Redesign of extensive protein–DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization

Ryo Takeuchi, Michael Choi, and Barry L. Stoddard

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Edited by David Baker, University of Washington, Seattle, WA, and approved February 13, 2014 (received for review November 8, 2013)

LAGLIDADG homing endonucleases (meganucleases) are sequence-specific DNA cleavage enzymes used for genome engineering. Recently, meganucleases fused to transcription activator-like effectors have been demonstrated to efficiently introduce targeted genome modifications. However, retargeting meganucleases to genomic sequences of interest remains challenging because it usually requires extensive alteration of a large number of amino acid residues that are situated in and near the DNA interface. Here we describe an effective strategy to extensively redesign such an extensive biomolecular interface. Well-characterized meganucleases are computationally screened to identify the best candidate enzyme to target a genomic region; that protein is then redesigned using iterative rounds of in vitro selections within compartmentalized aqueous droplets, which enable screening of extremely large numbers of protein variants at each step. The utility of this approach is illustrated by engineering three different meganucleases to cleave three human genomic sites (found in two exons and one flanking intron in two clinically relevant genes) and a fourth endonuclease that discriminates between single-nucleotide polymorphism variants of one of those targets. Fusion with transcription activator-like effector DNA binding domains significantly enhances targeted modification induced by meganucleases engineered in this study. Simultaneous expression of two such fusion endonucleases results in efficient excision of a defined genomic region.

Significance

Specific gene-targeting nucleases are critical for genetic engineering and therapy. Meganucleases have been shown to harbor many properties, including small monomeric folds and high cleavage specificities, that are appropriate for this purpose. However, they are rarely used for genome modification because their DNA specificity requires an extensive network of up to 50 amino acids, making redesign difficult. Because extensively reprogramming molecular recognition across such a complex interface is a fundamental biotechnology challenge, meganucleases are used here as a model system for the development of an efficient approach to alter their specificity. The nucleases produced by this method, when tethered to transcription activator-like effector domains, resulted in highly active DNA targeting reagents with notable recognition properties.

Author contributions: R.T., M.C., and B.L.S. designed research; R.T. and M.C. performed research; R.T. and M.C. contributed new reagents/analytic tools; R.T. and B.L.S. analyzed data; and R.T. and B.L.S. wrote the paper.

Conflict of interest statement: B.L.S. is a founder of a biotechnology startup company (Pregenen, Inc.) that manufactures and applies engineered meganucleases for research and medicine. None of the reagents or materials in this study are related to the work at that company.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: bstoddar@fhcrc.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321030111/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1321030111

PNAS | March 18, 2014 | vol. 111 | no. 11 | 4061–4066

protein engineering | gene modification | gene therapy

Genome engineering is a discipline in which specific chromosomal loci are altered at precisely defined sites, producing cells and organisms with heritable DNA sequence modifications (1). Many tools have been developed for this purpose, including zinc finger nucleases (ZFNs) (2), transcription activator-like effector nucleases (TALENs) (3), CRISPR/Cas9 endonucleases (CRISPRs) (4, 5), and meganucleases (6). ZFNs can be purchased or (with effort) engineered using publicly available resources (7, 8) but often display measurable off-target activity (9, 10). In contrast, TALENs and CRISPRs can be more easily reprogrammed to a wider range of DNA sequences (11), but their lengthy reading frames may complicate packaging and delivery in certain contexts and applications (12). A recent study indicated that guide RNA–Cas9 complexes may induce significant off-target mutagenesis (13–16); however, the simultaneous generation of single-stranded breaks at closely spaced sites on opposing DNA strands, using a Cas9 nickase and multiple guide RNAs, may reduce off-target mutagenesis (17, 18).

Meganucleases are naturally occurring, compact DNA cleavage enzymes that recognize long (~20 base pairs) DNA targets (19). Several studies have demonstrated that these nucleases can be retargeted to DNA sequences of interest, albeit with considerable effort (20). The challenge in altering their specificity lies in the same property that makes them excellent reagents for genome engineering: their DNA binding and cleavage functions are intimately coupled to one another, and a significant portion of their structure is engaged in a complex network of direct and indirect DNA contacts involving up to 50 separate amino acid residues. Although an industrial-scale engineering pipeline for altering meganuclease specificity has been developed (6), the specialized knowledge and technology required for that approach generally precludes their use by academic laboratories. In addition, a recent study has demonstrated that MegaTALs (fusions of TAL effector DNA binding regions to meganucleases) can induce a high level of targeted gene modification in human cells (21). However, the utility of this platform still depends upon the development of efficient strategies to engineer meganucleases with desired specificity.

Meganucleases and zinc finger arrays for ZFNs have been generally redesigned using cell-based selection systems, where cell culturing is time-consuming and the efficiency of transformation limits the number of mutants to be screened (6, 7, 22). In contrast, selection systems such as in vitro compartmentalization (IVC) and mRNA display (in which protein-coding libraries are readily generated by PCR-based approaches) allow for screening much more complex libraries (10^9 to 10^12 of variant genes) (23, 24). IVC has been exploited to isolate restriction enzymes from bacterial genome libraries (25) but not to extensively modify gene-targeting nucleases by multiple cycles of mutagenesis and selection. Here we describe a strategy for reprogramming meganuclease cleavage specificity, wherein (1) a genomic target region to be modified is computationally scanned to identify sites that are at least 50% identical to target sequences for well-characterized meganucleases and (2) variant meganucleases with
desired specificity are then engineered using IVC (Fig. 1A and Fig. S1). This approach is facilitated by a growing collection of meganucleases that have been crystallized in complex with their DNA targets (26).

**Results**

**Identification of Genomic Sequences to Be Targeted by Engineered Meganucleases.** To develop a strategy to efficiently retarget multiple types of meganucleases to genomic targets of interest, we selected two genome loci from the human genome for targeted modification on the basis of our research interest in potential therapeutic applications for hemoglobinopathies and cystic fibrosis (Fig. 1B): the exon 4 of *bcl11a* (which encodes a transcription regulator involved in B lymphopoiesis, erythropoiesis, and regulation of globin expression (27)) and the exon 11 in *cfr* (which encodes a transmembrane conductance regulator which is dysfunctional in individuals with cystic fibrosis, most often as a result of deletion of codon 508 within that exon (28)).

Using a publically available bioinformatic meganuclease Web server (LAHEDES; www.homingendonuclease.net) (26), we searched each of these regions for 22 base pair sequences that displayed significant identity to DNA targets for wild-type, monomeric meganucleases that have been crystallized in complex with their DNA targets (Fig. S2). An increasing number of such structures are now available, including I-Onul (PDB ID 3QQY), I-Ltrl (3RT7), I-PanMI (4JEE), I-GzeMI (4EFJ), I-HjeMI (4UVE), and I-Trw1 (4LQO), and I-SmaMI (4LOX). We excluded candidate target sites that contained multiple nucleotide mismatches within their central four base pairs compared with the wild-type endonuclease targets because base substitutions at those base positions may lead to DNA backbone distortion that significantly compromises meganuclease specificity and activity (29).

**Materials and Methods**

**Fig. 1.** Schematic of the IVC method used for selection of retargeted meganucleases (A) and human genome target sites for engineered meganucleases (B). (A) Individual PCR fragments containing variant meganuclease genes (blue box) and target sites (red boxes) are compartmentalized with in vitro transcription/translation components in oil-surfactant mixture. Successfully expressed endonuclease variants can only cleave target sites coupled to their own genes, allowing for maintenance of genotype-phenotype linkage during selection. Variant endonuclease genes coupled to cleaved target sites can be recovered by ligation to DNA adaptors containing a complementary sticky end to that generated by engineered meganuclease, followed by PCR amplification with a set of primers, one of which is specific to an adaptor sequence (green arrow). DNA fragments containing cleaved target sites are amplified from DNA libraries subjected to in vitro selection (+ lanes) but not from naive (unselected) ones (− lanes) (see Fig. S1 for more detail). (B) Nucleotide sequences that are substituted from the original target sites recognized by wild-type meganucleases are highlighted in red. The asterisk indicates a position of the SNP observed in the *cfr* exon 11 of the human population, where an A:T or G:C base pair has been observed at an ~50:50 ratio.

We ultimately chose three target sites within the two target loci that exhibited 50–70% identity against the target sequences recognized by different wild-type meganucleases (I-Onul, I-PanMI, and I-HjeMI) (Fig. 1B). Two of the three sites are situated in exons, and the other is in a flanking intron region. A site for a redesigned I-PanMI meganuclease (CFTR2) (19) was a well-known human single-nucleotide polymorphism (SNP) (A/G in *cfr* exon 11, rs213950; www.ensembl.org). We therefore also sought to create two variant endonucleases that each act respectively at only one of the SNP variants (to assess whether this approach can produce reagents to discriminate between such closely related DNA sequences).

