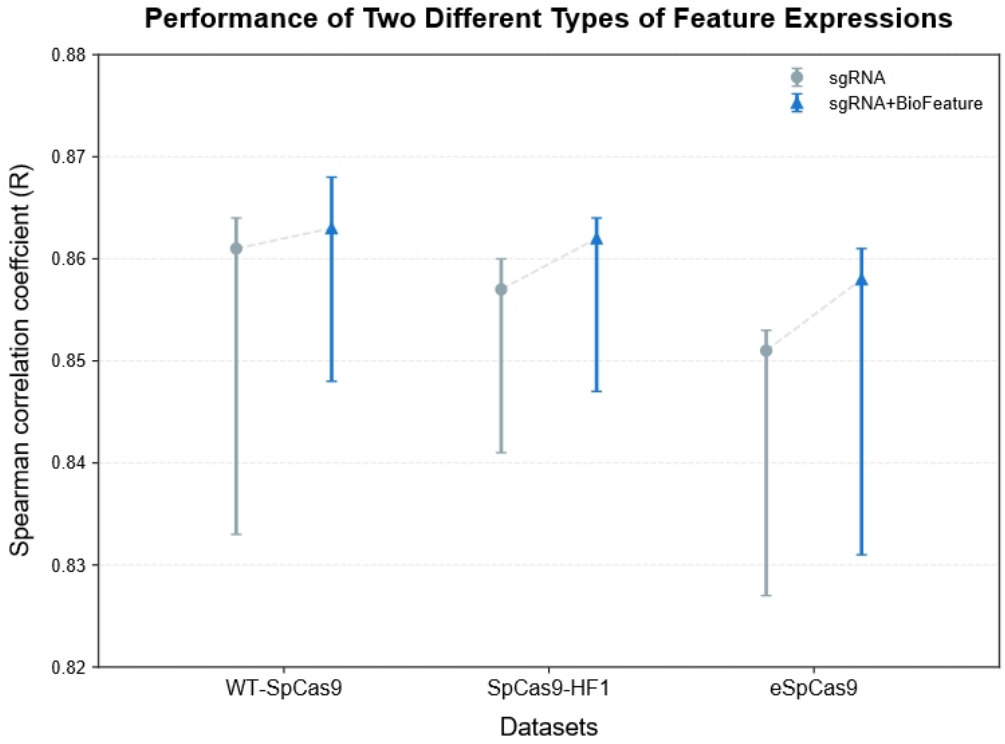



**Fig. S1. Performance evaluation of feature expressions across three datasets.** Comparison of Spearman correlation coefficient (R) for the baseline model (sgRNA, gray, circle) and the integrated model (sgRNA+BioFeature, blue, triangle) across three datasets. Biological features include melting temperature, secondary structure, minimum free energy, and GC content. Vertical error bars represent the [Min, Max] range, and the circles and triangles within the bars represent the median correlation coefficients of the pure sequence model and the model integrated with biological features, respectively.

.

