nature biotechnology

BRIEF COMMUNICATION

https://doi.org/10.1038/s41587-021-01030-2

Check for updates

Compact RNA editors with small Cas13 proteins

Soumya Kannan[1,2,3,4,5,9], Han Altae-Tran[1,2,3,4,5,9], Xin Jin[1,2,3,4,5,6,7], Victoria J. Madigan[1,2,3,4,5], Rachel Oshiro[1,2,3,4,5], Kira S. Makarova[8], Eugene V. Koonin[8] and Feng Zhang[1,2,3,4,5] ✉

CRISPR–Cas13 systems have been developed for precise RNA editing, and can potentially be used therapeutically when temporary changes are desirable or when DNA editing is challenging. We have identified and characterized an ultrasmall family of Cas13b proteins—Cas13bt—that can mediate mammalian transcript knockdown. We have engineered compact variants of REPAIR and RESCUE RNA editors by functionalizing Cas13bt with adenosine and cytosine deaminase domains, and demonstrated packaging of the editors within a single adeno-associated virus.

RNA-targeting CRISPR–Cas13 systems have been harnessed for a variety of applications[1], including programmable RNA editing[2,3]. RNA editing is a promising therapeutic strategy that allows for installation of temporary, non-heritable edits. However, therapeutic delivery of Cas13-based RNA editing systems remains challenging, in part because the size of Cas13-based RNA editors developed so far exceeds the packaging capacity of adeno-associated virus (AAV), the most widely used viral vector for gene delivery[4,5].

To overcome this limitation, we performed an iterative HMM profile search of Cas13 enzymes in prokaryotic and viral genomes and metagenomes, identifying 5,843 candidate systems. Phylogenetic analysis revealed novel groups of ultracompact Cas13 proteins that form distinct branches within the Cas13b and Cas13c subtypes (Fig. 1a,b and Supplementary Fig. 1a,b), hereafter referred to as Cas13bt and Cas13ct, respectively. Unlike other CRISPR–Cas13b systems[6], the genomic loci encoding the Cas13bt subfamily lack any accessory genes (Fig. 1c). Relative to BzoCas13b (1,224 amino acids (aa)), the smallest Cas13bt proteins (775–804 aa) have 26 large (>5 aa) deletions that total 408 aa (Supplementary Fig. 1c). As Cas13b proteins are more active than Cas13c proteins in mammalian systems and support programmable RNA editing[2], we focused our analysis on a group of 16 ultracompact Cas13bt proteins (Cas13bt1 to Cas13bt16; Supplementary Table 1).

To experimentally characterize Cas13bt, we first identified the required CRISPR RNA (crRNA) components. We transformed *Escherichia coli* with a plasmid containing one of the *cas13bt* loci, *cas13bt2*, with its CRISPR array truncated to two direct repeats (DRs) and performed small RNA sequencing to determine the configuration of the mature crRNA. As for previously characterized Cas13b proteins, the crRNA of Cas13bt2 was also found to have a 3′ DR (Fig. 1d). To determine whether Cas13bt proteins are also capable of mediating crRNA-guided RNA targeting, we performed an RNA interference screen using a library of crRNAs that were programmed to target essential gene transcripts in *E. coli* (Fig. 1e, Supplementary Fig. 2a)[6]. Two of the three tested members of the Cas13bt subfamily (Supplementary Tables 2 and 3), Cas13bt1 and Cas13bt3, mediated depletion of targeting spacers in *E. coli* (Fig. 1f and Supplementary Fig. 2b–d). Analysis of the sequences flanking sites targeted by depleted crRNAs revealed that both Cas13bt1 and Cas13bt3 have a permissive 5′ D (A/G/T) protospacer flanking sequence (PFS) preference (Fig. 1f and Supplementary Fig. 3). Additionally, crRNAs targeting the 5′ untranslated region and beginning of the coding sequence (CDS) were more depleted (Fig. 1f).

Cas13 enzymes have previously been reported to exhibit collateral RNA cleavage activity upon crRNA-guided binding of their single-stranded RNA target[7]. In vitro evaluation of Cas13bt3 showed that Cas13bt also performs target-activated collateral RNA cleavage, and that this collateral activity is mediated by the HEPN domains (Supplementary Fig. 4b). This target-activated collateral activity may render this subfamily amenable for use in diagnostic platforms such as SHERLOCK[8].

To evaluate the efficacy of Cas13bt-mediated knockdown of RNA in human cells, we tested Cas13bt1 and Cas13bt3 using a set of 20 crRNAs targeting a *Gaussia* luciferase (Gluc) mRNA. We found that both proteins promoted crRNA-guided Gluc knockdown in HEK293FT cells (Fig. 1g). Catalytically inactivating the HEPN domains of Cas13bt1 and Cas13bt3 abolished their RNA knockdown activity (Supplementary Fig. 4a,c). Further, we found that the PFS preference detected in *E. coli* was not required for crRNA-guided RNA knockdown in HEK293FT cells, similar to other Cas13 enzymes (Fig. 1g)[2]. Both Cas13bt1 and Cas13bt3 also mediated crRNA-guided knockdown of endogenous transcripts in HEK293FT cells (Fig. 1h and Supplementary Fig. 5).

To develop Cas13bt proteins for RNA editing, we fused catalytically inactive versions of Cas13bt1 and Cas13bt3 with a hyperactive mutant of the RNA adenosine deaminase ADAR2 catalytic domain (ADAR2dd-E188Q) to construct REPAIR.t1 and REPAIR.t3, respectively. We evaluated the ability of these programmable adenosine deaminases to revert a W85X mutation in the *Cypridina* luciferase (Cluc) mRNA by introducing a specific A-to-I mutation. Targeted RNA editing is achieved when ADAR2 deaminates a specific mismatched adenosine on the target RNA within the RNA duplex formed between the target RNA and the crRNA spacer (Fig. 2a)[2,9]. Previous studies have shown that the location of the mismatched target adenosine within the target RNA–crRNA duplex can dramatically affect the RNA editing efficiency[2,3]. To establish optimal parameters for crRNA design, we tested a range of mismatch positions within the target RNA–crRNA duplex. We found that for the crRNA used to target Cluc, both REPAIR.