**Reprogramming of Meganuclease Specificity.** Structural studies of meganucleases have demonstrated that two to four consecutive nucleotide base pairs within a DNA target site are generally recognized by a cluster of six to nine amino acid residues (26). These protein regions are termed “contact modules” and are cataloged for each meganuclease in the LAHEDES Web server. We randomized amino acid residues within those modules to alter the specificity of each meganuclease at altered base positions and constructed variant endonuclease libraries by overlap extension PCR using degenerate primers. Oligonucleotides used for redesign of meganucleases and amino acid positions subjected to randomization are shown in **Dataset S1**.

Shuffling protein residues at up to nine positions gives rise to a large theoretical number of unique variant endonucleases (as large as 209, or ~5 × 1011). To screen as many unique protein variants as possible from randomized libraries, we modified a previously described in vitro compartmentalization system (23, 25), where DNA fragments encoding variant endonucleases are individually expressed in compartmentalized, aqueous droplets that contain an in vitro transcription/translation reaction mixture (Fig. 1A and Fig. S1). In each droplet, a variant endonuclease has the opportunity to bind and cleave a target site that is placed in the same DNA construct as for its own gene, allowing us to couple the cleavage activity against the target site to the corresponding meganuclease gene.

To identify variant enzymes that are capable of cleaving a target site containing a cluster of altered base pairs, we carried out two to three rounds of selection under increasing levels of stringency (by reducing time periods for in vitro protein synthesis and DNA strand cleavage and/or by elevating temperatures during each round; see **Materials and Methods** for more detail). We continued with an iterative approach, in which variant genes enriched after the first rounds of site-directed mutagenesis were used as templates for further randomization of additional amino acids and subsequent selections against a target site containing more base pair substitution(s) (as illustrated in Fig. S3). Using this strategy, we redesigned the N- and C-terminal half domains of meganucleases to engineer variant enzymes that cleaved chimeric target sites composed of one half of human genomic target sites and one half of the original meganuclease target sites. We then shuffled together the engineered N- and C-terminal half domains, and identified a collection of variant endonucleases that recognized human genomic target sequences in the IVC selection system.

We then filtered these populations of active meganucleases through cleavage assays in bacteria (30) to select individual clones that displayed substantial activity in living cells. In this assay, meganucleases that are expressed in individual bacterial cells can rescue the cell growth only when they recognize and cleave their target sequences on reporter plasmids that bare such cells harbor, leading to elimination of the reporter plasmid that can induce expression of a toxic gene. If necessary, random mutagenesis was introduced to incorporate additional point mutations in the entire ORFs of variant endonuclease genes that might increase overall activity and/or stability.

The final engineered endonucleases all efficiently increased the survival rate of bacterial cells that harbored reporter plasmids.
containing their target sites but not those containing the original target sites recognized by their parental meganucleases (Fig. 2A). Sequencing of the retargeted meganucleases indicated that 21–28 amino acid substitutions were introduced into each wild-type endonuclease scaffold, with a small number of substitutions (0–5) found outside of the protein–DNA interface (Fig. S4). The two I-PanMI–derived meganucleases (termed “CFTRpan/A” and “CFTRpan/G”) that targeted two SNP variants within the CFTR2 site displayed excellent discrimination between those targets. Out of total of nine amino acid residues that differ between those redesigned meganucleases, six were situated on the DNA interface adjacent to the SNP base position.

**Targeted Mutagenesis by Redesigned Meganucleases at Their Endogenous Target Sites in Human Cells.** We next examined if the retargeted meganucleases could induce targeted modification at their endogenous sites in human embryonic kidney (HEK) 293T cells. Sequencing of the cftr exon 11 indicated that this cell line harbored only the SNP variant A; we therefore excluded the CFTRpan/G endonuclease from this analysis. All three engineered endonucleases were expressed at similar levels (Fig. S5A). We carried out T7 endonuclease I assays (31) to estimate the accumulation of small nucleotide insertions and deletions (indels) induced by the respective meganucleases. These assays revealed that expression of each endonuclease (either alone or in concert with the Trex2 exonuclease (32)) resulted in indels at the endogenous target sites in unsorted cell populations, with frequencies ranging from 2.5 to 34% (Fig. 2B).

The range of indel frequencies induced by the three redesigned enzymes at their corresponding target sites displayed considerably greater variation than their corresponding activities in the bacterial gene elimination assay (Fig. 2A). Most notably, the two cftr-targeting meganucleases (CFTRonu and CFTRpan/A) induce indels at their respective human genomic target sites (that are only 250 base pairs apart) with at least 10-fold different efficiencies. Additional assays in the same cell line, using an Episomal Direct Repeat (DR)-GFP reporter (33), also indicated that CFTRonu was more active (Fig. S5B), suggesting that the wide range of activities displayed by these redesigned meganucleases in human cells is not caused solely by differences in chromatin structures and/or nucleosome positioning at their target sites (34). Bioinformatic search for sites that were similar to the on-targets for the engineered meganucleases indicated that the human genome contains far more closely related sites (which contain no more than three base pair differences) for CFTRpan/A (a total of 32 sequences) than for CFTRonu (only eight such sites) (Table S1). It is therefore possible that a subset of closely related off-target sites may trap meganucleases and thereby act as competitive inhibitors because meganucleases have been shown to display more promiscuous binding specificity than DNA strand cleavage specificity (22, 35). However, we cannot rule out the possibility that these two meganucleases may have significant differences in activity that are not fully recapitulated in bacterial nuclease assays.

The same assays conducted using episomal DR-GFP reporter plasmids verified that the two I-PanMI–derived endonucleases discriminated between the SNP variants of the CFTR2 site (Fig. S5B), although the signal to noise ratio in these assays is relatively low because of spontaneous homologous recombination occurring within the episomal reporters.

**Impact of Tethered TAL Effectors on Targeted Mutagenesis by Retargeted Meganucleases.** We next examined whether tethering engineered meganucleases to TAL effectors (as additional DNA binding domains) would improve the efficiency of targeted genome modification in human cells as recently described (21). The two redesigned meganucleases that targeted the cftr locus were each tethered to a TAL effector DNA binding domain (termed “NΔ148/C+63,” a construct that spans from residue 149 of the canonical TAL effector N-terminus to residue +63 beyond the end of the TAL effector repeat sequences). The TAL effectors were individually designed to target a 10- or 12-nucleotide sequence that was six base pairs upstream of each meganuclease target (Fig. 2C). The frequencies of indels generated by both engineered meganucleases at the same endogenous target sites were increased after incorporation of the nuclease into the MegaTAL architecture, and at least a 10-fold improved efficiency was observed in the case of CFTRpan/A (Fig. 2B).

**Targeted Deletion by Multiplexed Coexpression of Two MegaTALs.** Previous studies have demonstrated that introduction of site-specific double strand breaks at two genomic loci promotes
targeted deletion and insertion (36–38), which expand the range of genome engineering applications. We therefore examined whether a pair of MegaTALs were capable of inducing targeted chromosomal deletions. Expression of two MegaTALs that targeted sites ∼250 base pairs apart gave rise to three DNA bands when a genomic region spanning their target sites was PCR-amplified using a single set of primers (Fig. 3A). Sequencing of the smallest DNA band revealed that approximately a half of the targeted deletions (DELI: 6 clones out of 12) were generated through three-base microhomology-mediated end joining between 3′, four-base overhangs generated by the two engineered meganucleases without any end processing (Fig. 3B). Sequencing of the larger DNA band identified two independent events: one was a chromosomal duplication, whereas the other was a targeted inversion and insertion of a small fragment derived from an expression plasmid for the two MegaTALs (Fig. S6A). Sequencing of clones harboring the middle-sized DNA band identified one inversion event (out of 30 clones sequenced: Fig. S6B). Deletion, duplication, and insertion promoted by the MegaTALs were substantially suppressed by coexpression of Trex2 (Fig. 3C). Sequencing analysis of deletion junctions (Fig. S6C) identified one inversion event (out of 30 clones sequenced: Fig. S6D). Se-

**Off-Target Cleavage Activity by an Engineered Meganuclease.** To assess off-target mutagenesis by engineered meganucleases, we assayed indels in cells expressing the CFTRonu meganuclease at 8 genomic sites that were the closest to its on-target sequence (all containing 3 base substitutions from the CFTR1 target site: Fig. 4A). When Trex2 was coexpressed, both the stand-alone CFTRonu and its corresponding MegaTAL promoted low but detectable levels of indels at two of the off-target sites (Fig. 4B). Comparable frequencies of off-target mutagenesis were induced by the two types of meganuclease constructs, suggesting that the fused TAL effector had little impact on binding and cleavage activity of an engineered meganuclease at off-target sites, presumably because of high binding affinity of the stand-alone nuclease toward the noncognate DNA target sequences. These results are consistent with our earlier study of MegaTALs (21). To reduce the level of undesired off-target mutagenesis, further introduction of mutagenesis into the engineered meganuclease and counterselection against the off-target sites might be useful. However, an alternative (and perhaps more straightforward) approach that could also improve specificity might be the simple incorporation of a point mutation in the nuclease domain that attenuates its overall cleavage activity, to increase the dependence of cleavage activity on simultaneous, cooperative recognition of the full-length target by the two DNA binding domains (i.e., the TAL effector and the meganuclease). Such point mutations can be easily identified based on crystal structures. This concept is illustrated by the results of experiments using the CFTRonu variant meganuclease that contains a single substitution (L49Q) in the vicinity of the active site: this point mutation suppressed the formation of indels at both the on- and off-target sites (Fig. 4B). Targeted modification was specifically enhanced at the on-target site by incorporating the CFTRonu L49Q meganuclease into the MegaTAL platform, although the efficiency of on-target mutagenesis was significantly compromised relative to that induced by constructs containing the original CFTRonu meganuclease.

