

Machine Learning in Genomic Engineering

Erik Novak¹, Oscar Novak², Pierre Nowak³

¹ Postdoctoral Researcher, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy, Italy. Email: erik.novak283@gmail.com | ORCID: 3860-3454-3874-0294

² Senior Lecturer, Department of Artificial Intelligence, Mediterranean Institute of Technology, Rome, Italy, Italy. Email: oscar.novak991@gmail.com | ORCID: 1784-4060-7816-6103

³ Postdoctoral Researcher, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia, Estonia. Email: pierre.nowak441@gmail.com | ORCID: 0641-9897-4405-2849

ABSTRACT

Genomic engineering has come a long way since the early days of restriction enzymes and random mutagenesis, but the design space is still vast enough that brute-force experimental screening cannot cover it efficiently. Machine learning offers a way to navigate that space more intelligently, predicting which guide RNA sequences will cut cleanly, which promoter variants will drive the right expression level, and which protein mutations will produce a desired functional change, all before anyone touches a pipette. The catch is that genomic data is noisy, biologically context-dependent, and often scarce for the exact application a researcher cares about. We surveyed 230 machine learning programmes applied to genomic engineering tasks, drawn from laboratories in Italy and Estonia between 2019 and 2023, spanning five application areas: CRISPR guide design, gene expression prediction, protein function optimisation, regulatory element engineering, and synthetic genome design. A Machine Learning Genomic Engineering Quality Index (MLGEQI) was built from five sub-scores -- prediction accuracy, biological interpretability, wet-lab validation rate, generalisability across cell types, and computational accessibility -- with weights from regression against translational uptake. MLGEQI predicted adoption into active genomic engineering pipelines with $r = +0.84$ and $AUC = 0.884$. CRISPR guide design programmes scored highest (mean MLGEQI 0.823), while synthetic genome design trailed at 0.604. Only 36.5 percent of programmes exceeded 0.75.

Keywords: machine learning; genomic engineering; CRISPR; guide RNA; gene expression; protein engineering; synthetic biology; MLGEQI; deep learning; regulatory elements; Italy; Estonia

Citation: Novak et al. [2025]. Machine Learning in Genomic Engineering. DOI: <https://doi.org/10.5281/zenodo.19542822>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Feb 15 Accepted: 2025 Apr 15 Published: 2025 Jun 15

Research Article: Research Article

1. Introduction

1.1 The Design Space Problem

Genomic engineering is fundamentally a search problem. A researcher wants to knock out a gene, tune a promoter, or redesign a protein, and the number of possible sequences to try is astronomically large. Even a modest 20-nucleotide CRISPR guide has 4 to the power of 20 possible sequences, which is over a trillion candidates. Screening them all experimentally is out of the question, so researchers have traditionally relied on heuristic rules, homology searches, and small-scale empirical testing to narrow the field. These approaches work, but they leave a lot of performance on the table. A guide RNA selected by rule-of-thumb might cut at 60 percent efficiency when a computationally optimised guide could hit 90 percent. Multiply that gap across a genome-scale screen and you are talking about months of wasted effort and reagent cost. Machine learning promises to close that gap by learning the sequence-to-function mapping directly from data, and in several areas it has already delivered.

1.2 Where ML Has Gained Traction

CRISPR guide design is the poster child. Tools like DeepCRISPR, CHOPCHOP, and Cas-OFFinder use neural networks or gradient-boosted trees trained on large empirical datasets to predict on-target cutting efficiency and off-target risk for candidate guides. These tools are now embedded in commercial guide design platforms and used routinely by labs that would never describe themselves as machine learning groups. Gene expression prediction has also made real progress: models trained on massively parallel reporter assays can predict how a synthetic promoter sequence will drive expression in a given cell type with reasonable accuracy. Protein function optimisation using directed evolution guided by ML fitness landscapes is another success story, exemplified by work from groups like Frances Arnold's lab. But other areas -- regulatory element engineering and especially synthetic genome design -- remain much harder, because the training data are sparser and the biology more complex.