[1]Howard Hughes Medical Institute, Cambridge, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]McGovern Institute for Brain Research at MIT, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Society of Fellows, Harvard University, Cambridge, MA, USA. [7]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [8]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. [9]These authors contributed equally: Soumya Kannan, Han Altae-Tran. ✉e-mail: zhang@broadinstitute.org

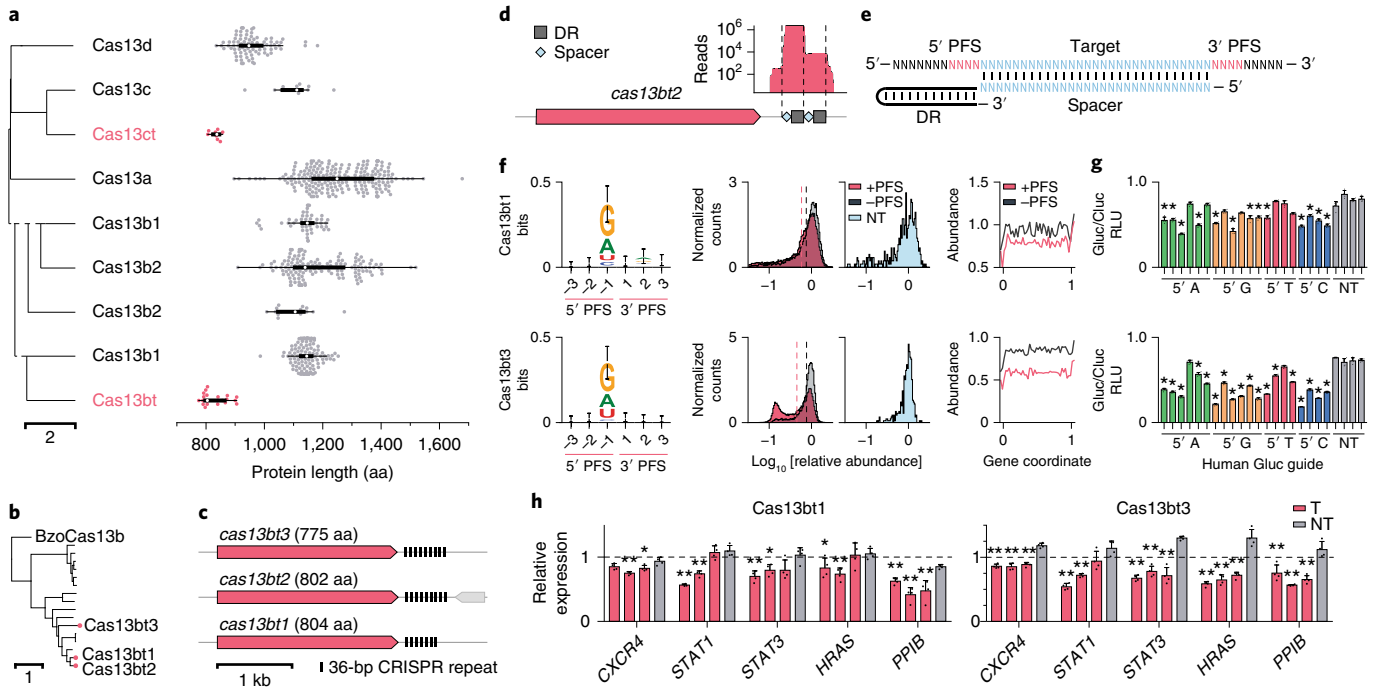


Fig. 1 | Cas13bt is a functional family of small Cas nucleases. a, Unweighted pair group method with arithmetic mean (UPGMA) dendrogram and protein size distribution of Cas13 variants (new subfamilies in red). Box plots show mean (white circle), lower and upper quartiles (thick line) and 1.5 times the interquartile range (thin line). From top to bottom $n = 87$, 12, 10, 223, 49, 152, 16, 123 and 22. **b**, Phylogenetic tree of Cas13bt proteins (experimentally studied proteins in red). **c**, *Cas13bt* locus organization. **d**, Cas13bt2 crRNA expression from *E. coli* heterologous expression. **e**, Schematic of PFS orientation relative to target sequence. **f**, *E. coli* essential gene screen shows Cas13bt1 (top) and Cas13bt3 (bottom) mediate interference with a weak 5′ D (A/G/T) PFS. Left: Weblogo of PFS preference based on top 1% of most depleted spacers (top $n = 304$, bottom $n = 313$). Total height corresponds to mean entropy; error bars as in ref. [15]. Middle: histograms displaying distribution of fold depletion ($\log_{10}$[relative abundance]) of both targeting and non-targeting spacers. Dashed lines represent the mean $\log_{10}$[relative abundance]. Right: line plots showing relative abundance in final library of spacers targeting regions across normalized positions in the target transcript. **g**, Evaluation of Cas13bt1 and Cas13bt3 for knockdown of luciferase reporter in HEK293FT cells. All values were normalized to a transfection control with crRNA alone. Data are presented as mean ± standard deviation (s.d.); $n = 4$. T, targeting crRNA; NT, non-targeting crRNA; RLU, relative luminescence units. $^*P < 0.05$ via two-tailed *t*-test incorporating Bonferroni correction. **h**, Quantitative PCR (qPCR) measurement of Cas13bt1- and Cas13bt3-mediated knockdown of endogenous transcripts in HEK293FT cells (mean ± s.d.; $n = 4$ biological replicates, each average of 4 technical replicates), normalized to a transfection control with crRNA alone via ddCt correction. Statistical significance was assessed using two-tailed *t*-tests compared to the negative control condition for each gene, respectively ($^*P < 0.05$, $^{**}P < 0.01$). See Supplementary Table 11 for *P* values where relevant.

t1 and REPAIR.t3 showed optimal editing when the mismatched target adenosine is positioned within 18–22 nucleotides of the 5′ end of the target site. Editing efficiency was comparable to that of the previously described REPAIRv1 and v2 systems, consisting of dPspCas13b fused to either ADAR2dd-E488Q or ADAR2-E488Q/T375G, respectively, and lower than that for dRanCas13b fused to ADAR2dd-E488Q (Fig. 2b)[2,3].

