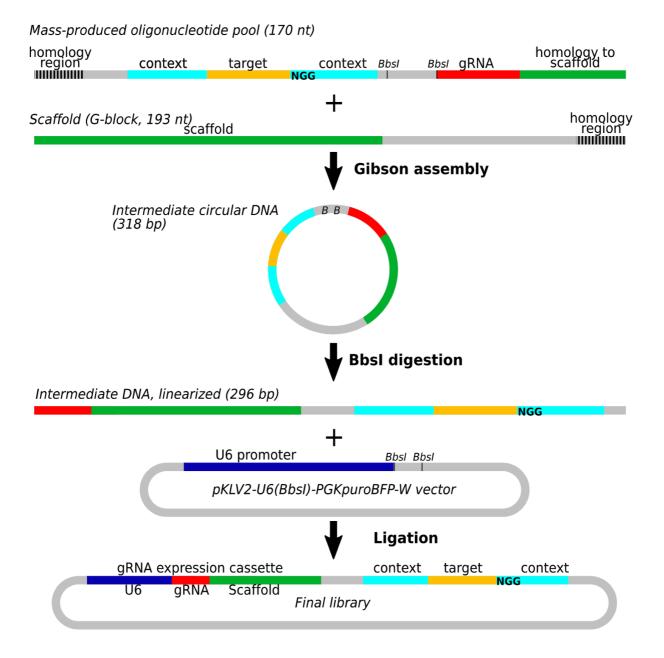
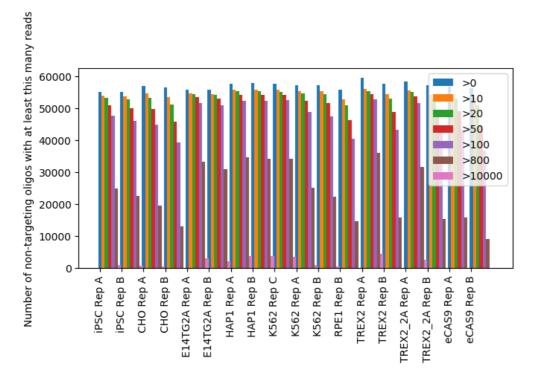
### Supporting Information

Figure S1. Construct design and cloning approach



Library cloning started by PCR amplification of the 170 nt oligonucleotide pool of designed sequences encoding gRNA and target sequence, separated by a spacer harbouring two BbsI restriction sites. Gibson assembly was employed to fuse the amplified pool to a 193 nt G- block fragment encoding either a conventional or improved version of the gRNA scaffold and a spacer. The resulting 318 bp circular DNA was linearised with BbsI and the 296 bp linear product was ligated into scaffold-less pKLV2-U6(BbsI)-PKGpuro2ABFP-W, to produce the complete library constructs encoding a functional gRNA expression cassette and its target sequence.





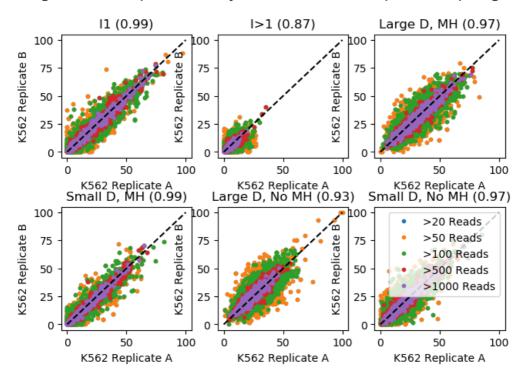
Number of constructs (y-axis) with a minimum number of mapped reads (color) in each experiment (x-axis).

