Engineered Cas12i2 is a versatile high-efficiency platform for therapeutic genome editing

Colin McGaw1,2, Anthony J. Garrity1,2, Gabrielle Z. Munoz1, Jeffrey R. Haswell1, Sejuti Sengupta1, Elise Keston-Smith1, Pratyusha Hunnewell1, Alexa Ornstein1, Mishti Bose1, Quinton Wessells1, Noah Jakimo1, Paul Yan1, Huaibin Zhang1, Lauren E. Alfonse1, Roy Ziblat1, Jason M. Carte1, Wei-Cheng Lu1, Derek Cerchione1, Brendan Hilbert1, Shanmugapriya Sothiselvam1, Winston X. Yan1, David R. Cheng1, David A. Scott1, Tia DiTommaso1✉ & Shaorong Chong1

The CRISPR-Cas type V-I is a family of Cas12i-containing programmable nuclease systems guided by a short crRNA without requirement for a tracrRNA. Here we present an engineered Type V-I CRISPR system (Cas12i), ABR-001, which utilizes a tracr-less guide RNA. The compact Cas12i effector is capable of self-processing pre-crRNA and cleaving dsDNA targets, which facilitates versatile delivery options and multiplexing, respectively. We apply an unbiased mutational scanning approach to enhance initially low editing activity of Cas12i2. The engineered variant, ABR-001, exhibits broad genome editing capability in human cell lines, primary T cells, and CD34+ hematopoietic stem and progenitor cells, with both robust efficiency and high specificity. In addition, ABR-001 achieves a high level of genome editing when delivered via AAV vector to HEK293T cells. This work establishes ABR-001 as a versatile, specific, and high-performance platform for ex vivo and in vivo gene therapy.
CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats–CRISPR associated proteins) systems are widely distributed in both archaea and bacteria and provide adaptive immunity against invading viruses. Class 2 CRISPR-Cas systems, notably type II Cas9 and type V Cpf1 (Cas12a), have been harnessed for therapeutic gene editing. To expand the CRISPR therapeutic toolbox, we recently identified a number of functionally diverse type V systems (subtypes V-G, V-H and V-I). The subtype V-I system (Cas12i) represents an evolutionarily distinct branch that closely resembles Cas12a in the fact that Cas12i can also functionally resemble Cas12a in the in vitro fluorescent reporter assay. Based on the GFP signal depletion relative to the wild-type controls, we identified a number of variants showing a significant increase in GFP depletion (with a Z-score>2) for at least one GFP target ( Supplementary Fig. 6a, b). A majority of arginine and glycine substitutions, however, did not enhance the GFP signal depletion activity (Supplementary Fig. 6a, Z score < =1). 14 of these in vitro hits were tested for indel activity at two genomic target sites in HEK293T cells (Fig. 1b). Three substitutions, D581R, I926R and V1030G, showed 1.5- to 2-fold improvement in indel activity relative to Cas12i WT at one or both target sites (Fig. 1b). The rest of the in vitro hits showed minimal improvement, or even decreased indel activity (Fig. 1b), suggesting underlying differences between cell-free in vitro activity and cellular editing function. Combining the top three substitutions resulted in a variant with at least 3-fold improvement in indel activity relative to Cas12i WT, suggesting that the effect of these single mutations was additive (Fig. 1c). This combination variant was chosen as Cas12i v1 and named ABR-001.

To test ABR-001 activity, we designed an experiment targeting a broad set of 18 genomic target loci in HEK293T cells. In this experiment, ABR-001 exhibited indel activity approaching SpCas9 efficiency (Fig. 1d). Notably, analysis of indel patterns formed by ABR-001 targeting these genomic loci in HEK293T cells identified prominent large deletions of 5–20 nucleotides (nts), in contrast to small deletions and +1 insertion observed with SpCas9 (Fig. 1e). Consistent with these data, ABR-001 showed increased dsDNA cleavage activity relative to Cas12i WT in in vitro biochemical assays (Supplementary Fig. 7). Deep sequencing analyses of DNA fragments from Cas12i WT in vitro cleavage reactions identified a number of cut sites in the PAM-distal DNA region on both the template strand (TS) and non-template strand (NTS), similar to those observed in Cas12a, but with a wide distribution of NTS cut sites outside of the R-loop (Supplementary Figs. 8–11). The extended distribution of cut sites does not seem to be caused by the substitutions in ABR-001 as the downstream cut sites were also found with Cas12i WT (Supplementary Fig. 11).

ABR-001 is a specific nuclease for genome editing. To comprehensively assess ABR-001 specificity, we employed an unbiased approach of tagmentation-based tag integration site sequencing (TTISS) which, similarly to GUIDE-seq, uses the integration of a known donor oligo to detect dsDNA breaks in an unbiased manner. TTISS allowed us to generate an empirical list of potential off-target sites for 18 targets across three genomic loci for both ABR-001 and SpCas9. For instance, at VEGFA target 1, no off-target was detected for ABR-001, whereas one major and six minor off-target sites were identified for SpCas9 (Fig. 2a). Aggregating all 18 targets, TTISS identified more potential off-target sites with a higher number of unique integration events for SpCas9 than ABR-001(Fig. 2b–d, Supplementary Fig. 12). Alternatively, we took a prediction-based approach to identify most likely off-target sequences via in silico prediction, the human genome...
was searched for sequences adjacent to the PAM for each of the 18 targets in the expanded set. Off-targets were predicted and ranked by ascending edit distance—the number of insertions, mismatches, and deletions between the on-target and off-target sequence. The editing efficiencies of the top 10 off-target sites were characterized using deep sequencing and background noise was corrected using maximum likelihood estimation (MLE). Using this approach, off-target indels greater than the limit of detection (>0.2%) were identified in 2 out of 18 targets for ABR-001 compared to 8 out of 18 targets for SpCas9 (Supplementary Fig. 13). However, the sparsity of identified off-targets could be due to limitations in the number of off-targets screened with this biased prediction-based approach. Taken together, these data demonstrate potential therapeutic utility of ABR-001 as an active and specific genome editing nuclease.