**Discussion**

The ability to reliably and efficiently redesign biomolecular interfaces and thereby drive predictable and controllable genotypic and/or phenotypic changes is a challenge of great importance in many areas of synthetic biology and bioengineering. The type of recognition mechanism used by meganucleases, involving a highly networked interface of up to 20% of the protein (i.e., as many as 50 residues) and a DNA target of 20 base pairs, is an

---

**Fig. 3.** Targeted genomic deletion by coexpression of the two MegaTALs. (A) A schematic of the cfr gene exon 11, its flanking introns, target sites for TALE-CFTRpanA and TALE-CFTRonu, and binding sites for primers used is shown in Upper. Targeted deletion and duplication/insertion were detected only in cells expressing the Trex2 exonuclease (lane 8). Asterisks indicate PCR fragments containing duplication/insertion or deletion. (B) Sequencing analysis of deletion junctions generated by coexpression of the two MegaTALs. Target sequences for CFTRpanA and CFTRonu are in red and orange, respectively, and their central four base positions are underlined. The DEL1 sequence was obtained six times, and the rest of the DEL sequences were found once.

**Fig. 4.** Analysis of off-target mutagenesis induced by the CFTRonu meganuclease. (A) Sequences of human genomic sites that are similar to the CFTR1 site. Base substitutions from the CFTR1 site are shown in red. (B) The T7EI assays (31) were carried out, and indels were detected at two closely related sites (off-targets #3 and #8) in cells coexpressing Trex2. The higher background of nonspecific cleavage products produced by T7 endonuclease I for the on-target site is due to a larger PCR fragment spanning the genomic target (∼850 base pairs) than those that were generated for the off-target sites (∼250 base pairs).
excellent representative system for the type of problem in molecular engineering that is difficult to address experimentally.

In this study, we retargeted three wild-type meganucleases to human genome sequences that differed from their original targets at up to 50% of the base positions (a process that eventually requires engineering of the central 4–8 bases of the target sites for variant endonuclease) and verified that the engineered meganucleases all induced targeted genome modification. Given that this approach allows for targeting any 22-bp sequence where the central 4-bp sequence is conserved and where up to 11 nucleotides are mismatched at any of the remainder of the base positions, the theoretical probability of finding sites that can be targeted using this strategy is 1 in 1,842 base pairs when using a single meganuclease. Using as few as six monomeric meganucleases that cleave target sites with entirely unique central 4–bp sequences (I-HjeMI, I-Onul, I-GzeMIII, I-PanMI I-SnaMI, and I-Ltrl/I-LtrWI), this probability is increased to 1 in 307 base pairs. This may be a conservative approximation because nearly the entire DNA interface of wild-type endonucleases (except regions adjacent to the central four base positions) appear to be amenable to randomization and selection (Fig. S4B).

The growing availability of crystal structures of wild-type meganucleases, combined with continued refinement of the IVT selection method, should further expand the sequence space that can be addressed using meganucleases and MegaTALs.

We showed that combining TAL effector DNA binding regions with engineered meganucleases into a single polypeptide led to an increase in the frequency of targeted modification. Not only could this MegaTAL platform significantly reduce the burden of optimizing meganuclease activity during engineering process, but it also might improve the off-target to on-target ratio, particularly when an engineered meganuclease that hardly induces targeted modification alone is fused instead of a highly active meganuclease that may raise an off-targeting issue.

Although the development of easily programmable genome engineering tools enables their routine use in a wide range of cell types and model organisms, concerns about off-target mutagenesis, particularly in therapeutic applications, must be resolved. In addition to DSBs at single sites that primarily lead to indels, those generated simultaneously at two chromosomal loci, albeit rare events, can promote even more significant changes of chromosomes including translocation (40). Although further experiments need to be conducted, this study suggests that an engineered meganuclease tethered to a sequence-specific DNA binding domain such as a TAL effector and zinc finger array (which recognizes extremely long target sequences) may be one of platforms that can be exploited for therapeutic applications requiring exceptionally high demands for sequence specificity.

Materials and Methods

IVC. The ORFs of meganucleases, I-Onul E178D, I-HjeMI, and I-PanMI, were cloned between NcoI and NotI sites of pET21d(+)/EMD Millipore). The sequences are shown in Dataset S1. To introduce site-directed saturation mutagenesis into the ORS, sequences containing their partial ORFs with regions (−20 base pairs) that overlapped adjacent PCR fragments at both ends were amplified in separate tubes, using primers that contained degenerate codon NNNK (coding all 20 amino acids) (Fig. S1A). The positions of amino acid residues shuffled and PCR primers used are also shown in Dataset S1. PCR products were purified by extraction from agarose gels and assembled in the subsequent round of PCR with a sequence containing two copies of target sites for variant endonucleases to construct DNA libraries (Fig. S1B). A successfully assembled DNA fragment was again purified by gel extraction (Fig. S1C).

Two to three rounds of IVC were conducted following each round of site-directed saturation mutagenesis to enrich active variant meganucleases. The oil-surfactant mixture [2% (vol/vol) ABIL EM 90 (gift from Evonik Industries AG Personal Care) and 0.05% Triton X-100 in light mineral oil] was thoroughly stirred at 1,400 r.p.m. for 3.5 min on ice (Fig. 1A and Fig. S1D). Eight nanograms of a DNA library were added in the aqueous phase in the first round of IVT and then heated at 75 °C for 15 min. After an addition of 170 μL of 10 mM Tris HCl (pH 8.0), emulsified droplets were collected by centrifugation at 16,000 × g for 15 min at 4 °C and broken by an addition of phenol/ chloroform/soyamyl alcohol. Nucleic acids were recovered by isopropanol precipitation and treated with 5 μL of RNase Mixture Enzyme Mix (Life Technologies) at 37 °C for 30 min. After purification using QiAquick PCR purification kit (Qiagen), a gene library was ligated to a more than 100-fold excess of a (unphosphorylated) DNA adaptor with a 4-base overhang sequence complementary to a sticky end of target sites generated by variant endonucleases during IVC (Fig. S1D) and added to PCR mixtures containing a pair of primers, one of which was specific to a DNA adaptor to enrich ORFs of variant genes coupled to cleaved target sites (Fig. S1F).

In the second round of selection, compartmentalized reaction mixtures were incubated at 42 °C for 75 min, and the third round of IVT was carried out at the same temperature for 30 min.

Two-Plasmid Cleavage Assay in Bacteria. The assays were carried out based on previous studies (30, 41). Briefly, the NovaXGF competent cells harboring a pCdbB reporter plasmid (that contained four copies of a target site) were transformed with the Endo plasmid encoding a meganuclease by electroporation. The transformants were grown in the 2 × YT medium (16 g/L tryptone, 10 g/L yeast extract, 5.0 g/L NaCl) supplemented with 0.02% L-arabinose at 37 °C for an hour to induce expression of meganuclease genes and screened on both the nonselective plates (1 × M9 salt, 1% glycerol, 0.8% tryptone, 1 mM MgSO4, 1 mM CaCl2, 2 μM thiamine, and 100 μg/mL kanamycin) and the selective plates (the nonselective plates supplemented with 0.02% L-arabinose and 0.4 mM isopropyl-β-D-thiogalactopyranoside) to induce expression of the toxic CcdB protein. After incubation at 37 °C for 18–24 h, colonies were counted to calculate the survival rates on the selective plates.