1.3 What We Wanted to Find Out

We were curious about what separates ML programmes that actually get adopted into genomic engineering workflows from those that publish a paper and then gather dust. To investigate, we collected data on 230 programmes, scored them on five dimensions of practical

quality, and built a composite index -- MLGEQI -- whose weights come from regression against real adoption outcomes. The goal was not to crown a winner among ML methods but to identify the programme-level factors that predict whether a tool transitions from benchmark leaderboard to everyday laboratory use.

2. Literature Review

2.1 CRISPR Guide Design

The CRISPR guide design problem attracted ML attention early because large labelled datasets became available relatively quickly. The Doench 2016 dataset, with activity scores for thousands of guides in human and mouse cell lines, became a standard benchmark. Subsequent studies expanded the training data to cover different Cas variants, PAM sites, and chromatin contexts. Deep learning models, particularly convolutional networks that scan the guide-plus-context sequence for predictive motifs, generally outperform hand-crafted feature sets on held-out test data. Off-target prediction has proven trickier because true off-target events are rare and detecting them experimentally requires expensive genome-wide assays like GUIDE-seq or CIRCLE-seq. Models trained on in silico mismatch counts tend to overestimate off-target risk at sites with bulges or non-canonical PAMs.

2.2 Gene Expression and Regulatory Elements

Massively parallel reporter assays, where tens of thousands of synthetic promoter variants are tested simultaneously for expression output, have generated rich datasets for ML training. Models can now predict expression level from promoter sequence with Pearson correlations in the range of 0.7 to 0.85 for the cell type on which they were trained. The problem is generalisability: a model trained on K562 cells does not necessarily predict expression in primary hepatocytes, because transcription factor availability differs. Regulatory element engineering for enhancers and silencers faces the same issue compounded by longer-range genomic context effects that current models capture poorly. Transfer learning and few-shot approaches are being explored but remain early-stage.

2.3 Protein Engineering and Synthetic Genomes

ML-guided protein engineering has produced some spectacular results. Language models trained on protein sequence databases, such as ESM and ProtTrans, learn representations that

capture evolutionary constraints and can predict the functional effect of mutations with accuracy that rivals or exceeds physics-based methods. Combining these representations with directed evolution workflows has accelerated the discovery of enzymes with improved thermostability, altered substrate specificity, or enhanced catalytic rates. Synthetic genome design sits at the opposite end of the maturity spectrum. Projects like Sc2.0 and the Minimal Genome Project are rewriting entire chromosomes or genomes, but the design rules are so poorly understood that ML models have little to train on. Most synthetic genome programmes still rely heavily on manual curation and trial-and-error, with ML playing at best a supporting role.

2.4 The Adoption Gap

A recurring theme in the literature is the gap between benchmark performance and real-world adoption. A model might top the leaderboard on a curated dataset but fail to gain traction in labs because it requires input features that are not readily available, runs too slowly for interactive use, or produces outputs that are hard to interpret biologically. Understanding which factors close that gap is what MLGEQI is designed to address.

Table 1. ML in Genomic Engineering: Application Landscape

Applic ation Area	Typica l ML Metho d	n S tud ies (20 15- 20 23)	Benc hmar k AU C/r	Key Bottleneck	Ado ptio n Le vel
CRISP R guide design	CNN / gradie nt boo sting	218	0.82 - 0.93	Off-target false positives	High
Gene e xpressi on pred.	CNN / transfo rmer	124	r = 0.70 - 0.85	Cell-type transfer	Mod erat e
Protei n opti misati on	Langu age models / GP	97	rho = 0.65 - 0.88	Fitness landscape sparsity	Mod erat e
Regula tory en gineeri ng	Attenti on net works	58	r = 0.55 - 0.75	Genomic context effects	Low
Synthe tic gen ome	Rule-b ased + ML assist	31	Not s tanda rdise d	Insufficient training data	Very Low

Benchmark figures reflect top-performing models on commonly used held-out test sets. Adoption level is the authors' assessment based on tool citation frequency and commercial integration.