Figure S3. Comparison of endogenous and synthetic repair profiles for Overbeek 25 gRNA

Overbeek25 (KL=3.2)

|    | ACTOACTOACTAGTOACATOCCTOCCATOATO           | nthetic Profile Rep A(16965 Reads)          |       |  |
|----|--|---|-------|--|
| -  | ACTOACTOACTACTOACATOCCTOCCATOATO           | Synthetic Profile Rep B(9573 Reads)         |       |  |
|    | ACT GAGT GAGT AGT CAGAT COOT GEOCCAT CAT G | Endogenous Profile(8364 Reads)              |       |  |
|    | ACTGAGTGAGTAGTCAGATCCGTGGCCCATC            | TTGGAAACGTTTTTGTCCCTGCTGCCCCTTAAA           | 7.4%  |  |
| D5 | ACTOAGTOAGTAGTOAGATOOGTOGCOCATO            | TTGGAAACGTTTTTGTCCCTGCTGCCCCTTAAA           | 23.6  |  |
|    | ACTOAGTOAGTAGTOAGATOOGTOGOCCATO            | TTGGAAACGTTTTTGTCCCTGCTGCCCCTTAAA           | 24.0  |  |
|    | ACTGAGTGAGTAGTCAGATCCGTGGCCCATCAT          | CTTGGAAACGTTTTTGTCCCTGCTGCCCCCTTAAA         | 6.4%  |  |
| D2 | ACTOAGTOAGTAGTOAGATOOGTOGCOCCATOAT         | CITEGAAACGITTTTGTCCCTCCTGCCCCCTTAAA         | 15.3% |  |
|    | ACTGAGTGAGTAGTCAGATCCGTGGCCCATCAT          | CITEGAAACGITTTTGTCCCTCCTGCCCCCTTAAA         | 16.0% |  |
|    | ACTOAGTOAGTAGTOAGATOCGTOGCCCATOATO         | GAAACGTTTTTCTCCCTCCTCCCCCCCTTAAA            | 6.9%  |  |
| D5 | ACTOAGTGAGTAGTCAGATCCGTCGCCCATCATC         | GAAACGITTTTGTCCCTCCTCCCCCCCTTAAA            | 13.6% |  |
|    | ACTGAGTGAGTAGTCAGATCCGTGGCCCATCATG         | GAAACGITTTTGTCCCTCCTGCCCCCTTAAA             | 12.3% |  |
|    | ACTOAGTGAGTAGTCAGATCOGTGGCCCATC            | TCTTGGAAACGTTTTTGTCCCTGCTGCCCCTTAAA         | 14.3% |  |
| D3 | ACTOAGTGAGTAGTCAGATCCGTGGCCCATC            | TCTTGGAAACGTTTTTCTCCCTCCTGCCCCCTTAAA        | 2.5%  |  |
|    | ACTOAGTGAGTAGTCAGATCCGTGGCCCATC            | T CTT CGAAACGTTTTT CT CCCT CCT CCCCCCCTTAAA | 2.2%  |  |
|    | ACTGAGTGAGTAGTCAGATCCGTGGCCCATCATG         | AAACGITTTTGTCCCTCCTGCCCCCTTAAA              | 9.5%  |  |
| D6 | ACTOAGTGAGTAGTCAGATCCGTGGGCCCATCATC        | AAACGTTTTTCTCCCTCCTCCCCCCCTTAAA             | 1.5%  |  |
|    | ACTOACTOACTAGTOACATCACGTOCCCATCATC         | AAACGTTTTTCTCCCTCCTCCCCCCCTTAAA             | 1.3%  |  |
|    |  |   | 0.0%  |  |
| D4 | ACTOAGTGAGTAGTCAGATCOGTGGCCCAT             |   | 2.5%  |  |
|    | ACTOAGTOAGTAGTOAGATOOGTOGCOCCAT            | TCTTCGAAACGTTTTTCTCCCTCCTCCCCCCCTTAAA       | 2.2%  |  |

Measured repair profile reproducibility for the outlier, Overbeek 25, gRNA-target pair. DNA sequence of the target (top) is edited to produce a range of outcomes in two synthetic replicates (green, blue bars) and one endogenous measurement (orange bars). The proportions (x-axis) of the four largest mutational outcomes (e.g. "D3" - deletion of three base pairs, "I1" - insertion of one base pair, etc.; y-axis) is consistent between the experiments. Stretches of microhomology (green) and inserted sequences (red) are highlighted at the cut site (dashed vertical line).



mFigure S4. Reproducibility of indel class frequencies per gRNA

Frequency of different indel classes (panels) in replicate A (x-axis) or B (y-axis) of K562 cell line measurements (markers), annotated by sequencing coverage (color).

