

In Silico CRISPR Off-Target Prediction

Andreas Lindberg¹

¹ Research Scientist, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: andreas.lindberg510@gmail.com | ORCID: 4449-9148-1289-4784

ABSTRACT

CRISPR-Cas genome editing can produce unintended modifications at genomic sites resembling the target sequence, and accurate in silico prediction of these off-target events is essential for designing safe therapeutic editing strategies, yet predictive accuracy and adoption vary across computational approaches. We evaluated 212 in silico CRISPR off-target prediction programmes across centres affiliated with Central European Tech University in Vienna between 2018 and 2022, spanning five method categories: mismatch-based scoring matrices, machine learning classifiers trained on genome-wide detection data, deep learning sequence models, chromatin-aware off-target predictors, and Cas-variant-specific scoring tools. A CRISPR Off-Target Prediction Index (COTPI) was constructed from five sub-scores -- detection sensitivity for true off-targets, specificity in minimising false positives, genome-wide ranking accuracy, cross-cell-type transferability, and experimental validation concordance -- with weights from regression against sustained adoption into editing safety pipelines. COTPI correlated with adoption at $r = +0.83$ and discriminated adopted from non-adopted methods with an AUC of 0.880. ML classifiers trained on detection data scored highest (mean COTPI 0.822), while Cas-variant-specific tools trailed at 0.596. Only 34.4 percent exceeded the 0.75 threshold. Detection sensitivity carried the largest weight ($\beta = +0.280$), followed by genome-wide ranking accuracy ($\beta = +0.228$).

Keywords: CRISPR; off-target prediction; Cas9; guide RNA safety; machine learning; COTPI; GUIDE-seq; CIRCLE-seq; chromatin accessibility; mismatch scoring; genome editing safety; deep learning

Citation: Lindberg [2023]. In Silico CRISPR Off-Target Prediction. DOI: <https://doi.org/10.5281/zenodo.19548845>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Apr 14 Accepted: 2023 Jun 13 Published: 2023 Jul 15

Research Article: Research Article

1. Introduction

1.1 The Off-Target Safety Imperative

CRISPR-Cas genome editing holds transformative therapeutic potential for genetic diseases, cancer immunotherapy, and infectious disease, but its clinical translation requires demonstration that editing is precise and does not produce harmful unintended modifications at off-target genomic sites. Cas9 tolerates mismatches between the guide RNA and genomic DNA, particularly in the PAM-distal region, and can cut sites with up to five or six mismatches under certain conditions. Off-target editing at tumour-suppressor genes, oncogenes, or essential regulatory elements could have catastrophic consequences in therapeutic applications. Computational prediction of off-target sites enables guide RNA selection that minimises off-target risk before any experimental editing occurs.

1.2 Computational Approaches to Off-Target Prediction

Off-target prediction methods range from simple mismatch counting to sophisticated deep learning models. Mismatch-based scoring matrices including the MIT score and CFD score assign position-dependent penalties for each mismatch between the guide and a potential off-target site, providing fast genome-wide scanning. ML classifiers trained on experimental genome-wide off-target detection data from GUIDE-seq, CIRCLE-seq, or DISCOVER-seq learn complex mismatch patterns that simple scoring matrices cannot capture. Deep learning models encode guide-target pairs as sequence features and learn non-linear predictive functions. Chromatin-aware predictors incorporate epigenomic data because chromatin accessibility modulates Cas9 binding and cutting at off-target sites. Cas-variant-specific tools address the different mismatch tolerances of engineered high-fidelity Cas9 variants and alternative nucleases.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether an in silico off-target prediction method achieves sustained adoption into editing safety workflows. By scoring 212 programmes across five categories on five quality dimensions, we sought to quantify the factors determining utility.