**ABR-001 ribonucleoproteins (RNPs) mediate ex vivo editing of human primary T cells and CD34+ HSPCs.** To test the therapeutic potential of ABR-001 for ex vivo cell therapy, we generated ABR-001-gRNA complexes and delivered the ribonucleoproteins (RNPs) into stimulated human CD3+ T cells by electroporation at various concentrations. We first targeted β2 microglobulin (B2M) with 4 different gRNAs and analyzed cells at 7 days post-electroporation for editing efficiency, cellular viability and B2M protein expression. All four B2M guides resulted in robust editing with positively correlated cellular viability and B2M protein expression. All four B2M knockdowns were observed along with maintenance of cellular viability (Fig. 3e). These data indicate that ABR-001 RNPs can be used for editing therapeutically relevant targets in human T cells at RNP concentrations that have no impact on cell viability.

To further characterize the indel pattern of ABR-001 editing in primary cells, we assessed the fraction of indels at a given indel size using data from the T cell experiments. Again, ABR-001 editing was biased toward large deletions in primary cells (Fig. 3g), similar to our initial observations in HEK293T cells (Fig. 1e). Further, the deletion length (indel size) and relative frequency appeared to be target dependent. Specifically, the most frequent deletions were >19 nt for TRAC, and 10-17 nt for...
CIITA (Fig. 3g). While deletion lengths vary between targets, deletion patterns were highly conserved across independent experiments (Fig. 3g), suggesting the editing was a non-random process.

To test the therapeutic potential of ABR-001 for ex vivo cell therapy, we delivered ABR-001 RNPs targeting the erythroid specific BCL11A enhancer in CD34+ hematopoietic stem and progenitor cells (HSPCs). Three guides were designed to generate indels within the BCL11A enhancer region to disrupt the GATAAA motif and induce fetal hemoglobin (HbF) in adult CD34+ HSPCs. All three guides generated robust indel rates comparable to SpCas9 within the enhancer region, two of which showed significant disruption of the GATAAA motif, albeit at lower levels than SpCas9 (Fig. 3h). The dual guide combination (multiplex target 1+2) maximized the GATAAA motif disruption as well as overall indel rates (Fig. 3h). Therefore, we selected the single guide (target 2) and dual guides (multiplex target 1+2) for further functional studies.

We used in vitro colony-forming cell (CFC) assays to test the function of ABR-001 edited within the BCL11a enhancer region in HSPCs and showed that editing with ABR-001 in vitro did not interfere with HSPC multilineage differentiation of these edited cells. Counts for the erythroid, myeloid, and mixed colonies were comparable among untreated samples and samples electroporated with ABR-001 protein or ABR-001 RNPs (Supplementary Fig. 15a). In the conditions containing ABR-001 RNPs, the indel rates detected in isolated colonies at day 15 matched or exceeded indel rates of the mixed input population at 72 h after electroporation suggesting that progenitor cells were successfully transfected in this in vitro assay (Supplementary Fig. 15b).

To evaluate indel persistence and HbF induction, CD34+ cells electroporated with ABR-001 and SpCas9 RNPs were expanded in erythroid differentiation medium. Cells were sampled at 3 and 20 days post-electroporation and analyzed for indel persistence and HbF expression. High levels of editing by ABR-001 were maintained throughout the 20-day erythroid differentiation period for both single and dual guides (Fig. 3i), demonstrating that ABR-001 edited cells persist long-term in vitro. The characteristic broad deletion profile of ABR-001 was retained in the HSPCs post differentiation (Fig. 3i). Further, the ABR-001 multiplex target 1+2 enabled a broader deletion profile, as compared to the single target 2 (Fig. 3i). HbF expression was assessed post 20-day erythroid differentiation period, showing a high level of fetal hemoglobin induction in the ABR-001 edited cells (Fig. 3k). Interestingly, ABR-001 edited cells using both single and dual guides yielded robust and SpCas9-equivalent HbF expression (Fig. 3k), despite lower levels of indel rates and GATAAA disruption relative to SpCas9 (Fig. 3i). These results
suggest that disruption of the GATAA motif and a broad region of the enhancer flanking the GATAA motif may be beneficial for inducing HbF expression, and by extension, that the broad indel profile could uniquely position ABR-001 for disrupting non-coding genetic elements to achieve positive therapeutic outcomes.

