

*Supplementary Materials*

# **RNA Editing with CRISPR-Cas13**

David B.T. Cox<sup>1,2,3,4,5,6</sup>\*, Jonathan S. Gootenberg<sup>1,2,3,4,7</sup>\*, Omar O. Abudayyeh<sup>1,2,3,4,6</sup>\*, Brian Franklin<sup>1,2,3,4</sup>, Max J. Kellner<sup>1,2,3,4</sup>, Julia Joung<sup>1,2,3,4</sup>, Feng Zhang<sup>1,2,3,4†</sup>

> <sup>1</sup> Broad Institute of MIT and Harvard Cambridge, MA 02142, USA

<sup>2</sup> McGovern Institute for Brain Research <sup>3</sup> Department of Brain and Cognitive Science <sup>4</sup> Department of Biological Engineering <sup>5</sup> Department of Biology <sup>6</sup> Harvard-MIT Division of Health Sciences and Technology Massachusetts Institute of Technology Cambridge, MA 02139, USA

> <sup>7</sup> Department of Systems Biology Harvard Medical School Boston, MA 02115, USA

\* These authors contributed equally to this work.

<sup>†</sup> Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

### **SUPPLEMENTARY TEXT**

As the knowledge of the protospacer flanking site (PFS) may be necessary for effective Cas13 targeting, we first sought to define (PFS) preferences for the recently described Cas13b family of RNases (*12, 13*). We heterologously expressed 15 Cas13b orthologs in *E. coli* and measured interference activity using an ampicillin resistance assay (fig. S1A, B). Sequencing of colonies revealed a mixture of PFS preferences, with Cas13b orthologs having either solely 5' PFS preferences or a dual 5' and 3' PFS (fig. S1C).

#### **SUPPLEMENTARY METHODS**

### Design and cloning of bacterial constructs

Mammalian codon optimized Cas13b constructs were cloned into the chloramphenicol resistant pACYC184 vector under control of the Lac promoter. Two corresponding direct-repeat (DR) sequences separated by BsaI restriction sites were then inserted downstream of Cas13b, under control of the pJ23119 promoter. Last, oligos for targeting spacers were phosphorylated using T4 PNK (New England Biolabs), annealed and ligated into BsaI digested vectors using T7 ligase (Enzymatics) to generate targeting Cas13b vectors. Guide sequences used are in Supplementary Table 6.

#### Bacterial PFS screens

Ampicillin resistance plasmids for PFS screens were cloned by inserting PCR products containing Cas13b targets with two 5' randomized nucleotides and four 3' randomized nucleotides separated by a target site immediately downstream of the start codon of the ampicillin resistance gene *bla* using NEB Gibson Assembly (New England Biolabs). 100 ng of ampicillin-resistant target plasmids were then electroporated with 65-100 ng chloramphenicolresistant Cas13b bacterial targeting plasmids into Endura Electrocompetent Cells (Lucigen). Plasmids were added to cells, incubated for 15 minutes on ice, electroporated using the manufacturer's recommended settings, and then 950 uL of recovery media was added to cells

before a one-hour outgrowth at 37° C. The outgrowth was plated onto chloramphenicol and ampicillin double selection plates. Serial dilutions of the outgrowth were used to estimate the cfu/ng DNA. 16 hours post plating, cells were scraped off plates and surviving plasmid DNA was harvested using the Qiagen Plasmid Plus Maxi Kit (Qiagen). Surviving Cas13b target sequences and their flanking regions were amplified by PCR and sequenced using an Illumina NextSeq. To assess PFS preferences, the positions containing randomized nucleotides in the original library were extracted, and sequences depleted relative to the vector only condition and that were present in both bioreplicates were extracted using custom python scripts. The -log<sub>2</sub> of the ratio of PFS abundance in the Cas13b condition compared to the vector only control was then used to calculate preferred motifs. Specifically, all sequences having  $-log_2(sample/vector)$ depletion ratios above a specific threshold were used to generate weblogos of sequence motifs (weblogo.berkeley.edu). The specific depletion ratio values used to generate weblogos for each Cas13b ortholog are listed in Supplementary table 2.

