CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Tools for experimental and computational analyses of off-target editing by programmable nucleases
Authors	Bao, X. Robert;Pan, Yidan;Lee, Ciaran M.;Davis, Timothy H.;Bao, Gang
Publication date	2020-12-07
Original Citation	Bao, X. R., Pan, Y., Lee, C. M., Davis, T. H. and Bao, G. (2020) 'Tools for experimental and computational analyses of off-target editing by programmable nucleases', Nature Protocols, 16, pp. 10-26. doi: 10.1038/s41596-020-00431-y
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1038/s41596-020-00431-y
Rights	© 2020, Springer Nature Limited. This is a post-peer-review, pre-copyedit version of an article published in Nature Protocols on 7 December 2020. The final authenticated version is available online at: https://doi.org/10.1038/s41596-020-00431-y
Download date	2026-11-15 02:34:41
Item downloaded from	https://hdl.handle.net/10468/10901



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

nature research

https://doi.org/10.1038/s41596-020-00431-y

Supplementary information

Tools for experimental and computational analyses of off-target editing by programmable nucleases

In the format provided by the authors and unedited

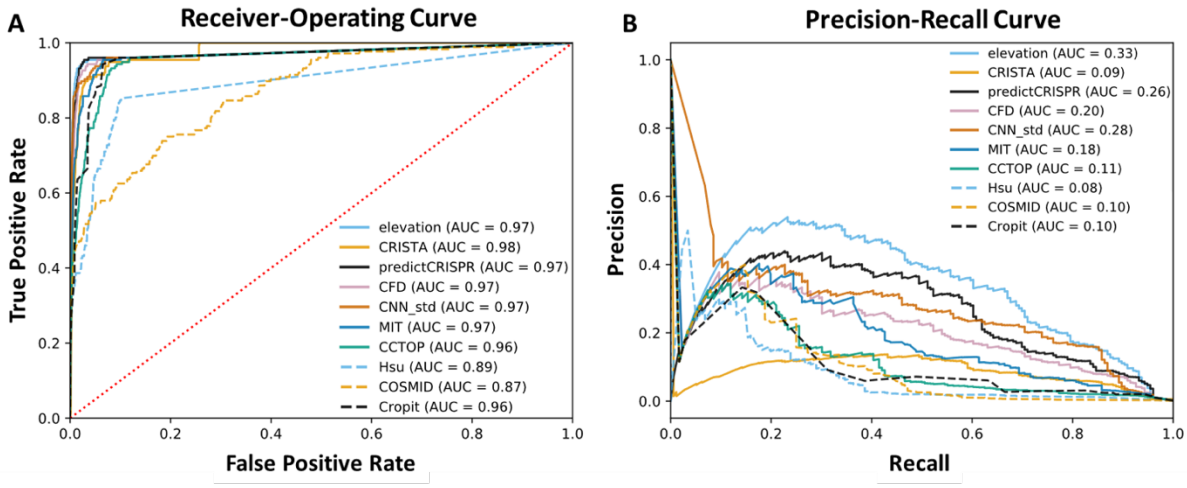
Experimental and Computational Analyses of Off-target Editing by Programmable Nucleases

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

Supplementary Information

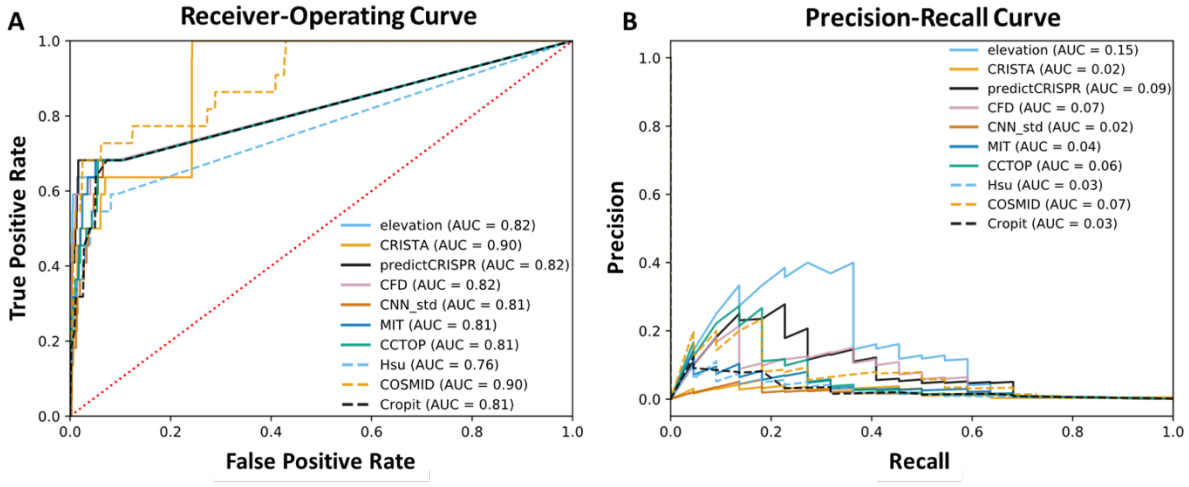
Supplementary Figures

Supplementary Figure 1



Supplementary Figure 1: ROC and PRC of various off-target scoring algorithms based on the true off-target sites identified. 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 3a) is provided in **Supplementary Table 4.** (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 2



Supplementary Figure 2: ROC and PRC of various off-target scoring algorithms based 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 3b**) is provided in **Supplementary Table 5**. (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.