These in vitro results prompted us to test the persistence of ABR-001 edited cells in vivo. CD34+ HSPCs from a single donor were electroporated with ABR-001 and SpCas9 RNPs. Three days following electroporation, edited cells were transplanted into irradiated immunodeficient NSG mice at a dose of 200,000 cells. Indel analyses at the time of adoptive cell
transfer and at 8 and 16 weeks post-engraftment showed persistent levels of editing over time (Supplementary Fig. 16a). ABR-001 edited human CD45+ cells harvested from the bone marrow showed SpCas9-equivalent HbF expression, although a significant HbF induction relative to the untransfected group was not observed (Supplementary Fig. 16b). Importantly, ABR-001 edited human cells showed similar and, in some cases, higher in vivo engraftment capacity than unedited or SpCas9 edited cells at 8 or 16 weeks post-transplantation (Supplementary Fig. 13c–f). Taken together, these data demonstrate potent long-term persistence and functionality of ABR-001-edited CD34+ HSPCs.

**ABR-001 achieves highly efficient genome editing when delivered via AAV vector.** With its compact size and lack of tracrRNA, ABR-001 is an ideal candidate for AAV delivery. There have been few demonstrations of efficient editing with type II systems when delivered via AAV vector25,26. To enable the AAV delivery, we constructed an AAV2 vector expressing the ABR-001 effector under a CMV promoter and gRNA under a U6 promoter (Fig. 4a). Indel levels were measured at 72 h following transduction of HEK293T cells at varying multiplicities of infection (MOI). We found increasing editing levels that correlated with increasing MOI. With the highest MOI, the indel rate reached up to 60%, equivalent to or exceeding the levels for SaCas9 at matched target sites (Fig. 4b). To test the editing persistence, cells were cultured for up to 7 days post-transduction. We found that cells maintained high indel levels after day 1, increasing up to >80% by day 4 (Fig. 4c) and indel patterns remained relatively
consistent after day 4 (Fig. 4d). The jump in indel activity seen between day 3 and day 4 is likely due to the expansion of cells that occurred on day 3. As observed with non-viral delivery methods, the indel pattern was target dependent (Fig. 4d). Under these conditions, our data show that ABR-001 can achieve high editing levels in vitro when delivered by AAV and therefore suggest that ABR-001 may be a promising candidate for development into an in vivo gene editor.

Discussion

Rapid growth in genomic and metagenomic sequence databases has led to discovery of an increasing number of CRISPR-Cas systems, including many functionally diverse type V-I nucleases systems. All CRISPR-Cas systems originate from bacteria, archaea, or bacteriophage, and therefore, are not evolved for efficient genome editing in mammalian cells. Prior to our work, Cas12i has not been utilized for mammalian genome editing. Here we demonstrate that by introducing arginine and glycine mutations at critical positions of the C-terminal RuvC domain-containing nuclease lobe, we can improve Cas12i editing efficiency to achieve comparable editing outcomes to those published for edited cell therapies. Our results with ABR-001, along with data from other studies suggest that at least one critical factor limiting therapeutic applications of Cas12i and perhaps other type V systems might be the weak interaction of a CRISPR-Cas effector or RNP with mammalian genomic target DNA. In agreement with this hypothesis, ABR-001 exhibited higher binding affinity on dsDNA target compared to Cas12i WT (Supplementary Fig. 17), suggesting that the combined substitutions of D581R, I926R and V1030G in ABR-001 enhanced the catalytic activity of the effector through tighter binding of DNA substrates. In addition, modeling of D581R and I926R in Cas12i structure publications subsequent to the generation of ABR-001 indicates that these WED and Nuc domain substitutions could strengthen the electrostatic interaction with the dsDNA backbone of the PAM region and non-template strand DNA, respectively (Supplementary Fig. 18). The residue V1030 is located in the RuvC domain which functions to bind and cleave DNA substrate. The exact mechanism of the V1030G substitution in enhancing indel activity is not clear due to the lack of local structural information. It is possible that the V1030G substitution might influence target DNA interaction in a way that favors DNA cleavage. Rational approaches to engineering using recently published Cas12i structural information may further improve Cas12i editing efficiency and targeting specificity.

Cas12i is an attractive type V CRISPR-Cas nuclease for genome editing because of its compact size that fits in AAV vector with short 43-mer gRNA, absence of tracrRNA, ability to process pre-crRNA, and high specificity. With the exception of the recently discovered type V-J Cas12j, no other CRISPR-Cas nuclease possesses this unique combination of favorable properties. Cas12i is an attractive type V CRISPR-Cas nuclease for genome editing because of its compact size that fits in AAV vector with short 43-mer gRNA, absence of tracrRNA, ability to process pre-crRNA, and high specificity. With the exception of the recently discovered type V-J Cas12j, no other CRISPR-Cas nuclease possesses this unique combination of favorable properties. In vitro fluorescence reporter assay for high throughput screening of Cas12i single-substitution variants. In vitro fluorescence reporter assay was designed to couple cell-free protein synthesis of a CRISPR-Cas system with a fluorescence reporter activity (detailed description in Supplementary Methods). The reconstituted cell-free protein synthesis system was chosen for its designability and minimal level of nucleosome contamination. The reconstituted cell-free system reagents containing the E. coli transcriptional and translational machinery were made according to protocols from recent studies. Briefly, the cell-free system (manufactured at Arbor Biotechnologies) was supplemented with E. coli RNA polymerase core enzyme (1 U/μL) and murine RNase inhibitor (40 μU/μL) (New England Biolabs #M0550S and #M0314S). To set up each 1.2 μL reaction in a 384-well plate, a MANTIS Microfluidic Liquid Handler (FOR-MATRIX) was used to dispense a master mix (1 μL) containing the cell-free protein synthesis reagent, 0.04 units E. coli RNA polymerase core enzyme, 1.6 units murine RNase inhibitor, 2 ng GFP plasmid DNA, and 1 ng RFP plasmid DNA to a well prefilled with 10 μL mineral oil. Then an Echo 650 liquid handler (Beckman Coulter) was used to dispense 4 ng linear DNA template of SpCas9, 2 plasmid, 0.2 ng target gRNA (40 nM), 0.2 ng target gRNA (80 nL), and 0.1 ng linear DNA template of the 3′8 subunit (40 nL) to each well. After being sealed with an adhesive film, the 384-well plate was centrifuged to mix all components at the bottom of the well with mineral oil covering on top to prevent vaporization. The reaction was initiated when the plate was placed in a microplate reader (TECAN) and incubated at 37 °C. The kinetic GFP and RFP fluorescence was measured at every 10 min for 12 hr. The endpoint fluorescence was measured after incubation at 37 °C for 12 h.