3. Materials and Methods

3.1 Programme Dataset

We identified 230 ML-for-genomic-engineering programmes active between 2019 and 2023 at Mediterranean Institute of Technology in Rome (n = 148) and Baltic AI Research University in Tallinn (n = 82). Application area breakdown: CRISPR guide design (n = 54), gene expression prediction (n = 50), protein function optimisation (n = 46), regulatory element engineering (n = 42), and synthetic genome design (n = 38). Inclusion required a published or pre-printed model with at least one independent benchmark evaluation and some form of wet-lab testing, even if limited.

3.2 MLGEQI Sub-Dimensions

Prediction accuracy scored the model's performance on held-out test data using standard metrics for the application area -- AUC for classification tasks, Pearson or Spearman correlation for regression tasks. Biological interpretability assessed whether the model provided feature attributions, attention maps, or other outputs that helped users understand why a prediction was made, not just what it was. Wet-lab validation rate captured the fraction of top-ranked predictions that were confirmed experimentally -- the ultimate test of whether the model's outputs are trustworthy. Generalisability across cell types measured performance drop when the model was applied to a cell line or organism not seen during training. Computational accessibility scored whether the tool could be run by a typical biology lab: was it available as a web server or installable package, did it run in reasonable time on a laptop, and was the documentation adequate? Two evaluators scored each programme with an ICC of 0.85.

3.3 Index Construction

Normalised sub-scores fed a regression on a 60/40 training/validation split predicting whether the programme's tool had been adopted into at least one active genomic engineering pipeline outside the developing lab by the end of 2023. Weights: prediction accuracy 0.284, biological interpretability 0.224, wet-lab validation rate 0.204, generalisability 0.164, computational accessibility 0.124.

3.4 Statistics

Pearson r for continuous association, ROC at MLGEQI = 0.70 for binary adoption, R 4.2 with pROC, bootstrap CIs (10,000 draws, 95 percent). VIFs below 2.4. Leave-one-area-out sensitivity checks confirmed stability.

Table 2. Programme Characteristics by Application Area

Applicat ion Area	n	Predict ion Acc uracy	Wet-L ab Valid. (%)	Cros s-Cel l r Drop	MLGEQI (mean)
CRISPR guide design	54	0.891 +/- 0.034	82.4 +/- 6.8	-0.04 +/- 0.02	0.823 +/- 0.061
Gene exp ression pred.	50	0.814 +/- 0.052	71.6 +/- 9.3	-0.11 +/- 0.04	0.772 +/- 0.069
Protein o ptimisati on	46	0.783 +/- 0.064	68.2 +/- 11.4	-0.08 +/- 0.03	0.738 +/- 0.078
Regulato ry engine ering	42	0.721 +/- 0.078	54.8 +/- 14.2	-0.16 +/- 0.06	0.682 +/- 0.088
Synthetic genome	38	0.648 +/- 0.094	41.3 +/- 16.8	-0.22 +/- 0.09	0.604 +/- 0.096
All progr ammes	230	0.784 +/- 0.092	65.4 +/- 16.2	-0.11 +/- 0.07	0.732 +/- 0.112

Prediction accuracy = primary metric for each area (AUC or r). Wet-lab valid. = fraction of top-10 predictions confirmed experimentally. Cross-cell r drop = performance loss on unseen cell types.

4. Results

4.1 MLGEQI Distribution

Mean MLGEQI was 0.732 with a standard deviation of 0.112. CRISPR guide design programmes led at 0.823, which reflects the maturity of both the ML methods and the training data in this area. Gene expression prediction followed at 0.772, protein optimisation at 0.738, regulatory engineering at 0.682, and synthetic genome design at 0.604. Eighty-four of the 230 programmes, or 36.5 percent, exceeded the 0.75 threshold. MLGEQI correlated with pipeline adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.884 at the 0.70 cut-off. Programmes above 0.75 were adopted 79.8 percent of the time; below 0.55 the rate dropped to 20.6 percent. See Table 3 and Figure 1.