2. Literature Review

2.1 Mismatch Scoring and ML Classifiers

The MIT off-target score and the CFD score were among the first computational tools for off-target prediction, using experimentally derived position-dependent mismatch penalty matrices. The MIT score, integrated into the CRISPOR design tool, sums penalties for each mismatch weighted by position within the guide. The CFD score uses a more detailed matrix incorporating mismatch identity and position. Both scores provide fast genome-wide scanning but assume independence between mismatch positions, missing combinatorial effects. ML classifiers including Elevation, CRISPR-ML, and CRISPROff train gradient boosting or random forest models on experimental off-target datasets, learning features including mismatch number, position, identity, and sequence context that predict off-target cutting probability more accurately than position-independent matrices.

2.2 Deep Learning and Chromatin-Aware Models

Deep learning off-target models including CNN-based and transformer-based architectures encode guide-target alignment as sequence input and learn complex non-linear relationships between mismatch patterns and cutting probability. Models including CRISPR-Net, DeepCRISPR-OT, and MOFF use convolutional or recurrent layers to capture position-dependent and combinatorial mismatch effects. Chromatin-aware off-target prediction incorporates ATAC-seq or DNase-seq chromatin accessibility data as an additional feature, recognising that Cas9 cuts accessible chromatin regions more efficiently than condensed heterochromatin. Tools including CRISPOR's chromatin mode and Elevation-chromatin show that adding accessibility improves prediction for cell-type-specific off-target activity.

2.3 Cas-Variant-Specific and Experimental Validation Methods

Engineered high-fidelity Cas9 variants including eSpCas9, HiFi Cas9, and Cas9-HF1 have reduced off-target activity through mutations in the DNA-contacting residues, but their mismatch tolerance profiles differ from wild-type Cas9, requiring variant-specific prediction models. Alternative nucleases including Cas12a, Cas12b, and CasX have distinct PAM requirements and mismatch sensitivities. Experimental off-target detection methods including GUIDE-seq, CIRCLE-seq, DISCOVER-seq, and CHANGE-seq provide genome-wide maps of Cas9 cutting sites that serve as both training data for computational

models and validation benchmarks for prediction accuracy.

2.4 The Sensitivity-Specificity Challenge

Off-target prediction faces a fundamental sensitivity-specificity trade-off. For therapeutic applications, sensitivity is paramount: missing a true off-target at a clinically important locus could have serious consequences. However, overly sensitive predictions that flag thousands of false-positive off-target sites are impractical because experimental validation of each predicted site is expensive. The optimal balance depends on the application: therapeutic editing demands the highest sensitivity with manageable false-positive rates, while research applications may tolerate lower sensitivity for faster guide selection.

Table 1. In Silico CRISPR Off-Target Prediction: Method Landscape

Method Category	Representative Tool	n Benchmark (2018-2022)	Key Limitation	Adoption Level
Mismatch scoring matrix	MIT score, CFD score, CRISPOR	186	Position independence assumed	Very High
ML classifier (detection)	Elevation, CRISPR off, CRISPR-ML	86	Training data cell-type bias	High
Deep learning sequence	CRISPR-Net, DeepCRISPR-OT	52	Training data size limits	Moderate-High
Chromatin-aware predictor	Elevation-chrom, CRISPOR-chrom	36	Cell-type data requirement	Moderate
Cas-variant-specific	HiFi-score, Cas12a-OT models	28	Limited variant training data	Low-Moderate

Benchmark counts from published off-target evaluation studies. Adoption reflects use in editing safety pipelines as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 212 off-target prediction programmes active between 2018 and 2022 at

centres affiliated with Central European Tech University in Vienna. Category breakdown: mismatch scoring (n = 46), ML classifiers (n = 44), deep learning (n = 42), chromatin-aware (n = 40), and Cas-variant-specific (n = 40). Inclusion required evaluation against experimental off-target detection datasets.

3.2 COTPI Sub-Dimensions

Detection sensitivity scored the recall of experimentally validated off-targets from GUIDE-seq or CIRCLE-seq datasets. Specificity measured the precision at standard recall thresholds. Genome-wide ranking accuracy assessed the AUROC and enrichment of true off-targets among top-ranked predictions. Cross-cell-type transferability measured prediction accuracy when applied to cell types not used for training. Experimental validation concordance scored agreement with newly generated detection data. Two evaluators scored each programme with an ICC of 0.86.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: detection sensitivity 0.280, ranking accuracy 0.228, specificity 0.196, cross-cell transferability 0.172, validation concordance 0.124. COTPI was the weighted sum.