Methods

Generation of DNA constructs for in vitro and mammalian cell screens and gene editing applications. For cell-free protein synthesis and the in vitro fluorescence reporter assay, the genes for SpCas9 and Cas12i WT effectors were cloned into pET28 vector (EMD Milpore #69864) and expressed under the T7 promoter. The genes for the GFP target and the RFP control were cloned into a pUC19 plasmid (New England Biolabs #N3041S) and expressed from a αβ8 promoter (8C). A linear DNA template for each effector variant was generated by PCR using a degenerate primer pair to amplify the region of the plasmid DNA that contained the open reading frame (ORF) of the effector, upstream T7 promoter and downstream T7 terminator. After PCR, the linear DNA templates (4–5 kb) were purified using CleanGNS SPRI beads (Bulldog Bio #CNGS500). Linear templates for 960 arginine and glycine single substitutions (Cas12i WT residues 575-1054) were generated by an overlapping PCR approach with mutagenic oligos, and then the open reading frame was cloned into the recombinant cell-free protein synthesis system was visualized via gel electrophoresis (Supplementary Methods and Supplementary Fig. 19a). For arginine or glycine residues present in Cas12i WT, the same overlapping PCR approach was used with no oligo for expirn oligos. Linear DNA templates for CRISPR-Cas systems might be the weak interaction of a CRISPR-Cas effector or RNP with mammalian genomic target DNA. In agreement with this hypothesis, ABR-001 exhibited higher binding affinity on dsDNA target compared to Cas12i WT (Supplementary Fig. 17), suggesting that the combined substitutions of D581R, I926R and V1030G in ABR-001 enhanced the catalytic activity of the effector through tighter binding of DNA substrates. In addition, modeling of D581R and I926R in Cas12i structure publications subsequent to the generation of ABR-001 indicates that these WED and Nuc domain substitutions could strengthen the electrostatic interaction with the dsDNA backbone of the PAM region and non-template strand DNA, respectively (Supplementary Fig. 18). The residue V1030 is located in the RuvC domain which functions to bind and cleave DNA substrate. The exact mechanism of the V1030G substitution in enhancing indel activity is not clear due to the lack of local structural information. It is possible that the V1030G substitution might influence target DNA interaction in a way that favors DNA cleavage. Rational approaches to engineering using recently published Cas12i structural information may further improve Cas12i editing efficiency and targeting specificity.

Cas12i is an attractive type V CRISPR-Cas nuclease for genome editing because of its compact size that fits in AAV vector with short 43-mer gRNA, absence of tracrRNA, ability to process pre-crRNA, and high specificity. With the exception of the recently discovered type V-J Cas12j, no other CRISPR-Cas nuclease possesses this unique combination of favorable properties (Supplementary Table 1). The robust activity and high specificity shown by the first engineered version, ABR-001, establishes Cas12i as a versatile genome editing platform with broad capabilities. Notably, ABR-001 demonstrated overall higher indel activity relative to the SaCas9 control with AAV delivery in HEK293T cells. Taken together, ABR-001 represents a high-performance genetic medicine platform with potential for both ex vivo and in vivo gene therapy.
according to the mean and standard deviation of the wild-type controls used within the same plate. Let $r_{x,y}$ be the depletion ratio measured for variant $x$, technical replicate $t$, and plate $p$. For a given plate $p$, set of wild-type variants $W_p$, and set of technical replicates $T_p$, let the plate control mean and standard deviation be given by:

$$\mu_p = \frac{1}{|W_p|} \sum_{w \in W_p} \sum_{t \in T_p} r_{w,t}$$

$$\sigma_p = \sqrt{\frac{1}{|W_p|} \sum_{w \in W_p} \sum_{t \in T_p} (r_{w,t} - \mu_p)^2}$$

The wild-type control-normalized Z-score for a variant $x$ screened in plate $p$ was computed as:

$$Z_{x,p} = \left( \frac{\sum_{w \in W_p} \sum_{t \in T_p} r_{x,w,t} - \mu_p}{\sigma_p} \right)$$

Only the data of variants with a coefficient of variation (CV) less than 20% were used. Those data of variants with CV > 20% were discarded. Furthermore, in cases where GFP or RFP fluorescence did not reach a minimum value of 60 rfU those data were discarded.