I-Onul variant endonucleases that were engineered through IVC were cloned between NcoI and NtNotI sites of pEndo and subjected to two rounds of selection under the conditions described above. The pEndo plasmids were extracted from individual colonies and sequenced. When engineered meganuclease derived from I-HjeMI and I-PanMI were screened, a pool of variant genes were subjected to selection under the conditions where meganuclease genes were expressed in the 2 × YT medium containing 0.02% L-arabinose and 100 μg/mL carbenicillin at 30 °C for 15 h before plating as described in our previous study (41). Surviving colonies were scraped off from the selective plates, and random mutagenesis was introduced into the variant genes using Gene Morph II Random Mutagenesis Kit (Agilent Technologies). The resultant PCR fragments were cloned into the same vector and screened with an increasing stringency through four rounds of selection. Expression of endonuclease genes was induced at 30 °C for 4 h before plating in the first two rounds (plates were incubated at 30 °C) and at 37 °C for an hour in the subsequent two rounds (plates were incubated at 37 °C). ORFs of variant endonucleases were PCR-amplified from survival clones and cloned into the pEndo vector containing the C-terminal tag (AANEDYAAA) that facilitates proteolytic degradation in Escherichia coli as described previously (42). The sequence of the plasmid (termed “pEndoDegron”) is shown in Dataset S1. Two more rounds of selection were carried out under the same conditions as for the last two rounds, and ORFs of retargeted meganuclease genes recovered from survival colonies were sequenced. We expected that incorporation of the C-terminal degradation tag would accelerate turnover of a meganuclease, facilitating identification of engineered enzymes with higher activity; however, relatively little improvement was observed.

Plasmids Used for Transformation of HEK 293T Cells. Meganuclease and MegaTAL genes with the N-terminal HA epitope tag and nuclear localization signal and the C-terminal 6×His tag were cloned into pEG3 (Addgene plasmid 39991) using Gibson Assembly Master Mix (New England Biolabs) or standard molecular cloning techniques. Double-stranded oligonucleotides containing each of the target sites for meganucleases and a stop
centrator-5 kit (Zymo Research), 200 μM TCA TGT GCC CCT TCT C-3

genomic target region was PCR-amplified using a pair of primers (5′-TCT TGG ATT TCC AGG AGT TTT TCC TTA TGG TTA TTT TCT C-3′). Three days after transfection, genomic DNA was extracted, and the HEK 293T cells were transfected with 0.2 μg of DNA of pExodus plasmid for meganucleases and a Dr-GFP reporter plasmid containing a target site and 0.08 μg of pExodus-mCherry, as described above. Two days after transfection, cells were analyzed by flow cytometry, and the transfection efficiency was normalized based on a fraction of mCherry-positive cells.

PCR fragments were cloned using TOPO TA cloning kit (Life Technology) and sequenced.

**Episomal GFP Reporter Gene Recombination Assay.** HEK 293T cells were transfected with 0.36 μg each of the pExodus plasmid for meganucleases and a DR-GFP reporter plasmid containing a target site and 0.08 μg of pExodus-mCherry, as described above. Two days after transfection, cells were analyzed by flow cytometry, and the transfection efficiency was normalized based on a fraction of mCherry-positive cells.

**Western Blotting.** Transfected cells were lysed in 20 mM Tris HCl (pH 8.8), 500 mM NaCl, 1 mM EDTA, and 1% Triton X-100 and centrifuged at 16,000 × g for 30 min at 4 °C following incubation on ice for 20 min. Three micrograms of each supernatant were separated on a 15% (wt/vol) polyacrylamide-SDS gel and transferred to a PVDF membrane. HA-tagged proteins and t-Jubulin were detected using anti-HA polyclonal antibody (1:200; Y-11; Santa Cruz) and anti-Jubulin monoclonal antibody (1:500; clone TUB 2.1; Sigma Aldrich) as primary antibodies, and ECL anti-rabbit IgG HRP-linked whole antibody (1:25,000) and ECL anti-mouse IgG HRP-linked whole antibody (1:4,000; GE Healthcare) as secondary antibodies. Protein-antibody complexes were detected using ECL Western Blotting Detection reagent (GE Healthcare Life Sciences) and X-ray films (Thermo Scientific).

**ACKNOWLEDGMENTS.** The I-PanMI homing endonuclease and target site were provided for this study by Dr. Abigail Lambert. Ms. Taylor contributed to IVC selection experiments against the crf target locus. The expression plasmid for the mouse Trex2 endonuclease, pExodus-CMV.Trex2 (Addgene plasmid 40210), was a gift from Kamila Gwiazda and Dr. Andrew Scharenberg (Seattle Children’s Research Institute). This work was funded in part by a fellowship from the Japan Science Foundation (to R.T.) and by a grant from the National Institutes of Health (R01 GM08537 to B.L.S.).

**codon** were inserted into the DR-GFP reporter plasmid (33). Sequences of individual plasmids used in this study are shown in Dataset S1.

**Chromosomal Modifications Generated by Coexpression of the Two MegaTALs.** HEK 293T cells were transfected with 0.2 μg each of the two pExodus plasmids expressing MegaTALs and 0.4 μg of pExodus.CMV.Trex2, as described above. Three days after transfection, genomic DNA was extracted, and the genomic target region was PCR-amplified using a pair of primers (5′-TCT TGG ATT TCC AGG AGT TTT TCC TTA TGG TTA TTT TCT C-3′). PCR fragments were separated on a 1.5% (wt/vol) agarose-TBE gel. Gel-extracted DNA fragments were separated on a 1.5% agarose-TBE gel. The frequency of indel was estimated using the following equation as described previously (43): (% indel) = 100 × (1 – (1 – fraction cleaved))^{2/3}.

**Table S2.** After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2. After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2. After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2. After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2.

**Western Blotting.** Transfected cells were lysed in 20 mM Tris HCl (pH 8.8), 500 mM NaCl, 1 mM EDTA, and 1% Triton X-100 and centrifuged at 16,000 × g for 30 min at 4 °C following incubation on ice for 20 min. Three micrograms of each supernatant were separated on a 15% (wt/vol) polyacrylamide-SDS gel and transferred to a PVDF membrane. HA-tagged proteins and t-Jubulin were detected using anti-HA polyclonal antibody (1:200; Y-11; Santa Cruz) and anti-Jubulin monoclonal antibody (1:500; clone TUB 2.1; Sigma Aldrich) as primary antibodies, and ECL anti-rabbit IgG HRP-linked whole antibody (1:25,000) and ECL anti-mouse IgG HRP-linked whole antibody (1:4,000; GE Healthcare) as secondary antibodies. Protein-antibody complexes were detected using ECL Western Blotting Detection reagent (GE Healthcare Life Sciences) and X-ray films (Thermo Scientific).

**ACKNOWLEDGMENTS.** The I-PanMI homing endonuclease and target site were provided for this study by Dr. Abigail Lambert. Ms. Taylor contributed to IVC selection experiments against the crf target locus. The expression plasmid for the mouse Trex2 endonuclease, pExodus-CMV.Trex2 (Addgene plasmid 40210), was a gift from Kamila Gwiazda and Dr. Andrew Scharenberg (Seattle Children’s Research Institute). This work was funded in part by a fellowship from the Japan Science Foundation (to R.T.) and by a grant from the National Institutes of Health (R01 GM08537 to B.L.S.).

**Codon** were inserted into the DR-GFP reporter plasmid (33). Sequences of individual plasmids used in this study are shown in Dataset S1.

**Chromosomal Modifications Generated by Coexpression of the Two MegaTALs.** HEK 293T cells were transfected with 0.2 μg each of the two pExodus plasmids expressing MegaTALs and 0.4 μg of pExodus.CMV.Trex2, as described above. Three days after transfection, genomic DNA was extracted, and the genomic target region was PCR-amplified using a pair of primers (5′-TCT TGG ATT TCC AGG AGT TTT TCC TTA TGG TTA TTT TCT C-3′). PCR fragments were separated on a 1.5% (wt/vol) agarose-TBE gel. Gel-extracted DNA fragments were separated on a 1.5% agarose-TBE gel. The frequency of indel was estimated using the following equation as described previously (43): (% indel) = 100 × (1 – (1 – fraction cleaved))^{2/3}.