### Design and cloning of mammalian constructs for RNA interference

To generate vectors for testing Cas13 orthologs in mammalian cells, mammalian codon optimized Cas13a, Cas13b, and Cas13c genes were PCR amplified and golden-gate cloned into a mammalian expression vector containing dual NLS sequences and a C-terminal msfGFP, under control of the EF1alpha promoter. For further optimization Cas13 orthologs were golden-gate cloned into destination vectors containing different C-terminal localization tags under control of the EF1alpha promoter.

The dual luciferase reporter was cloned by PCR amplifying *Gaussia* and *Cypridinia* luciferase coding DNA, the EF1alpha and CMV promoters and assembled using the NEB Gibson Assembly (New England Biolabs).

For expression of mammalian guide RNAs for Cas13a, Cas13b, or Cas13c orthologs, the corresponding direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into the corresponding expression backbones for each ortholog by golden-gate cloning. All Cas13 plasmids are listed in Supplementary Table 5. All Cas13 guide sequences for knockdown experiments are listed in Supplementary Tables 6-8.

### Measurement of Cas13 expression in mammalian cells

Dual-NLS Cas13-msfGFP constructs were transfected into HEK293FT cells with targeting and non-targeting guides. GFP fluorescence was measured 48 hours post transfection in the nontargeting guide condition using a plate reader.

## Cloning of pooled mismatch libraries for Cas13 interference specificity

Pooled mismatch library target sites were created by PCR using a forward primer containing the semi-degenerate target sequences and a constant reverse primer off of a *Gluc* template. The semi-degenerate forward oligo had at each position of the Cas13 target, plus the 5' and 3' three flanking bases, a nucleotide mixture containing 94% of the correct base and 2% of each incorrect base. The mismatch library amplicon was then cloned into the dual luciferase reporter in place of wild-type *Gluc* using NEB Gibson assembly (New England Biolabs).

## Design and cloning of mammalian constructs for RNA editing

PspCas13b was made catalytically inactive (dPspCas13b) via two histidine to alanine mutations (H133A/H1058A) at the catalytic site of the HEPN domains. The deaminase domains of human ADAR1 and ADAR2 were synthesized and PCR amplified for Gibson cloning into pcDNA-CMV vector backbones and were fused to dPspCas13b at the C-terminus via GS or GSGGGGS linkers. For the experiment in which we tested different linkers we cloned the following additional linkers between dPspCas13b and  $ADAR2<sub>DD</sub>$ : GGGGSGGGGSGGGGS, EAAAK, GGSGGSGGSGGSGGSGGS, and SGSETPGTSESATPES (XTEN). Specificity mutants were generated by Gibson cloning the appropriate mutants into the dPspCas13b-GSGGGGS backbone.

The luciferase reporter vector for measuring RNA editing activity was generated by creating a W85X mutation (TGG>TAG) in the luciferase reporter plasmid used for knockdown experiments. This reporter vector expresses functional *Gluc* as a normalization control, but a defective *Cluc* due to the addition of the W85X pretermination site. To test ADAR editing motif preferences, we cloned every possible motif around the adenosine at codon 85 (XAX) of *Cluc*. All plasmids are listed in Supplementary Table 5.

### Testing PFS preferences for dCas13b

For testing PFS preference of REPAIR, we cloned a pooled plasmid library containing a 6 basepair degenerate PFS sequence upstream of a target region and adenosine editing site. The library was synthesized as an ultramer from Integrated DNA Technologies (IDT) and was made double stranded via annealing a primer and using the Klenow fragment of DNA polymerase I (New England Biolabs) to fill in the sequence. This dsDNA fragment containing the degenerate sequence was then Gibson cloned into the digested reporter vector and this was then isopropanol precipitated and purified. The cloned library was then electroporated into Endura competent *E. coli* cells (Lucigen) and plated on 245mm x 245mm square bioassay plates (Nunc). After 16 hours, colonies were harvested and midiprepped using endotoxin-free MACHEREY-NAGEL midiprep kits. Cloned libraries were verified by next-generation sequencing.

#### Cloning pathogenic G>A mutations for assaying REPAIR activity

For cloning disease-relevant mutations for testing REPAIR activity, 34 G>A mutations related to disease pathogenesis as defined in ClinVar were selected and 200-bp regions surrounding these mutations were golden-gate cloned between mScarlett and EGFP under a CMV promoter. Two additional G>A patient mutations in *AVPR2* and *FANCC* and their cDNA sequences were synthesized and Gibson cloned under expression of EF1alpha.

