Supplementary Materials for

RNA-Guided Human Genome Engineering via Cas9

Prashant Mali, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, George M. Church*

*To whom correspondence should be addressed. E-mail: gchurch@genetics.med.harvard.edu

Published 3 January 2013 on Science Express
DOI: 10.1126/science.1232033

This PDF file includes
Materials and Methods
Supplementary Text
Figs. S1 to S11
Full References

Other Supplementary Material for this manuscript includes the following:
(available at http://arep.med.harvard.edu/human_crispr)

Table S1. Bioinformatically computed genome-wide resource of candidate unique gRNA targets in human exons.
Table S2. Incorporation of gRNA targets in Table 1 into a 200-bp format suitable for multiplex DNA array based synthesis.
Table S3. 12k gRNA targets from Table 1 in a 200-bp format synthesized by CustomArray Inc.

Corrections: The authors fixed some minor typographical errors, made bold the DNA sequences in Figs. S1 and S10 to improve readability, and included the link to their Addgene plasmid deposit. The findings have not changed.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Type II CRISPR-Cas System</td>
<td>3</td>
</tr>
<tr>
<td>Material and Methods</td>
<td>7</td>
</tr>
<tr>
<td>Supplementary Fig. S1. The engineered type II CRISPR system for human cells.</td>
<td>14</td>
</tr>
<tr>
<td>Supplementary Fig. S2. RNA-guided genome editing requires both Cas9 and guide RNA for successful targeting.</td>
<td>16</td>
</tr>
<tr>
<td>Supplementary Fig. S3. Analysis of gRNA and Cas9 mediated genome editing.</td>
<td>17</td>
</tr>
<tr>
<td>Supplementary Fig. S4. RNA-guided genome editing is target sequence specific.</td>
<td>19</td>
</tr>
<tr>
<td>Supplementary Fig. S5. Guide RNAs targeted to the GFP sequence enable robust genome editing.</td>
<td>21</td>
</tr>
<tr>
<td>Supplementary Fig. S6. RNA-guided genome editing is target sequence specific, and demonstrates similar targeting efficiencies as ZFNs or TALENs.</td>
<td>22</td>
</tr>
<tr>
<td>Supplementary Fig. S7. RNA-guided NHEJ in human iPS cells.</td>
<td>24</td>
</tr>
<tr>
<td>Supplementary Fig. S8. RNA-guided NHEJ in human K562 cells.</td>
<td>26</td>
</tr>
<tr>
<td>Supplementary Fig. S9. RNA-guided NHEJ in human 293T cells.</td>
<td>28</td>
</tr>
<tr>
<td>Supplementary Fig. S10. HR at the endogenous AAVS1 locus using either a dsDNA donor or a short oligonucleotide donor.</td>
<td>30</td>
</tr>
<tr>
<td>Supplementary Fig. S11. Methodology for multiplex synthesis, retrieval and U6 expression vector cloning of guide RNAs targeting genes in the human genome.</td>
<td>32</td>
</tr>
<tr>
<td>Supplementary Table S1. Bioinformatically computed genome-wide resource of candidate unique gRNA targets in human exons.</td>
<td>*</td>
</tr>
<tr>
<td>Supplementary Table S2. Incorporation of gRNA targets in Table 1 into a 200bp format suitable for multiplex DNA array based synthesis.</td>
<td>*</td>
</tr>
<tr>
<td>Supplementary Table S3. 12k gRNA targets from Table 1 in a 200bp format synthesized by CustomArray Inc.</td>
<td>*</td>
</tr>
</tbody>
</table>

*Due to size constraints, the Supplementary Tables S1, S2 and S3 are available on: [http://arep.med.harvard.edu/human_crispr](http://arep.med.harvard.edu/human_crispr).
The Type II CRISPR-Cas System

Bacteria and archaea have evolved adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems that use short RNA to direct degradation of foreign nucleic acids (DNA/RNA). CRISPR defense involves acquisition and integration of new targeting “spacers” from invading virus or plasmid DNA into the CRISPR locus, expression and processing of short guiding CRISPR RNAs (crRNAs) consisting of spacer-repeat units, and cleavage of nucleic acids (most commonly DNA) complementary to the spacer.