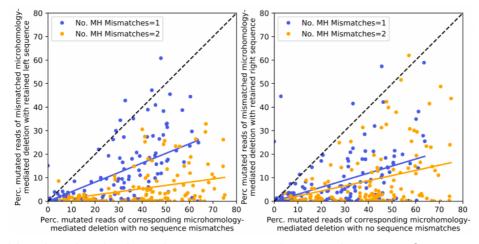


Figure S5. No bias in the side of sequence selected for MMEJ outcomes

Mutations in microhomology sequence reduce repair outcome frequency, but corresponding deletions are still present. For matched pairs of guides, with and without mutations in the microhomologous sequence, the fraction of mutated reads associated with the particular microhomology (y-axis) is smaller than without mismatches (x-axis) for most gRNAs (markers; blue: one mismatch, yellow: two mismatches). The rates of repair are not different depending on whether sequence was retained from PAM-distal (left panel) or PAM-proximal (right panel) side of the cut.

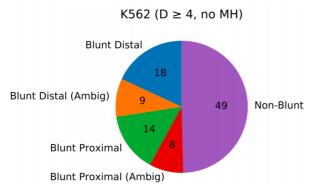


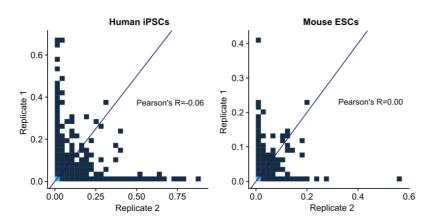
Figure S6. No bias in the side of deletions for NHEJ outcomes

Percent of alternative outcomes for large deletions without microhomology.