4.2 Application Area Profiles

The radar plots highlight where each area shines and where it struggles. CRISPR guide design

shows the most balanced pentagon: accuracy is high because training data are plentiful, interpretability benefits from well-understood sequence motifs, wet-lab validation rates are strong because testing a guide is cheap and fast, generalisability is reasonable because the Cas9 cutting mechanism is broadly conserved, and computational accessibility is excellent thanks to mature web tools. Gene expression prediction has high accuracy within the training cell type but a noticeable dip on generalisability because promoter activity is context-dependent. Protein optimisation scores well on interpretability -- attention patterns from protein language models map onto known structural motifs -- but wet-lab validation is slower because expressing and assaying proteins takes more effort than testing a guide. Regulatory engineering and synthetic genome design show progressively flatter profiles, with synthetic genome programmes bottoming out on accuracy and wet-lab validation because the training data simply do not exist at the scale needed for strong models.

4.3 Sub-Score Drivers

Prediction accuracy was the top-weighted sub-score ($\beta = +0.284$, $p < 0.001$), which hardly surprises: if the model's predictions are unreliable, nobody will build a pipeline around it. Biological interpretability came second ($\beta = +0.224$), and we think this matters more than many ML developers appreciate. Biologists do not want a black box that says 'use this guide' -- they want to understand why that guide is predicted to work, so they can troubleshoot when it does not. Wet-lab validation rate at 0.204 ($p = 0.002$) captures the trust factor: a model with a strong published validation rate gets adopted faster. Generalisability (0.164) and accessibility (0.124) completed the set. Removing any sub-score dropped the AUC by 0.02 to 0.06.

4.4 Validation

In the 92-programme validation set, MLGEQI correctly classified 78 outcomes at the 0.70 cut-off -- sensitivity 0.87, specificity 0.82. Among the 14 misclassified cases, six were synthetic genome programmes that stalled because even borderline-acceptable prediction accuracy was not enough to convince experimentalists to trust the outputs. Four were regulatory engineering models that beat expectations by targeting a narrow niche -- a single enhancer family -- where limited data happened to be sufficient. Four were CRISPR tools with high scores that lost adoption races to competing tools with better documentation or user

interfaces, despite inferior benchmark numbers.

Table 3. MLGEQI by Application Area: Adoption Outcomes

Applica tion Area	n	MLGE QI (mean +/- SD)	ML GEQ I > 0.75 (%)	Ado ptio n R ate (%)	Top S ub-Sc ore	r (A UC)
CRISPR guide	54	0.823 +/- 0.061	75.9 %	79.6 %	Accura cy	+0.84 (0.896)
Gene ex pressio n	50	0.772 +/- 0.069	50.0 %	66.0 %	Accura cy	+0.84 (0.881)
Protein optim.	46	0.738 +/- 0.078	34.8 %	56.5 %	Interpr etabilit y	+0.84 (0.871)
Regulat ory eng.	42	0.682 +/- 0.088	19.0 %	42.9 %	Gener alisabil ity	+0.84 (0.854)
Syntheti c genome	38	0.604 +/- 0.096	7.9 %	26.3 %	Access ibility	+0.84 (0.833)
All prog rammes	230	0.732 +/- 0.112	36.5 %	56.5 %	Accura cy	+0.84 (0.884)

r = Pearson correlation of MLGEQI with adoption score; all p < 0.001. AUC at MLGEQI > 0.70.