Three classes of CRISPR systems have been described thus far (Type I, II and III). Here we focus on the Type II CRISPR system, which utilizes a single effector enzyme, Cas9, to cleave dsDNA, whereas Type I and Type III systems require multiple distinct effectors acting as a complex (for a detailed review of CRISPR classification, see reference (29)). As a consequence, Type II systems are more likely to function in alternative contexts such as eukaryotic cells. The Type II effector system consists of a long pre-crRNA transcribed from the spacer-containing CRISPR locus, the multifunctional Cas9 protein, and a tracrRNA important for gRNA processing. The tracrRNAs hybridize to the repeat regions separating the spacers of the pre-crRNA, initiating dsRNA cleavage by endogenous RNase III, which is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9. Jinek et al. demonstrated that a tracrRNA-crRNA fusion, termed a guide RNA (gRNA) in this work (Fig. 1), is functional in vitro, obviating the need for RNase III and the crRNA processing in general (4).
Type II CRISPR interference is a result of Cas9 unwinding the DNA duplex and searching for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a “protospacer” sequence in the target DNA and the remaining spacer sequence in the crRNA. Importantly, Cas9 cuts the DNA only if a correct protospacer-adjacent motif (PAM) is also present at the 3’ end. Different Type II systems have differing PAM requirements. The *S. pyogenes* system utilized in this work requires an NGG sequence, where N can be any nucleotide. *S. thermophilus* Type II systems require NGGNG (30) and NNAGAAW (31), respectively, while different *S. mutans* systems tolerate NGG or NAAR (32). Bioinformatic analyses have generated extensive databases of CRISPR loci in a variety of bacteria that may serve to identify new PAMs and expand the set of CRISPR-targetable sequences (33, 34). In *S. thermophilus*, Cas9 generates a blunt-ended double-stranded break 3bp upstream of the protospacer (5), a process mediated by two catalytic domains in the Cas9 protein: an HNH domain that cleaves the complementary strand of the DNA and a RuvC-like domain that cleaves the non-complementary strand (refer Fig. 1A, fig. S1). While the *S. pyogenes* system has not been characterized to the same level of precision, DSB formation also occurs towards the 3’ end of the protospacer. If one of the two nuclease domains is inactivated, Cas9 will function as a nickase *in vitro* (4) and in human cells (fig. S3).

As a genome engineering tool, the specificity of gRNA-directed Cas9 cleavage will be of the utmost importance. Significant off-target activity could cause unwanted double-strand breaks at other regions of the genome, resulting in toxicity and possibly oncogenesis in gene therapy applications. The *S. pyogenes* system tolerates mismatches in the first 6 bases out of the 20bp mature spacer sequence *in vitro*. However, it is entirely possible that greater
stringency is required in vivo given the low toxicity we observed in human cell lines, as potential off-target sites matching (last 14 bp) NGG exist within the human reference genome for our gRNAs. Mismatches towards the 3’ end of the spacer, known as the “seed sequence” (6), are less well tolerated. Jinek et al. found that single mismatches in the PAM at positions -3 through -7 abolished interference in vitro, though a mismatch at position -10 did not (4). In S. thermophilus, single mutations in the PAM or at positions -1, -3 through -5, and -7 through -8 abolished interference. When transplanted into E. coli, the S. thermophilus system did not tolerate single mutations in the PAM or in positions -3, -6, or -8. As a caveat, Garneau et al. found that spacers acquired from plasmid DNA tolerated greater degeneracy in both the PAM and seed sequence while sufficing to block plasmid acquisition in S. thermophilus (35); however, similar degeneracy was not sufficient to block phage infection (31), emphasizing the importance of the assay utilized. Taken together, these results point towards the urgent need to assay specificity in the context of interest.

