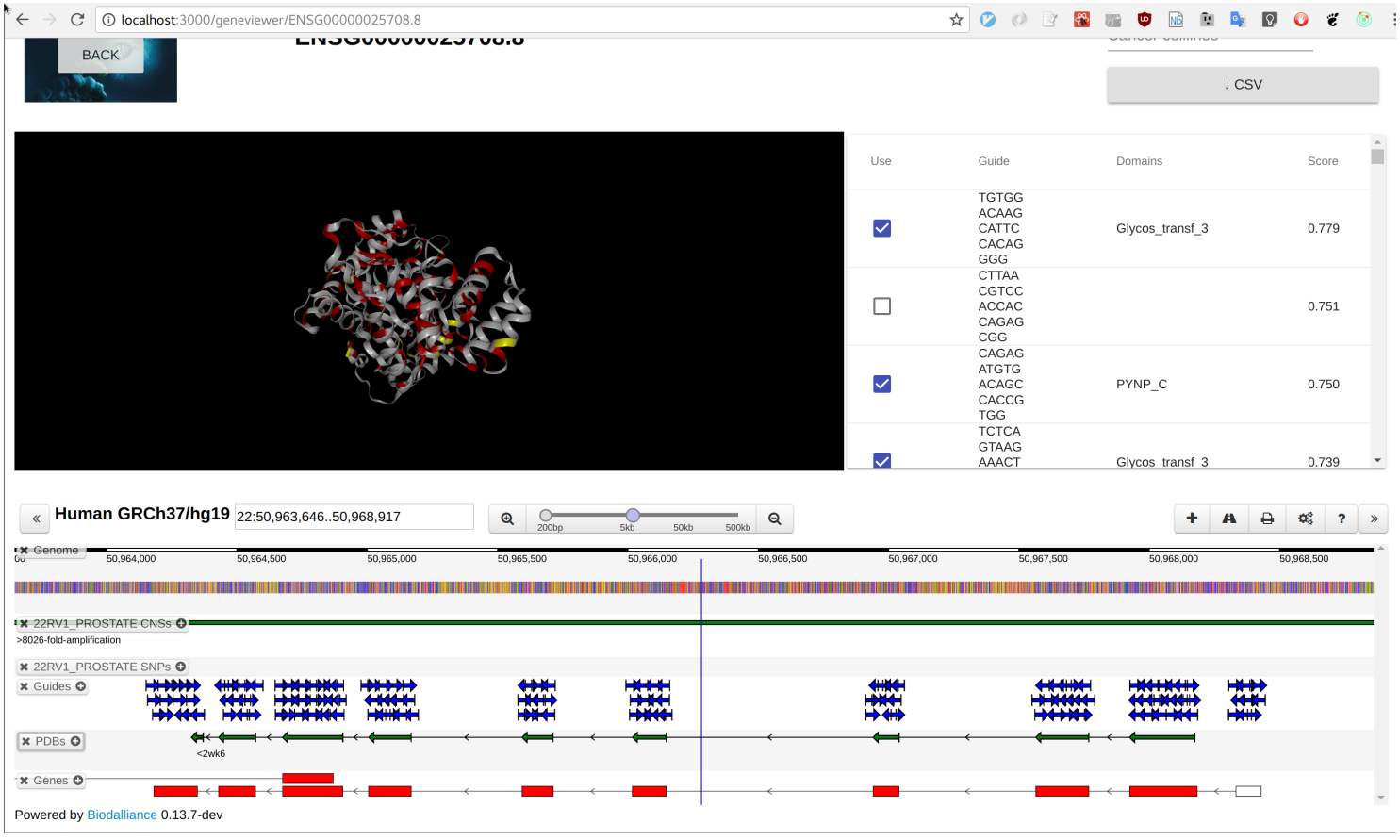
Application overview



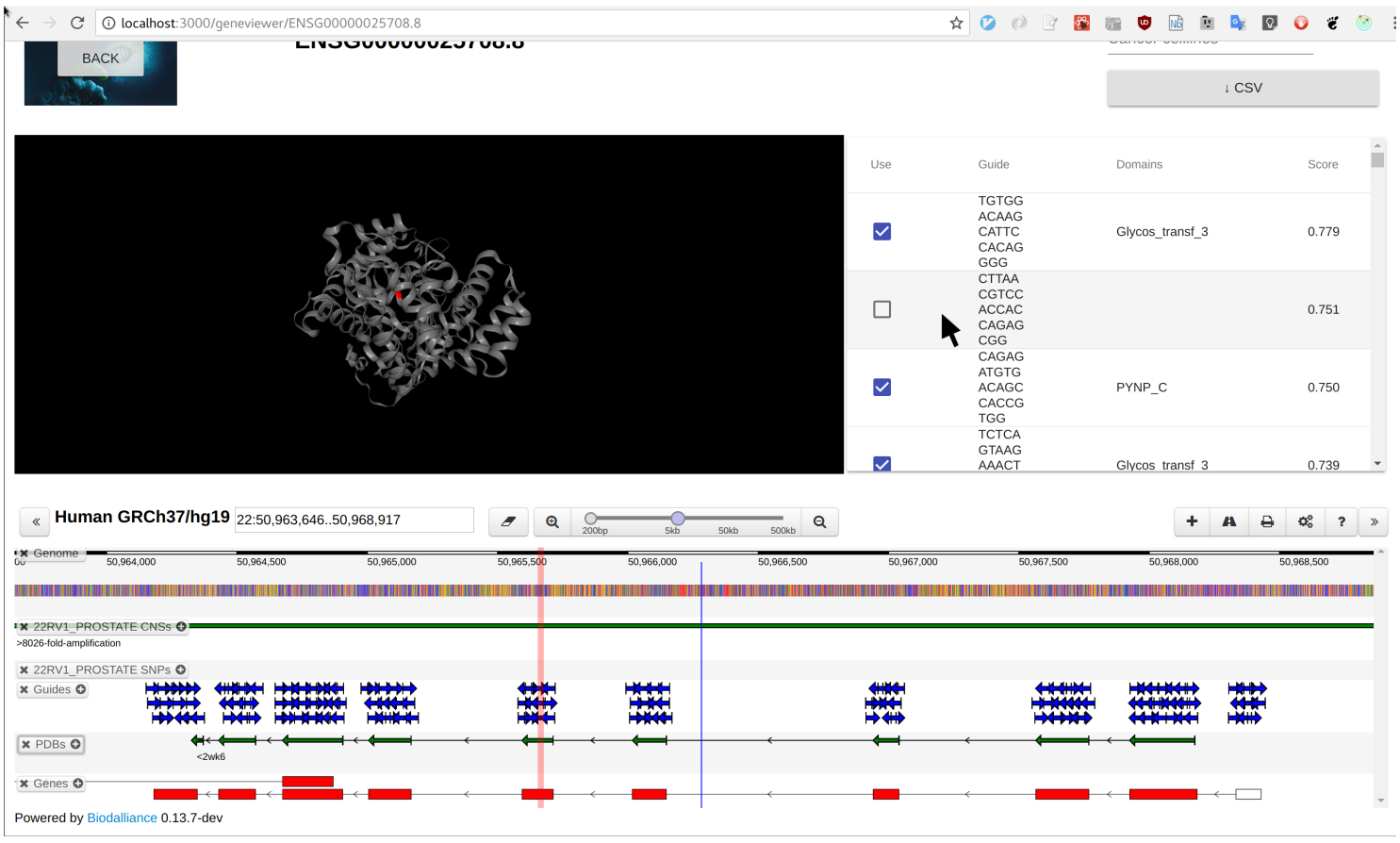


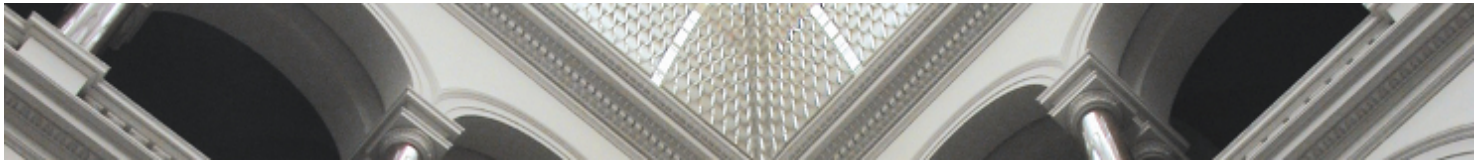
Technische Universität Berlin

Application overview



Application overview





Technische Universität Berlin

Acknowledgements

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