#### Guide cloning for REPAIR

For expression of mammalian guide RNAs for REPAIR, the PspCas13b direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into this expression backbone by golden-gate cloning. Guide sequences for REPAIR experiments are listed in Supplementary Table 9.

## Mammalian cell culture

Mammalian cell culture experiments were performed in the HEK293FT line (American Type Culture Collection (ATCC)), which was grown in Dulbecco's Modified Eagle Medium with high glucose, sodium pyruvate, and GlutaMAX (Thermo Fisher Scientific), additionally supplemented with  $1\times$  penicillin–streptomycin (Thermo Fisher Scientific) and 10% fetal bovine serum (VWR Seradigm). Cells were maintained at confluency below 80%. The U2OS specificity experiment was performed using the U2OS cell line from ATCC and cells were cultured in ATCCformulated McCoy's 5a Medium Modified.

Unless otherwise noted, all transfections were performed with Lipofectamine 2000 (Thermo Fisher Scientific) in 96-well plates coated with poly-D-lysine (BD Biocoat). Cells were plated at approximately 20,000 cells/well 16 hours prior to transfection to ensure 90% confluency at the time of transfection. For each well on the plate, transfection plasmids were combined with Opti-MEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 µl. Separately, 24.5 µl of Opti-MEM was combined with 0.5 µl of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and incubated for 5 minutes, after which they were pipetted onto cells. The U2OS transfections were performed using Lipofectamine 3000 according to the manufacturer's protocol.

### Mammalian cell RNA knockdown assays

To assess RNA targeting in mammalian cells with reporter constructs, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid and 12.5 ng of the knockdown reporter construct. 48 hours post-transfection, media containing secreted luciferase was removed from cells, diluted 1:5 in PBS, and measured for activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

For targeting of endogenous genes, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid. 48 hours post-transfection, cells were lysed and RNA was harvested and reverse transcribed using a previously described(*33*) modification of the Cells-to-Ct kit (Thermo Fisher Scientific). cDNA expression was measured via qPCR using TaqMan qPCR probes for the *KRAS* transcript (Thermo Fisher Scientific), *GAPDH* control probes (Thermo Fisher Scientific), and Fast Advanced Master Mix (Thermo Fisher Scientific). qPCR reactions were read out on a LightCycler 480 Instrument II (Roche), with four 5 μl technical replicates in 384-well format.

#### Evaluation of RNA specificity using pooled libraries of mismatched targets

The ability of Cas13 to interfere with the mismatched target library was tested using HEK293FT cells seeded in 6-well plates. ~70% confluent cells were transfected using 2400 ng Cas13 vector, 4800 ng of guide, and 240 ng of mismatched target library. 48 hours post-transfection, cells were harvested and RNA was extracted using the QIAshredder (Qiagen) and the Qiagen RNeasy Mini Kit. 1 μg of extracted RNA was reverse transcribed using the qScript Flex cDNA synthesis kit (Quantabio) following the manufacturer's gene-specific priming protocol with a *Gluc* specific RT primer. cDNA was then amplified and sequenced on an Illumina NextSeq.

Sequencing was analyzed by counting reads per sequence and depletion scores were calculated by determining the  $log_2$ -read count ratio) value, where read count ratio is the ratio of read counts in the targeting guide condition versus the non-targeting guide condition. This score represents the level of Cas13 activity on the sequence, with higher values representing stronger depletion and thus higher Cas13 cleavage activity. Separate distributions for the single mismatch and double mismatch sequences were determined and plotted as heatmaps with a depletion score for each mismatch identity. For double mismatch sequences the average of all possible double mismatches at a given position were plotted.

#### Transcriptome-wide profiling of Cas13 in mammalian cells by RNA sequencing

For measurement of transcriptome-wide specificity, 150 ng of Cas13 construct, 300 ng of guide expression plasmid, and 15 ng of the knockdown reporter construct were co-transfected; for shRNA conditions, 300 ng of shRNA targeting plasmid, 15 ng of the knockdown reporter construct, and 150 ng of EF1-alpha driven mCherry (to balance reporter load) were cotransfected. 48 hours post-transfection, RNA was purified with the RNeasy Plus Mini kit (Qiagen), mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs), and prepared for sequencing with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs). RNA sequencing libraries were then sequenced on a NextSeq (Illumina).