**Table S2.** After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2. After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2. After purification using DNA clean and concentration kit (Zymo Research), and genomic regions spanning a target site were PCR-amplified using Phusion DNA polymerase (New England Biolabs). Primers used are shown in Table S2.
Fig. S1. Schematic of in vitro selection using in vitro compartmentalization. Sequences of primers used during selection are shown in Dataset S1. (A) Using the pET21d(+) plasmid containing parental meganuclease genes (dark gray box) as a template, the T7 promoter (light gray box) and endonuclease genes are PCR-amplified with multiple sets of primers, a portion of which contain wobble bases for randomized residues (see Dataset S1 for more details). In parallel, a fragment containing two consecutive target sites is amplified using the Down1 primer and the MidFwd primer (which contains tandem two copies of target sites (black boxes)). (B) All PCR fragments with sequence overlaps are purified by gel extraction and assembled by overlap extension PCR using the Up2 and Down2 primers. (C) A sequence containing variant endonuclease genes and tandem two copies of target sites is gel-purified and quantified. (D) Individual DNA library fragments are compartmentalized with an in vitro protein synthesis reaction mixture in an oil–surfactant mixture. (E) After extraction from emulsion, DNA molecules with target sites cleaved by variant endonuclease are ligated to DNA adaptors containing complementary overhangs (purple). Adaptors used are described in Dataset S1. (F) ORFs of variant meganucleases coupled to adaptors are PCR-amplified using an adaptor-specific primer (purple arrow) and the Up3 primer. (G) The resultant PCR fragment is gel-purified and used for the next round of PCR using the Up4 and MidRev primers to eliminate an adaptor sequence. In parallel, an upstream sequence of meganuclease genes (corresponding to the entirely red line in H) is amplified using the Up1 and UpRev primers. (H) A library is resynthesized by assembling a variant gene-containing fragment (a region between the Up4 and MidRev primers) with a sequence amplified using the Up1 and UpRev primers (entirely red line) and one amplified using the MidFwd and the Down1 primers (black line and boxes).
Fig. S2. Search for sites similar to the target sequences recognized by wild-type meganucleases in the human genome target loci. Sites that displayed high identities to the original targets for four wild-type meganucleases (I-HjeMI, I-LtrI, I-OnuI, and I-PanMI) were explored using the Central Four PWM Search in the LAHEDES Web server (1) and are shown. Arrows indicate sites chosen for redesign of meganucleases. (A) The flanking 250-bp regions of the F508-coding triplet (red) of the human cftr gene were used as a query. The cftr exon 11 and its flanking intronic regions are shown in capital and lowercase letters, respectively. The target sites for the redesigned meganucleases derived from I-OnuI and I-PanMI (CFTRonu and CFTRpan/G) are highlighted in green and cyan, respectively. (B) Target search was carried out in a 501-bp sequence encoding the proline-rich region of the human Bcal11a protein. The target site for the engineered I-HjeMI variant endonuclease (BCL11Ahje) is highlighted in green.

Sequential mutagenesis and in vitro selection were carried out to retarget the wild-type l-ONuI meganuclease to the cftr intronic site (CFTR1) that differed from its original target at eight base positions (Fig. 1B). The DNA bands corresponding to the populations of active variant meganucleases (indicated by asterisks) were PCR-amplified from libraries screened in emulsified droplets (right lanes in each round of selection) but not from naive (unselected) libraries (left lanes).

Two to three rounds of selection in compartmentalized aqueous droplets were first performed to isolate variant endonuclease genes that cleaved target sites containing two altered base pairs in either half of the l-ONuI original target site. Note that the efficiencies of the PCR amplification were greatly varied not only by levels of stringency (that were increased in subsequent rounds of selection) but also by sequences of DNA adaptors and adaptor-specific primers used to capture cohesive ends generated in compartmentalized droplets. Then, additional base pair substitution(s) and site-directed mutagenesis were introduced into target sites and selected variant endonuclease genes, respectively, and a few rounds of selection experiments were performed again. Upon obtaining ORFs encoding variant enzymes that cleaved chimeric sites containing one half of the CFTR1 site and the other half of the original target, the N- and C-terminal half domains that were redesigned (ribbon diagrams colored red and blue) were shuffled, and two more rounds of selection were conducted to obtain engineered meganucleases that targeted the full CFTR1 site.

Fig. S3. Representative results of adaptor-specific PCR to recover endonuclease variant genes coupled to target sites cleaved during in vitro selection. Sequential mutagenesis and in vitro selection were carried out to retarget the wild-type l-ONuI meganuclease to the cftr intronic site (CFTR1) that differed from its original target at eight base positions (Fig. 1B). The DNA bands corresponding to the populations of active variant meganucleases (indicated by asterisks) were PCR-amplified from libraries screened in emulsified droplets (right lanes in each round of selection) but not from naive (unselected) libraries (left lanes). Two to three rounds of selection in compartmentalized aqueous droplets were first performed to isolate variant endonuclease genes that cleaved target sites containing two altered base pairs in either half of the l-ONuI original target site. Note that the efficiencies of the PCR amplification were greatly varied not only by levels of stringency (that were increased in subsequent rounds of selection) but also by sequences of DNA adaptors and adaptor-specific primers used to capture cohesive ends generated in compartmentalized droplets. Then, additional base pair substitution(s) and site-directed mutagenesis were introduced into target sites and selected variant endonuclease genes, respectively, and a few rounds of selection experiments were performed again. Upon obtaining ORFs encoding variant enzymes that cleaved chimeric sites containing one half of the CFTR1 site and the other half of the original target, the N- and C-terminal half domains that were redesigned (ribbon diagrams colored red and blue) were shuffled, and two more rounds of selection were conducted to obtain engineered meganucleases that targeted the full CFTR1 site.
Fig. S4. Comparison between wild-type and engineered meganucleases. (A) Sequence alignment of redesigned meganucleases and their parental enzymes. The residues highlighted in green (Middle) indicate those that differ between the two I-PanMI variant endonucleases that discriminate circulating SNPs in the CFTR2 site. (B–E) The original residues (orange spheres) substituted in the four retargeted meganucleases, CFTRonu (B), CFTRpan/A (C), CFTRpan/G (D), and BCL11Ahe (E), are mapped on the crystal structures of their parental meganucleases, and amino acid positions subjected to saturation mutagenesis are colored blue. The central four base pairs and mismatched nucleotides between the human genome target sites and the original target sites are highlighted in yellow and red, respectively.
Fig. S5. Expression of the engineered meganucleases and MegaTALs and homologous recombination induced on episomal DR-GFP reporter plasmids (1) containing target sites for the wild-type and retargeted meganucleases in HEK 293T cells. (A) Cell lysates were prepared 2 d after transfection and separated on a 15% polyacrylamide-SDS gel. Western blotting was carried out to detect HA-tagged proteins (meganucleases and MegaTALs) and β-tubulin (as a loading control). Coexpression of Trex2 had little effect on expression of the retargeted meganucleases and MegaTALs. (B) HEK 293T cells were transfected with the pExodus plasmid encoding meganucleases, a DR-GFP plasmid containing a target site, and pExodus-mCherry (as a transfection marker) and analyzed by flow cytometry 2 d after transfection. Each bar represents an increase in a fraction of GFP-positive cells relative to the background that was measured by transfection with the nonexpression plasmid (pUC19) instead of meganuclease expression plasmids. Spontaneous double strand breaks stimulate homologous recombination on a reporter plasmid even in cells cotransfected with pUC19 and a DR-GFP reporter. The percentage of GFP-positive cells in transfected cells (judged based on expression of mCherry) is shown on each bar. Error bars refer to SDs of three to four independent experiments. Asterisks indicate statistically significant difference compared with the background (P < 0.05).

**A**

>Wild_type

CCAGACTTCACTTCATAGATTATGGGAGAACATGGAGCCTTCAGAGGGTAAAATTAAGCACATGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGTAAGAACTATGTGAAAACTTTTTGATTATGCATATGAACCCTTCACACTACCCAAATTATATATTGGCTCCATATTCAATCGGTTA

>Duplication

CCAGACTTCACTTCATAGATTATGGGAGAACATGGAGCCTTCAGAGGGTAAAATTAAGCACATGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGTAAGAACTATGTGAAAACTTTTTGATTATGCATATGAACCCTTCACACTACCCAAATTATATATTGGCTCCATATTCAATCGGTTA

>Insertion

CCAGACTTCACTTCATAGATTATGGGAGAACATGGAGCCTTCAGAGGGTAAAATTAAGCACATGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGTAAGAACTATGTGAAAACTTTTTGATTATGCATATGAACCCTTCACACTACCCAAATTATATATTGGCTCCATATTCAATCGGTTA

>B

>Inversion

CCAGACTTCACTTCATAGATTATGGGAGAACATGGAGCCTTCAGAGGGTAAAATTAAGCACATGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGTAAGAACTATGTGAAAACTTTTTGATTATGCATATGAACCCTTCACACTACCCAAATTATATATTGGCTCCATATTCAATCGGTTA

---

Fig. S6. Sequences of genomic mutations induced by expression of the two MegaTALs. The sites targeted by CFTRpan/A and CFTRonu are colored red and orange, respectively, and their central four base positions are underlined. (A) Homologous chromosomes, which are cleaved either by TALE-CFTRpan/A (highlighted in yellow) or by TALE-CFTRonu (highlighted in gray), are joined, leading to duplication (Duplication). The genomic region excised by the two MegaTALs (highlighted in green) is inversed and incorporated together with a DNA fragment derived from a plasmid expressing MegaTALs (highlighted in cyan) (Insertion). (B) The inverted region is highlighted in green.