# Figure S7. Example of profile measured across different cell lines (most frequent 4 indels only)

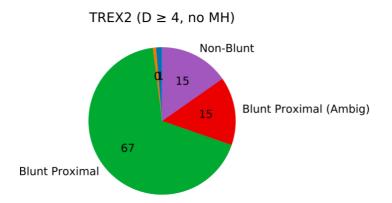
|          | RPE1 K562   | HAP1 — Mouse ESC _ CHO                  | Human iPSC         |
|----------|---|---|--------------------|
| Target   | GAGGCCTTGECGCTCGGGGGACTGCTTCACGGTCA   | ACGTCBBCCBCAECCECTTCGTCCTCTCBCAECAG     | Mutational Profile |
| Outcomes | GAGCCCTTGECCECTCGEGGACTGCTTCACG   | TGGGGGGAGCGCTTCGTGCTCTCGCAGCAG          | 12.0%              |
|          | GAECCCTTGECECTCGEGEACTCCTTCACG  | TEEECERAECCECTTCETECTCTCECAECAG         | 23.2%              |
|          | GAGCCCTTGECGCTCGGGGACTGCTTCACG  | TEEECEECAECCECTTCGTCCTCTCECAECAG        | 25.3%              |
|          | GAGCCCTTGECGCTCGEGGACTGCTTCACG  | TEEEEEEAECCECTTEEFECTCTCECAECAG         | 25.7%              |
|          | GAGCCCTTGECGCTCGEGGACTGCTTCACG  | TGGGGGGAGCGGCTTGGTGCTCTGGCAGCAG         | 23.5%              |
| 07       | GAGCCCTTGECECTCGEGGACTGCTTCACG  | TEEECEECAECCECTTCGTCCTCTCCECAECAG       | 33.1%              |
|          | GAGCCCTTGECCCTCCEGGACTCCTTCACG  | TGGGGGGAGCGGCTTGGTGCTCTGGCAGCAG         | 41.6%              |
|          | GAGCCCTTGECCCTGEGGACTGCTTCAGG   | TGGGGGGGAGCGGCTTGGTGCTCTGGGAGCAG        | 17.1%              |
|          | GAGCCCTTGECCCTCGEGGACTCCTTCACG  | TGGGGGGAGCGGCTTGGTGCTCTGGCAGCAG         | 16.5%              |
|          | GAGCCCTTGECCECTCGEGGACTGCTTCACG   | TGGGGGGAGCCGCTTGGTGCTCTGGCAGCAG         | 24.6%              |
|          | GAGCCCTTGECCCTCGEGGACTCCTTCACG  | TGGGGGGCAGCCGCTTGGTGCTCTGGCAGCAG        | 35.9%              |
|          | GAGCCCTTG2CCCTCC2GGACTGCTTCACCGTC4  | AACGTGEEGECAECCECTTGGTGCTCTGEAGCA       | 51.7%              |
|          | GAGCCCTTGEOCCTGEGGACTGCTTCACCGTC4   | I<br>AACGTGEEGEGAECCECTTOGTGCTCTCGEAECA | 19.7%              |
|          | GAGCCCTTGGCCTGCGGGGACTGCTTCAGCGTC4  | AACGTG33333AGCG3CTTG3TGCTCTG3CAGCA      | 9.0%               |
|          | GAGGOCITIGEOGCTOGEGGACTCCITICACCEGC4  | AACGTG33CC36CAGCC3CTTCGTGCTCTCG6AGCA    | 28.4%              |
|          | GAGCCCTTGECCTCGEGGACTGCTTCAGCETC4   | AACGTCCCCCCCCTCCGTCCTCCCCACCA           | 27.2%              |
| 1        | GAGCCCTTGECCETCGEGEGACTGCTTCAGCEGTC4  | AACGTCGGGCGCAGCCCTTCGTCCTCTCGCAGCA      | 8.4%               |
|          | GAGCCCTTGECCCTCGEGGACTGCTTCACCEGTC4   | AACGTCCCCCCACCACCCCTTCCGTCCTCTCCCACCA   | 8.5%               |
|          | GAGCCCTTGECCCTGCGGGGACTGCTTCAGCGGC4   | AACGTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 17.6%              |
|          | GAECCCTTGECECTCGEGEACTCCTTCACCETCA  | AACGTCCCCCCACCACCCCTTCCTCCTCCCCACCA     | 17.9%              |
|          | GAGCCCTTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   | AAQGTQQQQQQAQQQQCTTQGTQCTCTQQQAQQA      | 23.4%              |
|          | GAGCCCTTGECCCTCGEGGACTCCTTCACCEGTC4   | AACGTCCCCCCACCACCACCA                   | 15.7%              |
|          | GAECCCTTGECECTCGEGEGACTCCTTCACCETC4   | QFTQGGCQGCAGCQCTTQGTQCTCTQQCAGCAG       | 4.4%               |
