Experimental and Computational Analyses of Off-target Editing by Programmable Nucleases

X. Robert Bao, Yidan Pan, Ciaran M. Lee, Timothy H. Davis and Gang Bao

Supplementary Information

Supplementary Figures

Supplementary Figure S1

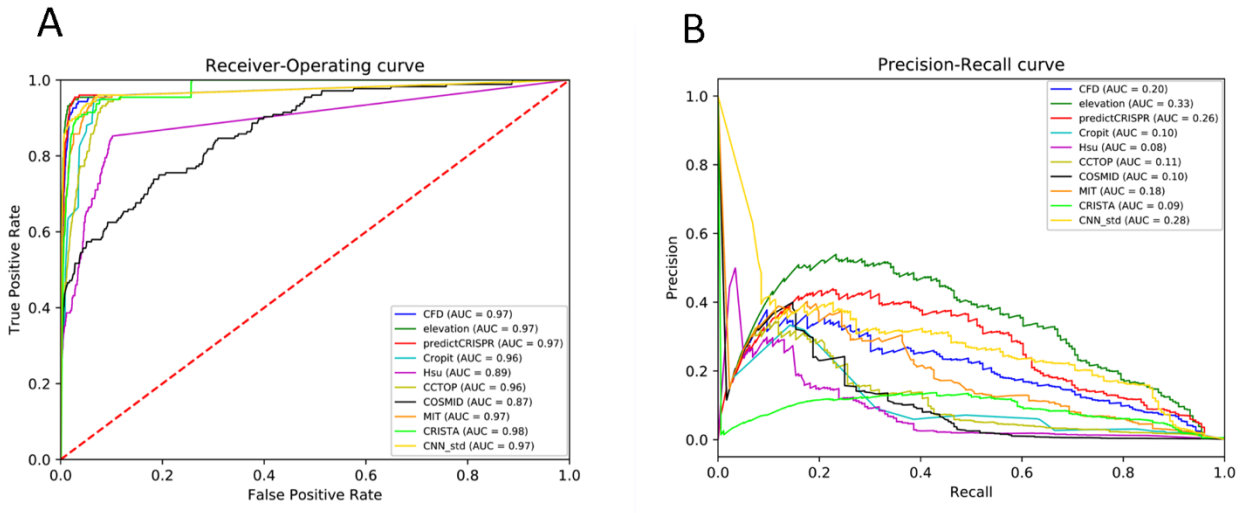

Figure S1. ROC and PRC of various off-target scoring algorithms on the true off-target list. Potential off-target sites of 27 gRNAs were screened by Cas-OFFinder and scored by each of the algorithms, and the classification results were compared with the experimental validated true off-target sites. The dataset used in this figure (also Figure 2A) is provided in Supplementary Table S2. (A) The visualization of ROC shows comparable AUCs for most of the algorithms, which is hard to interpret due to severe data imbalance (176 positive sites out of 123,383 total off-target sites). (B) Precision (True Positives events / (True Positives events + False Positives events) is not impacted by a large number of total true negative events, which reveals the ability to classify true off-targets better. PRC shows clear over-performance of elevation to the other algorithms.

Supplementary Figure S2

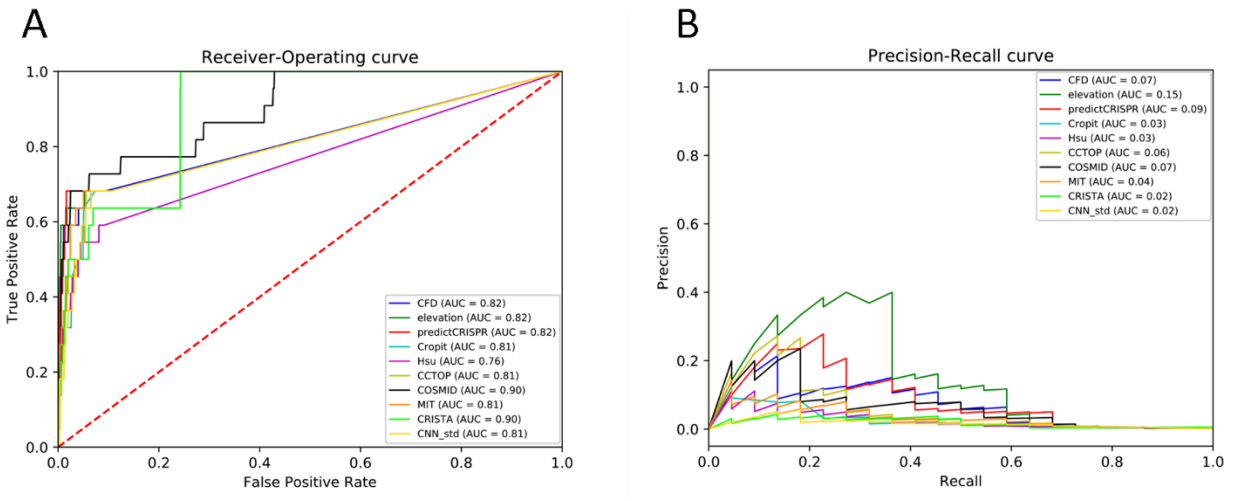

Figure S2. ROC and PRC of various off-target scoring algorithms on novel gRNA off-target datasets. Potential off-target sites of 4 gRNAs that were not included in any machine learning tools' training set were screened by Cas-OFFinder and scored by each of the algorithms, and the classification results were compared with the experimental validated true off-target sites. The dataset used in this figure (also Figure 2B) is provided in Supplementary Table S3. (A)ROC shows that CRISTA has the best performance since it was capable of capturing the off-target sites with DNA/RNA bulges. The data imbalance is still severe (22 positive sites out of 17,485 total off-target sites). Elevation showed the top performance among the algorithms that can only score mismatches. (B) Despite the fact that only CRISTA and COSMID were able to score the off-target sites with DNA/RNA bulges, PRC shows clear over-performance of elevation to all the other algorithms.

Supplementary Tables

Supplementary Table S1. Components of the true positive list used in the analysis, including nine studies that used amplicon-specific experimental techniques to detect off-target editing rates.

Supplementary Table S2. The full off-target dataset used in the performance assessment. Raw data for generating Figure 2A and Figure S1. Potential off-target sites of 27 gRNAs from 9 studies shown in Table S1 were screened by Cas-OFFinder allowing up to 4 mismatches and 1 base DNA/RNA bulges and scored by each of the algorithms. "noind": DNA/RNA sequences before alignments. Note that after introducing bulges, one locus might be called by Cas-OFFinder multiple times with different alignment patterns,

leading to different scores. These sites were treated the same as other off-target sites without manipulation.

Supplementary Table S3. The off-target dataset of novel gRNAs used in the performance assessment. Raw data for generating Figure 2B and Figure S2. Potential off-target sites and scores of 4 gRNAs that were not in the CRISPOR dataset[1].

References

1 Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, doi:10.1186/s13059-016-1012-2 (2016).