**DL-CRISPR: a deep learning method for off-target activity prediction in CRISPR/Cas9 with data augmentation**

**Supplementary Tables**

**Content**

**Tables:**

[Table S1. Overview of off-target datasets. 2](#_Toc6219519)

[Table S2. Acc, Sn and Sp for three kinds of training data with 5 times running 2](#_Toc6219520)

[Table S3. 13 out of 28 sgRNAs in our dataset predictted by Elevation 3](#_Toc6219521)

**Supplementary Table S1.** Overview of all off-target datasets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | Technique | sgRNA number | off-target sites number (only mutation type) | PCR validated or not |
| Hsu et al. 2013 [1] | targeted PCR | 4 | 46 | yes |
| Cho et al. 2014 [2] | targeted PCR | 10 | 106 | yes |
| Wang et al 2015 [3] | targeted PCR | 2 | 13 | yes |
| Ran et al. 2015 [4] | BLESS | 2 | 19 | yes |
| Tsai et al. 2015 [5] | GUIDE-seq | 10 | 411 | no |
| Kim et al. 2015 [6] | Digenome-seq | 2 | 257 | yes |
| Frock et al. 2015 [7] | HTGTS | 4 | 86 | no |
| Kim et al. 2016 [8] | Multiplex Digenome-seq | 11 | 953 | yes |
| Tsai et al. 2017 [9] | CIRCLE-seq | 17 | 6937 | no |
| Yan et al. 2017 [10] | BLISS | 2 | 28 | no |
| Cameron et al 2017 [11] | SITE-seq | 8 | 501 | no |
| Tang et al 2018 [12] | TEG-seq | 2 | 137 | no |
| Kim et al 2018 [13] | DIG-seq | 8 | 133 | yes |

**Supplementary Table S2.** Acc, Sn and Sp for three kinds of training data with 5 times running.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Positive + Negative |  | Acc | Sn | Sp |
|  | Run1 | 78.33 | 73.45 | 83.20 |
|  | Run 2 | 74.26 | 66.81 | 81.71 |
| reliable + ”detected” | Run 3 | 78.92 | 80.97 | 76.86 |
|  | Run 4 | 76.91 | 70.35 | 83.46 |
|  | Run 5 | 77.39 | 77.43 | 77.34 |
|  | Average | 77.16 | 73.80 | 80.51 |
|  | Run1 | 82.81 | 78.32 | 87.29 |
|  | Run 2 | 82.40 | 76.11 | 88.69 |
| “detected” + negative | Run 3 | 82.65 | 78.32 | 86.97 |
|  | Run 4 | 82.51 | 80.09 | 84.92 |
|  | Run 5 | 81.81 | 74.78 | 88.84 |
|  | Average | 82.44 | 77.52 | 87.34 |
|  | Run1 | 94.03 | 92.04 | 96.02 |
|  | Run 2 | 92.48 | 94.25 | 90.71 |
| reliable + negative | Run 3 | 93.36 | 96.02 | 90.71 |
|  | Run 4 | 94.03 | 97.35 | 90.71 |
|  | Run 5 | 94.69 | 94.25 | 95.13 |
|  | Average | 93.72 | 94.78 | 92.66 |

**Supplementary Table S3.** 13 out of 28 sgRNAs in our dataset predictted by Elevation.

|  |  |  |
| --- | --- | --- |
| Gene | Ensemble ID | sgRNA |
| EMX1 | ENSG00000135638 | GAGTCCGAGCAGAAGAAGAAGGG |
|  |  | GACATCGATGTCCTCCCCATTGG |
|  |  | GCGCCACCGGTTGATGTGATGGG |
| HBB | ENSG00000244734 | CTTGCCCCACAGGGCAGTAACGG |
| C4BPB | ENSG00000123843 | AATGACCACTACATCCTCAAGGG |
| CCR5 | NSG00000160791 | TGACATCAATTATTATACATCGG |
|  |  | GACACCGAAGCAGAGTTTTTAGG |
|  |  | GGTCCTGCCGCTGCTTGTCATGG |
|  |  | CCAGTGAGTAGAGCGGAGGCAGG |
|  |  | ATGAACACCAGTGAGTAGAGCGG |
|  |  | AACACCAGTGAGTAGAGCGGAGG |
|  |  | GAATCCTAAAAACTCTGCTTCGG |
| RAG1 | ENSG00000166349 | GCCTCTTTCCCACCCACCTTGGG |

# **References**

|  |  |
| --- | --- |
| [1] | P. Hsu, D. Scott, J. Weinstein, F. Ran, S. Konermann, V. Agarwala, Y. Li, E. Fine, X. Wu, O. Shalem and T. Cradick, "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nature biotechnology,* pp. 31(9), p.827, 2013. |
| [2] | S. Cho, S. Kim, Y. Kim, J. Kweon, H. Kim, S. Bae and J. Kim, "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases," *Genome research,* vol. 24(1), pp. pp.132-141, 2014. |
| [3] | X. Wang, Y. Wang, X. Wu, J. Wang, Y. Wang, Z. Qiu, T. Chang, H. Huang, R. Lin and J. Yee, "Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors," *Nature biotechnology,* pp. 33(2), p.175, 2015. |
| [4] | F. Ran, L. Cong, W. Yan, D. Scott, J. Gootenberg, A. Kriz, B. Zetsche, O. Shalem, X. Wu, K. Makarova and E. Koonin, "In vivo genome editing using Staphylococcus aureus Cas9," *Nature,* pp. 520(7546), p.186, 2015. |
| [5] | S. Tsai, Z. Zheng, N. Nguyen, M. Liebers, V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. Iafrate, L. Le and M. Aryee, "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases," *Nature biotechnology,* pp. 33(2),p.187, 2015. |
| [6] | D. Kim, S. Bae, J. Park, E. Kim, S. Kim, H. Yu, J. Hwang, J. Kim and J. Kim, "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells," *Nature methods,* pp. 12(3), p.237, 2015. |
| [7] | R. Frock, J. Hu, R. Meyers, Y. Ho, E. Kii and F. Alt, "Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases," *Nature biotechnology,* pp. 33(2), p.179, 2015. |
| [8] | D. Kim, S. Kim, S. Kim, J. Park and J. Kim, "Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq," *Genome research,* pp. 26(3), pp.406-415, 2016. |
| [9] | S. Tsai, N. Nguyen, J. Malagon-Lopez, V. Topkar, M. Aryee and J. Joung, "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets," *Nature methods,* pp. 14(6), p.607, 2017. |
| [10] | W. Yan, R. Mirzazadeh, S. Garnerone, D. Scott, M. Schneider, T. Kallas, J. Custodio, E. Wernersson, Y. Li, L. Gao and Y. Federova, "BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks," *Nature communications,* pp. 8,p.15058, 2017. |
| [11] | P. Cameron, C. Fuller, P. Donohoue, B. Jones, M. Thompson, M. Carter, S. Gradia, B. Vidal, E. Garner, E. Slorach and E. Lau, "Mapping the genomic landscape of CRISPR–Cas9 cleavage," *Nature methods,* pp. 14(6), p.600, 2017. |
| [12] | P. Tang, B. Ding, L. Peng, V. Mozhayskiy, J. Potter and J. Chesnut, "TEG-seq: an ion torrent-adapted NGS workflow for in cellulo mapping of CRISPR specificity," *BioTechniques,* pp. 65(5), pp.259-267, 2018. |
| [13] | D. Kim and J. Kim, "DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA," *Genome research,* pp. 28(12), pp.1894-1900, 2018. |