We additionally constructed cytosine RNA editors with an evolved ADAR2dd capable of cytidine deamination[3] (RESCUE.t1 and RESCUE.t3) and directed both editors to reporter and endogenous transcripts in HEK293FT cells. We found that these fusion proteins were capable of mediating C-to-U editing of all tested targets (Fig. 2c–f and Supplementary Fig. 6) and can edit when the mismatch is positioned between 14–28 nucleotides of the 5′ end of the target site.

To demonstrate the ability of Cas13bt-REPAIR to edit functionally relevant targets, we targeted RNA regions corresponding to previously characterized phosphorylation sites[3]. In particular, we attempted to alter activation of the Wnt–beta-catenin pathway by editing the T41 codon of *CTNNB1*, a site known to promote degradation of beta-catenin when phosphorylated[10]. We found that REPAIR.t1 achieved 40% editing at this site, converting the codon to alanine and leading to a 52-fold increase in beta-catenin signaling as measured by a beta-catenin-driven (TCF/LEF) luciferase reporter (see Methods), which may be relevant for applications such as promoting regeneration after acute liver failure[11,12] (Fig. 2c,e). REPAIR.t1 was also able to efficiently edit sites corresponding to phosphorylated residues in the *STAT1*, *STAT3* and *LATS1* transcripts as measured by targeted RNA sequencing (Fig. 2c).

To assess the potential of using Cas13bt fusion proteins in a single viral vector, we packaged REPAIR.t1 along with a crRNA expression cassette in a recombinant AAV2 vector and delivered the system to HEK293FT cells (Fig. 2g). Immunofluorescence staining demonstrated that REPAIR.t1 delivered by AAV was expressed and localized to the cytoplasm with the help of a nuclear export signal (Fig. 2h and Supplementary Fig. 7). RNA sequencing showed site-specific A-to-I editing of 7.5% ± 0.8% at the *CTNNB1* T41 site following enrichment for transduced cells (see Methods; Fig. 2i). This demonstrates the feasibility of using a single AAV genome for delivery. However, as editing rates achieved using AAV2 delivery are lower than for plasmid DNA transfection, further optimization would likely be needed to boost the editing efficiency, such as targeting the editing machineries to specific subcellular compartments to achieve the optimal editing rate for specific target transcripts. When attempting in vivo delivery to specific tissues, AAV serotypes should be chosen to maximize delivery and expression