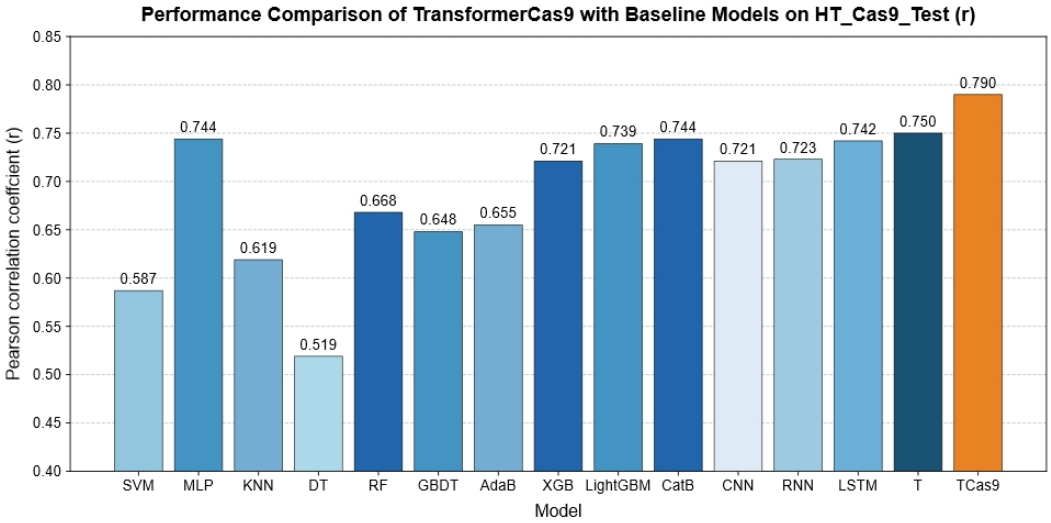



**Fig. S2.** Performance comparison of TransformerCas9 with various traditional machine learning and deep learning models on the same dataset (HT_Cas9_Test, $n$ = 542), measured by Pearson correlation coefficient ($r$). (AdaB: AdaBoost; XGB: XGBoost; CatB: CatBoost; T: Transformer; TCas9: TransformerCas9)

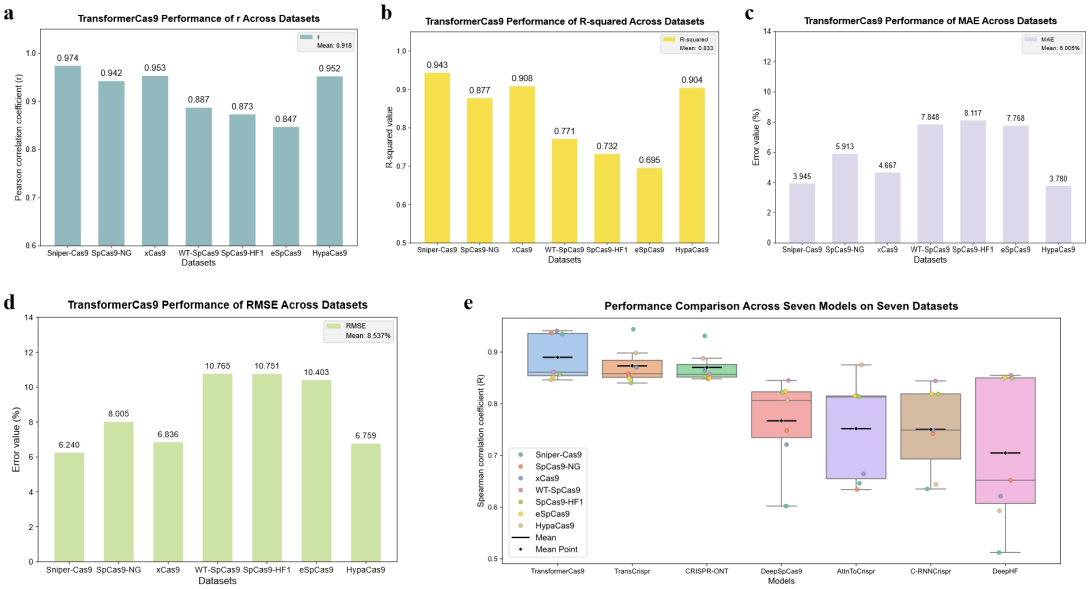



**Fig. S3. Performance evaluation of TransformerCas9 on different metrics across SpCas9 variants.** **a–d,** Pearson correlation coefficient (*r*), R-squared ($R^2$), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) of TransformerCas9 across seven datasets, with the average performance for each metric shown in the bottom-right legend of the corresponding panel. **e,** Comparison of Spearman correlation coefficients across seven models evaluated on seven datasets. The horizontal line within each box indicates the median; the box spans the interquartile range (IQR, 25th–75th percentile); whiskers extend to 1.5× IQR; points beyond the whiskers represent outliers; the solid line and filled diamond within each box denote the mean and mean point, respectively. TransformerCas9 achieved the highest median Spearman correlation and the narrowest IQR among all compared models, demonstrating superior and more consistent predictive performance across diverse datasets.