To analyze transcriptome-wide sequencing data, reads were aligned to the RefSeq GRCh38 assembly using Bowtie and RSEM version 1.2.31 with default parameters(*34*). Transcript expression was quantified as  $log_2(TPM + 1)$ , genes were filtered for  $log_2(TPM + 1) > 2.5$ . For selection of differentially expressed genes, only genes with differential changes of  $>2$  or  $< .75$ were considered. Statistical significance of differential expression was evaluated using a Student's t-test on three targeting replicates versus non-targeting replicates, and filtered for a false discovery rate of <0.01% by the Benjamini-Hochberg procedure.

### REPAIR editing in mammalian cells

To assess REPAIR activity in mammalian cells, we transfected 150 ng of REPAIR vector, 300 ng of guide expression plasmid, and 40 ng of the RNA editing reporter. After 48 hours, RNA from cells was harvested and reverse transcribed using a method previously described(*33*) with a gene specific reverse transcription primer. The extracted cDNA was then subjected to two rounds of PCR to add Illumina adaptors and sample barcodes using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs). The library was then subjected to next generation sequencing on an Illumina NextSeq or MiSeq. RNA editing rates were then evaluated at all adenosines within the sequencing window.

In experiments where the luciferase reporter was targeted for RNA editing, we also harvested the media with secreted luciferase prior to RNA harvest. In this case, because corrected *Cluc* might be at low levels, we did not dilute the media. We measured luciferase activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

#### PFS binding mammalian screen

To determine the contribution of the PFS to editing efficiency in mammalian cells, 625 ng of PFS target library, 4.7 μg of guide, and 2.35 μg of REPAIR were co-transfected in HEK293FT cells plated in 25 cm<sup>2</sup> flasks. Plasmids were mixed with 33  $\mu$ l of PLUS reagent (Thermo Fisher Scientific), brought to 533 μl with Opti-MEM, incubated for 5 minutes, combined with 30 μl of Lipofectamine 2000 and 500 μl of Opti-MEM, incubated for an additional 5 minutes, and then pipetted onto cells. 48 hours post-transfection, RNA was harvested with the RNeasy Plus Mini kit (Qiagen), reverse transcribed with qScript Flex (Quantabio) using a gene specific primer, and amplified with two rounds of PCR using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) to add Illumina adaptors and sample barcodes. The library was sequenced on an Illumina NextSeq, and RNA editing rates at the target adenosine were mapped to PFS identity. To increase coverage, the PFS was computationally collapsed to 4 nucleotides adjacent to the 5' end of the target sequence. REPAIR editing rates were calculated for each PFS, averaged over biological replicates with non-targeting rates for the corresponding PFS subtracted.

### Whole-transcriptome sequencing to evaluate ADAR editing specificity

For analyzing off-target RNA editing sites across the transcriptome, we harvested total RNA from cells 48 hours post-transfection using the RNeasy Plus Miniprep kit (Qiagen). The mRNA fraction was then enriched using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and this RNA was then prepared for sequencing using an NEBNext Ultra RNA Library Prep Kit for Illumina (NEB). The libraries were then sequenced on an Illumina NextSeq and loaded such that there were at least 5 million reads per sample.

## RNA editing analysis for targeted and transcriptome-wide experiments

Analysis of the transcriptome-wide editing RNA sequencing data was performed on the FireCloud computational framework (https://software.broadinstitute.org/firecloud/) using a custom workflow we developed:

https://portal.firecloud.org/#methods/m/rna\_editing\_final\_workflow/rna\_editing\_final\_workflow /1. For analysis, unless otherwise denoted, sequence files were randomly downsampled to 5 million reads. For the high-coverage sequencing analysis, samples were randomly downsampled to 5 million, 15 million, or 50 million reads. An index was generated using the RefSeq GRCh38 assembly with *Gluc* and *Cluc* sequences added, and reads were aligned and quantified using Bowtie/RSEM version 1.3.0. Alignment BAMs were then sorted and analyzed for RNA editing sites using REDitools (*35, 36*) with the following parameters: -t 8 -e -d -l -U [AG or TC] -p -u m20 -T6-0 -W -v 1 -n 0.0. Any significant edits found in untransfected or EGFP-transfected conditions were considered to be SNPs or artifacts of the transfection and filtered out from the analysis of off-targets. Off-targets were considered significant if the Fisher's exact test yielded a p-value less than 0.05 after multiple hypothesis correction by Benjamini Hochberg correction and at least 2 of 3 biological replicates identified the edit site. Overlap of edits between samples was calculated relative to the maximum possible overlap, equivalent to the fewer number of edits between the two samples. The percentage of overlapping edit sites was calculated as the number of shared edit sites divided by minimum number of edits of the two samples, multiplied by 100. For the high-coverage sequencing analysis, an additional layer of filtering for known SNP positions was performed using the Kaviar (*37*) method for identifying SNPs.