**Table S1.** The number of sites closely related to targets for retargeted meganucleases in the human genome

<table>
<thead>
<tr>
<th>Meganuclease</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFTRonu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,311</td>
</tr>
<tr>
<td>CFTRpan/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120</td>
<td>1,311</td>
</tr>
<tr>
<td>BCL11Ahe</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>413</td>
<td>4,548</td>
</tr>
</tbody>
</table>

*These sites were identified using the TagScan Web server (http://ccg.vital-it.ch/tagger/tagscan.html).
Table S2. Primers used to amplify human genomic regions

<table>
<thead>
<tr>
<th>Target site</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFTR1</td>
<td>5′-ACTTGCCAATGTAGCTGTAC-3′</td>
<td>5′-TACCACATATCACTTTATATGCATGC-3′</td>
</tr>
<tr>
<td>CFTR2</td>
<td>5′-CTTTCCATTCTTCCTCACAAC-3′</td>
<td>5′-TCATGTGTTTGCAAGCTTCTTA-3′</td>
</tr>
<tr>
<td>BCL11A</td>
<td>5′-GCTGGAAATGGTTGCAAGTAAC-3′</td>
<td>5′-CTGCCAGCTCTCATAGTCTCC-3′</td>
</tr>
<tr>
<td>Off-target #1</td>
<td>5′-GGAAGTTGGAAGAGAGAATAC-3′</td>
<td>5′-GGTTGTTCTTCAAGTAAGGAAAC-3′</td>
</tr>
<tr>
<td>Off-target #2</td>
<td>5′-GACAAGTACAGTGCCATTAAAGA-3′</td>
<td>5′-GGACAAAGATGGCCCAATAACTC-3′</td>
</tr>
<tr>
<td>Off-target #3</td>
<td>5′-ATTACAAGCTTGGACCCACTTG-3′</td>
<td>5′-AGCCACTTTGGACCTTATTAGTG-3′</td>
</tr>
<tr>
<td>Off-target #4</td>
<td>5′-CCATATCTCACAGCACCACATCTACC-3′</td>
<td>5′-CTACTTCCAGCAATGCTTCTCC-3′</td>
</tr>
<tr>
<td>Off-target #5</td>
<td>5′-CAGGAAATTCTCAAGGTGCCTCT-3′</td>
<td>5′-GTGACTAGACTAGCCACAGC-3′</td>
</tr>
<tr>
<td>Off-target #6</td>
<td>5′-ACCAGTTGTTAAGGTATTTGGTGGA-3′</td>
<td>5′-GGATGAGTGGAGGTCAAGGCA-3′</td>
</tr>
<tr>
<td>Off-target #7</td>
<td>5′-CCCTGTGTGCTTTCAAGCCC-3′</td>
<td>5′-TGTTGAGGGAAGGACTAGGAAAG-3′</td>
</tr>
<tr>
<td>Off-target #8</td>
<td>5′-ATGGTCTGGCAATACCTTCCCTC-3′</td>
<td>5′-TGAGGCAAATACGCGGTTTCT-3′</td>
</tr>
</tbody>
</table>

Other Supporting Information Files

Dataset S1 (PDF)
Sequences of primers specific to each adaptor are shown in bold. We avoided the use of the identical adaptor in two consecutive rounds of selection.

Sequences of the 3’ overhang used to redesign the N-terminal and C-terminal half domains of I-HjeMI were 5’-GACA-3’ and 5’-GAAA-3’, respectively.

Four-base, 3’ overhangs (red) are complementary to those of target sites generated by variant endonucleases. Adaptors containing 5’-GAAT-3’ in the 3’ overhang region were used for engineering of CFTRonu.

---

**Dataset 1. Oligonucleotides used for in vitro compartmentalization**

<table>
<thead>
<tr>
<th>Parental enzyme</th>
<th>Final target</th>
<th>Sequence, 5’ to 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/A</td>
<td>3’-CTGATATAAGAGGTTAGAGCCTCGTTTCCCCGAGCG-5’</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/A</td>
<td>5’-GACTATAT</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/A</td>
<td>5’-ACT</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/A</td>
<td>5’-GCTCGAGACT</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/A</td>
<td>3’-GGATCTGCCTATTGCGCATGAGAAAGGAGGCTAACC-5’</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>5’-CCTAGACGGATAAC</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>3’-TCGATTGCAACCAGGTTTGTCCTATGGACGCCACT-5’</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>5’-AGCTAA</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>3’-GTTTGCAGACTTGTAGTTACGCCGGTTTAGAAGTAAGG-5’</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>5’-CAAACGTC</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>3’-CTAGAATGGCGACAACTCTAGGTCAAGCTACATTGG-5’</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>5’-GATCTTAC</td>
</tr>
<tr>
<td>I-OnuI E178D</td>
<td>CFTR2/3’</td>
<td>3’-CACTCGTGCACCCAACTGATCTTC</td>
</tr>
<tr>
<td>Residue</td>
<td>Sequence</td>
<td>Position</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>1</td>
<td>CAGCTGTCGTTCTGTT</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>GAAGCAACCTTCAGCTTCGATGAA</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>TTCATCGAAGCTGAAGGTTGCTTCNNKACCTACAAANNKNNKAAAGAtNNKNNKTACCTGNNKGCTNNKTTC</td>
<td>S154,L158,N159,D162,D163,I166,S168</td>
</tr>
<tr>
<td>2</td>
<td>GAATTTAACCACGTTTTTAACAGAACGTACACCMNNCACMNNCAG</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>CGTTCTGTTAAAAACGTGGTTAAATTC</td>
<td>2</td>
</tr>
</tbody>
</table>

* Residue numbers correspond to those assigned in the deposited PDB files (3QQY for I-OnuI, 4JEE for I-PanMI, and 3UVF for I-JneI).