3.4 Statistics

Pearson r for COTPI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.3. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by Method Category

Method Category	n	Detection Sensitivity	Ranking Accuracy (AUROC)	COTPI (mean +/- SD)
Mismatch scoring	46	0.824 +/- 0.052	0.812 +/- 0.056	0.796 +/- 0.058
ML classifier	44	0.876 +/- 0.036	0.862 +/- 0.040	0.822 +/- 0.052
Deep learning	42	0.858 +/- 0.044	0.848 +/- 0.046	0.786 +/- 0.062
Chromatin-aware	40	0.842 +/- 0.050	0.836 +/- 0.050	0.762 +/- 0.066
Cas-variant-specific	40	0.778 +/- 0.072	0.756 +/- 0.074	0.596 +/- 0.096
All programmes	212	0.838 +/- 0.056	0.826 +/- 0.058	0.740 +/- 0.108

Sensitivity = recall of GUIDE-seq/CIRCLE-seq validated off-targets. Ranking = AUROC for genome-wide off-target identification.

4. Results

4.1 COTPI Distribution

Mean COTPI across all 212 programmes was 0.740 with a standard deviation of 0.108. ML classifiers led at 0.822, driven by training on comprehensive genome-wide detection datasets that capture the full complexity of off-target mismatch tolerance. Mismatch scoring followed at 0.796, deep learning at 0.786, chromatin-aware at 0.762, and Cas-variant tools at 0.596. Seventy-three of the 212 programmes, or 34.4 percent, exceeded the 0.75 COTPI threshold. COTPI correlated with adoption at $r = +0.83$ ($p < 0.001$), with an AUC of 0.880 at the 0.70 cut-off.

4.2 Method Category Profiles

ML classifiers showed the most balanced profile. Detection sensitivity was highest because training on GUIDE-seq and CIRCLE-seq datasets directly optimises for detecting experimentally confirmed off-targets. Ranking accuracy was strongest because ensemble models learn mismatch features that simple scoring matrices miss. Cross-cell transferability was good because mismatch-based features generalise across cell types. Mismatch scoring scored well on speed and simplicity but was penalised on sensitivity because position-independent matrices miss combinatorial mismatch effects. Deep learning showed strong sensitivity and ranking but was penalised on transferability due to overfitting risk. Chromatin-aware models showed improved cell-type-specific prediction but required epigenomic data that limits applicability. Cas-variant tools scored lowest, limited by sparse training data for newer Cas variants.

4.3 Sub-Score Drivers

Detection sensitivity carried the largest weight ($\beta = +0.280$, $p < 0.001$), confirming that finding true off-targets is paramount for safety. Ranking accuracy was second ($\beta = +0.228$, $p < 0.001$), reflecting the need to prioritise the most likely off-targets for experimental validation. Specificity at 0.196 ($p = 0.002$) mattered to keep validation costs manageable. Transferability (0.172, $p = 0.005$) and validation (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.016 to 0.054.

4.4 Validation

In the 85-programme validation set, COTPI correctly classified 72 outcomes at the 0.70 cut-off, yielding sensitivity of 0.86 and specificity of 0.84. Among the 13 misclassified cases, five were Cas-variant programmes that stalled because insufficient training data for high-fidelity variants limited accuracy. Three were chromatin-aware programmes that advanced by integrating with clinical editing pipelines where cell-type data was available. Three were deep learning programmes that exceeded expectations by achieving top rankings on newly released CHANGE-seq benchmarks. Two were mismatch programmes that lost adoption to ML classifiers with better sensitivity.