For analyzing the predicted variant effects of each off-target, the list of off-target edit sites was analyzed using the variant annotation integrator (https://genome.ucsc.edu/cgi-bin/hgVai) as part of the UCSC genome browser suite of tools using the SIFT and PolyPhen-2 annotations. To predict whether the off-target genes are oncogenic, a database of oncogenic annotations from the COSMIC catalogue of somatic mutations in cancer was used to characterize off-target genes (cancer.sanger.ac.uk).

For analyzing whether the REPAIR constructs perturbed RNA levels, the transcript per million (TPM) values output from the RSEM analysis were used for expression counts and transformed to log-space by taking the  $log_2(TPM+1)$ . To find differentially regulated genes, a Student's t-test was performed on three targeting guide replicates versus three non-targeting guide replicates. The statistical analysis was only performed on genes with  $log_2(TPM+1)$  values greater than 2.5 and genes were only considered differentially regulated if they had a fold change greater than 2 or less than 0.8. Genes were reported if they had a false discovery rate (Benjamini Hochberg correction) of less than 0.01.

## **SUPPLEMENTARY FIGURES**

**Figure S1**



## **Figure S1: Bacterial screening of Cas13b orthologs for** *in vivo* **efficiency and PFS determination.**

- A) Schematic of bacterial assay for determining the PFS of Cas13b orthologs. Cas13b orthologs with beta-lactamase targeting spacers are co-transformed with beta-lactamase expression plasmids containing randomized PFS sequences and subjected to dual antibiotic selection. PFS sequences that are depleted during co-transformation with Cas13b suggest targeting activity and are used to infer PFS preferences.
- B) Quantification of interference activity of Cas13b orthologs targeting beta-lactamase as measured by colony forming units (cfu). Values represent mean +/– S.D.
- C) PFS weblogos for Cas13b orthologs as determined by depleted sequences from the bacterial assay. PFS preferences are derived from sequences depleted in the Cas13b condition relative to empty vector controls. Depletion values used to calculate PFS weblogos are listed in table S2.





## **Figure S2: Relative expression of Cas13 orthologs in mammalian cells and correlation of expression with interference activity.**

- A) Expression of Cas13 orthologs as measured by msfGFP fluoresence. Cas13 orthologs Cterminally tagged with msfGFP were transfected into HEK293FT cells and their fluorescence measured 48 hours post transfection.
- B) Correlation of Cas13 expression to interference activity. The average RLU of two *Gluc* targeting guides for Cas13 orthologs, separated by subfamily, is plotted versus expression as determined by msfGFP fluoresence. The RLU for targeting guides are normalized to RLU for a non-targeting guide, whose value is set to 1. The non-targeting guide is the same as in Figure 1B for Cas13b.



## **Figure S3: Optimization of Cas13b knockdown and further characterization of mismatch specificity.**

- A) Gluc knockdown with two different guides is measured using the top two Cas13a and top four Cas13b orthologs fused to a variety of C-terminal nuclear localization and nuclear export tags.
- B) Knockdown of *KRAS* is measured for LwaCas13a, RanCas13b, PguCas13b, PspCas13b and shRNA with four position-matched guides. Non-targeting guide is the same as in Figure 1B. shRNA non-targeting guide sequence is listed in table S6.
- C) Schematic of the single and double mismatch plasmid libraries used for evaluating the specificity of LwaCas13a and PspCas13b knockdown. Every possible single and double mismatch is present in the target sequence as well as in three positions directly flanking the 5' and 3' ends of the target site.
- D) The depletion levels of transcripts with the indicated single mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. The wildtype base is outlined by a green box.
- E) The depletion levels of transcripts with the indicated double mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. Each box represents the average of all possible double mismatches for the indicated position.