**Purification of ABR-001 protein.** The genes for Cas1212 WT and ABR-001 were cloned in a pET28-derived expression vector that introduced an N-terminal hexaHis tag and a C-terminal nuclear localization signal (NLS) tag. The plasmids were transformed into E. coli BL21 (DE3) (New England Biolabs #70326-3) cells for protein expression. Cells were grown at 37 °C in Terrific Broth in 2.5 l. Thomsom Ultra Yield flasks (Thomson #931136-B) until OD600 reached ~0.7–0.8. The temperature was then lowered to 20 °C. Protein expression was induced with 0.5 mM IPTG for 16–20 h before harvesting and freezing cells at −80 °C. Cell paste was suspended in a lysis buffer (25 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol, and 0.5 mM TCEP) (Thermo Fisher #77720). Cells were lysed using a cell disruptor (Constant Systems CFI). Polyethyleneimine (PEI) (MilliporeSigma #408727) (5%) was slowly added to the cleared lysate to a final concentration of 0.2% to precipitate nucleic acids. After centrifugation to remove the precipitates, the supernatant was loaded on a Sulfate 650F column equilibrated with Sulfate A buffer (20 mM Bis-Tris, pH 6.5, 5% glycerol, and 0.5 mM TCEP) and the bound proteins are eluted with 100-2000 mM NaCl gradient. The pooled peak fractions were diluted to −250 mM NaCl and loaded onto a Q Sepharose FF column equilibrated with Q buffer (25 mM Tris-HCl, pH 8.0, 250 mM NaCl, and 0.5 mM TCEP). The flow-through fractions were pooled and concentrated to a final concentration of 400 μM. Fractions were visualized via gel electrophoresis and the purified proteins were stored at −80 °C (Supplementary Fig. 20).

**In vitro RNA synthesis.** All mature crRNAs were purchased from IDT or generated by in vitro transcription of dsDNA PCR templates. All pre-crRNAs were generated by in vitro transcription of dsDNA PCR templates. In vitro transcription was performed using the HighScript 17 High Yield RNA synthesis kit (New England Biolabs #E2040S) at 37 °C for 3 h followed by treatment with Turbo DNase (Thermo Fisher #AM2238). The RNA was purified using Zymo RNA clean and concentrator (Zymo Research #R1013) using the manufacturer’s instructions.

**In vitro cleavage reactions and electrophoretic mobility shift analysis (EMSA).** For in vitro cleavage analyses, DNA substrates were generated by PCR amplification using IR800 and IR700 labeled forward and reverse primers, respectively, resulting in dsDNA targets with the same protocol as in PCR1 with 25 cycles of amplification. PCR2 samples were then pooled together and purified using DNA Clean & Concentrator—25 (Zymo Research #D4033). Libraries were sequenced on an Illumina NextSeq 550 system using v2.5 chemistry and a 1 × 150 read length.

**Sequencing library preparation.** Target-specific amplification (PCR1) was performed by combining 2 μl of cell lysate with 5 μl NEB Ultra II Q5 Master Mix (New England Biolabs #M0544S), 0.05 μl each of target-specific forward and reverse primer stock at 100 μM, and 3 μl of water. Primers were delivered to wells using an Echo 650 liquid handler. The reaction was mixed well and amplification was performed on a thermocycler (Initial Denaturation: 60 s at 98 °C, Amplification: 10 s at 98 °C, 15 s at 60 °C, 20 s at 72 °C, cycle count = 25. Final extension: 120 °C, 20 min). Sample indexing (PCR2) was performed by combining 0.5 μl of the PCR1, 5 μl NEB Ultra II Q5 Master Mix, 0.5 μl each of forward and reverse primer stock at 2 μM, and 3 μl of water. Amplification was performed using the same protocol as in PCR1 with 12 cycles of amplification. PCR2 samples were then pooled together and purified using DNA Clean & Concentrator—25 (Zymo Research #D4033). Libraries were sequenced on an Illumina NextSeq 550 system using v2.5 chemistry and a 1 × 150 read length.

**RNP complexing and delivery.** ABR-001 RNA complexation reactions were made by mixing purified ABR-001 (400 μM) with individual crRNAs (1 mM in 250 mM NaCl sequences in Supplementary Table 5) at a 1:1 (effector:crRNA) volume ratio (2.5:1 crRNA:effector molar ratio). For no guide control, ABR-001 was mixed with 250 mM NaCl at the same volume ratio as the crRNA. SprCas9 RNP complexation reactions were made by mixing purified SprCas9 (Aldevron; 62 μM) with sgRNA (1 mM in water; sequences in Supplementary Table 5) at a 6:4:5:1 (effector:sgRNA) volume ratio (2:5:1 sgRNA:effector molar ratio). Complexations were incubated on ice for 30–60 min prior to T cells or CD34+ HSPCs electroporation. During incubation, T cells or CD34+ HSPCs were harvested and electroporated as described below.