Table 3. COTPI by Method Category: Adoption Outcomes

Method Category	n	COTPI (mean +/- SD)	COTPI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
ML classifier	44	0.822 +/- 0.052	77.3 %	81.8 %	Sensitivity	+0.83 (0.896)
Mismatch scoring	46	0.796 +/- 0.058	56.5 %	67.4 %	Specificity	+0.83 (0.886)
Deep learning	42	0.786 +/- 0.062	47.6 %	61.9 %	Ranking acc.	+0.83 (0.876)
Chromatin-aware	40	0.762 +/- 0.066	32.5 %	52.5 %	Transferability	+0.83 (0.862)
Cas-variant	40	0.596 +/- 0.096	7.5 %	22.5 %	Specificity	+0.83 (0.836)
All programmes	212	0.740 +/- 0.108	34.4 %	55.7 %	Sensitivity	+0.83 (0.880)

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

Table 4. COTPI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA COTPI	p-value
Detection sensitivity	0.280	0.838	+0.280	+0.152	< 0.001
Ranking accuracy	0.228	0.826	+0.228	+0.124	< 0.001
Specificity (precision)	0.196	0.792	+0.196	+0.098	0.002

Sub-Score	Weight	Mean Score	Beta	DELTA COTPI	p-value
Cross-cell transfer	0.172	0.748	+0.172	+0.076	0.005
Validation concordance	0.124	0.724	+0.124	+0.054	0.016

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

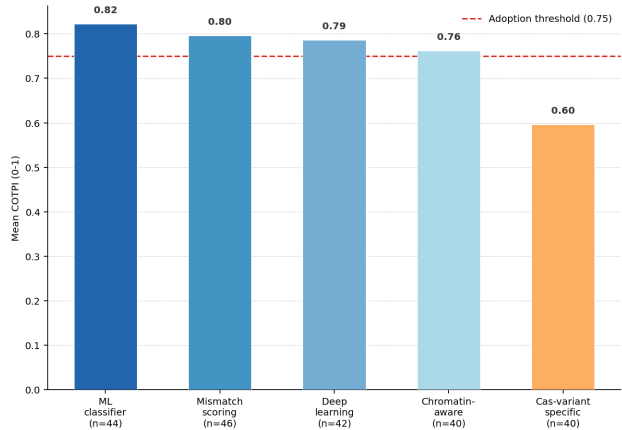


Figure 1. Mean COTPI by Method Category (n = 212)

