DEEPPRIME7: PREDICTING PE7 PRIME EDITING EFFICIENCY ACROSS PAM VARIANTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Prime editing efficiency depends critically on pegRNA design and PAM compatibility, but the emergence of Prime Editor 7 (PE7) systems with engineered PAM variants has introduced an additional layer of complexity. Selecting the appropriate editor configuration for each target now requires more than pegRNA ranking. Existing predictors, largely trained on PE2/NGG data, do not explicitly model PE7-or PAM-dependent activity differences.

Here, we present DeepPrime7, a PE7-optimized predictor trained on PAM-diverse pooled screening data that jointly models pegRNA sequence features and PE7 system identity. Using a multi-task formulation and strict guide-disjoint evaluation, DeepPrime7 accurately predicts editing efficiency across NGG-, NRCH-, and NRTH-compatible PE7 systems and enables both PE-system selection and pegRNA prioritization.

1 INTRODUCTION

Prime editing (PE) enables precise genome modification without double-strand breaks, but editing efficiency varies widely depending on pegRNA design, cellular context, and PAM compatibility (Anzalone et al., 2019). While recent deep learning–based predictors have improved pegRNA prioritization, most are trained on PE2 systems with canonical NGG PAMs and do not explicitly model emerging PE7 systems or engineered PAM variants (Yu et al., 2023; Mathis et al., 2023; 2025).

Prime Editor 7 (PE7) improves pegRNA stability through La-protein fusion and exhibits activity profiles distinct from those of PE2 (Yan et al., 2024). In parallel, PAM-relaxed Cas9 variants such as NRCH and NRTH expand the targetable genomic space but introduce PAM-dependent efficiency trade-offs (Miller et al., 2020). As a result, efficient prime editing increasingly requires not only pegRNA ranking but also rational selection among multiple PE7 systems, a task that remains experimentally costly to evaluate exhaustively.

Here, we present DeepPrime7, a PE7-optimized predictor trained on PAM-diverse pooled screening datasets that jointly models pegRNA sequence features and PAM-dependent system behavior. DeepPrime7 enables both pegRNA prioritization and PE-system selection under a strict guide-disjoint generalization regime that mirrors realistic deployment on unseen targets.

2 DATASETS

We constructed pooled PE7 screening libraries spanning substitutions, insertions, and deletions to enable joint modeling of pegRNA sequence features and PAM-dependent system behavior. The libraries include both fixed-guide PAM sweeps and randomized guide–PAM combinations, allowing systematic assessment of PAM-dependent activity while preserving sequence diversity.

Screens were carried out in DLD1 (MMR-deficient) and A549 (MMR-proficient) cell lines expressing PE7 variants. In total, approximately 24k pegRNAs were profiled across NGG-, NRCH-, and NRTH-compatible PE7 systems. Editing efficiencies were quantified by targeted sequencing with background correction and standard quality filtering.

We adopted a strict guide-disjoint split, ensuring that pegRNAs sharing the same guide sequence do not overlap between training and test sets. Library composition is summarized in Table 1.

Table 1: Dataset design summary: pegRNA library composition.

Group	Subgroup	Edit Type	Design strategy / purpose	# pegRNAs
Group 1	Substitution	Sub (1–3 bp)	Broad profiling of point mutations with balanced DeepPrime scores	6,000
Group 2	Indel	Ins/Del (1–20 bp)	Edit length-dependent PE7 activity	11,700
Group 3	Multibase	Multibase	MMR evasion and multi-base editing efficiency	1,000
Group 4	Fixed-guide PAM sweep	PAM variant	12 guides $\times$ 64 PAMs	640
	Random guide–PAM variants	PAM variant	Random guide $\times$ PAM variants	2,913
Group 5	Positive control	Various	Previously validated pegRNAs (Yu et al., 2023; Mathis et al., 2023; 2025)	1,309
Total				**23,562**

3 METHODS

3.1 PROBLEM FORMULATION

We predict prime editing efficiency for a pegRNA–target pair, conditioned on the PE7 system (NGG-, NRCH-, or NRTH-compatible). Because not every pegRNA is assayed in every system, we frame this as a multi-condition regression problem with partially observed labels.

3.2 MODEL

DeepPrime7 is a PE7-optimized multi-task model with a shared sequence encoder and system-specific output heads (Fig. 1). The shared backbone learns pegRNA sequence features common across PE7 systems, while the independent heads capture PAM-dependent activity differences. The sequence input encodes paired wild-type and edited targets spanning the guide, PBS, RT template, and local target context, and is processed by a compact convolutional architecture adapted from prior prime editing predictors. Auxiliary thermodynamic features are concatenated with the sequence-derived representations.