## **Figure S4: Characterization of design parameters for REPAIRv1.**

- A) Knockdown efficiency of Gluc with wild-type Cas13b or catalytically inactive H133A/H1058A Cas13b (dCas13b).
- B) Quantification of luciferase activity restoration by dCas13b fused to either the wild-type

ADAR2 deaminase domain (ADAR2<sub>DD)</sub> or the hyperactive E488Q mutant ADAR2<sub>DD</sub>(E488Q) deaminase domain, tested with tiling *Cluc* targeting guides.

- C) Guide design and sequencing quantification of A to I editing for 30-nt guides targeting *Cluc* W85X.
- D) Guide design and sequencing quantification of A to I editing for 50-nt guides targeting *PPIB*.
- E) Influence of linker choice on luciferase activity restoration by REPAIRv1. Values represent mean  $+/-$  S.E.M.



### **Figure S5: Comparison of RNA editing activity of dCas13b and REPAIRv1.**

A) Schematic of guides used to target the W85X mutation in the *Cluc* reporter.

- B) Sequencing quantification of A to I editing for indicated guides transfected with dCas13b. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Nontargeting guide is the same as in Fig2C.
- C) Sequencing quantification of A to I editing for indicated guides transfected with REPAIRv1. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Nontargeting guide is the same as in Fig2C.
- D) Comparison of on-target A to I editing rates for dCas13b and dCas13b-ADAR2<sub>DD</sub>(E488Q) for guides tested in panel B and C.
- E) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1. Values represent mean +/– S.E.M.





## **Figure S6: ClinVar motif distribution for G>A mutations.**

The number of each possible triplet motif observed in the ClinVar database for all G>A mutations.



## **Figure S7: Truncations of dCas13b support functional RNA editing.**

N-terminal and C-terminal truncations of dCas13b allow for RNA editing as measured by restoration of luciferase signal for the *Cluc* W85X reporter. Values represent mean +/– S.E.M. The construct length refers to the coding sequence of the REPAIR constructs.



## **Figure S8: REPAIRv1 editing activity evaluated without a guide and in comparison to ADAR2 deaminase domain alone.**

- A) Quantification of A to I editing of the *Cluc* W85X mutation by REPAIRv1 with and without guide as well as the ADAR2 deaminase domain only without guide. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig2C.
- B) Number of differentially expressed genes in the REPAIRv1 and  $ADAR2<sub>DD</sub>$  conditions from panel A.
- C) The number of significant off-targets from the REPAIRv1 and  $ADAR2<sub>DD</sub>$  conditions from panel A.
- D) Overlap of off-target A to I editing events between the REPAIRv1 and  $ADAR2<sub>DD</sub>$  conditions from panel A. The values plotted are the percent of the maximum possible intersection of the two offtarget data sets.





#### **Figure S9: Comparison of REPAIRv1 to other programmable ADAR systems.**

A) Schematic of two programmable ADAR schemes: BoxB-based targeting (top) and full length ADAR2 targeting (bottom). For BoxB-based targeting,  $ADAR<sub>po</sub>(E488Q)$  is fused to the viral protein lambda  $N$  (BoxB- $\lambda$ ), and the fusion protein is recruited to target adenosines by a guide RNA containing homology to the target site and hairpins that BoxB- $\lambda$  binds to. Full length ADAR2 targeting utilizes a guide RNA with homology to the target site and a motif recognized by the double strand RNA binding domains of ADAR2.

- B) Transcriptome-wide sites of significant RNA editing by BoxB-ADAR2  $_{\text{DD}}(E488Q)$  with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange.
- C) Transcriptome-wide sites of significant RNA editing by full length ADAR2 with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange.
- D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange. The nontargeting guide is the same as in Fig2C.
- E) Quantification of on-target editing rate percentage for BoxB-ADAR2  $_{\text{DD}}(E488Q)$ , ADAR2, and REPAIRv1 for targeting guides against *Cluc.*
- F) Overlap of off-target sites between different targeting and non-targeting conditions for programmable ADAR systems. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.