** Nucleotides that differ from the original target sites recognized by parental meganucleases are highlighted in red.
Dataset 2. Sequences of ORFs and plasmids used in this study. To construct libraries for IVC, parental meganuclease genes inserted between the NcoI and NotI sites in pET21d(+) were used as PCR templates. To express the wild type and engineered meganucleases and MegaTALs with the N-terminal HA epitope tag and nuclear localization signal, the sequence between XmaI and XbaI sites of pExodus-I-OnuI was replaced with ORFs of meganucleases and MegaTALs shown below.

```
> I-OnuI_E178D_cloned_in_pET21d(+)
ATGGGATCCGCTACATGTCCGGTACCCATTAACCCGTGGATTTTCTACCCGTGCCTGAGCAGCGGAAGGCTCCCTTTGCTGCGCATCCGCAACAACAACAACAAGTCCAGCGTCGGCTACTCCACTGAGCTCGGCTTCCAAATTACACTTCATAACAAGGACAAGAGCATTCTTGAGAACATCCAGTCAACATGGAAGGTGGGCGTGATCGCCAACAGCGGTGACAACGCCGTGTCGCTGAAGGTCACGCGTTTTGAGGACCTGAAGGTCATTATCGACCATTTTGAAAAATACCCACTGATTACGCAGAAGCTCGGCTGACTACATGCTGTTTAAGCAGGCGTTTTGCGTCATGGAGAACAAGGAGCATTTGAAGATTAATGGTATCAAGGAGCTGGTGCGCATTAAGGCAAAGCTCAATTGGGGTCTGACGGATGAGCTGAAGAAGGCCTTTCCGGAGATCATCTCGAAGGAGCGCTCCCTCATCAACAAGAACATCCCTAATTTCAAGTGGCTGGCGGGTTTTACCTCGGGCGATGGTTGCTTCTTTGTTAACCTGATCAAGTCAAAGTCGAAGCTAGGTGTCCAGGTGCAGCTGGTGTTCAGCATTACCCAACATCAAGGATAAAGACTCAAGACTCAAGACGAACTCTGTCAATATCTGCAGAAGTTGACGATTACCTGATCGCTTCTTTCGACATCGCTCAGAAAGACGGTGACATCCTGATCTCTGCTATCCACAAATACTGTCTTTCTTGGATGGTGTAAGGTCGCTAAGCTGATCGAGGAGAAAAAGCACCTGACAGAAAGTGGCCTGGACGAGATCAAGAAGATTAAGCTGAACATGAACAAGGGCAGAGTATTCTAA
```

```
> I-PanMI_cloned_in_pET21d(+)
ATGGAATCAAAACTGAACCCGAGCTACATTTCTGGGTTTGTTGCACTCAGGCTCGGCTCTCCTCTACACCTGCAAAGACTCAAGACTCAAGACGAACTCTGTCAATATCTGCAGAAGTTGACGATTACCTGATCGCTTCTTTCGACATCGCTCAGAAAGACGGTGACATCCTGATCTCTGCTATCCACAAATACTGTCTTTCTTGGATGGTGTAAGGTCGCTAAGCTGATCGAGGAGAAAAAGCACCTGACAGAAAGTGGCCTGGACGAGATCAAGAAGATTAAGCTGAACATGAACAAGGGCAGAGTATTCTAA
```

```
> I-HjeMI_cloned_in_pET21d(+)
ATGGGATCCCACATGGACCTGACCTACGCTTACCTGGTTGGTCTGTTCGAAGGTGACGGTTACTTCTCTATCACCAAAAAAGGTAAATACCTGACCTACGAACTGGGTATCGAACTGTCTATCAAAGACGTTCAGCTGATCTACAAAATCAAAGACATCCTGGGTGTTGGTAAAGTTTCTTTCCGTAAACGTAACGAAATCGAAATGGTTTCTCTGCGTATCCGTGACAAGAATCACCTGAAAAACTTCATCCTGCCGATCTTCGACAAATACCCGATGCTGTCTAACAAGCAGTACGACTACCTGCGTTTCAAAGACGTCTCCTGTCTAACATTATCTACTCTGACGATCTGCCGGAATACGCTCGTTCTAACGAATCTATCAACTCTGTTGACTCTATTTATCAACACCTCTTACTTCTCTGCTTGGCTGGTTGGTTTCATCGAAGCTGAAGGTTGCTTCTCTACCTACAAACTGAACAAAGATGACGATTACCTGATCGCTTCTTTCGACATCGCTCAGAAAGACGGTGACATCCTGATCTCTGCTATCCACAAATACTGTCTTTCTTGGATGGTGTAAGGTCGCTAAGCTGATCGAGGAGAAAAAGCACCTGACAGAAAGTGGCCTGGACGAGATCAAGAAGATTAAGCTGAACATGAACAAGGGCAGAGTATTCTAA
```

```
LOCUS pEndo/Devron-BCL11Ahje 5093 bp DNA circular
FEATURES Location/Qualifiers
rep_origin 783..1328 /label=p15A ori
CDS complement(1856..2785) /label=araC
```
promoter complement(2885..2913)
/label=araC promoter
promoter 3010..3037
/label=arabinose promoter
RBS 3088..3094
/label=RBS
misc_feature 3102..3104
/label=Start codon
misc_feature 3100..3105
/label=NcoI
CDS 3102..3869
/label=BCL11Ahje ORF
misc_feature 3871..3878
/label=NotI
sig_peptide 3882..3914
/label=Degron
misc_feature 3915..3917
/label=Stop codon
promoter 4781..4879
/label=bla promoter
CDS 4880..615
/label=bla

ORIGIN
1 GTTGACGCC GGCAGGACCA AACTCGTGCC CGCATACTG ATTCTCAAGA TGACTTGGTT
61 GAGTACTCAC CAGTACAGAA AAGCATTGTC ACGGATGCGA TGCACATGAG AGAATTATGC
121 AGTACGTCGCA TAACGATGAG TGATAACTGT TGGCGACTAC ACTACTCTTG AACATGGGG
181 GACCGAGGAC AGCTACACAC TCTCGGTGCT TCTCGTGCTG CAGGACCGGA ATCAGACTGT
241 CGTGGGAAAG TCAGGCGATA AAAAAAACCA ACGGGCGGCT ATTGACCTGC TGCCTGGCTG
301 GACGGCCGAT CGCGCGTCTTG TCTCTGCGCT TAAATTGCGT CAGGACACGC CAGGGCGTAC
361 CGCGACACAA TTCTGCGCTG AGGCGGATTG ACTCTCTCGG GAAAAATGAC GTTGGGACCT
421 GGGGTTGCGC CAGCGCGCAT CGCGCGTCTG CAGACGCTCA TGCAGGGGCT ACAGGCGCTG
481 GGTATCATTG CAGCAGAGTT CACGTGTGCT TGCAGATGAC CAGCAGGGTC ACAGGCGCTG
541 ACCTTCAGATCG CGAGAAGACG ACCTCAGGTT CAGACGCTCA TGCAGGGGCT ACAGGCGCTG
601 CTGATGTTAG ACAGTGACAC AAGTACCATG TCTCTGCTCG TGCAGGCGCT TCCGTTGCGA
661 AAGTCACTAC TTTTTTTTGA AAGGGAATCG GTGGATTTCG TCTCTGCTCG TGCAGGCGCT
721 AAAATCTTCTT CACTCGGATG GACGGCTGAG AACATGGGGC TCTCGCTGGT AGGGGGCTG
781 TCTTGGTACT ATTCCTGCAG TCTGCTTGGG TCGTTGCTCG TGCAGGCGCT TCCGTTGCGA
841 GGGGTTTCTT TCTCTGCGCT CAGAAGTACG AGCTGGGTTT GGTGGTGGTA GTGGTTGTGC
901 GGGGTTTCTT TCTCTGCGCT CAGAAGTACG AGCTGGGTTT GGTGGTGGTA GTGGTTGTGC
961 AAGAAGCTAG AGGTGTGGTT TTGCTTGTGC TGATGGTGGT TTGCTTGTGC TGATGGTGGT
1021 TCTCGTTGCG CAGCTCCGCA GACTAGCTAC CAGCAAGGAG GCAGGTTGCG CAGCAAGGAG
1081 GGTGTTTCTT TTTTATTTTA AAAAGCCCTG TTTTGAGAAG TTGCTGTGAG TTTTGAGAAG
1141 TAGTGGATCA GAGAAGCAGG TTTTGGGTGT AAGGGAATCG GACGGCTGAG AACATGGGGC
1201 CAGGAGAAGG CAGCAGGGGA CACGTGTAAGG TCTCTGCTCG TGCAGGCGCT TCCGTTGCGA
1261 GTGTTGTTTA CAAAGTGATG GACGGCTGAG TTTTGAGAAG TTGCTGTGAG TTTTGAGAAG
1321 TCTGCGTCTT TCTCGCTGGT TCTCGCTGGT TCTCGCTGGT TCTCGCTGGT TCTCGCTGGT
1381 TCTGCGGCTG GACGGCTGAG TTTTGAGAAG TTGCTGTGAG TTTTGAGAAG TTGCTGTGAG
1441 GCCCTGTTGA GCACAGTGACG AAGGGAATCG GACGGCTGAG AACATGGGGC TCTCGCTGGT
1501 GCCCTGTTGA GCACAGTGACG AAGGGAATCG GACGGCTGAG AACATGGGGC TCTCGCTGGT
1561 ACCACAGATT GCAGGGAGCC TCTCTGCTCG TGCAGGCGCT TCCGTTGCGA TTGCTGTGAG
1621 GCCCTGTTGA GCACAGTGACG AAGGGAATCG GACGGCTGAG AACATGGGGC TCTCGCTGGT
1681 GCCCTGTTGA GCACAGTGACG AAGGGAATCG GACGGCTGAG AACATGGGGC TCTCGCTGGT
1741 GCCCTGTTGA GCACAGTGACG AAGGGAATCG GACGGCTGAG AACATGGGGC TCTCGCTGGT
1801 TCTCTGCTCG TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
LOCUS       pExodus-I-OnuI      5396 bp      DNA     circular
FEATURES             Location/Qualifiers
promoter        232..819
     /label=CMV
misc_feature    982..984
     /label=Start codon
misc_feature    988..1014
     /label=HA tag
misc_feature    1015..1038
     /label=SV40 NLS
misc_feature    1039..1044
     /label=XmaI
misc_feature    1945..1950
     /label=XbaI
polyA_signal    1982..2206
     /label=BGH pA
rep_origin      2252..2680
     /label=f1 ori
polyA_signal    3072..3202
     /label=SV40 pA
rep_origin      3585..4255
     /label=pUC ori
CDS             complement(4400..5260)
     /label=bla
promoter        complement(5261..5359)
     /label=bla promoter
CDS             1045..1941
     /label=I-OnuI ORF