There exist at least four possible ways to improve specificity. First, it is possible that interference is sensitive to the melting temperature of the gRNA-DNA hybrid, in which case AT-rich target sequences are likely to have fewer off-target sites. Second, carefully choosing target sites to avoid pseudo-sites with at least 14bp matching sequences elsewhere in the genome of interest. Third, the use of a Cas9 variant requiring a longer PAM sequence will reduce the set of potential targetable sequences, but should similarly reduce the frequency of off-target sites. Finally, directed evolution might be utilized to improve Cas9 specificity to a level sufficient to completely preclude off-target activity, ideally requiring a perfect 20bp gRNA match with a minimal PAM. Such a project is likely to require extensive modifications to the Cas9 protein, as
the small genomes of bacteria are unlikely to select for high specificity in natural variants. As such, novel methods permitting many rounds of evolution in a short timeframe (36) may be warranted. For more detailed reviews of CRISPR systems, see references (1, 37).
Material and Methods

Plasmid construction

The Cas9 gene sequence was human codon optimized and assembled by hierarchical fusion PCR assembly of 9 500bp gBlocks ordered from IDT (sequence in fig. S1A). Cas9_D10A was similarly constructed. The resulting full-length products were cloned into the pcDNA3.3-TOPO vector (Invitrogen). The target gRNA expression constructs were directly ordered as individual 455bp gBlocks from IDT (sequence in fig. S1B) and either cloned into the pCR-BluntII-TOPO vector (Invitrogen) or pcr amplified. The vectors for the HR reporter assay involving a broken GFP were constructed by fusion PCR assembly of the GFP sequence bearing the stop codon and 68bp AAVS1 fragment (or mutants thereof; refer fig. S4), or 58bp fragments from the DNMT3a and DNMT3b genomic loci (refer fig. S6) assembled into the EGIP lentivector from Addgene (plasmid #26777). These lentivectors were then used to establish the GFP reporter stable lines. TALENs used in this study were constructed using the protocols described in (11). The dsDNA donor for HR at the native AAVS1 locus is described in (13). All DNA reagents developed in this study are available at Addgene (http://www.addgene.org/crispr/church/).

Cell culture

PGP1 iPS cells were maintained on Matrigel (BD Biosciences)-coated plates in mTeSR1 (Stemcell Technologies). Cultures were passaged every 5–7 d with TrypLE Express (Invitrogen). K562 cells were grown and maintained in RPMI (Invitrogen) containing 15% FBS. HEK 293T cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM, Invitrogen) high glucose supplemented with 10% fetal bovine serum (FBS, Invitrogen), penicillin/streptomycin
(pen/strep, Invitrogen), and non-essential amino acids (NEAA, Invitrogen). All cells were maintained at 37°C and 5% CO₂ in a humidified incubator.

**Gene targeting of PGP1 iPS, K562 and 293Ts**

PGP1 iPS cells were cultured in Rho kinase (ROCK) inhibitor (Calbiochem) 2h before nucleofection. Cells were harvest using TrypLE Express (Invitrogen) and 2×10⁶ cells were resuspended in P3 reagent (Lonza) with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid, and nucleofected according to manufacturer’s instruction (Lonza). Cells were subsequently plated on an mTeSR1-coated plate in mTeSR1 medium supplemented with ROCK inhibitor for the first 24h. For K562s, 2×10⁶ cells were resuspended in SF reagent (Lonza) with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid, and nucleofected according to manufacturer’s instruction (Lonza). For 293Ts, 0.1×10⁶ cells were transfected with 1μg Cas9 plasmid, 1μg gRNA and/or 1μg DNA donor plasmid using Lipofectamine 2000 as per the manufacturer’s protocols. The DNA donors used for endogenous AAVS1 targeting were either a dsDNA donor (Fig. 2C) or a 90mer oligonucleotide. The former has flanking short homology arms and a SA-2A-puromycin-CaGGS-eGFP cassette to enrich for successfully targeted cells.

**Assess the targeting efficiency**

Cells were harvested 3 days after nucleofection and the genomic DNA of ~1 X 10⁶ cells was extracted using prepGEM (ZyGEM). PCR was conducted to amplify the targeting region with genomic DNA derived from the cells and amplicons were deep sequenced by MiSeq Personal Sequencer (Illumina) with coverage >200,000 reads. The sequencing data was analyzed to estimate NHEJ efficiencies. The reference AAVS1 sequence analyzed is:
The PCR primers for amplifying the targeting regions in the human genome are:

**AAVS1-R**
CTCGGCATTCTGTGAAACCGCTCTTCCGATCTAcaggaggtgggggttagac

**AAVS1-F.1**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATtatattcccagggccgtta

**AAVS1-F.2**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGtatattcccagggccgtta

**AAVS1-F.3**
ACACTCTTTCCCCTACACGACGCTCTTCCGATCTGCCTAAtatattcccagggccgtta

**AAVS1-F.4**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAtatattcccagggccgtta

**AAVS1-F.5**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTGTtatattcccagggccgtta

**AAVS1-F.6**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGGCtatattcccagggccgtta

**AAVS1-F.7**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTGtatattcccagggccgtta

**AAVS1-F.8**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAAGTtatattcccagggccgtta

**AAVS1-F.9**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATCtatattcccagggccgtta

**AAVS1-F.10**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCTAtatattcccagggccgtta

**AAVS1-F.11**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGCCtatattcccagggccgtta

**AAVS1-F.12**
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAAGtatattcccagggccgtta

To analyze the HR events using the DNA donor in Fig. 2C the primers used were:

**HR_AAVS1-F**
CTGCCGTCTCTCCTCTGAGT

**HR_Puro-R**
GTGGGCTTGTACTCGGTCAT
Bioinformatics approach for computing human exon CRISPR targets and methodology for their multiplexed synthesis

We sought to generate a set of gRNA gene sequences that maximally target specific locations in human exons but minimally target other locations in the genome. Maximally efficient targeting by a gRNA is achieved by 23nt sequences, the 5’-most 20nt of which exactly complement a desired location, while the three 3’-most bases must be of the form NGG. Additionally, the 5’-most nt must be a G to establish a pol-III transcription start site. However, according to (4), mispairing of the six 5’-most nt of a 20bp gRNA against its genomic target does not abrogate Cas9-mediated cleavage so long as the last 14nt pairs properly, but mispairing of the eight 5’-most nt along with pairing of the last 12 nt does, while the case of the seven 5-most nt mispairs and 13 3’ pairs was not tested. To be conservative regarding off-target effects, we therefore assumed that the case of the seven 5’-most mispairs is, like the case of six, permissive of cleavage, so that pairing of the 3’-most 13nt is sufficient for cleavage. To identify CRISPR target sites within human exons that should be cleavable without off-target cuts, we therefore examined all 23bp sequences of the form 5’-GBBBBB BBBBB BBBBB BBBBB NGG-3’ (form 1), where the B’s represent the bases at the exon location, for which no sequence of the form 5’-NNNNN NNBBBB BBBBB BBBBB NGG-3’ (form 2) existed at any other location in the human genome. Specifically, we (i) downloaded a BED file of locations of coding regions of all RefSeq genes the GRCh37/hg19 human genome from the UCSC Genome Browser (38-40). Coding exon locations in this BED file comprised a set of 346089 mappings of RefSeq mRNA accessions to the hg19 genome. However, some RefSeq mRNA accessions mapped to multiple genomic locations (probable gene duplications), and many accessions mapped to subsets of the same set of exon
locations (multiple isoforms of the same genes). To distinguish apparently duplicated gene instances and consolidate multiple references to the same genomic exon instance by multiple RefSeq isoform accessions, we therefore (ii) added unique numerical suffixes to 705 RefSeq accession numbers that had multiple genomic locations, and (iii) used the mergeBed function of BEDTools (41) (v2.16.2-‐zip-87e3926) to consolidate overlapping exon locations into merged exon regions. These steps reduced the initial set of 346089 RefSeq exon locations to 192783 distinct genomic regions. We then downloaded the hg19 sequence for all merged exon regions using the UCSC Table Browser, adding 20bp of padding on each end. (iv) Using custom perl code, we identified 1657793 instances of form 1 within this exonic sequence. (v) We then filtered these sequences for the existence of off-‐target occurrences of form 2: For each merged exon form 1 target, we extracted the 3’-‐most 13bp specific (B) “core” sequences and, for each core generated the four 16bp sequences 5’-‐BBB BBB BBBB NGG-‐3’ (N = A, C, G, and T), and searched the entire hg19 genome for exact matches to these 6631172 sequences using Bowtie version 0.12.8 (42) using the parameters -l 16 -v 0 -k 2. We rejected any exon target site for which there was more than a single match. Note that because any specific 13bp core sequence followed by the sequence NGG confers only 15bp of specificity, there should be on average ~5.6 matches to an extended core sequence in a random ~3Gb sequence (both strands). Therefore, most of the 1657793 initially identified targets were rejected; however 189864 sequences passed this filter. These comprise our set of CRISPR-‐targetable exonic locations in the human genome. The 189864 sequences target locations in 78028 merged exonic regions (~40.5% of the total of 192783 merged human exon regions) at a multiplicity of ~2.4 sites per targeted exonic region. To assess targeting at a gene level, we clustered RefSeq mRNA
mappings so that any two RefSeq accessions (including the gene duplicates we distinguished in (iii)) that overlap a merged exon region are counted as a single gene cluster, the 189864 exonic specific CRISPR sites target 17104 out of 18872 gene clusters (~90.6% of all gene clusters) at a multiplicity of ~11.1 per targeted gene cluster. (Note that while these gene clusters collapse RefSeq mRNA accessions that represent multiple isoforms of a single transcribed gene into a single entity, they will also collapse overlapping distinct genes as well as genes with antisense transcripts.) At the level of original RefSeq accessions, the 189864 sequences targeted exonic regions in 30563 out of a total of 43726 (~69.9%) mapped RefSeq accessions (including our distinguished gene duplicates) at a multiplicity of ~6.2 sites per targeted mapped RefSeq accession.