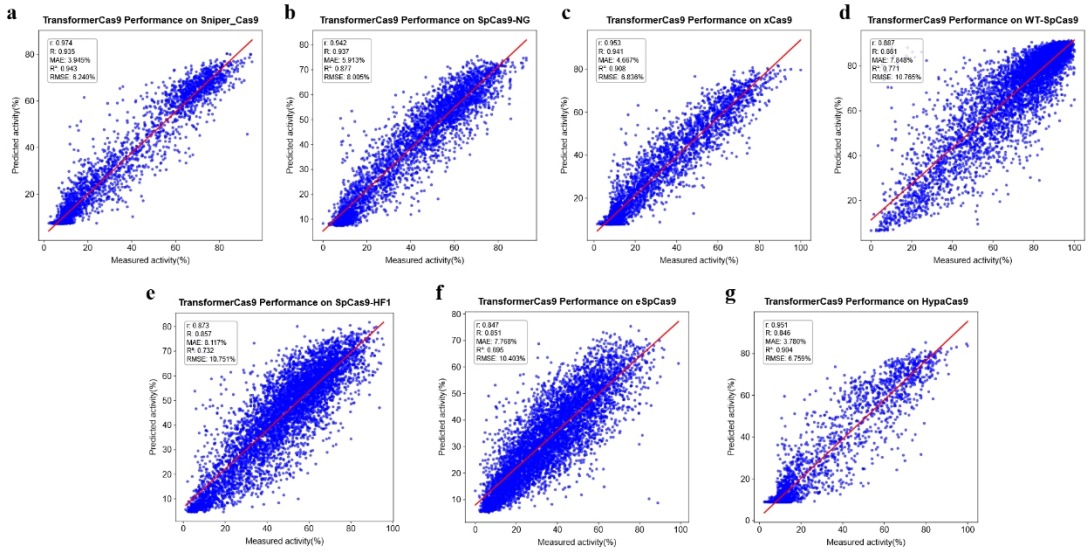



**Fig. S4 Correlation between predicted and measured sgRNA activities across seven SpCas9 variant datasets. a-g,** Scatter plots illustrate the correlation between measured and predicted activities for TransformerCas9 across seven datasets: (**a**) Sniper_Cas9, (**b**) SpCas9-NG, (**c**) xCas9, (**d**) WT-SpCas9, (**e**) SpCas9-HF1, (**f**) eSpCas9, and (**g**) HypaCas9. The red solid line in each plot represents the linear regression line. Statistical metrics, including Pearson correlation coefficient (*r*), Spearman correlation coefficient (*R*), mean absolute error (MAE), coefficient of determination ($R^2$), and root mean square error (RMSE), are indicated within each panel.

