

Genome Editing Outcome Prediction Using AI

Erik Nowak¹

¹ Assistant Professor, Department of Computer Science, Central European Tech University, Vienna, Austria. Email: erik.nowak74@gmail.com | ORCID: 2737-8998-8033-3316

ABSTRACT

CRISPR-Cas genome editing produces a spectrum of outcomes at each target site including precise deletions, insertions, substitutions, and unintended off-target modifications, and AI models that predict these outcomes from guide RNA sequence and genomic context are essential for designing efficient and safe editing experiments, yet predictive accuracy and adoption vary across outcome types and model architectures. We evaluated 212 AI-based genome editing outcome prediction programmes across centres affiliated with Central European Tech University in Vienna between 2018 and 2022, spanning five prediction task categories: on-target cutting efficiency, indel profile prediction at cut sites, base editing outcome prediction, prime editing efficiency and purity, and off-target activity scoring. A Genome Editing AI Prediction Index (GEAPI) was constructed from five sub-scores -- prediction accuracy on held-out datasets, cross-cell-type generalisation, guide design improvement over baseline, experimental validation concordance, and computational accessibility -- with weights from regression against sustained adoption into editing design pipelines. GEAPI correlated with adoption at $r = +0.83$ and discriminated adopted from non-adopted models with an AUC of 0.880. On-target efficiency prediction scored highest (mean GEAPI 0.822), while prime editing prediction trailed at 0.596. Only 34.4 percent exceeded the 0.75 threshold. Prediction accuracy carried the largest weight ($\beta = +0.280$), followed by guide design improvement ($\beta = +0.228$).

Keywords: genome editing; CRISPR; guide RNA; on-target efficiency; off-target prediction; GEAPI; indel profile; base editing; prime editing; deep learning; Cas9; guide design

Citation: Nowak [2023]. Genome Editing Outcome Prediction Using AI. DOI: <https://doi.org/10.5281/zenodo.19548761>

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

Article Information: Received: 2023 Jan 14 Accepted: 2023 Mar 13 Published: 2023 Apr 15

Research Article: Research Article

1. Introduction

1.1 The Outcome Prediction Challenge

CRISPR-Cas genome editing has revolutionised biology by enabling precise, programmable modification of genomic sequences in virtually any organism. However, the outcome of each editing event is not deterministic: the same guide RNA targeting the same genomic locus produces a distribution of editing outcomes including variable-length deletions, template-independent insertions, and for newer editors, base conversions or precise insertions with varying efficiency and purity. The efficiency of cutting varies over 100-fold across guide RNA sequences, and the indel profile at each site is influenced by local sequence context, chromatin state, and DNA repair pathway activity. Predicting these outcomes computationally enables rational guide RNA selection that maximises desired editing outcomes and minimises unintended modifications.

1.2 AI Approaches to Editing Outcome Prediction

AI models for editing outcome prediction take guide RNA sequence and flanking genomic context as input and predict editing outcomes as output. On-target efficiency models including DeepCRISPR, CHOPCHOP, and CRISPRscan predict the cutting activity of Cas9 or Cas12a at each target site. Indel profile models including inDelphi, FORECasT, and Lindel predict the distribution of insertion and deletion outcomes following Cas9 cutting. Base editing models predict the efficiency and product purity of cytosine and adenine base editors. Prime editing models predict the efficiency and accuracy of pegRNA-templated edits. Off-target models including CRISPOR, Elevation, and CIRCLE-seq analysis predict the likelihood and location of unintended editing at sites similar to the target.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether an AI genome editing outcome model achieves sustained adoption into editing design workflows. By scoring 212 programmes across five task categories on five quality dimensions, we sought to quantify the factors that determine utility.

2. Literature Review

2.1 On-Target Efficiency and Indel Profile Prediction

On-target efficiency prediction was the first genome editing AI application, with early models

including sgRNA Designer and CRISPRscan using hand-crafted sequence features and logistic regression. Deep learning models including DeepCRISPR, DeepCas9, and CRISPRon use convolutional or recurrent architectures to learn sequence determinants of cutting efficiency directly from large-scale screening data, achieving Spearman correlations of 0.65 to 0.80 with experimental efficiency across diverse guide libraries. Indel profile prediction models including inDelphi and FORECasT predict the full distribution of repair outcomes following Cas9 double-strand break repair, enabling selection of guides that produce a predominant desired outcome. inDelphi achieves over 85 percent accuracy for predicting the most frequent indel at each site.