|          | GAGCCCTTGCCCCTCCCCCTCCACCCTCCCCCTCCACCCTCCCCCTCCACCCTCCACCCTCCCCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCTCCACCCCCC  | QFTQGGCGGCAGCCGCTTQGTQCTCTQGCAGCAG      | 11.9%              |
|          | GAGCCCTTGECCCTCCCCGCGACTGCTTCACCGTC4  | QFTQGGCGGCAGCCGCTTQGTQCTCTQGCAGCAG      | 11.7%              |
|          | GAGCCCTTGECCCTCGEGGACTGCTTCACCETCA  | QFTQGGCQGCAGCQCCTTQGTQCTCTQQCAGCAG      | 6.1%               |
|          | GAGCCCTTGECECTCGEGGACTGCTTCACEGTCA  | QTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 8.4%               |
| 91       | GAGCCCTTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   | QFTQGGCGGCAGCCGCTTQGTQCTCTQGCAGCAG      | 22.2%              |
|          | GAGCCCTTGECECTCGEGGACTGCTTCACEGTCA  | QTQ32CACCCCTTQ5TQCTCTQ3CAQCAG           | 18.3%              |
|          | GAGCCCTTGGCGCTGGGGGACTGCTTCAGCGTC4  | CGTCGGGCGGCAGCCGCTTCGTCCTCTCGCAGCAG     | 12.5%              |
|          | GAGCCCTTGGCCCTCGGGGACTGCTTCACGGTC4  | QGTQGGQQCAGQQCTTQGTQCTCTQGQAQCAG        | 11.2%              |
|          | GAGCCCTTGGCGCTGGGGGACTGCTTCACGGTC4  | QTQ32Q3CAQCQCTTQ5TQCTCTQ3CAQCAG         | 8.4%               |
|          | GAGCCCTTGECCCTGEGGGACTGCTTCACCETC4  | CGTCCCCCCACCCCCTTCCGTCCTCCTCCCACCAG     | 12.1%              |
|          | GAGCCETTIGECCETCOEGGACTCCETTICACCG  | CAGCCGCTTCGTGCTCTCGCAGCAG               | 1.0%               |
|          | GAGCCETTIGECCETTIGECGETTIGECCETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECCETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECCETTIGECCETTIGECGETTIGECCETTIGECGETTIGECCETTIGECGETTIGECCETTIGECGETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECET | CAGCCCCTTCGTCCTCTCCCAGCAG               | 1.8%               |
|          | GAGCCETTGEECETGEEGACTGETTCACCG  | CAGCCCCTTCGTGCTCTCCCAGCAG               | 1.8%               |
|          | GAGCCETTIGECCETTIGECGETTIGECCETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECGETTIGECETTIGECETTIGECETTIGECETTIGECGETTIGECETTIGECCETTIGECCETTIGECGETTIGECCETTIGECCETTIGECGETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECCETTIGECETT | CAGCCCCTTCGTCCTCTCCCAGCAG               | 2.5%               |
|          | GAGCCETTGEECETGEEGGACTGETTCACCG   | CAGCCGCTTCGTGCTCTCGCAGCAG               | 3.7%               |
| 013      | GAGCCETTIGECCETTIGECGEGACTGETTICACCG  | CAGCCCCTTCGTCCTCTCCCAGCAG               | 1.9%               |
|          | GAGGOCTTGEGGECTGGEGGACTGCTTCAGGG  | CAGCOGETTOGTGETCTCOGEAGEAG              | 1.9%               |
|          | GAGCCETTGEEGECTGGEEGACTGETTCAGGG  | CAGCCCCTTCGTCCTCTCCCAGCAG               | 2.4%               |
|          | GAGCCTTGECCCTCGEGGACTGCTTCACCG  | CABCCCECTTCGTCCTCTCCCABCAG              | 0.8%               |
|          | GAGCCETTGECCETCGECGGACTGETTCACCG  | CAGCCCCTTCGTCCTCTCCCAGCAG               | 3.1%               |
|          | GAGCCCTTGECCECTGECGACTGCTTGACGG   | CAGCCCCTTCGTCCTCTCCCAGCAG               | 0.8%               |