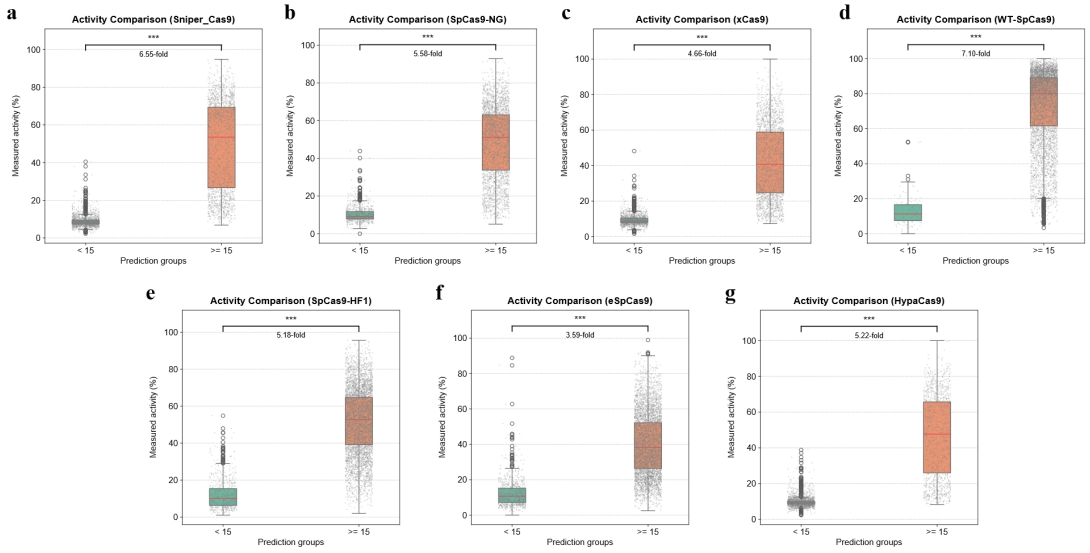



**Fig. S5. Comparison of measured activities between sgRNAs predicted to have high (≥ 15%) and low (< 15%) editing scores across seven SpCas9 variant datasets. a–g,** Activity distributions for Sniper-Cas9, SpCas9-NG, xCas9, WT-SpCas9, SpCas9-HF1, eSpCas9, and HypaCas9, respectively. The red horizontal line within each box indicates the median; the box spans the interquartile range (IQR, 25th–75th percentile); whiskers extend to 1.5× IQR; and circles beyond the whiskers represent outliers. Overlaid gray dots represent individual sgRNA measurements. The fold-change in median activity between the high and low prediction groups is annotated within each panel. Statistical significance was assessed using the Mann–Whitney U test (***$p < 0.001$).

