## Improved prediction of bacterial CRISPRi guide efficiency through data integration and automated machine learning

This manuscript was automatically generated on February 8, 2023.

## **Authors**

- · John Doe
  - **©** XXXX-XXXX-XXXX · **©** johndoe · **⋑** johndoe · **@** @johndoe@mastodon.social Department of Something, University of Whatever · Funded by Grant XXXXXXXX
- ∙ Jane Roe <sup>⊠</sup>

Department of Something, University of Whatever; Department of Whatever, University of Something

☑ — Correspondence possible via GitHub Issues or email to Jane Roe <jane.roe@whatever.edu>.

## **Abstract**

This paper proposes a novel approach to improve the prediction of bacterial CRISPRi guide efficiency through data integration and automated machine learning. We combine the existing gRNA design rules with the features generated from the machine learning algorithms to form a hybrid prediction model. This hybrid model is then trained with a large-scale dataset to optimize the gRNA selection. Our results demonstrate that the hybrid model significantly outperforms the state-of-the-art models in predicting bacterial CRISPRi guide efficiency.

This work presents a novel machine learning model that improves the prediction of CRISPRi guide efficiency in bacteria. By leveraging feature engineering, data integration, interpretable AI, and automated machine learning, the model is able to systematically investigate the influential factors that attribute to the extent of depletion in CRISPRi genome-wide essentiality screens in Escherichia coli. This approach allows for the segregation of confounding gene-specific effects and provides a better estimate of guide efficiency, resulting in improved performance compared to existing tools. Additionally, the model is able to extract design rules for robust gene silencing, such as the preference for cytosine and disfavoring for guanine and thymine within and around the PAM sequence. This research is made freely accessible through a web-based tool at www.ciao.helmholtz-hiri.de.

This paper presents a machine learning approach to improve the prediction of bacterial CRISPRi guide efficiency. The workflow includes three steps: 1) accommodating the feature set for the CRISPR-Cas system or technique; 2) optimizing a machine learning model using automated machine learning; and 3) explaining the model using interpretable AI. The approach was applied to analyze three CRISPR-Cas genome-wide screens, revealing design rules for robust antimicrobial activity across different organisms and providing a predictive algorithm for gRNA design. The results demonstrate that the proposed approach can effectively predict guide efficiency across different CRISPR-Cas systems, enabling the development of more reliable prediction algorithms.

This thesis presents a machine learning approach for accurately predicting CRISPRi guide efficiency in bacteria, which provides insights into the determinants of efficient silencing and guide designs. The approach was systematically explored and developed into a robust model for use in other bacteria and CRISPR-Cas systems. Applying the approach in the analysis of independent CRISPR-Cas screens reveals not only the design rules but also the mechanisms of the CRISPR-Cas systems. The results demonstrate that machine learning can be applied to gain a deeper understanding and broader application of CRISPR-Cas systems.

## References