Figure S8. Lack of reproducibility in large insertions in human and mouse stem cells



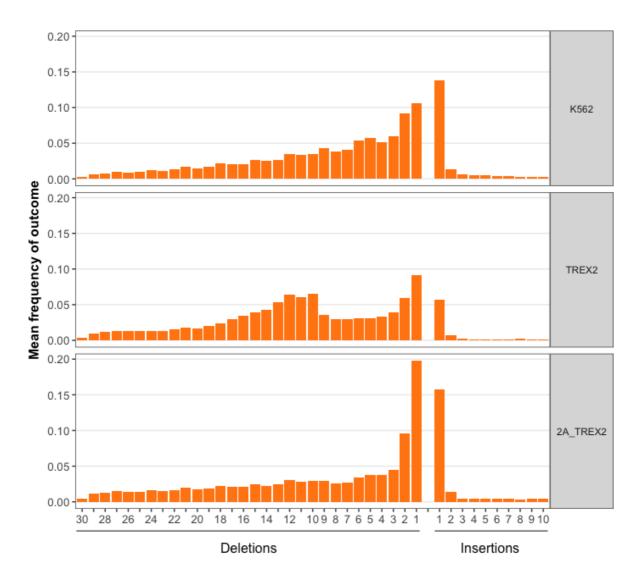
Proportion of individual indels in two replicates of human iPSCs (left) and mouse ESCs (right).

Figure S9. Repair outcomes from Cas9-TREX2 fusion favour blunt end joins



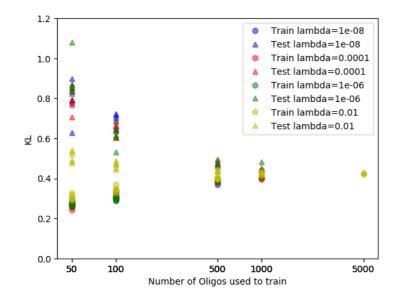
Percent of alternative outcomes for large deletions without microhomology.

Figure S10. Cas9-2A-TREX2 has smaller influence on repair outcomes than Cas9-TREX2



The mean frequency (y-axis) of deletion or insertion size (x-axis) across genomic sequence targets for three alternative Cas9 effector constructs (panels).

Figure S11. Training success depends on hyperparameters and dataset size



# Table S1. Primer sequences (5' > 3')

| Cloning                             | Cloning of pKLV2-U6(BbsI)-PKGpuro2ABFP-W                   |  |  |
|-------------------------------------|--|--|--|
| P1                                  | GGCAGCACTGCATAATTCTCTTAC                                   |  |  |
| P2                                  | CCTACCCGGTAGAATTGGATCCAAACGTGTCTTCTCGAAGACCC               |  |  |
| P3                                  | GTAAGAGAATTATGCAGTGCTGCC                                   |  |  |
| P4                                  | GGGTCTTCGAGAAGACACGTTTGGATCCAATTCTACCGGGTAGG               |  |  |
| Amplifi                             | Amplification of oligo pools for library cloning           |  |  |
| P5                                  | GGAAACTACACTTGCCTGGC                                       |  |  |
| P6                                  | AACTTGCTATTTCTAGCTCTAAAAC                                  |  |  |
| P7                                  | GACGTCCAGAGCACAGATGG                                       |  |  |
| P8                                  | GCTGTTTCCAGCATAGCTCTTAAAC                                  |  |  |
| Preparation of sequencing libraries |  |  |  |
| P9                                  | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTGGAAAGGACGAAACA      |  |  |
| P10                                 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAAACTACACTTGCCTGGC      |  |  |
| P11                                 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACGTCCAGAGCACAGATGG      |  |  |
| P12                                 | TCGGCATTCCTGCTGAACCGCTCTTCCGATCTACCCGGTAGAATTGGATCCAAAC    |  |  |
| P13                                 | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT |  |  |
| P14*                                |  |  |  |
| Sequencing primers                  |  |  |  |
| P15                                 | TCTTCCGATCTCTTGTGGAAAGGACGAAACACCG                         |  |  |
| P16                                 | CTCTTCCGATCTGACGTCCAGAGCACAGATGG                           |  |  |
| P17                                 | GCTCTTCCGATCTGGAAACTACACTTGCCTGGC                          |  |  |
| P18                                 | CGCTCTTCCGATCTACCCGGTAGAATTGGATCCAAAC                      |  |  |
| . NI in                             |  |  |  |

\*: N<sub>10</sub>, index for multiplexed sequencing.