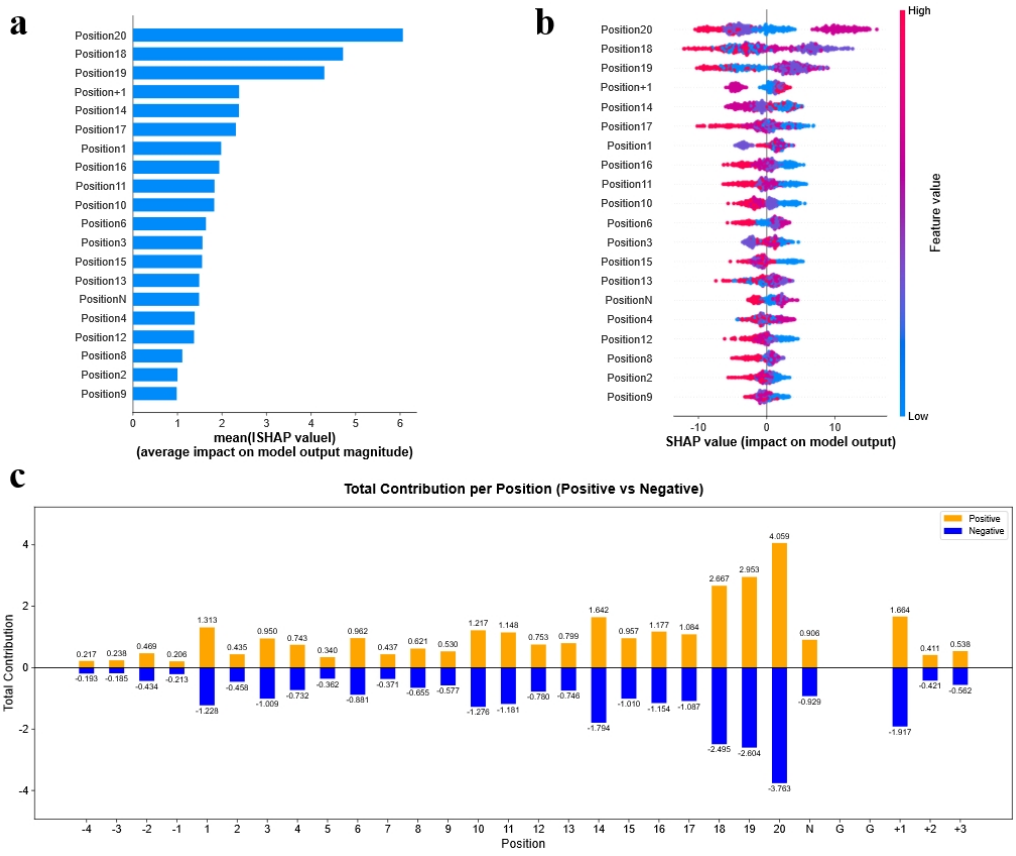



**Fig. S6. Post-hoc interpretability via SHAP and LIME. a–b**, SHAP-based interpretability analysis for the top 20 contributing positions: (**a**) mean absolute SHAP values quantifying global importance, and (**b**) summary plot illustrating impact distribution across individual samples (blue: low feature value; red: high feature value). **c**, LIME analysis of feature contributions per position. Orange and blue bars represent positive and negative contributions, respectively; the total contribution for each position is derived from the sum of its positive and negative contributions.

## Supplementary Table 1

**Sequence Overlap Between HT_Cas9_Train and All Test Datasets.**

| Test Datasets | N | Identical (d=0) | Identical (%) | >90% (N) | >90% (%) |
|---|---|---|---|---|---|
| HT_Cas9_Test | 542 | 15 | 2.768 | 15 | 2.768 |
| Sniper-Cas9 | 3708 | 8 | 0.216 | 49 | 1.321 |
| SpCas9-NG | 3670 | 8 | 0.218 | 49 | 1.335 |
| xCas9 | 3709 | 8 | 0.216 | 49 | 1.321 |
| WT-SpCas9 | 5561 | 37 | 0.665 | 41 | 0.737 |
| SpCas9-HF1 | 5689 | 52 | 0.914 | 53 | 0.932 |
| eSpCas9 | 5862 | 46 | 0.785 | 46 | 0.785 |
| HypaCas9 | 3715 | 8 | 0.215 | 49 | 1.319 |

**Supplementary Table 1. Sequence Overlap Between HT_Cas9_Train and All Test Datasets.** Test Datasets: names of test datasets; N: total number of sgRNAs in each test set; Identical (d=0): count of sgRNA sequences identical to training sequences (Hamming distance = 0); Identical (%): percentage of identical sgRNA sequences relative to total sgRNAs; >90% (N): count of sgRNA sequences with >90% similarity to training sequences; >90% (%): percentage of sgRNA sequences with >90% similarity relative to total sgRNAs. The similarity is derived from Hamming distance..

**Supplementary Table 2.**
**Hyperparameter space and optimal framework of the benchmark model.**[1]

| Hyperparameter | Alternative values | Optimal value |
|---|---|---|
| Batch size | {2048, 4096} | 2048 |
| Feed-forward dimension | — | 2048 |
| Embedding dimension | — | 32 |
| Number of attention heads | {1, 2, 4} | 1 |
| Number of layers | {1, 2, 3, 4, 5} | 5 |
| Dropout rate | {0.1, 0.2} | 0.1 |
| Learning rate | {0.001, 0.002, ..., 0.010} | 0.007 |
| Fully connected layers | {32, 64, 128, 256, 512, 1024} | [256, 128, 64] |

**Table S2. Hyperparameter space and optimal framework of the benchmark model.** Detailed parameter search space and optimal configuration of TransformerCas9 trained on the HT_Cas9_Train dataset. Selected hyperparameters were optimized using a systematic grid search; structural hyperparameters including embedding and feed-forward dimensions were strictly held constant throughout all experiments (indicated by "—").