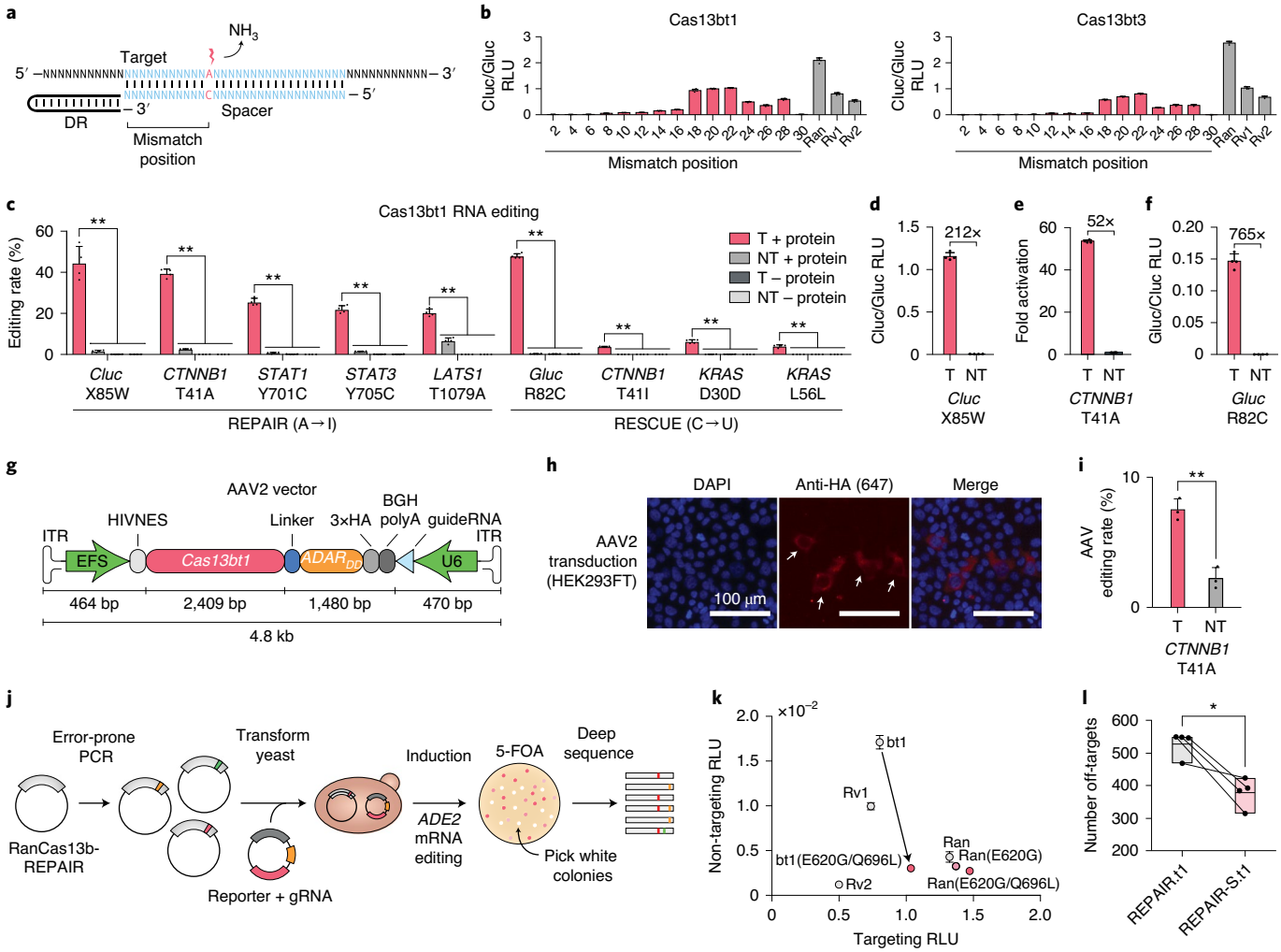


Fig. 2 | RNA editing with Cas13bt. a, Schematic of crRNA design for RNA editing. Mismatch at target adenosine shown in red. Mismatch position is measured from the 5′ end of the target. **b**, Evaluation of RNA editing of a W85X Cluc reporter in HEK293FT cells. Data are presented as mean ± s.d.; $n = 4$ for REPAIR.t1 and $n = 3$ for REPAIR.t3. Rv1, REPAIRv1; Rv2, REPAIRv2[2]; Ran, RanCas13b-REPAIR[3]. **c**, Quantification of RNA editing by REPAIR.t1 and RESCUE.t1 by deep sequencing. Statistical significance for interaction of ±protein and targeting versus non-targeting crRNA was assessed by two-way analysis of variance (ANOVA) followed by post hoc analysis using Games–Howell pairwise comparison. $^{**}P < 0.01$. For this panel and **d**–**f**, values show mean ± s.d.; $n = 4$; and T, targeting crRNA; NT, non-targeting crRNA. **d**, Quantification of Cluc restoration by REPAIR.t1. **e**, Quantification of beta-catenin reporter activation by REPAIR.t1. **f**, Quantification of Gluc restoration by RESCUE.t1. **g**, Schematic of AAV2 vector. ITR, inverted terminal repeat; EFS, short *EF1alpha* promoter. **h**, Representative immunofluorescence staining of HEK293FT cells transduced with AAV2 encoding REPAIR.t1 ($n = 2$). **i**, Quantification of RNA editing of *CTNNB1* in HEK293FT cells by AAV2-mediated expression of REPAIR.t1. T, targeting crRNA; NT, non-targeting crRNA. Data are presented as mean ± s.d.; $n = 3$; statistical significance was assessed with a two-tailed *t*-test; $^{**}P < 0.01$. **j**, Schematic of a directed evolution approach for engineering specific ADAR2dd variants. 5-FOA, 5-fluoroorotic acid. **k**, Evaluation of specificity-enhancing ADAR2dd mutants applied to REPAIR.t1 targeting the W85X mutation in a Cluc reporter. bt1, REPAIR.t1; other construct abbreviations are as in **b**. Data are presented as mean ± s.d. of non-targeting RLU; $n = 3$. **l**, Quantitative comparison of off-target editing between REPAIR.t1 variants. REPAIR-S, ADAR2dd-E488Q/E620G/Q696L; REPAIR, ADAR2dd-E488Q. Data are presented as mean with top and bottom boundaries denoting minimum and maximum; $n = 4$. Statistical significance was assessed using a two-tailed paired *t*-test; $^{*}P < 0.05$. Paired samples are shown by connected lines between the two conditions. See Supplementary Table 11 for exact *P* values where relevant.

efficiency (for example, AAV8 for liver[13] and PHP.eB for brain in C57BL/6 mice[14]).

Finally, we quantified the transcriptome-wide specificity of REPAIR.t1 and found the number of off-target edits introduced by this system was comparable to that for REPAIRv1 (Supplementary Fig. 8). Although off-target edits in RNA are transient, they may lead to deleterious protein products that may cause adverse effects. To improve the specificity of REPAIR, we used a yeast-based directed evolution approach to identify two promising mutations in ADAR2dd (E620G and Q696L; Fig. 2j and Supplementary Figs. 9 and 10) that reduce promiscuous deamination activity. We incorporated these two mutations in REPAIR.t1 and found that the number of off-target edits decreased without reducing the on-target activity (Fig. 2k,l and Supplementary Fig. 8). With additional mutagenesis and optimization, it may be possible to even further reduce off-target activity for off-target sensitive applications.

The small size of Cas13bt proteins provides new opportunities for programmable RNA modulation, particularly in vivo. We have shown here that Cas13bt1 and 3 can be used to generate compact REPAIR and RESCUE constructs compatible with delivery by AAV, thereby further advancing the development of programmable RNA editing technologies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-01030-2.

Received: 20 June 2020; Accepted: 20 July 2021;
Published online: 30 August 2021

References

1. Terns, M. P. CRISPR-based technologies: impact of RNA-targeting systems. *Mol. Cell* **72**, 404–412 (2018).
2. Cox, D. B. T. et al. RNA editing with CRISPR-Cas13. *Science* **1027**, 1019–1027 (2017).
3. Abudayyeh, O. O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382–386 (2019).
4. Dong, J. Y., Fan, P. D. & Frizzel, R. A. Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Hum. Gene Ther.* **7**, 2101–2112 (1996).
5. Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).
6. Smargon, A. A. et al. Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell* **65**, 618–630 (2017).
7. Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
8. Gootenberg, J. S. et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science* **360**, 439–444 (2018).
9. Matthews, M. M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).
10. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17**, 9–26 (2009).
11. Apte, U. et al. Beta-catenin activation promotes liver regeneration after acetaminophen-induced injury. *Am. J. Pathol.* **175**, 1056–1065 (2009).
12. Bhushan, B. et al. Pro-regenerative signaling after acetaminophen-induced acute liver injury in mice identified using a novel incremental dose model. *Am. J. Pathol.* **184**, 3013–3025 (2014).
13. Kattenhorn, L. M. et al. Adeno-associated virus gene therapy for liver disease. *Hum. Gene Ther.* **27**, 947–961 (2016).
14. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
15. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Data curation and search pipeline. The bioinformatics analysis in this study was performed on contigs from public sequence databases including the National Center for Biotechnology Information's prokaryotic and virus databases (http://ncbi.nlm.nih.gov) and Whole Genome Shotgun database (https://www.ncbi.nlm.nih.gov/genbank/wgs/), MG-RAST (https://www.mg-rast.org/)[16] and the Department of Energy Joint Genome Institute's online resources (http://jgi.doe.gov)[17,18], totaling 3.17 trillion base pairs (bp). All open reading frames larger than 80 aa were annotated, resulting in 10 billion putative proteins for further analysis. Previously developed Cas13 profiles[19] were used to identify Cas13 family proteins with HMMER3.2[20] using a minimum bit-score threshold of 25. A group of small (~800 aa) but divergent Cas13b proteins was identified and used to generate a new profile for a second HMMER search with the same settings to retrieve additional members of this subfamily. In total, 5,843 *cas13* loci were identified.

