**Supplementary Table S1.** The information of all datasets used in this study.

|  |  |  |
| --- | --- | --- |
| **Datasets** | **Brief introduction** | **Download link** |
| Datasets for building the seq2seq-attention model | The datasets from Jost *et al*.[1] including relative activity of 1978 DNA targets and corresponding 26,248 MMsgRNAs in the fitting step. | https://static-content.springer.com/esm/art%3A10.1038%2Fs41587-019-0387-5/MediaObjects/41587\_2019\_387\_MOESM3\_ESM.zip |
| CRISPR/Cas9 target DNA sequences from Human GeCKOv2 Library [2] in the feedback and prediction step. | https://media.addgene.org/cms/filer\_public/a4/b8/a4b8d181-c489-4dd7-823a-fe267fd7b277/human\_geckov2\_library\_a\_09mar2015.csv  https://media.addgene.org/cms/filer\_public/2d/8b/2d8baa42-f5c8-4b63-9c6c-bd98f333b29e/human\_geckov2\_library\_b\_09mar2015.csv |
| Datasets for building the off-target activity prediction model | The dataset [3] containing the absolute on-target activity for 55,604 DNA targets and its corresponding PMsgRNAs was used for building the absolute on-target activity prediction model. This model was then used to obtain the absolute off-target activity for the datasets from Jost *et al*.[1], and the resulting dataset was used to build the off-target activity prediction model. | https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-019-12281-8/MediaObjects/41467\_2019\_12281\_MOESM3\_ESM.xlsx |
| Datasets for validating the off-target activity prediction model | The off-target activities measured by Guide-Seq[4] for sgRNAs in their off-target sites. Only the data of 4 sgRNAs with more than 20 off-targets were used. | https://static-content.springer.com/esm/art%3A10.1038%2Fnbt.3117/MediaObjects/41587\_2015\_BFnbt3117\_MOESM22\_ESM.xlsx |

1. Jost, M., et al., *Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs.* Nature Biotechnology, 2020. **38**(3): p. 355-364.

2. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nature Methods, 2014. **11**(8): p. 783-784.

3. Wang, D., et al., *Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning.* Nature Communications, 2019. **10**(1): p. 4284.

4. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-197.