#### Table S2. Screen conditions

| Cell line          | gRNA Scaffold | Cells<br>(x10 <sup>6</sup> ) | Multiplicity of infection | Coverage<br>(cells per<br>construct) | Replicates |
|--------------------|---------------|------------------------------|---------------------------|--------------------------------------|------------|
| K562-Cas9          | Improved      | 70                           | 0.6                       | 800                                  | 2          |
| K562-Cas9          | Improved      | 140                          | 0.6                       | 1600                                 | 2          |
| K562-Cas9          | Conventional  | 32                           | 0.5                       | 1600                                 | 1          |
| K562               | Improved      | 70                           | 0.6                       | 800                                  | 1          |
| K562               | Conventional  | 16                           | 0.5                       | 800                                  | 1          |
| K562-eCas9         | Improved      | 70                           | 0.6                       | 800                                  | 2          |
| K562-eCas9         | Conventional  | 16                           | 0.5                       | 800                                  | 2          |
| K562-Cas9-TREX2    | Improved      | 70                           | 0.6                       | 800                                  | 2          |
| K562-Cas9-TREX2    | Conventional  | 16                           | 0.5                       | 800                                  | 2          |
| K562-Cas9-2A-TREX2 | Improved      | 70                           | 0.6                       | 800                                  | 2          |
| K562-Cas9-2A-TREX2 | Conventional  | 16                           | 0.5                       | 800                                  | 2          |
| RPE-1-Cas9         | Improved      | 52                           | 0.5                       | 500                                  | 2          |
| HAP1-Cas9          | Improved      | 83                           | 0.5                       | 800                                  | 2          |
| CHO-Cas9           | Improved      | 83                           | 0.5                       | 800                                  | 2          |
| iPSC-Cas9          | Improved      | 83                           | 0.5                       | 800                                  | 2          |
| E14TG2a-Cas9       | Improved      | 83                           | 0.5                       | 800                                  | 2          |

## Table S3. Most important features for prediction

| Feature Symbol      | θ value/s     | Description   |
|---------------------|---------------|---|
| I1Rpt               | 1.461         | Single nucleotide insertion repeating the PAM distal nucleotide adjacent to the cut site                                |
| IL-11, IL-22        | 1.432, 0.838  | Insertion at the cut site   |
| PW_I1_T_vs_I1Rpt    | 0.692         | Single nucleotide insertion repeating the PAM<br>distal nucleotide adjacent to the cut site in<br>which a T is inserted |
| PW_I1_A_vs_I1Rpt    | 0.544         | Single nucleotide insertion repeating the PAM distal nucleotide adjacent to the cut site in which an A is inserted      |
| PW_No MH_vs_DL-11   | 0.355         | No microhomology, blunt-end deletion on PAM-proximal side   |
| 11                  | 0.354         | Any single nucleotide insertion   |
| PW_I1_C_vs_I1Rpt    |               | Single nucleotide insertion repeating the PAM<br>distal nucleotide adjacent to the cut site in<br>which a C is inserted |
| PW_L0_NT=C_vs_DL-11 | 0.311         | Blunt-end deletion on PAM-proximal side with a C adjacent to the cut site.  |
| PW_CS0_NT=G_vs_I1   | 0.303         | Any single nucleotide insertion when there is a G adjacent to the cut site on the PAM-proximal side                     |
|                     |               |   |
|                     |               |   |
| I2NonRpt            | -0.195        | Double nucleotide insertion that is not just an additional repeat of the repeat single nucleotide insertion             |
| IL>=0, IL<-3        | -0.249,-0.244 | Any insertion away from the cut site  |
| 12                  | -0.250        | Any double nucleotide insertion   |
| PW_D>12_vs_DR0-0    | -0.364        | Blunt-end deletion on PAM-distal side of cut of size greater than 12  |
| I1_G                | -0.473        | Single nucleotide insertion of a G  |