As we gather information on CRISPR performance at our computationally predicted human exon CRISPR target sites, we plan to refine our database by correlating performance with factors we expect to be important, such as base composition and secondary structure of both gRNAs and genomic targets (43, 44), and the epigenetic state of these targets in human cell lines for which this information is available (45).

Finally, we also incorporated these target sequences into a 200bp format that is compatible for multiplex synthesis on DNA arrays (14, 46). Our design allows for targeted retrieval of a specific or pools of gRNA sequences from the DNA array based oligonucleotide pool and its ready cloning into a common expression vector (fig. S11A, Supplementary Table 2). Specifically we tested this approach by synthesizing a 12k oligonucleotide pool from CustomArray Inc. (Supplementary Table 3). Furthermore, as per our approach we were able to
successfully retrieve gRNAs of choice from this library (fig. S11B). We observed an error rate of ~4 mutations per 1000bp of array synthesized DNA.
The engineered type II CRISPR system for human cells. (A) Expression format and full sequence of the cas9 gene insert. The RuvC-like and HNH motifs (4), and the C-
terminus SV40 NLS are respectively highlighted by blue, brown and orange colors. (B) U6 promoter based expression scheme for the guide RNAs and predicted RNA transcript secondary structure. The use of the U6 promoter constrains the 1st position in the RNA transcript to be a ‘G’ and thus all genomic sites of the form GN_{20}GG can be targeted using this approach. (C) The 7 gRNAs used in this study are listed.
Supplementary Fig. S2. RNA-guided genome editing requires both Cas9 and guide RNA for successful targeting. Using the GFP reporter assay described in Fig. 1B, all possible combinations of the repair DNA donor, Cas9 protein, and gRNA were tested for their ability to effect successful HR (in 293Ts). GFP+ cells were observed only when all the 3 components were present, validating that these CRISPR components are essential for RNA-guided genome editing. Data are shown as mean ± SEM (N=3).
Supplementary Fig. S3. Analysis of gRNA and Cas9 mediated genome editing. We closely examined the CRISPR mediated genome editing process using either (A) a GFP reporter assay as described earlier, and (B) deep sequencing of the targeted loci (in 293Ts). As comparison we
also tested a D10A mutant for Cas9 that has been shown in earlier reports to function as a nickase in *in vitro* assays. Our data shows that both Cas9 and Cas9D10A can effect successful HR at nearly similar rates. Deep sequencing however confirms that while Cas9 shows robust NHEJ at the targeted loci, the D10A mutant has significantly diminished NHEJ rates (as would be expected from its putative ability to only nick DNA). Also, consistent with the known biochemistry of the Cas9 protein, our NHEJ data confirms that most base-pair deletions or insertions occurred near the 3’ end of the target sequence: the peak is ~3-4 bases upstream of the PAM site, with a median deletion frequency of ~9-10bp. Data are shown as mean ± SEM (N=3).