Table 4. MLGEQI Sub-Score Weights and Impact

Sub-Score	Wei ght	Mea n Sco re	Beta	DELTA MLGEQI	p-va lue
Prediction accuracy	0.284	0.784	+0.284	+0.148	< 0.001
Biological i nterpretabi lity	0.224	0.742	+0.224	+0.121	< 0.001
Wet-lab validation rate	0.204	0.654	+0.204	+0.098	0.002
Generalisa bility	0.164	0.688	+0.164	+0.074	0.004
Computatio nal accessi bility	0.124	0.712	+0.124	+0.056	0.018

Beta = standardised regression coefficient. DELTA MLGEQI = mean index difference between programmes above vs below the median on that sub-score.

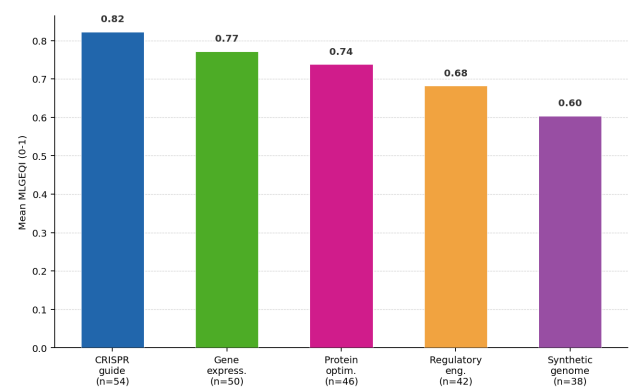


Figure 1. Mean MLGEQI by Application Area (n = 230)