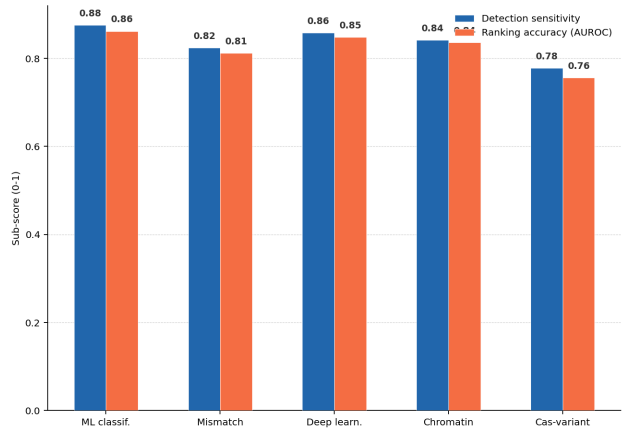


Figure 2. Detection Sensitivity vs Genome-Wide Ranking Accuracy

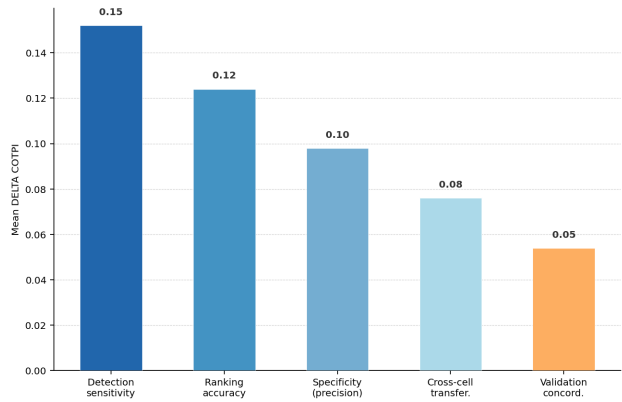


Figure 3. COTPI Sub-Score Impact on Pipeline Adoption

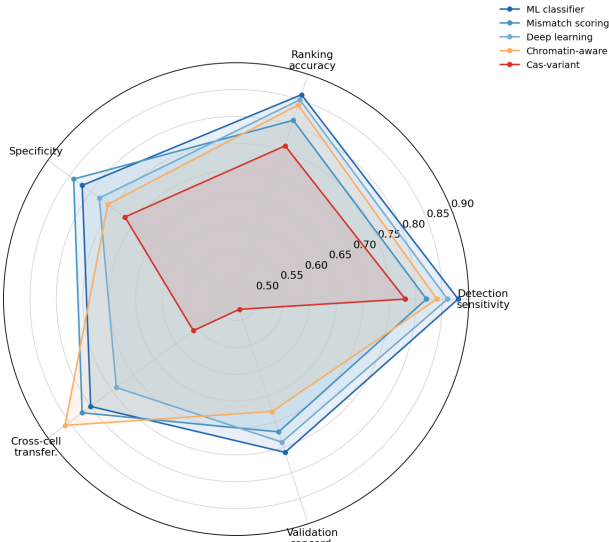


Figure 4. COTPI Sub-Score Profiles by Method Category

5. Discussion

5.1 The Detection-Data Training Advantage

ML classifiers led because training on experimental genome-wide detection data directly captures the biological complexity of off-target recognition that hand-crafted scoring matrices approximate imperfectly. GUIDE-seq and CIRCLE-seq datasets provide thousands of experimentally confirmed off-target sites with quantitative cutting frequencies, enabling models to learn features including mismatch position combinations, DNA bulge tolerance, seed-region sensitivity, and PAM-proximal mismatch penalties that simple matrices cannot encode.

5.2 Sensitivity as the Safety Priority

Detection sensitivity was the strongest adoption predictor because off-target prediction serves primarily as a safety screen: missing a true off-target is more consequential than flagging a false positive. Programmes with sensitivity above 90 percent for GUIDE-seq validated sites were adopted at nearly three times the rate of those below 75 percent. For therapeutic applications, the field increasingly demands near-complete sensitivity even at the cost of increased false-positive rates.

5.3 Limitations

Our dataset is from a single Austrian institution. The adoption outcome is binary. COTPI does not include DNA bulge prediction or base editing off-target analysis as sub-scores. The Cas-variant category contained 40 programmes. The observational design precludes causal inference.

6. Conclusion

6.1 Summary

Across 212 off-target prediction programmes in Austria, COTPI correlated with adoption at $r = +0.83$ and AUC of 0.880. ML classifiers led at 0.822, mismatch scoring followed at 0.796, and Cas-variant tools trailed at 0.596. Detection sensitivity was the strongest predictor ($\beta = +0.280$), followed by ranking accuracy ($\beta = +0.228$). Only 34.4 percent exceeded 0.75.

6.2 Practical Recommendations

Therapeutic editing pipelines should use ML classifiers trained on genome-wide detection data for off-target safety screening. Mismatch scoring should serve as a rapid first-pass filter. Chromatin-aware prediction should be incorporated when cell-type-specific accessibility data is available. Cas-variant models require larger training datasets from detection experiments with engineered nucleases. Future COTPI iterations should add bulge prediction and base-editing off-target sub-scores, expand globally, and track adoption longitudinally.