**Ex vivo editing of T cells.** Frozen human Peripheral Blood Mononuclear Cells (PBMCs) (StemCell Technologies #70025) from a donor were revived and counted using an automated cell counter. T cells were isolated from PBMCs using the EasySep Human T Cell Isolation Kit (StemCell Technologies #17951). Following isolation, a sample was collected and stained for CD3 for flow cytometry analysis of surface expression, to determine T cell purity of the isolated cells. Cell density was adjusted to 1×10^6 cells/mL and cells were stimulated for 3 days with a cocktail of anti-CD3/CD28 antibodies. Cells were cultured in fresh complete ImmunoCult-XF Cell
Expansion Medium (StemCell Technologies #10981) with 10 ng/mL IL-2 and 2 mM L-glutamine and supplemented with 25 μL/mL of ImmuNoCell Human CD3/ CD28 (StemCell Technologies). RNP complex reaction mixtures were made as described above. Diluted complexes were dispensed at 2 μL per well into Lonza 16-well nucleocuvette strips. Activated T cell suspensions were collected and counted using an automated cell counter. A sample of cells was collected and stained for CD25 for flow cytometry analysis, to determine activation efficiency. Cell density was adjusted to 1.1e7 cells/mL in P3 buffer (Lonza) and was dispensed at 2e5 cells/reaction (18 μL) into the Nucleocuvette strips. Cell density was adjusted to 1.1e7 cells/mL in P3 buffer (Lonza) and was dispensed at 2e5 cells/reaction (18 μL) into the Nucleocuvette strips. Strips were electroporated using an electroporation device (program EO-115, Lonza 4D-nucleofector), excluding the unelectroporated conditions. Immediately following electroporation, added 40 μL pre-warmed ImmuNoCell-XF+IL-2+L-Glutamine to cells and mixed gently by pipetting. For each technical replicate plate, plated 15 μL (~30,000 cells) of diluted nucleofected cells into pre-warmed 96-well plate with wells containing 200 μL ImmuNoCell-XF+IL-2+L-Glutamine. Editing plates were incubated for 7 days at 37 °C with 100 μL media replacement at day 4. After 7 days, wells were mixed by pipetting and 150 μL from each well was transferred to a fresh 96-well plate for staining. Pellets were spun down @ 400 x g for 10 min. Pellets were then resuspended in 200 μL of PBS. 100 μL of sample was collected and stained with either a fluorescently tagged antibody or LIVE/DEAD stain (Thermo Fisher #L34964) to assess viability. For genomic DNA extraction, we pelleted the remaining 50 μL of cell suspension by centrifugation, removed the supernatant, and resuspended pellets in 30 μL QuickExtract. The QuickExtract solution was subjected to the appropriate thermal profile (15 min at 65°C, 15 min at 68°C, 10 min at 98°C) to complete cellular and protein degradation.

**Ex vivo editing of CD34+ HSPCs.** Frozen bone marrow CD34+ cells were thawed and assessed for cell number and viability. CD34+ cells were maintained in culture for about 2 days prior to electroporation. For the in vitro experiments, about 100,000 cells were used per electroporation reaction in a 20 μL format using the P3 primary cell 4D-nucleofector kit S (Lonza #VXP-3032J). For the in vivo experiments, about 1,000,000 cells from a single donor were used per electroporation reaction in a 100 μL format using the P3 primary cell 4D-nucleofector kit L (Lonza #VXP-3024J). The cells were mixed with transfection enhancer oligo (~4 μM final concentration) and 100 μg/mL RNP complex (transfected cells). Pellet of cells were mixed with transfection enhancer oligo (~4 μM final concentration) at 100 μg/mL RNP complex (transfected cells). The cells were assessed for intracellular HbF staining as described above. Ex vivo edited bone marrow cells were collected post-transplantation and washed twice with FACS buffer (PBS + 2% FBS + 1 mM EDTA). Cells were resuspended in Mouse or Human FC Blocking solution (BD) and incubated at room temperature for 10 min. Cells were incubated with fluorescently-conjugated antibodies for 30 min at 4°C, then washed twice with FACS buffer. Cells were analyzed using the CytoFLEX (Beckman Coulter).

For viability staining, diluted 7-AAD was added to each well and incubated for 10 min at room temperature. Flow cytometry was performed using a CytoFLEX S instrument (Beckman Coulter), and data were analyzed using FlowJo v10.8.1 (BD). A representative gating example can be found in Supplementary Fig. 21.

**In vivo engraftment of ex vivo edited CD34+ HSPCs.** This work was covered under an ethics protocol reviewed and approved by the Institutional Animal Care Committee at the University of British Columbia under protocol A#A18-0276. During the study the care, housing and use of animals was performed in accordance with the Canadian Council on Animal Care Guidelines. Edited and control cells were resuspended at 1,000,000 cells/mL and prepared in PBS + 2% FBS for each of the conditions. 200 μL of cell suspension was then injected into the tail vein of groups of sub-lethally irradiated Il2rgtm1Wjl/SzJ (NSG) mice (4–5 mice/group) at a density of 200,000 cells/mouse. In addition 7 days prior to irradiation, and for the duration of the study, acidified water containing antibiotic was provided to the mice. At 8 and 16 weeks after injection, mice were sacrificed and the peripheral blood (PB) and bone marrow (BM) from each individual mouse was harvested and the engraftment of human cells was assessed by flow cytometry with both anti-human and anti-mouse CD45 antibodies. Flow cytometry data were analyzed using FlowJo v10.8.1 (BD). A representative gating example can be found in Supplementary Fig. 22. The number of cells was assessed for intracellular fetal hemoglobin staining as described above. The number of cells assessed by flow cytometry were counted and assessed by intracellular fetal hemoglobin staining as described above. The number of cells assessed by flow cytometry were counted and assessed by intracellular fetal hemoglobin staining as described above.