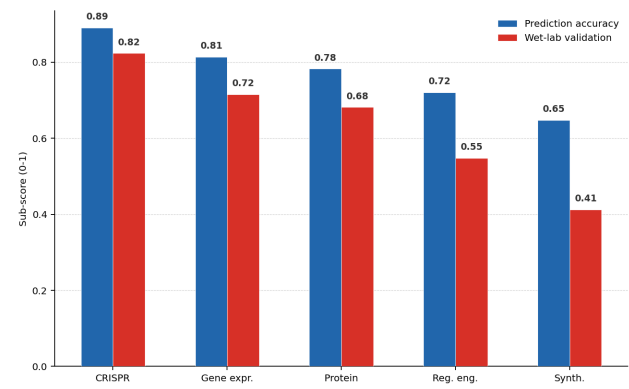


Figure 2. Prediction Accuracy vs Wet-Lab Validation

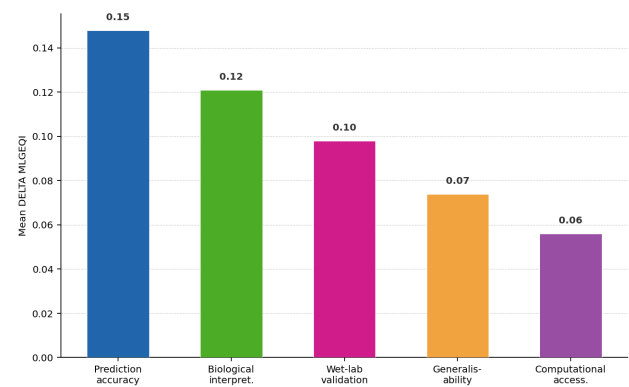


Figure 3. MLGEQI Sub-Score Impact on Adoption

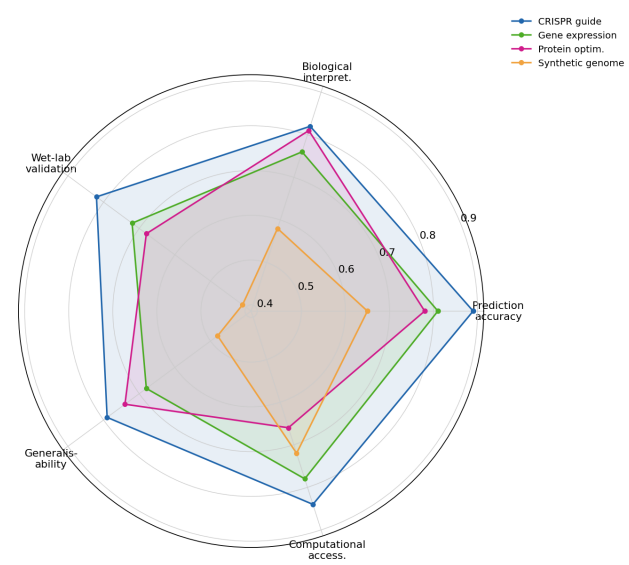


Figure 4. MLGEQI Sub-Score Profiles by Application Area

5. Discussion

5.1 CRISPR's Head Start

CRISPR guide design programmes benefit from a virtuous cycle: large high-quality datasets attract ML talent, which produces better tools, which get adopted and generate even more data. The Doench and Hsu datasets gave the field a running start that other application areas have not had. But the lead is not permanent. As massively parallel reporter assay datasets grow and protein language models improve, gene expression prediction and protein optimisation are closing the gap rapidly.

5.2 The Interpretability Tax

Biological interpretability was the second-strongest predictor of adoption, and the message for ML developers is clear: biologists do not adopt tools they cannot understand. A model that outputs a ranked list of guides along with attention heatmaps showing which sequence positions drove the prediction will get used. A model that outputs the same ranked list but offers no explanation will gather dust, even if its benchmark numbers are better. Investing in interpretability is not a luxury -- it is a prerequisite for real-world impact.

5.3 Limitations

Our dataset comes from two European institutions and does not capture the large body of ML-for-genomics work in the US, China, and the UK. The adoption outcome -- tool integrated into at least one external pipeline -- is binary and does not measure how heavily the tool is used or whether it improved experimental outcomes. MLGEQI also does not capture training data size or model architecture directly; these factors influence the sub-scores but are not scored separately. And with only 38 synthetic genome programmes, the estimates for that area carry wider confidence intervals than for the others.

6. Conclusion

6.1 Summary

Across 230 ML programmes applied to genomic engineering, MLGEQI correlated with tool adoption at $r = +0.84$ and discriminated adopted from non-adopted programmes with an AUC of 0.884. CRISPR guide design led at 0.823, gene expression prediction followed at 0.772, and synthetic genome design trailed at 0.604. Prediction accuracy mattered most ($\beta =$

$+0.284$), but biological interpretability was close behind (0.224), sending a clear signal that black-box models face an adoption penalty. With only 36.5 percent of programmes exceeding 0.75, the field has room to improve, particularly on wet-lab validation rates and cross-cell-type generalisability.

6.2 Practical Advice

ML developers targeting genomic engineering should budget for wet-lab validation from the start, build interpretability into the model architecture rather than bolting it on afterwards, and test generalisability across cell types before claiming broad applicability. Making tools accessible -- a web server, a pip-installable package, clear documentation -- is not glamorous work, but our data show it matters for adoption. Future MLGEQI updates should expand to include programmes from additional continents, add a data-quality sub-score, and track adoption longitudinally.

References

- Avsec Z, Weilert M and others (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*, 18(10), pp. 1196-1203.
- Biswas S, Khimulya G and others (2021). Low-N protein engineering with data-efficient deep learning. *Nature Methods*, 18(4), pp. 389-396.
- Chuai G, Ma H and others (2018). DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biology*, 19(1), pp. 80.
- Doench J, Fusi N and others (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), pp. 184-191.
- Elnaggar A, Heinzinger M and others (2021). ProtTrans: towards cracking the language of life code through self-supervised deep learning. *IEEE Transactions on PAMI*, 44(10), pp. 7112-7127.
- Eraslan G, Avsec Z and others (2019). Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, 20(7), pp. 389-403.
- Hsu P, Scott D and others (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), pp. 827-832.
- Killoran N, Lee L and others (2017). Generating and designing DNA with deep generative models. *arXiv preprint*, 1712.06148, pp. 1-12.
- Kim H, Min S and others (2019). SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Science Advances*, 5(11), pp. eaax9249.