3.3 TRAINING AND EVALUATION

Models are trained under a strict guide-disjoint split. We first train the shared backbone and PE7 system heads on the larger DLD1 dataset, then fine-tune an A549-specific head while keeping the shared parameters fixed to limit overfitting. Losses are computed only for available labels using a validity mask. Performance is assessed on held-out guide-disjoint test sets using rank-based and linear correlation metrics, along with accuracy in predicting the higher-activity PE7 system for a given target.

4 RESULTS

4.1 PAM-DEPENDENT ACTIVITY LANDSCAPES ACROSS PE7 SYSTEMS

We characterized PAM-dependent activity landscapes of PE7, NRCH-PE7, and NRTH-PE7 using pooled screening data. NRCH-PE7 and NRTH-PE7 both expand targetability beyond canonical NGG sites but show distinct and complementary motif preferences. NRCH-PE7 favors NRCN- and NGAN-like motifs, whereas NRTH-PE7 exhibits higher activity at NRTN- and NAAN-like PAMs (Fig. 2).

These results indicate that PAM-variant PE7 systems define system-specific activity regimes. Selecting an appropriate PE7 system is therefore a meaningful design decision rather than a fixed experimental choice.

4.2 PREDICTIVE PERFORMANCE UNDER GUIDE-DISJOINT GENERALIZATION

We next evaluated DeepPrime7 on held-out guide-disjoint test sets to assess generalization to unseen targets. Across substitutions, insertions, and deletions, DeepPrime7 achieves strong predictive

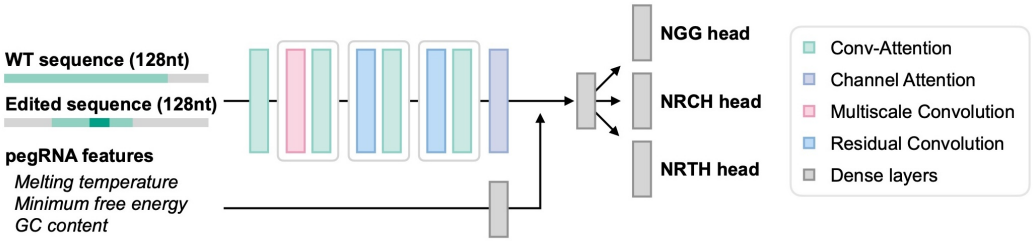
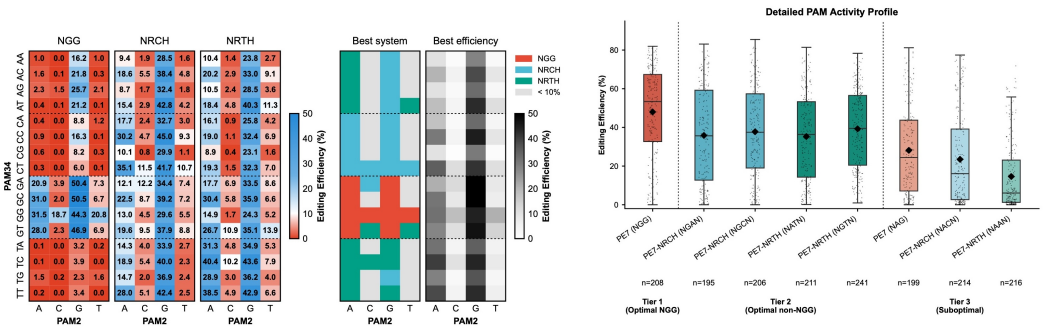


Figure 1: DeepPrime7 model architecture. The overall design is inspired by PrimeNet (Liao et al., 2025), with a shared sequence encoder and system-specific output heads for NGG-, NRCH-, and NRTH-compatible PE7 systems.



(a) PAM-dependent activity landscape.

(b) Activity tiers across PAMs.

Figure 2: PAM-dependent activity landscape and activity tiers of PE7 systems. NRCH-PE7 and NRTH-PE7 expand PAM compatibility beyond NGG-PE7 and show complementary motif-specific strengths (e.g., NRCH-PE7 favoring NRCN/NGAN; NRTH-PE7 favoring NRTN/NAAN).

performance for all PE7 systems (Spearman's $\rho = 0.80$–0.92; Pearson's $r = 0.58$–0.93) in both DLD1 and A549 cells (Fig. 3).

Compared with prior predictors trained primarily on PE2/NGG data, including DeepPrime, DeepPrime-FT, and PRIDICT2.0, DeepPrime7 consistently shows improved rank correlation on PE7 datasets. This performance gain is most pronounced for PAM-variant systems, supporting the importance of explicitly modeling PAM-dependent activity rather than relying on implicit transfer from NGG-only training.