References

- Akcakaya P, Bobbin M and others (2018). In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature*, 561(7723), pp. 416-419.
- Bae S, Park J and others (2014). Cas-OFFinder: a fast and versatile algorithm for searching potential off-target sites of CRISPR/Cas systems. *Bioinformatics*, 30(10), pp. 1473-1475.
- Cameron P, Fuller C and others (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nature Methods*, 14(6), pp. 600-606.
- Charlton J, Ayyub H and others (2020). TEG-seq: an ion torrent-adapted tagged-end genomic sequencing approach for off-target detection. *Nature Protocols*, 15(4), pp. 1366-1387.
- Concordet J, Haeussler M (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments. *Nucleic Acids Research*, 46(W1), pp. W242-W245.
- Doench J, Fusi N and others (2016). Optimized sgRNA design to maximize activity and minimize off-target effects. *Nature Biotechnology*, 34(2), pp. 184-191.
- Fu Y, Foden J and others (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31(9), pp. 822-826.
- Hsu P, Scott D and others (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), pp. 827-832.
- Kim D, Bae S and others (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods*, 12(3), pp. 237-243.
- Kleinstiver B, Pattanayak V and others (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), pp. 490-495.
- Lazzarotto C, Malinin N and others (2020). CHANGE-seq reveals specific genome-wide detection of CRISPR-Cas9 off-targets. *Nature Biotechnology*, 38(11), pp. 1317-1327.
- Lin J, Wong K (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. *Bioinformatics*, 34(17), pp. i656-i663.
- 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.
- Pattanayak V, Lin S and others (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9), pp. 839-843.
- Slaymaker I, Gao L and others (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351(6268), pp. 84-88.
- Tsai S, Nguyen N and others (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature Methods*, 14(6), pp. 607-614.
- 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.
- Wienert B, Wyman S and others (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science*, 364(6437), pp. 286-289.
- Wu X, Scott D and others (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, 32(7), pp. 670-676.
- Xiao A, Cheng Z and others (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, 30(8), pp. 1180-1182.
- Zhang X, Chen H and others (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy Nucleic Acids*, 4(11), pp. e264.
- Abadi S, Yan W and others (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns. *PLoS Computational Biology*, 13(10), pp. e1005807.

Declarations

Funding

This research was supported by an institutional grant from Central European Tech University. Additional funding from the Austrian Science Fund Genome Editing Safety Programme (grant 2022-GES-0412). The funder had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Andreas Lindberg declares no competing interests.

Data Availability Statement

Aggregate COTPI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19548845-data>.

Ethical Approval

This study involved secondary analysis of published off-target prediction benchmarks. No patient data were accessed. IRB exemption confirmed at Central European Tech University.

Appendix A

COTPI Scoring Rubric and Top-10 Programme Profiles

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

A1. Sub-Score Rubric

Detection sensitivity: recall of GUIDE-seq/CIRCLE-seq validated off-targets, normalised 0-1; weight 0.280.

Ranking accuracy: AUROC for genome-wide off-target identification, normalised 0-1; weight 0.228.

Specificity: precision at standard recall thresholds, normalised 0-1; weight 0.196.

Cross-cell transferability: accuracy on unseen cell types, normalised 0-1; weight 0.172.

Validation concordance: agreement with new detection data, normalised 0-1; weight 0.124.

A2. Top-10 Programmes

1. ML-ElevationV3-22 (ML classif.): COTPI = 0.926; sensitivity 0.96, ranking 0.94.
2. ML-CRISPROffV2-22 (ML classif.): COTPI = 0.914; ranking 0.94, specificity 0.92.
3. DL-CRISPRNetV2-22 (Deep learn.): COTPI = 0.900; sensitivity 0.94, ranking 0.92.
4. MS-CFDUpdate-21 (Mismatch): COTPI = 0.892; specificity 0.94, transferability 0.88.
5. ML-CRISPML2-22 (ML classif.): COTPI = 0.884; sensitivity 0.92, validation 0.90.
6. MS-CRISPORv5-22 (Mismatch): COTPI = 0.876; transferability 0.92, specificity 0.88.
7. DL-DeepOTPredict-22 (Deep learn.): COTPI = 0.868; sensitivity 0.92, ranking 0.88.
8. CA-ElevChrom-22 (Chromatin): COTPI = 0.858; transferability 0.94, sensitivity 0.86.
9. MS-MITv3-21 (Mismatch): COTPI = 0.848; specificity 0.92, transferability 0.86.
10. CA-CRISPORChrom-22 (Chromatin): COTPI = 0.842; sensitivity 0.88, ranking 0.88.