[1] Note: "—" indicates that the corresponding hyperparameter was fixed and not included in the search process

**Supplementary Table 3.**
**Attention weights and nucleotide composition at each sgRNA position.**

| Position | Attention Weight | Spearman | A% | T% | C% | G% |
|---|---|---|---|---|---|---|
| -4 | 0.006036 | -0.008654 | 39.334538 | 43.291569 | 40.927970 | 40.920408 |
| -3 | 0.002525 | 0.042492 | 41.748073 | 39.161556 | 41.769959 | 41.410297 |
| -2 | 0.002556 | 0.049797 | 41.964702 | 38.835871 | 41.696417 | 41.289102 |
| -1 | 0.007229 | -0.014467 | 41.500621 | 39.146768 | 40.690599 | 42.452721 |
| 1 | 0.031099 | 0.032871 | 43.139779 | 40.983919 | 38.759517 | 41.384280 |
| 2 | 0.017271 | -0.102230 | 43.351496 | 36.283022 | 44.390235 | 39.971149 |
| 3 | 0.011039 | 0.001853 | 43.379943 | 39.589616 | 41.282106 | 39.618145 |
| 4 | 0.011971 | -0.005811 | 42.594108 | 38.443139 | 42.432815 | 40.927586 |
| 5 | 0.006071 | -0.018092 | 42.182420 | 41.672240 | 40.119258 | 40.681223 |
| 6 | 0.008951 | 0.012973 | 40.641452 | 35.925114 | 42.172596 | 43.984962 |
| 7 | 0.024292 | 0.004869 | 42.960490 | 39.546509 | 42.065446 | 39.969240 |
| 8 | 0.041120 | 0.038321 | 40.471046 | 39.601911 | 41.612020 | 42.247056 |
| 9 | 0.047592 | -0.001000 | 43.650733 | 37.953868 | 42.431898 | 40.482096 |
| 10 | 0.027298 | 0.023861 | 44.989510 | 37.637552 | 41.166660 | 40.546697 |
| 11 | 0.012145 | 0.068625 | 44.989892 | 40.925366 | 39.801240 | 39.101429 |
| 12 | 0.023921 | -0.008320 | 42.705442 | 41.300005 | 41.725572 | 39.040198 |
| 13 | 0.050370 | -0.108630 | 38.429020 | 38.495687 | 42.898660 | 43.303341 |
| 14 | 0.046919 | -0.080419 | 46.845418 | 41.886932 | 40.354206 | 36.738059 |
| 15 | 0.067537 | -0.031286 | 44.719085 | 42.428045 | 38.565981 | 39.506811 |
| 16 | 0.093515 | -0.105609 | 46.873856 | 35.288420 | 41.899586 | 39.497582 |
| 17 | 0.047691 | -0.144348 | 43.433487 | 36.448131 | 41.777267 | 41.766656 |
| 18 | 0.028605 | -0.230300 | 41.177795 | 37.524969 | 46.273669 | 38.904672 |
| 19 | 0.037950 | -0.117800 | 37.298513 | 34.082192 | 44.425954 | 44.979474 |
| 20 | 0.068487 | 0.291374 | 41.770596 | 36.542190 | 33.800612 | 50.413224 |
| N | 0.070752 | 0.037638 | 41.863427 | 36.774478 | 41.681961 | 44.427567 |
| G | 0.082455 | 0.036306 | 0 | 0 | 0 | 41.059345 |
| G | 0.036649 | 0.071028 | 0 | 0 | 0 | 41.059345 |
| +1 | 0.015475 | -0.061580 | 42.498502 | 44.397254 | 40.926589 | 38.258528 |
| +2 | 0.026695 | 0.013633 | 41.252170 | 42.360531 | 40.328906 | 40.642532 |
| +3 | 0.045782 | 0.060152 | 40.371851 | 38.870060 | 40.924881 | 43.344029 |

**Table S3. Attention weights and nucleotide composition at each sgRNA position.** The attention weights assigned by TransformerCas9 to each position of the sgRNA sequence, along with the corresponding Spearman correlation coefficients and nucleotide frequency distributions (A%, T%, C%, G%). Positions with higher attention weights indicate greater model-assigned importance to editing activity prediction.