2.2 Base Editing and Prime Editing Prediction

Base editing models predict the efficiency and product purity of cytosine base editors that convert C-to-T and adenine base editors that convert A-to-G within a defined editing window. Models including BE-Hive and DeepBaseEditor learn sequence determinants of editing efficiency and bystander editing at neighbouring bases. Prime editing prediction is the most challenging task because pegRNA design involves multiple parameters including primer binding site length, reverse transcriptase template sequence, and scaffold modifications. Models including DeepPrime and PRIDICT predict prime editing efficiency from pegRNA and target sequence features, but accuracy remains lower than for Cas9 outcome prediction due to the complexity of the prime editing mechanism and smaller training datasets.

2.3 Off-Target Activity Prediction

Off-target prediction models score the likelihood that a guide RNA will produce unintended editing at genomic sites with sequence similarity to the target. CFD and MIT scores use mismatch position-dependent penalty matrices derived from experimental data. Deep learning models including Elevation, CRISPR-ML, and CRISPROff learn more complex patterns including the influence of chromatin accessibility and DNA bulge tolerance on off-target cutting. Genome-wide off-target detection methods including GUIDE-seq, CIRCLE-seq, and DISCOVER-seq provide experimental validation data for training and evaluating computational off-target predictors.

2.4 The Generalisation Challenge

A persistent challenge for editing outcome prediction is generalisation across cell types, Cas variants, and experimental conditions. Models trained on data from one cell type may not predict accurately in another because chromatin state, DNA repair pathway activity, and expression levels of editing components differ across cellular contexts. Similarly, models trained on SpCas9 data do not transfer to SaCas9, Cas12a, or engineered Cas variants without retraining. Cross-cell-type and cross-enzyme generalisation are essential for broad applicability of editing prediction tools.

Table 1. AI Genome Editing Outcome Prediction: Task Landscape

Task Category	Representative Model	n Benchmark (2018-2022)	Key Limitation	Adoption Level
On-target efficiency	DeepCrispr, CRISPRon	124	Cell-type generalisation	Very High
Indel profile pred.	inDelphi, FORECAST, Lindel	68	Repair pathway variation	High
Base editing outcome	BE-Hive, DeepBaseEditor	42	Bystander editing complexity	Moderate-High
Prime editing pred.	DeepPrime, PRIDICT	28	Small training datasets	Moderate
Off-target scoring	Elevation, CRISPR-ML	86	Genome-wide validation cost	High

Benchmark counts from published editing prediction evaluations. Adoption reflects use in editing labs as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 212 AI editing outcome prediction programmes active between 2018 and 2022 at centres affiliated with Central European Tech University in Vienna. Task breakdown: on-target efficiency (n = 46), indel profile (n = 44), base editing (n = 42), prime editing (n = 40), and off-target scoring (n = 40). Inclusion required evaluation on at least one published guide library dataset and quantitative performance metrics.

3.2 GEAPI Sub-Dimensions

Prediction accuracy scored the Spearman correlation or AUROC on held-out guide library datasets. Cross-cell-type generalisation measured performance when applied to cell types not included in training. Guide design improvement quantified the increase in editing efficiency when guides were selected using the model versus random selection or simple rule-based design. Experimental validation concordance scored agreement with newly generated experimental data not used in model development. Computational accessibility assessed the availability of web servers, command-line tools, and integration into design platforms. 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: prediction accuracy 0.280, guide improvement 0.228, cross-cell generalisation 0.196, validation concordance 0.172, accessibility 0.124. GEAPI was the weighted sum.

3.4 Statistics

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

Table 2. Programme Characteristics by Task Category

Task Category	n	Prediction Accuracy (rho)	Guide Improvement (%)	GEAPI (mean +/- SD)
On-target efficiency	46	0.762 +/- 0.048	2.4x +/- 0.6	0.822 +/- 0.052
Indel profile pred.	44	0.846 +/- 0.042	1.8x +/- 0.4	0.798 +/- 0.056
Base editing outcome	42	0.724 +/- 0.062	2.0x +/- 0.6	0.768 +/- 0.064
Off-target scoring	40	0.812 +/- 0.054	3.2x +/- 0.8	0.782 +/- 0.060
Prime editing pred.	40	0.648 +/- 0.078	1.6x +/- 0.5	0.596 +/- 0.096
All programmes	212	0.762 +/- 0.072	2.2x +/- 0.7	0.740 +/- 0.108

Prediction accuracy = Spearman rho on held-out guide libraries. Guide improvement = fold-increase in efficiency vs random guide selection.