## Figure S10: Efficiency and specificity of dCas13b-ADAR2<sub>DD</sub>(E488Q) mutants.

- A) Quantification of luciferase activity restoration by dCas13b-ADAR2<sub>DD</sub>(E488Q) mutants for *Cluc*targeting and non-targeting guides. Non-targeting guide is the same as in Fig2C.
- B) Relationship between the ratio of targeting and non-targeting guide RLU and the number of RNA-editing off-targets as quantified by transcriptome-wide sequencing
- C) Quantification of transcriptome-wide off-target RNA editing sites versus on-target *Cluc* editing efficiency for  $dCas13b-ADAR2_{\text{DD}}(E488Q)$  mutants.



Figure S11: Transcriptome-wide specificity of RNA editing by dCas13b-ADAR2 <sub>DD</sub>(E488Q).

- A) Transcriptome-wide sites of significant RNA editing by  $dCas13b-ADAR2_{\text{nn}}(E488Q)$  mutants with a guide targeting *Cluc*. The on-target *Cluc* site (254 A>I) is highlighted in orange.
- B) Transcriptome-wide sites of significant RNA editing by  $dCas13b-ADAR2_{po}(E488Q)$  mutants with a non-targeting guide.



## **Figure S12: Characterization of motif biases in the off-targets of dCas13b-ADAR2 DD(E488Q) editing.**

- A) For each dCas13b-ADAR2  $_{\text{DD}}(E488Q)$  mutant, the motif present across all A>I off-target edits in the transcriptome is shown.
- B) The distribution of off-target A $>$ I edits per motif identity is shown for REPAIRv1 with targeting

and non-targeting guide.

C) The distribution of off-target A>I edits per motif identity is shown for REPAIRv2 with targeting and non-targeting guide.

**Figure S13**



## **Figure S13: Further characterization of REPAIRv1 and REPAIRv2 off-targets.**

A) Histogram of the number of off-targets per transcript for REPAIRv1.

- B) Histogram of the number of off-targets per transcript for REPAIRv2.
- C) Variant effect prediction of REPAIRv1 off targets.
- D) Distribution of REPAIRv1 off targets in cancer-related genes. TSG, tumor suppressor gene.
- E) Variant effect prediction of REPAIRv2 off targets.
- F) Distribution of REPAIRv2 off targets in cancer-related genes.



### **Figure S14: Evaluation of off-target sequence similarity to the guide sequence.**

- A) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv1 with a Cluc targeting guide.
- B) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv2 with a Cluc targeting guide.



## **Figure S15: Comparison of REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting at two different doses of vector (150ng and 10ng effector).**

- A) Quantification of RNA editing activity at the *Cluc* W85X (254 A>I) on-target editing site by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting or non-targeting guide. Values shown are the mean of the three replicates.
- B) Quantification of RNA editing off-targets by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting guide for the *Cluc* W85X (254 A>I) site or non-targeting guide. For REPAIR constructs, nontargeting guide is the same as in Fig. 2C.



## **Figure S16: RNA editing efficiency and genome-wide specificity of REPAIRv1 and REPAIRv2.**

- A) Quantification of RNA editing activity at the *PPIB* guide 1 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean  $+/-$  S.E.M.
- B) Quantification of RNA editing activity at the *PPIB* guide 2 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M.
- C) Quantification of RNA editing off-targets by REPAIRv1 or REPAIRv2 with *PPIB* guide 1, *PPIB*

guide 2, or non-targeting guide.

D) Overlap of off-targets between REPAIRv1 for *PPIB* targeting, Cluc targeting, and non-targeting guides. The values plotted are the percent of the maximum possible intersection of the two offtarget data sets.





**Figure S17: High coverage sequencing of REPAIRv1 and REPAIRv2 off-targets.** 

- A) Quantitation of off-target edits for REPAIRv1 and REPAIRv2 as a function of read depth with a total of 5 million reads (12.5x coverage), 15 million reads (37.5x coverage) and 50 million reads (125x coverage) per condition.
- B) Overlap of off-target sites at different read depths of the following conditions: REPAIRv1 versus REPAIRv1 (left), REPAIRv2 versus REPAIRv2 (middle), and REPAIRv1 versus REPAIRv2 (right). The values plotted are the percent of the maximum possible intersection of the two offtarget data sets.
- C) Editing rate of off-target sites compared to the coverage (log2(number of reads)) of the off-target for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.
- D) Editing rate of off-target sites compared to the log2(TPM+1) of the off-target gene expression for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.