**Supplementary Table S2:** Details of the 37 predicted mismatch sgRNA and Dual-luciferase reporter Experiment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **DNA target** | **Seq2seq predicted sgRNA** | **MRE** | **PM1** | **PM2** | **PM3** | **A.PM** | **MM1** | **MM2** | **MM3** | **A.MM** | **Ctrl1** | **Ctrl2** | **Ctrl3** | **A.Ctrl** |
| *Experiment batch 1* | | | | | | | | | | | | 1.51 | 1.51 | 1.48 | 1.50 |
| 1 | TGTGCCAGAGTCCTTCGATAGG | GGTGCCAGAGTCCTTCGAT | 0.79 | 0.16 | 0.15 | 0.15 | 0.15 | 0.46 | 0.42 | 0.41 | 0.43 | - | - | - | - |
| 2 | ATTCTAAAACGGATTACCAGGG | ACTCTAAAACGGATTACCA | 0.99 | 0.39 | 0.37 | 0.43 | 0.40 | 0.47 | 0.38 | 0.39 | 0.41 | - | - | - | - |
| 3 | CGAAGAAGGAGAATAGGGTTGG | CGGAGAAGGAGAATAGGGT | 1.04 | 0.27 | 0.25 | 0.27 | 0.27 | 0.22 | 0.22 | 0.22 | 0.22 | - | - | - | - |
| 4 | GCTATTCTGATTACACCCGAGG | GCTGTTCTGATTACACCCG | 0.99 | 0.18 | 0.16 | 0.20 | 0.18 | 0.18 | 0.19 | 0.19 | 0.19 | - | - | - | - |
| 5 | CCAGCGGATAGAATGGCGCCGG | CCGGCGGATAGAATGGCGC | 0.90 | 0.42 | 0.45 | 0.40 | 0.42 | 0.54 | 0.54 | 0.50 | 0.53 | - | - | - | - |
| 6 | CCAAGAGGTTCCATCTGCCAGG | CCCGGAGGTTCCATCTGCC | 1.01 | 0.23 | 0.25 | 0.24 | 0.24 | 0.24 | 0.22 | 0.24 | 0.23 | - | - | - | - |
| 7 | CTGGATACCGGGAAAACGCTGG | CGGGGTACCGGGAAAACGC | 0.74 | 0.45 | 0.43 | 0.44 | 0.44 | 0.73 | 0.72 | 0.72 | 0.72 | - | - | - | - |
| *Experiment batch 2* | | | | | | | | | | | | 3.08 | 2.11 | 2.76 | 2.65 |
| 8 | ACAGATGCACATATCGAGGTGG | CCGGATGCACATATCGAGG | 0.94 | 0.03 | 0.11 | 0.10 | 0.08 | 0.07 | 0.39 | 0.25 | 0.24 | - | - | - | - |
| 9 | TCACGCAGGCAGTTCTATGAGG | CCCCGCGGGCAGTTCTATG | 0.94 | 0.07 | 0.09 | 0.10 | 0.09 | 0.25 | 0.25 | 0.20 | 0.23 | - | - | - | - |
| 10 | GTACATCGACTGAAATCCCTGG | GCGCGTCGACTGAAATCCC | 0.87 | 0.11 | 0.05 | 0.11 | 0.09 | 0.42 | 0.39 | 0.47 | 0.43 | - | - | - | - |
| 11 | CGCTGGAAGATGGAACCGCTGG | GCCTGGAGGATGGAACCGC | 0.86 | 0.19 | 0.08 | 0.08 | 0.12 | 0.41 | 0.47 | 0.56 | 0.48 | - | - | - | - |
| 12 | AAACATTCCAAAACCGTGATGG | GCCCGTTCCAAAACCGTGA | 0.63 | 0.15 | 0.30 | 0.10 | 0.18 | 0.60 | 0.52 | 2.18 | 1.10 | - | - | - | - |
| 13 | TAAATAACGCGCCCAACACCGG | GCCGTAACGCGCCCAACAC | 0.59 | 0.11 | 0.11 | 0.12 | 0.11 | 1.12 | 1.03 | 1.31 | 1.15 | - | - | - | - |
| 14 | TGAAGAAGAGCTGTTTCTGAGG | GCCGGAAGAGCTGTTTCTG | 0.40 | 0.32 | 0.27 | 0.24 | 0.28 | 2.12 | 2.44 | 0.56 | 1.71 | - | - | - | - |
| 15 | GCTATGAAGAGATACGCCCTGG | CCCGTGAGGAGATACGCCC | 0.64 | 0.19 | 0.12 | 0.07 | 0.12 | 1.07 | 1.06 | 1.00 | 1.04 | - | - | - | - |
| *Experiment batch 3* | | | | | | | | | | | | 24.86 | 23.61 | 20.38 | 22.95 |
| 16 | CCAGCGGATAGAATGGCGCCGG | CCGGCGGGTAGAATGGCGC | 0.31 | 4.02 | 4.41 | 5.56 | 4.66 | 16.56 | 17.21 | 18.33 | 17.37 | - | - | - | - |
| 17 | TCACATCTCATCTACCTCCCGG | CCGCGTCTCATCTACCTCC | 0.93 | 3.65 | 4.16 | 3.94 | 3.92 | 6.68 | 4.77 | 4.55 | 5.33 | - | - | - | - |
| 18 | GTGCCAGAGTCCTTCGATAGGG | GCTCCAGGGTCCTTCGATA | 0.32 | 2.03 | 2.47 | 2.57 | 2.36 | 14.08 | 21.63 | 13.57 | 16.42 | - | - | - | - |
| 19 | TGAACTCCTCTGGATCTACTGG | GGCGCTCCTCTGGATCTAC | 0.85 | 2.85 | 2.31 | 1.97 | 2.38 | 6.65 | 4.03 | 5.55 | 5.41 | - | - | - | - |
| 20 | ATGATAAACCGGGCGCGGTCGG | GCGGTAAACCGGGCGCGGT | 0.51 | 4.39 | 4.19 | 5.24 | 4.61 | 13.23 | 12.79 | 15.03 | 13.68 | - | - | - | - |