Phylogenetic analysis. For phylogenetic analysis and classification, the 5,843 candidate genes were clustered using MMseqs2 with a minimum sequence identity of 50% and minimum coverage of 70%[21,22]. Proteins within each cluster were clustered at 90% identity and 80% minimum coverage for redundancy reduction. Each redundancy-reduced cluster was aligned using MAFFT[23] with default parameters. Proteins identified as truncated or partial on the basis of partial match to a larger protein in an alignment, and/or clusters entirely composed of partial/truncated proteins, were removed from the analysis.

The aligned redundancy-reduced clusters were converted into HHsuite profiles using all columns with less than 50% gaps, and each of these profiles was searched against each other with profile–profile alignment using HHsearch[24]. The resulting pairwise bit-scores between clusters, s_{ij}, where i and j denote clusters i and j, respectively, were used to construct a classification dendrogram. First, the asymmetric bit-scores were symmetrized by setting $s_{ij} = (s_{ij} + s_{ji})/2$. Then, pseudo-distances were calculated by setting $d_i = -(\log[s_{ij}] - \log[\min(s_{ii}, s_{jj})])/2$ to generate a distance matrix[25]. A UPGMA dendrogram for classification was constructed using these distances. Branches and the subtrees of the dendrogram were contracted without modifying their topology to highlight known subtypes and subgroups within each subtype. Lengths in amino acids of the redundancy-reduced proteins from each subtree were used to generate protein size distributions.

Non-redundancy-reduced Cas13bt proteins, along with one Cas13b protein sequence (BzoCas13b) as an outgroup, were aligned using MAFFT. A phylogenetic tree was constructed using FastTree[26], and the tree was rooted using the outgroup.

Design and cloning of bacterial expression plasmid constructs. All cloning in this study was performed using chemically competent Stbl3 *E. coli* (NEB) unless otherwise noted. All PCR for cloning was performed using 2X Phusion Flash High-Fidelity Master Mix (Thermo Fisher) unless otherwise noted.

The *cas13bt2* full locus was synthesized and cloned into the BamHI site of pACYC184 by GenScript.

To clone bacterial expression plasmids for the PFS screen, Cas13bt protein CDSs were human codon optimized using GeneArt GeneOptimizer (Thermo Fisher) and synthesized by GenScript into a pcDNA3.1(+) backbone. Genes were amplified by PCR to add a pLac promoter and cloned into a pBR322 backbone (NEB) digested with EcoRV (Thermo Fisher) by Gibson assembly.

crRNA expression cassettes for each DR corresponding to each Cas13bt of interest were synthesized by IDT, amplified by PCR, and cloned into a pACYC184 backbone digested with EcoRV and BamHI (Thermo Fisher) by Gibson assembly. All primers are listed in Supplementary Table 4 and final constructs in Supplementary Table 10.

Design and cloning of mammalian expression plasmid constructs. Mammalian crRNA expression cassettes were amplified from pC0048 (Addgene plasmid no. 103854; http://n2t.net/addgene:103854; RRID:Addgene_103854)[2] using primers to add the DR for each Cas13bt ortholog of interest and cloned into pC0048 digested with LguI and KpnI (Thermo Fisher) using Gibson assembly.

Mammalian protein expression cassettes were cloned by amplifying previously mentioned synthesized Cas13bt genes by PCR and cloning into pC0053 (Addgene plasmid no. 103869; http://n2t.net/addgene:103869; RRID:Addgene_103869)[2] digested with HindIII and NotI (Thermo Fisher), either alone or with ADAR2dd-E488Q amplified from pC0053 for REPAIR constructs and pC0078 (Addgene plasmid no. 130661; http://n2t.net/addgene:130661; RRID:Addgene_130661)[3] for RESCUE constructs. Site-directed mutagenesis was used to create catalytically inactivated Cas13bt proteins. ADAR2 mutants derived from directed evolution screens were cloned by introduction of mutations via PCR primers. All primers are listed in Supplementary Table 4 and final constructs in Supplementary Table 10.

crRNA spacers were cloned into expression backbones by Golden Gate assembly as previously described[27]. Spacer sequences are listed in Supplementary Tables 6 and 8.

Bacterial RNA sequencing. Bacterial RNA sequencing was performed as previously described[28]. Briefly, 5 ml overnight culture of a Stbl3 *E. coli* colony transformed with a plasmid containing the locus of interest was spun down and resuspended in 1 ml of TRI Reagent (Zymo Research). After a 5-min room temperature incubation, 250 µl of 0.5-mm Zirconia beads was added and the Trizol resuspension was vortexed vigorously for 30 s to 1 min. A 200-µl volume of chloroform was added, and samples were inverted gently, incubated at room temperature for 3 min, and then spun down at 12,000*g* for 5 min at 4 °C. Following centrifugation, the aqueous fraction was used as input to the Qiagen miRNeasy kit, as per the manufacturer's instructions.

Purified RNA was treated with DNase I (NEB), purified again using RNA Clean & Concentrate-25 (Zymo Research) and treated with T4 polynucleotide kinase (PNK; NEB). PNK-treated RNA was again purified using RNA Clean & Concentrate-25 (Zymo Research), and ribosomal RNA was removed using the Ribominus Transcriptome Isolation Kit (Yeast and Bacteria) (Thermo Fisher). Samples were subsequently treated with RNA 5′ polyphosphatase (Epicentre) and purified again using an RNA Clean & Concentrate-5 kit (Zymo Research). Purified RNA was used as input to the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB). Library preparation was performed as per the manufacturer's instructions, except with a final PCR of 20 cycles. Libraries were quantified by qPCR using the KAPA Library Quantification Kit for Illumina (Roche) on a StepOnePlus Real-Time PCR System (Thermo Fisher) and sequenced on an Illumina NextSeq. Reads were mapped using BWA.

***E. coli* essential gene PFS screen.** Libraries were designed and cloned as previously described[6]. Briefly, the library of spacers was cloned into each Cas13bt pJ23119–spacer–DR backbone containing a chloramphenicol-resistance gene using Golden Gate assembly with a 5:1 ratio of spacer library to predigested backbone with 210 cycles. Libraries were transformed into Endura Electrocompetent Cells (Lucigen) by electroporation and plated over five 22.7 cm×22.7 cm chloramphenicol LB agar plates. At 12 h after plating, libraries were scraped from plates and DNA was extracted using the Macherey-Nagel Nucleobond Xtra Maxiprep Kit (Macherey-Nagel).