4.3 Prediction enables PE7 system selection

Given the observed PAM-dependent activity differences, we evaluated whether DeepPrime7 can correctly predict which PE7 system yields higher activity for a given pegRNA–target pair. Among pegRNAs with measurable activity ($\geq 1\%$ in at least one system), DeepPrime7 correctly identifies the higher-activity system in 89.1% of PE7 vs. NRCH-PE7 comparisons, 90.6% of PE7 vs. NRTH-PE7 comparisons, and 73.8% of NRCH-PE7 vs. NRTH-PE7 comparisons (Fig. 4).

These results show that DeepPrime7 supports PE-system selection as a practical decision-making task, reducing the need for exhaustive experimental screening.

4.4 ClinVar-based in silico pegRNA design

To assess practical utility in therapeutic design settings, we performed an in silico study on 1,000 randomly sampled pathogenic or likely pathogenic ClinVar variants. For each variant, candidate pegRNAs were generated and ranked using either DeepPrime or DeepPrime7.

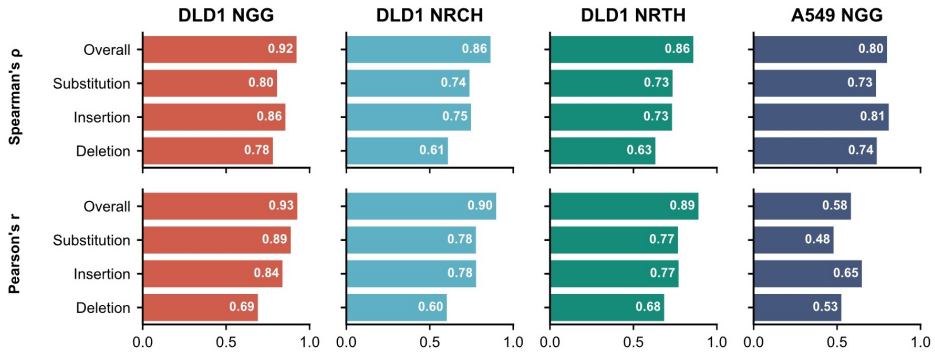


Figure 3: Predictive performance across PE systems. Correlation between predicted and measured editing efficiencies on held-out test data for DLD1-PE7, DLD1-NRCH, DLD1-NRTH, and A549-PE7 across substitutions, insertions, and deletions.

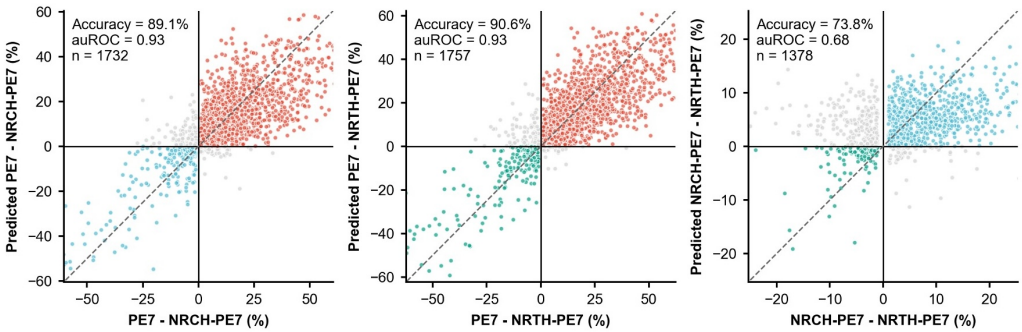


Figure 4: Pairwise PE system selection accuracy. DeepPrime7 predicts the higher-activity system for a given pegRNA with accuracy of 89.1% (PE7 vs. NRCH-PE7), 90.6% (PE7 vs. NRTH-PE7), and 73.8% (NRCH-PE7 vs. NRTH-PE7) among pegRNAs with measurable activity.

DeepPrime7 increased predicted editing efficiency on NGG sites (43.4% $\rightarrow$ 58.0%) compared with DeepPrime and further improved performance when NRCH-PE7 and NRTH-PE7 systems were allowed (Fig. 5). The fraction of variants exceeding 50% predicted mean efficiency increased from 19.0% (DeepPrime) to 61.6% (DeepPrime7 on NGG) and to 74.8% when all PAM-variant systems were included. These results highlight how PAM-aware prediction can substantially expand actionable design space prior to experimental validation.

5 Discussion and Limitations

DeepPrime7 provides PE7- and PAM-aware efficiency prediction that supports both pegRNA ranking and PE-system selection. Although the model generalizes well under guide-disjoint splits, it is trained in specific cellular contexts and assay designs. Therefore, transfer to other cell types or experimental settings may require recalibration. By explicitly incorporating PAM identity and learning sequence-dependent activity from PAM-diverse PE7 screens, DeepPrime7 improves prediction accuracy and broadens the actionable design space for therapeutic prime editing.