## **Figure S18: Quantification of REPAIRv2 activity and off-targets in the U2OS cell line.**

- A) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a guide targeting *Cluc*  in the U2OS cell line. The on-target *Cluc* site (254 A>I) is highlighted in orange.
- B) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a non-targeting guide in the U2OS cell line.
- C) The on-target editing rate at the *Cluc* W85X (254 A>I) by REPAIRv2 with a targeting guide or non-targeting guide in the U2OS cell line.
- D) Quantification of off-targets by REPAIRv2 with a guide targeting *Cluc* or non-targeting guide in the U2OS cell line.



### **Figure S19: RNA editing efficiency and specificity of REPAIRv1 and REPAIRv2.**

- A) Quantification of percent editing of *KRAS* with *KRAS-*targeting guide 1 at the targeted adenosine (blue triangle) and neighboring sites for REPAIRv1 and REPAIRv2. For each guide, the region of duplex RNA is outlined in red. Values represent mean  $+/-$  S.E.M. Non-targeting guide is the same as in Fig. 2C.
- B) Quantification of percent editing of *KRAS* with *KRAS-*targeting guide 3 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 2C.
- C) Quantification of percent editing of *PPIB* with *PPIB-*targeting guide 2 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 2C.



## **Figure S20: Demonstration of all potential codon changes with an A>I RNA editor.**

- A) Table of all potential codon transitions enabled by A>I editing.
- B) A codon table demonstrating all the potential codon transitions enabled by A>I editing. Adapted and modified based on (*38*).
- C) Model of REPAIR A to I editing of a precisely encoded nucleotide via a mismatch in the guide sequence. The A to I transition is mediated by the catalytic activity of the ADAR2 deaminase

domain and will be read as a guanosine by translational machinery. The base change does not rely on endogenous repair machinery and is permanent for as long as the RNA molecule exists in the cell.

- D) REPAIR can be used for correction of Mendelian disease mutations.
- E) REPAIR can be used for multiplexed A to I editing of multiple variants for engineering pathways or modifying disease. Multiplexed guide delivery can be achieved by delivering a single CRISPR array expression cassette since the Cas13b enzyme processes its own array.
- F) REPAIR can be used for modifying protein function through amino acid changes that affect enzyme domains, such as kinases.
- G) REPAIR can modulate splicing of transcripts by modifying the splice acceptor site.



# **Supplementary Table 1: Cas13 Orthologs used in this study.**



## **Supplementary Table 2: PFS cutoffs in bacterial screens**



**Supplementary Table 3: dCas13b-ADAR linker sequences used in this study for RNA editing in mammalian cells.**





## **Supplementary Table 4: Disease information for disease-relevant mutations**





## **Supplementary Table 5: Key plasmids used in this study**





## **Supplementary Table 6: Guide/shRNA sequences used in this study for knockdown in mammalian cells**







## **Supplementary Table 7: Guide sequences used for** *Gluc* **knockdown**









# **Supplementary Table 8: Guide sequences used for** *Cluc* **knockdown**





## **Supplementary Table 9: Guide sequences used in this study for RNA editing in mammalian cells.** Mismatched base flips are capitalized



























- 31. Y. Li *et al.*, Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J Invest Dermatol* **130**, 2768-2772 (2010).
- 32. R. C. Ferreira *et al.*, Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. *Nat Genet* **42**, 777-780 (2010).
- 33. J. Joung *et al.*, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).
- 34. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- 35. E. Picardi, A. M. D'Erchia, A. Montalvo, G. Pesole, Using REDItools to Detect RNA Editing Events in NGS Datasets. *Curr Protoc Bioinformatics* **49**, 12 12 11-15 (2015).
- 36. E. Picardi, G. Pesole, REDItools: high-throughput RNA editing detection made easy. *Bioinformatics* **29**, 1813-1814 (2013).
- 37. G. Glusman, J. Caballero, D. E. Mauldin, L. Hood, J. C. Roach, Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216-3217 (2011).
- 38. J. D. Watson, *Molecular biology of the gene*. (Pearson, Boston, ed. Seventh edition, 2014), pp. xxxiv, 872 pages.