**Supplementary Fig. S4.** RNA-guided genome editing is target sequence specific. Similar to the GFP reporter assay described in Fig. 1B, we developed 3 293T stable lines each bearing a distinct GFP reporter construct. These are distinguished by the sequence of the AAVS1
fragment insert (as indicated in the figure). One line harbored the wild-type fragment while the
two other lines were mutated at 6bp (highlighted in red). Each of the lines was then targeted by
one of the following 4 reagents: a GFP-ZFN pair that can target all cell types since its targeted
sequence was in the flanking GFP fragments and hence present in along cell lines; a AAVS1
TALEN that could potentially target only the wt-AAVS1 fragment since the mutations in the
other two lines should render the left TALEN unable to bind their sites; the T1 gRNA which can
also potentially target only the wt-AAVS1 fragment, since its target site is also disrupted in the
two mutant lines; and finally the T2 gRNA which should be able to target all the 3 cell lines since
unlike the T1 gRNA its target site is unaltered among the 3 lines. Consistent with these
predictions, the ZFN modified all 3 cell types, the AAVS1 TALENs and the T1 gRNA only targeted
the wt-AAVS1 cell type, and the T2 gRNA successfully targets all 3 cell types. These results
together confirm that the guide RNA mediated editing is target sequence specific. Data are
shown as mean ± SEM (N=3).
Supplementary Fig. S5. Guide RNAs targeted to the GFP sequence enable robust genome editing. In addition to the 2 gRNAs targeting the AAVS1 insert, we also tested two additional gRNAs targeting the flanking GFP sequences of the reporter described in Fig. 1B (in 293Ts). These gRNAs were also able to effect robust HR at this engineered locus. Data are shown as mean ± SEM (N=3).
**Supplementary Fig. S6.** RNA-guided genome editing is target sequence specific, and demonstrates similar targeting efficiencies as ZFNs or TALENs. Similar to the GFP reporter assay described in Fig. 1B, we developed 2 293T stable lines each bearing a distinct GFP reporter construct. These are distinguished by the sequence of the fragment insert (as indicated in the
figure). One line harbored a 58bp fragment from the DNMT3a gene while the other line bore a homologous 58bp fragment from the DNMT3b gene. The sequence differences are highlighted in red. Each of the lines was then targeted by one of the following 6 reagents: a GFP-ZFN pair that can target all cell types since its targeted sequence was in the flanking GFP fragments and hence present in along cell lines; a pair of TALENs that potentially target either DNMT3a or DNMT3b fragments; a pair of gRNAs that can potentially target only the DNMT3a fragment; and finally a gRNA that should potentially only target the DNMT3b fragment. Consistent with these predictions, the ZFN modified all 3 cell types, and the TALENs and gRNAs only their respective targets. Furthermore the efficiencies of targeting were comparable across the 6 targeting reagents. These results together confirm that RNA-guided editing is target sequence specific and demonstrates similar targeting efficiencies as ZFNs or TALENs. Data are shown as mean ± SEM (N=3).
endogenous ‘native’ hAAVS1 locus sequence
(PGP1 iPS)

..TTATCTGTCCCTCCACCCACAGTGGGGCCACTAGGGACAGGATGTTGA..