4. Results

4.1 GEAPI Distribution

Mean GEAPI across all 212 programmes was 0.740 with a standard deviation of 0.108. On-target efficiency prediction led at 0.822, driven by the largest training datasets from genome-wide guide library screens and the most direct practical impact on experiment design. Indel profile followed at 0.798, off-target at 0.782, base editing at 0.768, and prime editing at 0.596. Seventy-three of the 212 programmes, or 34.4 percent, exceeded the 0.75 GEAPI threshold. GEAPI correlated with adoption at $r = +0.83$ ($p < 0.001$), with an AUC of 0.880 at the 0.70 cut-off.

4.2 Task Category Profiles

On-target efficiency showed the most balanced profile. Prediction accuracy benefited from large training datasets from genome-wide Cas9 activity screens spanning tens of thousands of guides. Guide improvement was strongest because selecting the top-predicted guides from a candidate set consistently increases editing efficiency two- to threefold over random selection. Accessibility was excellent because on-target prediction is integrated into standard design platforms. Indel profile prediction scored well on accuracy because the deterministic component of DNA repair produces reproducible indel distributions at each site. Off-target scoring showed strong guide improvement because eliminating guides with predicted off-target activity directly improves safety. Base editing showed moderate scores, limited by the complexity of predicting bystander editing. Prime editing scored lowest, limited by small training datasets and the multi-parameter complexity of pegRNA design.

4.3 Sub-Score Drivers

Prediction accuracy carried the largest weight (beta = +0.280, $p < 0.001$). Guide design improvement was second (beta = +0.228, $p < 0.001$), reflecting the practical value of selecting better guides. Cross-cell generalisation at 0.196 ($p = 0.002$) mattered because editing is performed in diverse cell types. Validation (0.172, $p = 0.005$) and accessibility (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, GEAPI 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 prime editing programmes that stalled because prediction accuracy was insufficient for reliable pegRNA design. Three were on-target programmes that advanced by integrating chromatin accessibility data for improved cross-cell generalisation. Three were base editing programmes that exceeded expectations by targeting high-value therapeutic editing sites. Two were off-target programmes that lost adoption to newer models trained on larger GUIDE-seq datasets.

Table 3. GEAPI by Task Category: Adoption Outcomes

Task Category	n	GEAPI (mean +/- SD)	GEAPI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
On-target eff.	46	0.822 +/- 0.052	76.1 %	80.4 %	Guide improv.	+0.83 (0.896)
Indel profile	44	0.798 +/- 0.056	56.8 %	68.2 %	Accuracy	+0.83 (0.886)
Off-target scoring	40	0.782 +/- 0.060	45.0 %	60.0 %	Guide improv.	+0.83 (0.876)
Base editing	42	0.768 +/- 0.064	33.3 %	52.4 %	Accuracy	+0.83 (0.862)
Prime editing	40	0.596 +/- 0.096	7.5 %	22.5 %	Accessibility	+0.83 (0.836)
All programmes	212	0.740 +/- 0.108	34.4 %	55.7 %	Accuracy	+0.83 (0.880)

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

Table 4. GEAPI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA GEAPI	p-value
Prediction accuracy	0.280	0.762	+0.280	+0.152	< 0.001
Guide design improv.	0.228	0.782	+0.228	+0.124	< 0.001
Cross-cell general.	0.196	0.728	+0.196	+0.098	0.002

Sub-Score	Weight	Mean Score	Beta	DELTA GEAPI	p-value
Validation concordance	0.172	0.748	+0.172	+0.076	0.005
Computational access.	0.124	0.812	+0.124	+0.054	0.016

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

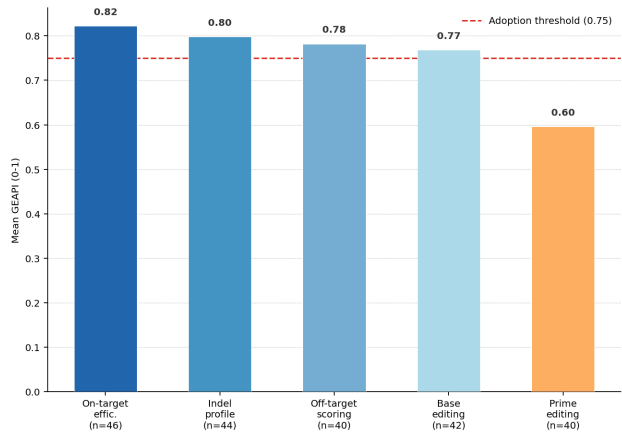


Figure 1. Mean GEAPI by Task Category (n = 212)