A 200-ng quantity of library plasmid and 200 ng Cas13bt gene plasmid containing an ampicillin-resistance gene were transformed into 100 µl of Endura Electrocompetent Cells (Lucigen) by electroporation as per the manufacturer's protocol and plated across four 22.7 cm × 22.7 cm ampicillin/chloramphenicol LB agar plates per biological replicate, with three biological replicates per condition. At 10–12 h post-transformation, libraries of transformants were scraped from the plates and DNA was extracted using the Macherey-Nagel Nucleobond Xtra Maxiprep Kit (Macherey-Nagel). Libraries were prepared from extracted DNA for next-generation sequencing using primers in Supplementary Table 5 with NEBNext High-Fidelity 2X PCR Master Mix (NEB) and sequenced on an Illumina NextSeq.

Spacers with 100 or more total reads across all replicates in the control sample were retained for analysis. Average spacer abundance (fraction of reads relative to all reads) was calculated as the average abundance of each spacer across all replicates. Relative abundance was calculated as the ratio of the spacer abundance in the experimental condition relative to the spacer abundance in the input library. To account for off-target cleavage activity, a two-component Gaussian mixture distribution was fit to the $\log_{10}$[relative abundance] distribution of non-targeting (negative control) spacers, and the Gaussian component with the higher mean, m, was used as the null distribution as it contains fewer non-targeting guides with off-target activity. $\log_{10}$[abundance] values for all spacers in the Cas13bt library were normalized by subtracting m (see Supplementary Fig. 2b). Depleted spacers were selected as those with normalized $\log_{10}$[relative abundance] below -5σ, where σ is the standard deviation of the null distribution. Weblogos were then generated (https://weblogo.berkeley.edu/logo.cgi)[15] using all depleted spacers and separately using the top 1% most depleted spacers of all spacers.

To analyze gene position preference, all spacers were searched against their respective CDS targets in *E. coli* using the genome NC_010473.1 to identify their coordinates relative to the respective CDS. Gene coordinates were normalized by the CDS length. Bins of normalized gene coordinates from −0.1 to 1.1 were generated with a step size of 0.025. For each bin, and each CDS, the average relative abundance of spacers with start positions that fall in the bin was computed. Then the overall abundance for the bin was taken to be the average relative abundance for each CDS in the bin averaged across all CDSs, excluding CDSs with no spacers in the given bin. This analysis was performed using Python and is available via GitHub.

Mammalian cell culture and transfection. Mammalian cell culture experiments were performed in the HEK293FT line (Thermo Fisher catalog no. R70007) grown in Dulbecco's modified Eagle medium with high glucose, sodium pyruvate and GlutaMAX (Thermo Fisher), additionally supplemented with 1× penicillin–streptomycin (Thermo Fisher), 10 mM HEPES (Thermo Fisher) and 10% fetal bovine serum (VWR Seradigm). All cells were maintained at confluency below 80%.

All transfections were performed with Lipofectamine 2000 (Thermo Fisher) in 96-well plates unless otherwise noted. Cells were plated at approximately 20,000 cells per well 16–20 h before transfection to ensure 90% confluency at the time of transfection. For each well on the plate, transfection plasmids were combined with OptiMEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 µl. Separately, 23 µl of OptiMEM was combined with 2 µl of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and pipetted onto cells.

Mammalian RNA knockdown assays. HEK293FT cells were transfected as described with 75 ng of a plasmid encoding either a Cas13b or bt protein or GFP expressed from a CMV promoter, 150 ng of a plasmid encoding a crRNA expressed from a human U6 promoter and, where relevant, 45 ng of reporter plasmid. After 48 h, RNA was collected as described previously[27] with 2× the amount of recommended DNase and a 20 min lysis step. RNA expression was measured by qPCR using commercially available TaqMan probes (Thermo Fisher; Supplementary Table 7) on a LightCycler 480 II (Roche) with GAPDH as an endogenous internal control in 5 µl multiplexed reactions. Probes and primer sets were generally selected to amplify across the Cas13 target site so as to minimize detection of cleaved transcripts. Each of the four biological replicates is the average of four technical qPCR replicates, and relative expression was calculated using the ddCt method[29] with a negative control condition consisting of the corresponding crRNA expression plasmid co-transfected with the GFP expression plasmid rather than the corresponding Cas13b or bt expression plasmid. Negative control values are the average of four biological replicates. Statistical significance was assessed using a two-tailed *t*-test.

For luciferase reporter assays, medium was aspirated from cells and Cluc and Gluc activity in the medium was measured as RLUs using *Gaussia* and *Cypridina* Luciferase Assay Kits (Targeting Systems) with an injection protocol on a Biotek Synergy Neo 2 (Agilent). Each experimental luciferase measurement was normalized to the appropriate control luciferase measurement (that is, if Cluc was targeted, the Gluc measurement was used as the control value and vice versa). For knockdown assays, normalized luciferase values were then again normalized to an average normalized luciferase measurement of four biological replicates of a negative control condition consisting of the corresponding crRNA expression plasmid co-transfected with a GFP expression plasmid rather than the corresponding Cas13b or bt expression plasmid. Error bars were calculated in GraphPad Prism 7 and represent the standard deviation of the luciferase values normalized to negative control transfection; $n=4$. Statistical significance was assessed by pairwise comparison of each targeting condition with each of the four non-targeting conditions using a two-tailed *t*-test. The maximum *P* value was taken from each of the four pairwise comparisons for each targeting condition, and significance was assessed by comparison using an experiment-wide critical value of $\alpha=0.05$ with Bonferroni correction using $n=4$.

Mammalian RNA editing assays. HEK293FT cells were transfected as described with 150 ng of plasmid encoding a dCas13b or bt protein–ADAR2dd variant fusion expressed from a CMV promoter, 300 ng of a plasmid encoding a crRNA expressed from a human U6 promoter and, where relevant, 45 ng of a reporter plasmid. After 48 h, RNA was collected as described above and reverse transcription was performed as described previously[27] using gene-specific primers for the relevant target transcript (Supplementary Table 9). cDNA was used as input for library preparation of next-generation sequencing libraries (Supplementary Table 5) using NEBNext High-Fidelity 2X PCR Master Mix (NEB), and amplicons were sequenced on an Illumina MiSeq. Editing was quantified by counting the number of reads at which the expected edited position in the amplicon was called as a G (for A-to-I editing) or T (for C-to-U editing) and dividing by the total number of reads in the sample using Python (available via GitHub). Unless otherwise noted, all reported data are the average of four biological replicates.