| 21 | ATCAGAGAGATCCTCATAAAGG | CCCGGAGAGATCCTCATAA | 0.25 | 9.96 | 13.13 | 12.55 | 11.88 | 18.78 | 24.07 | 17.83 | 20.23 | - | - | - | - |
| 22 | ATAGCTTCTGCCAACCGAACGG | GCGGCTTCTGCCAACCGAA | 0.03 | 5.03 | 9.38 | 8.49 | 7.63 | 20.73 | 23.80 | 22.93 | 22.49 | - | - | - | - |
| 23 | ATCCATGATAATAATTTTTTGG | GCCCGTGATAATAATTTTT | 0.44 | 8.42 | 7.82 | 7.99 | 8.08 | 16.04 | 18.61 | 14.69 | 16.45 | - | - | - | - |
| 24 | GTACTTAATCAGAGACTTCAGG | GCGCTTAGTCAGAGACTTC | 0.62 | 1.64 | 1.91 | 2.00 | 1.85 | 10.09 | 10.24 | 9.58 | 9.97 | - | - | - | - |
| 25 | AAAAACATAAAGAAAGGCCCGG | CCGGACATAAAGAAAGGCC | 0.58 | 1.33 | 2.70 | 2.56 | 2.20 | 11.06 | 10.19 | 11.39 | 10.88 | - | - | - | - |
| 26 | TTTACAGATGCACATATCGAGG | CCCGCAGATGCACATATCG | 0.70 | 1.38 | 1.59 | 4.67 | 2.55 | 8.39 | 9.38 | 8.14 | 8.64 | - | - | - | - |
| 27 | ATCTGCCAGGTATCAGGCAAGG | CCCCGTCAGGTATCAGGCA | 0.26 | 2.03 | 1.99 | 2.31 | 2.11 | 15.48 | 18.35 | 19.07 | 17.63 | - | - | - | - |
| 28 | ATTAAGTACAAAGGCTATCAGG | CCCGAGTACAAAGGCTATC | 0.11 | 5.15 | 4.56 | 4.48 | 4.73 | 19.74 | 22.07 | 20.88 | 20.90 | - | - | - | - |
| 29 | TCTTCGACGCAGGTGTCGCAGG | CCCTTGACGCGGGTGTCGC | 0.40 | 3.74 | 3.13 | 3.55 | 3.47 | 17.39 | 11.39 | 16.73 | 15.17 | - | - | - | - |
| 30 | AAGTACCGAAAGGTCTTACCGG | GCCTGCCGAAAGGTCTTAC | 0.50 | 3.28 | 1.93 | 2.27 | 2.49 | 7.03 | 15.76 | 15.67 | 12.82 | - | - | - | - |
| 31 | GAGGCAGAGCGACACCTTTAGG | GGCGTAGGGCGACACCTTT | 0.14 | 2.61 | 4.24 | 3.32 | 3.39 | 19.44 | 20.39 | 20.53 | 20.12 | - | - | - | - |
| 32 | TGATTTGATTGCCAAAAATAGG | GCGTTTGGTTGCCAAAAAT | 0.14 | 4.15 | 4.13 | 4.38 | 4.22 | 18.32 | 20.84 | 21.89 | 20.35 | - | - | - | - |
| 33 | TTGGCGAAGAAGGAGAATAGGG | CCGGTGAGGAAGGAGAATA | 0.77 | 2.09 | 2.39 | 2.35 | 2.27 | 8.17 | 8.50 | 4.42 | 7.03 | - | - | - | - |
| 34 | TTTGGCGAAGAAGGAGAATAGG | CCCGGCGGAGAAGGAGAAT | 0.29 | 3.71 | 3.14 | 3.24 | 3.36 | 15.93 | 16.70 | 19.01 | 17.21 | - | - | - | - |
| 35 | TTGCCTGATACCTGGCAGATGG | CCGCTTGGTACCTGGCAGA | 0.61 | 0.74 | 0.99 | 0.65 | 0.79 | 9.39 | 9.16 | 9.53 | 9.36 | - | - | - | - |
| 36 | TTAACGCCCAGCGTTTTCCCGG | GCGGCGCCCAGCGTTTTCC | 0.59 | 2.80 | 2.57 | 2.77 | 2.71 | 10.17 | 12.01 | 11.14 | 11.10 | - | - | - | - |
| 37 | CATAACCGGACATAATCATAGG | ACCGACCGGACATAATCAT | 0.56 | 1.02 | 1.09 | 0.86 | 0.99 | 10.48 | 11.17 | 10.31 | 10.65 | - | - | - | - |

**Note:** Since the microplate luminometer-machine (Promega GloMax Navigator) could only run 96 measurements at one time, we have devided our measurement into different batches. **MRE:** Measured relative activity; **PM:** Perfect matched sgRNA Luciferase activity; **A.PM:** Average perfect matched sgRNA Luciferase activity; **MM:** Mismatched sgRNA Luciferase activity; **A.MM:** Average mismatched sgRNA Luciferase activity; **Ctrl:** Control Luciferase activity; **A.Ctrl:** Average control Luciferase activity.

**Supplementary Table S3:** Validation of absolute off-target activity predictor in GUIDE-seq data .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cell;targeted gene** | **off-target size** | **pearson** | | **spearman** | |
| **R** | **P** | **R** | **P** |
| U2OS;VEGFA\_site2 | 152 | 0.36 | **4.09×10-6** | 0.38 | **1.30×10-6** |
| 293;HEK293\_sgRNA4 | 134 | 0.11 | 0.22 | 0.19 | **0.02** |
| U2OS;VEGFA\_site3 | 60 | 0.35 | **6.78×10-3** | 0.39 | **2.21×10-3** |
| U2OS;VEGFA\_site1 | 22 | 0.64 | **1.45×10-3** | 0.61 | **2.52×10-3** |