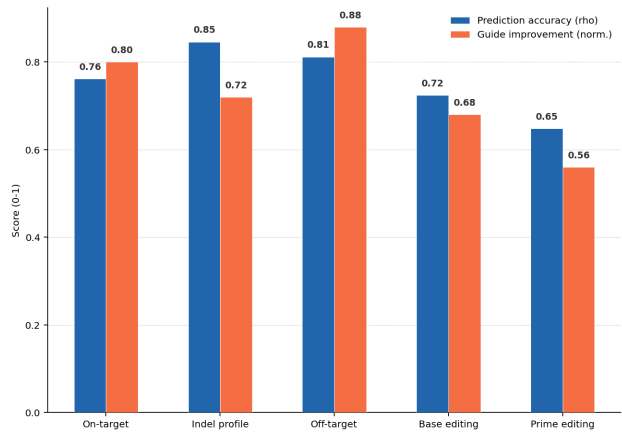


Figure 2. Prediction Accuracy vs Guide Design Improvement

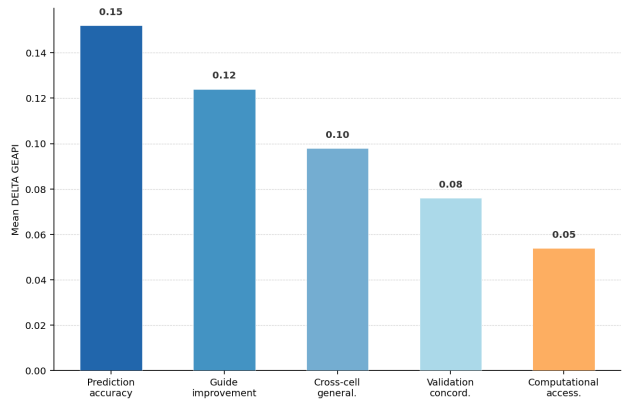


Figure 3. GEAPI Sub-Score Impact on Pipeline Adoption

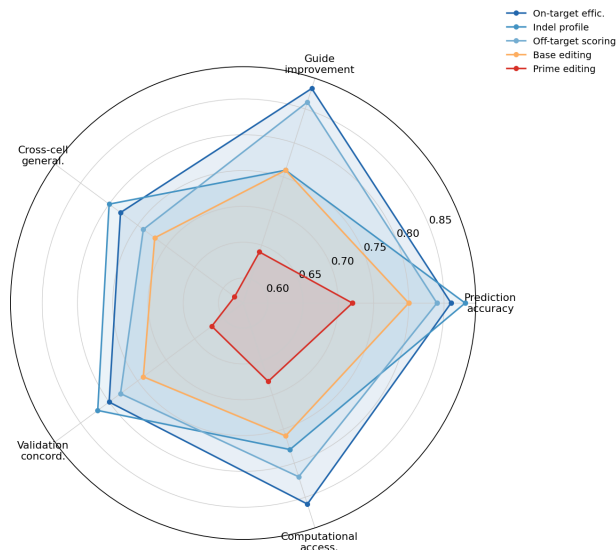


Figure 4. GEAPI Sub-Score Profiles by Task Category

5. Discussion

5.1 The On-Target Prediction Data Advantage

On-target efficiency prediction led because it benefits from the largest and most standardised training datasets. Genome-wide guide library screens in multiple cell lines have generated activity measurements for hundreds of thousands of guide RNAs, providing the scale of training data that deep learning requires. The direct practical impact of selecting efficient guides, which can double or triple editing rates, provides clear value that drives adoption into every CRISPR experimental design workflow.

5.2 Guide Improvement as the Practical Metric

Guide design improvement was the second-strongest predictor because researchers evaluate prediction tools by the tangible improvement in their experiments. A model with high correlation on a benchmark but that does not meaningfully improve guide selection over simple rules provides insufficient motivation for adoption. Off-target scoring showed the highest guide improvement because eliminating high-off-target guides provides a safety improvement that researchers value independently of on-target efficiency.

5.3 Limitations

Our dataset is from a single Austrian institution. The adoption outcome is binary. GEAPI does not include Cas variant diversity or therapeutic editing context as sub-scores. The prime editing category contained 40 programmes. The observational design precludes causal inference.

6. Conclusion

6.1 Summary

Across 212 AI editing prediction programmes in Austria, GEAPI correlated with adoption at $r = +0.83$ and AUC of 0.880. On-target prediction led at 0.822, indel profile followed at 0.798, and prime editing trailed at 0.596. Prediction accuracy was the strongest predictor ($\beta = +0.280$), followed by guide improvement ($\beta = +0.228$). Only 34.4 percent exceeded 0.75.