References

Andrew V. Anzalone, Peyton B. Randolph, Jessie R. Davis, Alexander A. Sousa, Luke W. Koblan, Jonathan M. Levy, Peter J. Chen, Christopher Wilson, Gregory A. Newby, Aditya Raguram, and David R. Liu. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785):149–157, December 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1711-4.

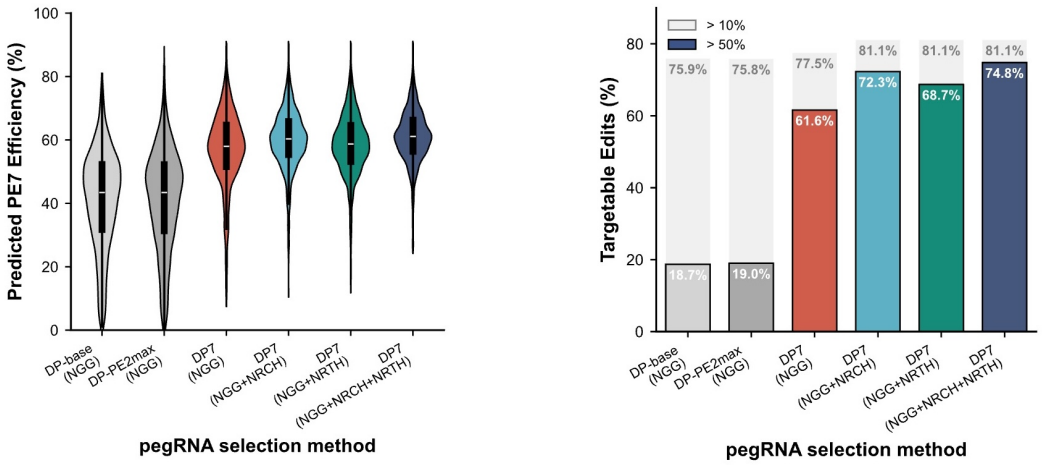


Figure 5: Impact of DeepPrime7 on therapeutic pegRNA design. Evaluation on 1,000 pathogenic or likely pathogenic ClinVar variants shows improved pegRNA prioritization compared with DeepPrime and increased fraction of high-efficiency designs when PAM-relaxed PE7 systems are enabled.

Xichen Liao, Qi Liu, and Guohui Chuai. PrimeNet: rational design of Prime editing pegRNAs by deep learning. *Briefings in Bioinformatics*, 26(3):bbaf293, May 2025. ISSN 1477-4054. doi: 10.1093/bib/bbaf293.

Nicolas Mathis, Ahmed Allam, Lucas Kissling, Kim Fabiano Marquart, Lukas Schmidheini, Cristina Solari, Zsolt Balázs, Michael Krauthammer, and Gerald Schwank. Predicting prime editing efficiency and product purity by deep learning. *Nature Biotechnology*, 41(8):1151–1159, August 2023. ISSN 1546-1696. doi: 10.1038/s41587-022-01613-7.

Nicolas Mathis, Ahmed Allam, András Tálas, Lucas Kissling, Elena Benvenuto, Lukas Schmidheini, Ruben Schep, Tanav Damodharan, Zsolt Balázs, Sharan Janjuha, Eleonora I. Ioannidi, Desirée Böck, Bas van Steensel, Michael Krauthammer, and Gerald Schwank. Machine learning prediction of prime editing efficiency across diverse chromatin contexts. *Nature Biotechnology*, 43(5): 712–719, May 2025. ISSN 1546-1696. doi: 10.1038/s41587-024-02268-2.

Shannon M. Miller, Tina Wang, Peyton B. Randolph, Mandana Arbab, Max W. Shen, Tony P. Huang, Zaneta Matuszek, Gregory A. Newby, Holly A. Rees, and David R. Liu. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology*, 38(4):471–481, April 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0412-8.

Jun Yan, Paul Oyler-Castrillo, Purnima Ravisankar, Carl C. Ward, Sébastien Levesque, Yangwode Jing, Danny Simpson, Anqi Zhao, Hui Li, Weihao Yan, Laine Goudy, Ralf Schmidt, Sabrina C. Solley, Luke A. Gilbert, Michelle M. Chan, Daniel E. Bauer, Alexander Marson, Lance R. Parsons, and Britt Adamson. Improving prime editing with an endogenous small RNA-binding protein. *Nature*, 628(8008):639–647, April 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07259-6.

Goosang Yu, Hui Kwon Kim, Jinman Park, Hyunjong Kwak, Yumin Cheong, Dongyoung Kim, Jiyun Kim, Jisung Kim, and Hyongbum Henry Kim. Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell*, 186(10):2256–2272.e23, May 2023. ISSN 00928674. doi: 10.1016/j.cell.2023.03.034.