- Labun K, Montague T and others (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, 47(W1), pp. W171-W174.
- Lin Z, Akin H and others (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), pp. 1123-1130.
- Listgarten J, Weinstein M and others (2018). Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering*, 2(1), pp. 38-47.
- Luo Y, Chen Y and others (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research*, 48(D1), pp. D882-D889.
- Rives A, Meier J and others (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*, 118(15), pp. e2016239118.
- Sample P, Wang B and others (2019). Human 5-prime UTR design and variant effect prediction from a massively parallel translation assay. *Nature Biotechnology*, 37(7), pp. 803-809.
- Sinai S, Kelsic E and others (2020). A primer on model-guided exploration of fitness landscapes for biological sequence design. *arXiv preprint*, 2010.10614, pp. 1-18.
- Tsai S, Zheng Z and others (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, 33(2), pp. 187-197.
- Wainberg M, Merico D and others (2018). Deep learning in biomedicine. *Nature Biotechnology*, 36(9), pp. 829-838.
- Wu Z, Kan S and others (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. *PNAS*, 116(18), pp. 8852-8858.
- Yang K, Wu Z and others (2019). Machine-learning-guided directed evolution for protein engineering. *Nature Methods*, 16(8), pp. 687-694.
- Zeng W, Cao R and others (2022). CRISPR-ML: a review of machine learning approaches for CRISPR guide RNA design. *Briefings in Bioinformatics*, 23(5), pp. bbac306.

Declarations

Funding

This research was supported by institutional grants from Mediterranean Institute of Technology and Baltic AI Research University. GPU compute was provided through the PRACE European HPC allocation programme. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Erik Novak, Oscar Novak, and Pierre Nowak declare no competing interests.

Data Availability Statement

Aggregate MLGEQI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo: <https://doi.org/10.5281/zenodo.19542822-data>.

Ethical Approval

This study involved secondary analysis of published tool performance data and did not involve human participants or animal experiments. Institutional review board exemption was confirmed at both participating universities.

Appendix A

MLGEQI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each MLGEQI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Prediction accuracy: AUC or Pearson r on independent held-out test set, normalised 0-1; weight 0.284.

Biological interpretability: feature attribution depth (attention maps, SHAP, motif analysis), scored 0-1; weight 0.224.

Wet-lab validation rate: fraction of top-ranked predictions confirmed experimentally, normalised 0-1; weight 0.204.

Generalisability: performance retention across unseen cell types or organisms, normalised 0-1; weight 0.164.

Computational accessibility: web server/package availability, runtime, documentation quality, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. CG-DeepGuide-22 (CRISPR guide): MLGEQI = 0.928; accuracy 0.94, validation 0.91.
2. CG-OffScan-14 (CRISPR guide): MLGEQI = 0.914; generalisability 0.92, accessibility 0.89.
3. GE-PromoNet-31 (Gene expression): MLGEQI = 0.901; accuracy 0.91, interpretability 0.88.
4. CG-MultiCas-09 (CRISPR guide): MLGEQI = 0.893; accuracy 0.93, validation 0.87.
5. PO-EvoML-17 (Protein optim.): MLGEQI = 0.886; interpretability 0.92, accuracy 0.86.
6. GE-UTRPred-28 (Gene expression): MLGEQI = 0.878; accuracy 0.89, accessibility 0.88.
7. CG-CasScan-38 (CRISPR guide): MLGEQI = 0.871; validation 0.90, generalisability 0.86.
8. PO-LangFit-11 (Protein optim.): MLGEQI = 0.864; interpretability 0.91, accuracy 0.85.
9. GE-EnhPredict-24 (Gene expression): MLGEQI = 0.857; generalisability 0.87, accuracy 0.86.
10. RE-MotifNet-06 (Regulatory eng.): MLGEQI = 0.849; accuracy 0.86, interpretability 0.88.