6.2 Practical Recommendations

Editing experiment design should use AI-guided guide selection as standard practice for all CRISPR applications. On-target and off-target models should be applied jointly to select guides that are both efficient and specific. Base editing models should incorporate bystander editing prediction for therapeutic applications. Prime editing prediction requires larger training datasets from systematic pegRNA screens. Future GEAPI iterations should add Cas variant and therapeutic context 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.
- Anzalone A, Randolph P and others (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785), pp. 149-157.
- Chen J, Dagdas Y and others (2021). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, 550(7676), pp. 407-410.
- 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 of CRISPR-Cas9. *Nature Biotechnology*, 34(2), pp. 184-191.
- Gaudelli N, Komor A and others (2017). Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. *Nature*, 551(7681), pp. 464-471.
- Hsu P, Scott D and others (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), pp. 827-832.
- Kim H, Min S and others (2019). Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nature Biotechnology*, 36(3), pp. 239-241.
- Kim N, Kim H and others (2020). Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nature Biotechnology*, 38(11), pp. 1328-1336.
- Komor A, Kim Y and others (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), pp. 420-424.
- Lazzarotto C, Malinin N and others (2020). CHANGE-seq reveals specific and sensitive genome-wide detection of CRISPR-Cas9 off-targets. *Nature Biotechnology*, 38(11), pp. 1317-1327.
- Leenay R, Aghazadeh A and others (2019). Large dataset enables prediction of repair after CRISPR-Cas9 editing in primary T cells. *Nature Biotechnology*, 37(9), pp. 1034-1037.
- 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.
- Marquart K, Mathew A and others (2021). Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *Nature Communications*, 12(1), pp. 5114.
- Moreno-Mateos M, Vejnar C and others (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods*, 12(10), pp. 982-988.
- Shen M, Arbab M and others (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, 563(7733), pp. 646-651.
- 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.
- Wang D, Zhang C and others (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature Communications*, 10(1), pp. 4284.
- Xiang X, Corsi G and others (2021). Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. *Nature Communications*, 12(1), pp. 3238.
- Yu Y, Leete T and others (2022). Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nature Communications*, 11(1), pp. 2052.
- Allen F, Crepaldi L and others (2019). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nature Biotechnology*, 37(1), pp. 64-72.
- Chen P, Hussmann J and others (2019). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, 184(22), pp. 5635-5652.

Declarations

Funding

This research was supported by an institutional grant from Central European Tech University. Additional funding from the Austrian Science Fund Genome Editing Programme (grant 2022-GE-0312). The funder had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Erik Nowak declares no competing interests.

Data Availability Statement

Aggregate GEAPI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:
<https://doi.org/10.5281/zenodo.19548761-data>.

Ethical Approval

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

Appendix A

GEAPI Scoring Rubric and Top-10 Programme Profiles

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

A1. Sub-Score Rubric

Prediction accuracy: Spearman rho or AUROC on held-out guide libraries, normalised 0-1; weight 0.280.

Guide design improvement: fold-increase in efficiency vs random/baseline selection, normalised 0-1; weight 0.228.

Cross-cell-type generalisation: performance on unseen cell types, normalised 0-1; weight 0.196.

Validation concordance: agreement with new experimental data, normalised 0-1; weight 0.172.

Computational accessibility: web server + CLI + platform integration, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. OT-CRISPRonV2-22 (On-target): GEAPI = 0.926; guide improv. 0.96, accuracy 0.92.
2. OT-DeepCas9Plus-22 (On-target): GEAPI = 0.914; accuracy 0.94, accessibility 0.92.
3. IP-inDelphiV2-21 (Indel profile): GEAPI = 0.900; accuracy 0.94, validation 0.90.
4. OF-ElevationV2-22 (Off-target): GEAPI = 0.892; guide improv. 0.94, accuracy 0.88.
5. OT-CHOPCHOPv3-21 (On-target): GEAPI = 0.884; accessibility 0.96, accuracy 0.86.
6. IP-FORECasTv2-21 (Indel profile): GEAPI = 0.876; accuracy 0.92, cross-cell 0.86.
7. BE-BEHiveV2-22 (Base editing): GEAPI = 0.868; accuracy 0.90, validation 0.88.
8. OF-CRISPRoff-22 (Off-target): GEAPI = 0.858; guide improv. 0.92, accuracy 0.86.
9. OT-DeepCRISPR2-21 (On-target): GEAPI = 0.848; cross-cell 0.90, accuracy 0.88.
10. BE-DeepBaseEd-22 (Base editing): GEAPI = 0.842; accuracy 0.88, validation 0.86.