T1 target  T2 target

NHEJ rates evaluated using NGS of targeted AAVS1 locus
Supplementary Fig. S7. RNA-guided NHEJ in human iPS cells. We nucleofected human iPS cells (PGP1) with constructs indicated in the left panel. 4 days after nucleofection, we measured NHEJ rate by assessing genomic deletion and insertion rate at double-strand breaks (DSBs) by deep sequencing. Panel 1: Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot the deletion incidence at each nucleotide position in black lines and we calculated the deletion rate as the percentage of reads carrying deletions. Panel 2: Insertion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot the incidence of insertion at the genomic location where the first insertion junction was detected in black lines and we calculated the insertion rate as the percentage of reads carrying insertions. Panel 3: Deletion size distribution. We plot the frequencies of different size deletions among the whole NHEJ population. Panel 4: Insertion size distribution. We plot the frequencies of different sizes insertions among the whole NHEJ population. iPS targeting by both gRNAs is efficient (2-4%), sequence specific (as shown by the shift in position of the NHEJ deletion distributions), and reaffirming the results of fig. S2, the NGS-based analysis also shows that both the Cas9 protein and the gRNA are essential for NHEJ events at the target locus.
Supplementary Fig. S8. RNA-guided NHEJ in K562 cells. We nucleofected K562 cells with constructs indicated in the left panel. 4 days after nucleofection, we measured NHEJ rate by assessing genomic deletion and insertion rate at DSBs by deep sequencing. **Panel 1**: Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot the deletion incidence at each nucleotide position in black lines and calculated the deletion rate as the percentage of reads carrying deletions. **Panel 2**: Insertion rate detected at targeting region. Red dash lines:
boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot
the incidence of insertion at the genomic location where the first insertion junction was
detected in black lines and we calculated the insertion rate as the percentage of reads carrying
insertions. Panel 3: Deletion size distribution. We plot the frequencies of different size deletions
among the whole NHEJ population. Panel 4: insertion size distribution. We plot the frequencies
of different sizes insertions among the whole NHEJ population. K562 targeting by both gRNAs is
efficient (13-38%) and sequence specific (as shown by the shift in position of the NHEJ deletion
distributions). Importantly, as evidenced by the peaks in the histogram of observed frequencies
of deletion sizes, simultaneous introduction of both T1 and T2 guide RNAs resulted in high
efficiency deletion of the intervening 19bp fragment, demonstrating that multiplexed editing of
genomic loci is also feasible using this approach.
Supplementary Fig. S9. RNA-guided NHEJ in 293T cells. We transfected 293T cells with constructs indicated in the left panel. 4 days after nucleofection, we measured NHEJ rate by assessing genomic deletion and insertion rate at DSBs by deep sequencing. **Panel 1:** Deletion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot the deletion incidence at each nucleotide position in black lines and calculated the deletion rate as the percentage of reads.
carrying deletions. Panel 2: Insertion rate detected at targeting region. Red dash lines: boundary of T1 RNA targeting site; green dash lines: boundary of T2 RNA targeting site. We plot the incidence of insertion at the genomic location where the first insertion junction was detected in black lines and we calculated the insertion rate as the percentage of reads carrying insertions. Panel 3: Deletion size distribution. We plot the frequencies of different size deletions among the whole NHEJ population. Panel 4: insertion size distribution. We plot the frequencies of different sizes insertions among the whole NHEJ population. 293T targeting by both gRNAs is efficient (10-24%) and sequence specific (as shown by the shift in position of the NHEJ deletion distributions).
Supplementary Fig. S10. HR at the endogenous AAVS1 locus using either a dsDNA donor or a short oligonucleotide donor. (A) PCR screen (refer Fig. 2C) confirmed that 21/24 randomly
picked 293T clones were successfully targeted. (B) Similar PCR screen confirmed 3/7 randomly picked PGP1-iPS clones were also successfully targeted. (C) Finally short 90mer oligos could also effect robust targeting at the endogenous AAVS1 locus. The pink bar in the histogram highlights the frequency of events where an ‘AA’ base modification by oligonucleotide mediated homology directed repair (HDR) was successfully effected (shown here for K562 cells).
Supplementary Fig. S11. Methodology for multiplex synthesis, retrieval and U6 expression vector cloning of guide RNAs targeting genes in the human genome. We established a resource of ~190k bioinformatically computed unique gRNA sites targeting ~40.5% of all exons of genes in the human genome (list in Supplementary Table 1). (A) We incorporated these into a 200bp format (list in Supplementary Table 2) that is compatible for multiplex synthesis on DNA arrays.
Specifically, our design allows for (i) targeted retrieval of a specific or pools of gRNA targets from the DNA array oligonucleotide pool (through 3 sequential rounds of nested PCR as indicated in the figure schematic); and (ii) its rapid cloning into a common expression vector which upon linearization using an AflIII site serves as a recipient for Gibson assembly mediated incorporation of the gRNA insert fragment. (B) We confirmed this methodology by targeted retrieval of 10 unique gRNAs from a 12k oligonucleotide pool synthesized by CustomArray Inc. (refer Methods; list in Supplementary Table 3).
References