**Delivery of AAV-001 via AAV vector.** Prior to transduction, HEK293T cells were plated at 25,000 cells per well in 100 μL of D10 media into wells of a 96-well plate and grown at 37 °C for 18–24 h to 70–90% confluence. AAV master mixes were prepared by diluting AAV preps in 25% D10 media to a concentration double that required for transduction. Cells were transduced by removing 50 μL of media and replacing dropwise with 50 μL AAV particles diluted in D10 to produce a series of twofold dilutions ranging from a AAV GC-Cell ratio of 5120001 down to 400001. Negative controls were prepared by replacing the media of control wells with D10 media only. Post-transduction, cells were harvested and sequencing libraries were prepared as described above.

**Indel pattern analysis.** Indel rates and patterns were analyzed from targeted deep sequencing data. The analysis pipeline sampled up to 50,000 reads and used a kmer-scanning algorithm to calculate the edit operations (match, mismatch, insertion, deletion) between each read and the amplicon reference sequence. The indel rate was calculated as the number of reads containing an insertion or deletion divided by the total number of reads analyzed. Indel patterns were analyzed by aggregating all indel reads for a sample, binning the reads by indel length (size) and calculating the fraction of indel reads at each indel length (size).

**TTISS off-target screen.** The TTISS off-target screen was conducted as previously described with the following modifications. 375 μL 100 ng/μL transposon DNA was mixed with 375 μL glycerol and 750 μL EZ-Tn5 Transposase (Lucigen #TNP2110). The solutions were then mixed by vortexing for 5 s and incubated at room temperature for 30 min to form the loaded Tn5 transposase. Transposase was stored at −20 °C until ready for use.

Prior to transfection, HEK293T cells were plated in 24-well tissue culture-treated plates at a density of 125,000 cells per well in 500μL D10 media. Cells were transfected approximately 15 h after plating using GeneJuice transfection reagent (Millipore Sigma #70967) for each well to be transfected, 500 ng donor oligo at ~1.25 ng/μL, 375 ng ABR-001 effector plasmid at 1000 ng/μL, 125 ng guide plasmid at 100 ng/μL were added to Opti-MEM media to a final volume of 125μL. In a separate vessel, 125μL Opti-MEM media was mixed with 2.5μL GeneJuice transfection reagent and 6–8 μL cold NON ID C2.00k at room temperature for 5 min. Following the incubation, the DNA + Opti-MEM and GeneJuice+Opti-MEM solutions were combined and incubated at room temperature for 5–15 min.
After incubation, the combined solution was added dropwise to a single well of a 24-well plate. Transfected cells were incubated for approximately 72 h. Cells were then dissociated from the plate by removing media, washing once with 200μL PBS (Thermo Fisher #10010023), adding 50μL of TrypLE dissociation reagent (Thermo Fisher #126040413), and incubating at 37 °C for 5 min. Cells were then resuspended by adding 150μL of D10 media and mixing well. Resuspended cells were then transferred to a 96-well PCR plate and spun down at 400g for 10 min. The supernant was removed and cell pellets were stored at −20 °C until DNA extraction.

DNA was extracted as described and gDNA was purified using the Zymo gDNA clean and concentrator-5 kit (Zymo Research #D4067) following the manufacturer’s instructions and eluting in 35μL of 10 mM Tris-Cl. gDNA was visualized on a gel and quantified by the Qubit high-sensitivity dsDNA kit (Thermo Fisher Q38251) following the manufacturer’s instructions. gDNA extracts were then normalized to 50 ng/μL in 10 mM Tris-Cl.

Genomic DNA was tagmented by TagMe Reagent (Thermo Fisher #12604013), and incubating at 37 °C for 5 min. Cells were then prepared a solution 24 μL of D10 media and mixing well. Resuspended cells were incubated for approximately 72 h.