For AAV-transduced cells, 250 µl of AAV2 was mixed with 750 µl medium, and medium on HEK293FT cells was replaced with AAV-containing medium solutions in 12-well plates. After approximately 40 h, RNA was collected as previously described[30]. Briefly, cells were dissociated from plates with TrypLE (Thermo Fisher) to a single-cell suspension and fixed with fixation buffer (4% paraformaldehyde, 0.1% saponin, 1:100 SuperaseIn RNAse inhibitor (Invitrogen) in phosphate buffered saline (PBS)) for 30 min at 4 °C. Cells were then stained with anti-HA primary antibodies (Roche Anti-HA High Affinity 3F10 rat IgG and Cell Signaling Technologies Anti-HA C29F4 rabbit monoclonal antibody) at 1:100 dilution each in staining buffer (0.1% saponin, 1% BSA, 1:100 SuperaseIn RNAse inhibitor in PBS), followed by staining with secondary antibodies (AlexaFluor 488 goat anti-rabbit and AlexaFluor 555 goat anti-rat, Invitrogen) at 1:2,000 dilution in staining buffer, each for 30 min at 4 °C. Cells were then sorted by FACS for HA-positive cells and de-crosslinked in RecoverAll digestion buffer with proteinase K (Invitrogen) for 3 h at 50 °C. RNA was extracted using a QuickRNA microprep kit (Zymo Research) and libraries were prepared as described above.

Luciferase reporter assays for RNA editing were performed as described above, with the modification that normalized luciferase values were not normalized to a GFP control condition. For *CTNNB1* targeting, we used a previously reported engineered luciferase reporter generated by replacing the *EF1alpha* promoter driving Gluc expression in our dual-luciferase reporter plasmid with a promoter derived from either an M50 Super 8X TOPFlash (TOP) or M51 Super 8X FOPFlash (FOP) reporter[3]. M50 Super 8x TOPFlash (Addgene plasmid no. 12456; http://n2t.net/addgene:12456; RRID:Addgene_12456) and M51 Super 8x FOPFlash (TOPFlash mutant) (Addgene plasmid no. 12457; http://n2t.net/addgene:12457; RRID:Addgene_12457) were gifts from Randall Moon[31]. 8x TOPFlash and FOPFlash reporters contain eight beta-catenin-binding sites either intact (TOP) or scrambled (FOP); the TOPFlash reporter thus provides a metric of beta-catenin activation when compared to background as measured by the FOPFlash reporter with otherwise similar conditions. Luciferase activity was measured for these custom dual-luciferase reporters for each protein/crRNA condition and normalized as described for a dual-luciferase reporter. Fold activation was calculated by taking the ratio of the average TOP measurement and dividing by the average FOP measurement, and error was calculated by a standard error propagation formula.

Optimal spacers for all target sites tested were determined by tiling spacers across the site of interest, varying the distance of the mismatch from the DR from 14 bp to 28 bp in intervals of 2 bp.

Statistical significance between ±protein and targeting/non-targeting crRNA conditions was assessed using two-way ANOVA followed by post hoc analysis using Games–Howell pairwise comparison.

RNA editing specificity. HEK293FT cells were transfected as described for mammalian RNA editing assays. After 48 h, RNA was collected using a QIAGEN RNeasy Plus 96 kit as per the manufacturer's protocol. The mRNA fraction was enriched using an NEBNext Poly(A) Magnetic Isolation Module (NEB). Libraries were prepared using an NEBNExt Ultra II Directional RNA library prep kit (NEB) as per the manufacturer's protocol and sequenced on an Illumina NextSeq. Each sample was sequenced with an average read depth of 8 million reads per sample and randomly downsampled to 5 million reads per sample. Data were analyzed using a previously described custom pipeline on the FireCloud computational framework and downstream analysis was performed using a previously described custom Python script[2,3]. Any significant edits found in eGFP-transfected conditions were considered to be single nucleotide polymorphisms (SNPs) or artifacts of the transfection and filtered out. An additional layer of filtering for known SNP positions was performed using the Kaviar method for identifying SNPs[32].

AAV production. Recombinant AAV2 vectors packaging base editing and GFP control cassettes described above were generated using triple plasmid transfection in HEK293T cells as described previously[33]. Briefly, 60 µg Adenoviral helper genes (Addgene plasmid no. 112867; http://n2t.net/addgene:112867; RRID:Addgene_112867, a gift from James M. Wilson), 50 µg AAV2 Rep/AAV2 Cap (Addgene plasmid no. 104963; http://n2t.net/addgene:104963; RRID:Addgene_104963, a gift from Melina Fan) and 30 µg ITR-flanked transgene were transfected into five 15-cm plates of HEK293T cells. Medium and cells were collected 5 days post transfection. Medium was then combined with PEG8000 in 2.5 M NaCl at a 4:1 ratio, incubated overnight at 4 °C, and centrifuged at 3,000 x *g* for 30 min at 4 °C to pellet the precipitate. Cell pellets were sonicated to lyse cells and release viral particles. Combined sonicated cell pellet and medium precipitate was digested with DNAse for 1 h at 37 °C before iodixanol gradient ultracentrifugation. Iodixanol fractions of 17%, 25%, 40% and 60% were prepared as previously described, and viral preparations were loaded on top of the gradient[34]. Following 20 h of ultracentrifugation at 28,000 r.p.m., gradients were fractionated, and peak fractions at the 40%/60% interface were subject to buffer exchange and concentration with Pierce Protein Concentration Columns according to the manufacturer's instructions (Thermo Fisher).