2. 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 (2020).
3. Netherton, M. A. et al. Cell-free transcription: real-time, repeat-based classification of CRISPR loci. CRISPR J. 4, 558–574 (2021).
4. Gillmore, J. D. et al. CRISPR-Cas9 in vivo gene editing for transhyretin amyloidosis. N. Engl. J. Med. 385, 493–502 (2021).
5. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278 (2014).
6. Doumda, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258906 (2014).
7. Yan, W. X. et al. Functionally diverse type V CRISPR-Cas systems. Science 363, 88–91 (2019).
8. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015).
9. Strehkendl, I., Safiuddin, F. A., Rybarski, J. I., Finkelstein, I. J. & Russell, R. Kinetic basis for DNA target specificity of CRISPR-Cas12a. Mol. Cell 71, 816–824 e813 (2018).
10. Kleinstiver, B. P. et al. Genome-wide specificities of CRISPR-Cas9 nucleases in human cells. Nat. Biotechnol. 34, 869–874 (2016).
11. Kuis, D. et al. Genome-wide analysis reveals specificities of Cas9 endonucleases in human cells. Nat. Biotechnol. 34, 863–868 (2016).
12. Streekler, J. et al. Engineering of CRISPR-Cas12b for human genome editing. Nat. Commun. 10, 219 (2022).
13. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. Mol. Cell 63, 633–646 (2016).
14. Cysky, J. et al. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. Elife 9, e55143 (2020).
15. Schmid-Burgk, J. L. et al. Highly parallel profiling of Cas9 variant specificity. Mol. Cell 78, 794–800 e798 (2020).
16. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas9 nucleases. Nat. Biotechnol. 33, 187–197 (2015).
17. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832 (2013).
18. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527, 192–197 (2015).
19. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. Cell 173, 430–442 e417 (2018).
20. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and nomenclature of CRISPR-Cas systems: where from here? CRISPR J. 1, 325–336 (2018).
21. Ran, F. A. et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191 (2015).
22. Xu, C. L., Ruan M. Z. C., Mahajan V. B. & Tsang S. H. Viral delivery systems for CRISPR. Viruses 11, 28 (2019).
23. Stadtmauer, E. A. et al. CRISPR-engineered T cells in patients with refractory cancer. Science 367, eaba7365 (2020).
24. Frangoul, H. et al. CRISPR-Cas9 gene editing for sickle cell disease and beta-thalassemia. N. Engl. J. Med. 384, 252–260 (2021).
25. Zhang, L. et al. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. Nat. Commun. 12, 3908 (2021).
26. Huang, X. et al. Structural basis for two metal-ion catalysis of DNA cleavage by Cas12a. Nat. Commun. 11, 5241 (2020).
27. Zhang, B. et al. Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12a. Nat. Commun. 12, 3476 (2021).
28. Cheng, J. S. et al. Enhancing CRISPR-Cas9 targeting accuracy. Nature 550, 407–410 (2017).
29. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84–88 (2016).
30. Pausch, P. et al. CRISPR-CasPhi from huge phages is a hypercompact genome editor. Science 369, 333–337 (2020).
31. Pausch, P. et al. DNA interference states of the hypercompact CRISPR-CasPhi effector. Nat. Struct. Mol. Biol. 28, 652–661 (2021).
32. McGaw, C. & Chong, S. Cell-free protein synthesis of CRISPR ribonucleoproteins (RNP). Methods Enzymol. 659, 371–389 (2021).
33. Hou, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional repression screening. Nat. Protoc. 12, 828–863 (2017).
34. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. Nat. Biotechnol. 19, 751–755 (2001).
39. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).

**Acknowledgements**

The authors would like to thank Kelly Morgan, John Murphy, Greg Hoffman, Craig Mak, and Devyn Smith for critical reading and comments on the manuscript. We thank Amy Biasella, Austin Jones and Greg Hoffman for helping analyzing in vitro screen data, Lauren Killingsworth and Erin Doherty for contributing to biochemical experiments, and entire Arbor Biotechnologies team for support.

**Author contributions**

S.C. and T.D. conceived and designed the project with input from P.H., W.X.Y., and D.A.S. C.M. established in vitro fluorescent reporter assay and performed in vitro screen experiments. G.Z.M., A.J.G., W.X.Y., C.M., W.L., D.C. performed mammalian cell screen experiments. A.J.G., G.Z.M., and Q.W. conducted off-target experiments, prediction and analysis. J.R.H. and S.S. conducted ex vivo primary T cell and CD34+HSPC experiments. A.J.G and M.B. performed AAV experiments. A.O. and P.H. performed biochemical and cut site characterization experiments. E.K.S., J.M.C., S.S., and B.H. performed protein purification and RNP production. E.K.S. conducted EMSA experiments. B.H. conducted structural analyses. P.Y. and H.Z. conducted sample preparation and deep sequencing experiments. L.E.A., R.Z., D.R.C., and D.A.S conducted characterization and phylogenetic analysis for Cas12i systems. Q.W., N.J., R.Z., and D.A.S analyzed all deep sequencing and indel data. S.C. and T.D. wrote the manuscript with input from all authors. S.C. and T.D. contributed equally to this work.

**Competing interests**

C.M., A.J.G., G.Z.M., J.R.H., S.S., E.K.S., M.B., A.O., Q.W., N.J., P.Y., H.Z., L.E.A., R.Z., D.R.C., and T.D. are current employees and shareholders of Arbor Biotechnologies (Arbor). W.X.Y. and D.A.S are co-founders and shareholders of Arbor. P.H., J.M.C., W.L., D.C., B.H., S.S., and S.C. are former Arbor employees and were employed when this work was conducted. Some of the content in the manuscript has been included in a patent application published as WO 2021202800 on October 7, 2021. S.C., B.H., Q.W., N.J., R.Z., J.M.C., T.D., J.R.H., A.J.G., C.M., D.A.S., and D.C. are listed as inventors on WO 2021202800.

**Additional information**

**Supplementary information** The online version contains supplementary material available at [https://doi.org/10.1038/s41467-022-30465-7](https://doi.org/10.1038/s41467-022-30465-7).

**Correspondence** and requests for materials should be addressed to Tia DiTommaso.

**Peer review information** *Nature Communications* thanks Songying Ouyang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at [http://www.nature.com/reprints](http://www.nature.com/reprints)

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

---

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

© The Author(s) 2022