For determination of viral titers, viral preparation samples were treated with DNAse (10 mg ml^{-1}) to remove unencapsidated viral genomes. Following inactivation of DNAse with 0.5 M EDTA, samples were digested with Proteinase K to release encapsidated viral genomes. Viral DNA was then measured against a vector standard (Addgene no. 37825-AAV2) by qPCR with primers against the ITRs (5′-AACATGCTACGCAGAGAGGGAGTGG-3′ and 5′-CATGAGAC AAGGAACCCCTAGTGATGGAG-3′) using a Lightcycler 480 II (Roche).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Deep sequencing data from whole-transcriptome sequencing were deposited as a BioProject under Project ID PRJNA641934. Data from the main figures are available in the Supplementary Information.

Code availability

All Python scripts used for data analysis are available in a GitHub repository found at https://github.com/fengzhanglab/Cas13bt-analysis.

References

16. Keegan, K. P., Glass, E. M. & Meyer, F. MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods Mol. Biol.* **1399**, 207–233 (2016).
17. Nordberg, H. et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* **42**, 26–31 (2014).
18. Chen, I. M. A. et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2020).
19. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl Acad. Sci. USA* **115**, E5307–E5316 (2018).

20. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
21. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
22. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018).
23. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
24. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 1–15 (2019).
25. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2019).
26. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
27. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
28. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
29. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).
30. Hrvatin, S., Deng, F., O'Donnell, C. W., Gifford, D. K. & Melton, D. A. MARIS: method for analyzing RNA following intracellular sorting. *PLoS ONE* **9**, e89459 (2014).
31. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish Prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).
32. Glusman, G., Caballero, J., Mauldin, D. E., Hood, L. & Roach, J. C. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216–3217 (2011).
33. Shen, S., Troupes, A. N., Pulicherla, N. & Asokan, A. Multiple roles for sialylated glycans in determining the cardiopulmonary tropism of adeno-associated Virus 4. *J. Virol.* **87**, 13206–13213 (2013).
34. Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* **1**, 1412–1428 (2006).

Acknowledgements

We thank E. Eloe-Fadrosh for generously sharing JGI data from study Ga0114919; D. L. Valentine and J. Tarn for generously sharing JGI data from study Ga0180434; E. Puccio for assistance with AAV production; F. E. Demircioglu for assistance with protein purification; members of the laboratory of F.Z. for advice and discussions, R. Macrae for discussion and editing of the manuscript, and R. Belliveau for technical support. S.K. is supported by a National Science Foundation Graduate Research Fellowship. F.Z. is supported by NIH grants (1R01-HG009761, 1R01-MH110049 and 1DP1-HL141201), HHMI, Open Philanthropy, The G. Harold and Leila Y. Mathers Charitable, Bill and Melinda Gates, and Edward Mallinckrodt, Jr Foundations, the Poitras Center for Psychiatric Disorders Research at MIT, the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT, the Yang-Tan Center for Molecular Therapeutics, and by L. Yang, the Phillips family and J. and P. Poitras.

Author contributions

S.K., H.A.-T. and F.Z. conceived the project and designed the computational pipeline and experiments. H.A.-T. performed computational searching and phylogenetic analysis with input from K.S.M. and E.V.K. S.K. and H.A.-T. performed the essential gene screen experiment and H.A.-T. analyzed the results. X.J. performed the transcriptome-wide specificity sequencing experiment, and S.K. analyzed the results. V.M. produced AAV and assisted with design of AAV constructs and experiments. S.K. performed and analyzed results from all other experiments with assistance from X.J., R.O. and V.M. S.K., H.A.-T. and F.Z. drafted the manuscript with input from all authors.

Competing interests

F.Z. is a co-founder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies and Sherlock Biosciences. S.K., H.A.-T. and F.Z. are co-inventors on US provisional patent application 62/905,645 relating to the Cas proteins described in this manuscript. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01030-2.

Correspondence and requests for materials should be addressed to F.Z.

Peer review information *Nature Biotechnology* thanks Magdy Mahfouz, Mitchell O'Connell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Feng Zhang

Last updated by author(s): Jun 9, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data was collected with Illumina MiSeq (amplicon sequencing) or NextSeq (small RNA and total RNA-seq). qPCR data was collected using a Roche LightCycler 480 II. Luciferase data and collateral activity fluorescence measurements were collected using an Agilent Biotek Synergy Neo2 with Gen5 software. Gel images were collected using a LiCor Odyssey Clx.
Data analysis	Publicly available software and platforms: Apache Spark, HMMer version 3.2, MMSeqs2, MAFFT version 7, HHsuite version 3.2, BWA version 0.7.17, BEDTools version 2.28.0, Samtools version 1.8, Python 3.7, Microsoft Excel version 16.16.23, GraphPad Prism 7, Terra FireCloud computational framework, Kaviar database version 160204-Public, Python packages matplotlib version 3.1.0, pandas version 0.24.2, seaborn version 0.9.0, numpy version 1.16.4, biopython version 1.74, scipy version 1.3.0. All custom Python scripts described are available on GitHub at: https://github.com/fengzhanglab/Cas13bt-analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Values plotted in main figures are available in Supplementary Information. Raw data from whole transcriptome sequencing is available as a BioProject with Project ID PRJNA641934.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on sample sizes used in the work of our and other groups generating reproducible results for similar experiments.
Data exclusions	For testing of ADAR mutants, biological replicates that were drastic outliers due to experimental variability were excluded.
Replication	Experiments were reproduced with cells from different batches/passages. All attempts to reproduce results were successful.
Randomization	Samples were not randomized. Controls were run alongside experimental conditions to control for experimental conditions.
Blinding	Blinding was not performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies: Anti-HA high affinity rat IgG1 (Roche 3F10 Lot # 42155800); anti-HA rabbit mAb (Cell Signaling Technologies C29F4 Lot # 9). Secondary antibodies: Goat anti-rat AlexaFluor 647 (Invitrogen Lot # 2156534); Goat anti-rat AlexaFluor 555 (Invitrogen Lot # 2147636); Goat anti-rabbit AlexaFluor 488 (Invitrogen Lot # 2110498)
Validation	HA-tag antibodies are species independent. Anti-HA high affinity rat IgG1 (Roche 3F10) was validated for immunofluorescence staining by Sinnamon et al. (2017). Site-directed RNA repair of endogenous Mecp2 RNA in neurons. PNAS, 114(44), E9395-E9402. Anti-HA rabbit mAb (Cell Signaling Technologies C29F4) is validated for FACS by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293FT cells were acquired from ATCC.
Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.