ORIGIN
  1 GACGGATCGG GAGATCTCCC GATCCCTAT GGTGCACTCT CAGTACAATC TGCTCTGATG
  61 CCCATAGTTG AAGGGGATAT CTTGCTCCTT CTTGCTGTGG GAGGTGTCGG TGGATGCGG
 121 CGAGCAAATT TTAAGCTACA ACAAGGCAAAG GCTTGACGGA CAATTGCTAG AGAATCTGC
 181 TTACGGCTAG ATGCTTTCGG CTGGTACGGG ATGTCAGCGG CAGATATACG CGTTGACATT
 241 GATTATGACC TAGTTTATAA TAGGATTTCA AATGGAATAA TTACGGGTCG ATTTGAGGTA
 301 TGAGCTTCTC GCCTTACATA ATCTAGCTAA ATGCGCCCGC ATGCGGTCCT ATGCGGCTCG
 361 CCGCCCCCAT TGGGACATTA ATGAGCGTTA GTCTGCTGGT CGGCGGGCCG CCCAGACCC
 421 ATGACGTGCA ATGGTGAGGT ATTTTCTGTT GTACCTGCTA CATCATGATG CATGATGATG
 481 ATCCTATGCC AACTTCCCTT CACATGAGCC GGTCTGGGCG GGTCTGGGCG GGTCTGGGCG
 541 ATGACGTGCA ATGGTGAGGT ATTTTCTGTT GTACCTGCTA CATCATGATG CATGATGATG
 601 TCTCATTGTC CATATGCGGT AAGCTTGGCC AGCTGAGCTA AGATATGAGT GTGCATTTTC
 661 ACTACGCCGGG ATCTCAAACT AAGGAAAAGG GCTGTTTCTG GTGATGGAGT GGTCTGGGCG
 721 AAATCAACC AATATCCGTA AAGAAGCAGA GACGGCGGTC AGTTTTGAGA GTGATGGAGT
 781 GTAGCCGCTG TAGTTTCGAG GTCTATTATG TGGCAGGGTA AGTCTGAGAA TTGAGCTGGG
 841 CTGCTTACTG CTGGGACTCA ATTAACTAGG GGTCTGGGCG GGTCTGGGCG GGTCTGGGCG
 901 GTTTAAATCTT ACACTGCTA GACGCTGGGA GACATGCTAA GTGCTGACCTT GATTGTGGT
 961 AGGATCTTGGG CAGGGGCTCC CATTGGAGAT CATATGAGTG AGTTTTTCTT TTGTGAGGTT
1021 AGAAGAAGAC CGAGAGCATC CGGCTTCTCG AGATATGAGT GTGCTGACCTT GATTGTGGT
1081 GGTGTTTGGG ACAGTCTAAG CTGTAGTGGG CAGGGGCTCC CATATGAGTG AGTTTTTCTT
1141 GTGAGGCTATA TGACCCAAGC AATTTTCTT TATGAGGTCG AAGAAGAAGAC CGAGAGCATC
1201 CTGACGCTTA TGGGGGCGGT TGGGCTTCTG ATGCTGACCTT GATTGTGGT GTGCTGACCTT
1261 GTGATGAGTG AGGTAGCTTG GTTGGATGCT ATGATGAGTG AGTTTTTCTT TTGTGAGGTT
1321 TACCTCTCAGA ATATTGTGAT CAGGGGCTCC CATATGAGTG AGTTTTTCTT TTGTGAGGTT
1381 ATGGAGAATA AGGAACATCT GAAAATCAAC GGCATCAAGG AGCTGGTGCG GATCAAGGCC
1441 AAGAAGACCG TTGACTAAAT GAGAAGTTGT GAGCTAAGT AAGTTCAAAG ATCCTGAAGCT
1501 GGCAGAGGGT GCTTCCTGTG GAACCTGATC AAGGAAAGCTT TTCCCGAGAT CATTTCAAAG
1561 GAAAGAAGCC TGATCAATAA GAACATCCCT AACTTCAAGT GGCTGGCTGG ATTCACAAGC
1621 GGCGAGGGGT GCTTCTTTGT GAACCTGATC AAGTCTAAGA GTAAGCTGGG CGTGCAGGTG
1681 CAGCTGGTGT TCAGCATCAC TCAGCACATC AAGGACAAGA ATCTGATGAA CTCCCTGATC
1741 ACCTACCTGG GATGTGGCTA TATCAAGGAG AAGAACAAAT CAGAATTCAG CTGGCTGGAT
1801 TTGTGGTGA CAAAGTTCAG CGACATCAAT GATAAAATCA TTCCCGTGTT TCAGGAGAAC
1861 ACTCTGATTG GCGTGAAGCT GGAGGACTTC GAAGATTGGT GTAAGGTGGC CAAACTGATC
1921 GAGGAAAAGA AACATCTGAC CGAGAGCGGC CTGGATGAAA TCAAGAAAAT TAAGCTGAAT
1981 ATGAACAAAG GGAGAGTGTT CTAATCTAGA GGGCCCGTTT AAACCGCTG ATCAGCCTCG
2041 ACTGTGCCTT CTAGTTGCCA GCCATCTGTT GTTTGCCCCT CCCCCGTGCC TTCCTTGACC
2101 CTGGAAGGTG CCACTCCCAC TGTCCTTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT
2161 TGGGAAGACA ATACAGGCTCA TGTCGCTGGG GGCAGGGCAG AGGAGAGAGC CTGGCCGCAA
2221 AGACACACGCT CACGCGCAGT CCCAGTTGAG ATGAAAAAGA AGGAGAGAGC CTGGCCGCAA
2281 GCACAGGATG TGGTTAGCAGC CGCCGCGAGT CTGGGCTGGA TAGAGGAGGT TCCGCTGCGG
2341 CTTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT GTCTCTGTGTT CTTCTGTGTT CTCTCTGTGTT
2401 ATCTGGATTT TGTCGTGTTTG TCAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2461 ATTACCTTG TGAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2521 TGGATCTGTT CACGCGCCCT GCCTGGTTGA TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2581 ACAAACCGCC GGGCCGCGAG TCGGCTGGG GGGCCGCGAG TCGGCTGGG GGGCCGCGAG
2641 ATCTCTCTGTG TGAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2701 ATCTCTCTGTG TGAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2761 GACATACGCT GTGCCGCTGG GGGCCGCGAG TCGGCTGGG GGGCCGCGAG TCGGCTGGG GGGCCGCGAG
2821 TGGGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2881 CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
2941 ATCTCTCTGTG TGAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3001 TGGGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3061 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT TCCAACTGG AACAACACTC
3121 ACCCTTGTTT ATCAGGACCA TACGCGCCCA TACGCGCCCA TACGCGCCCA TACGCGCCCA
3181 TGATCTCTTG TTATCATGTC TGTATACCGT CGACCTCTAG CTAGAGCTTG CGGTAATCAT
3241 TGATCTCTTG TTATCATGTC TGTATACCGT CGACCTCTAG CTAGAGCTTG CGGTAATCAT
3301 ATCTCTCTGTG TGAGGAGGTT TCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3361 TGGTGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3421 TGGTGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3481 TGGTGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3541 ATCTCTCTTG TTATCATGTC TGTATACCGT CGACCTCTAG CTAGAGCTTG CGGTAATCAT
3601 GGCTCTCTTG TTATCATGTC TGTATACCGT CGACCTCTAG CTAGAGCTTG CGGTAATCAT
3661 CCCCTCTGCA GCACTCAGCT TATGCAGTTA ACAAAAATTT AACCCGAGAT GCTCCAAAAA
3721 CTTTGTTTCTG TATGCAGTTA ACAAAAATTT AACCCGAGAT GCTCCAAAAA
3781 TGGTGCTGCTG GCCGCGCCCG GTGGGTGTGT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3841 GCTGCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3901 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
3961 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4021 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4081 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4141 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4201 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4261 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4321 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
4381 ATCTCTCTTG AACTCTCTTG GGCAGAGGTT CTCTCTGTGTT CTCTCTGTGTT CTCTCTGTGTT
GGACACAGGCCAACCTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCATGCATCGCGCAATGCAC
TGACGGGTGCCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAA
AGGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACG
ATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAG
cGTCCGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTG
GCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTG
GCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGC
GGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACG
ATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAG
cGTCCGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTG
GCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTG
GCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGC
GGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACG
ATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAG
cGTCCGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTG
GCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTG
GCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGC
GGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACG
ATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAG
cGTCCGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTG
GCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTG
GCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGC
GGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACG
ATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAG
<table>
<thead>
<tr>
<th>LOCUS</th>
<th>pExodus-mCherry</th>
<th>5149 bp</th>
<th>DNA</th>
<th>circular</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEATURES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>promoter</td>
<td>Location/Qualifiers</td>
<td>232..819</td>
<td>label=CMV</td>
<td></td>
</tr>
<tr>
<td>CDS</td>
<td></td>
<td>953..1660</td>
<td>label=mCherry</td>
<td></td>
</tr>
<tr>
<td>misc_feature</td>
<td></td>
<td>953..955</td>
<td>label=Start codon</td>
<td></td>
</tr>
<tr>
<td>polyA_signal</td>
<td></td>
<td>1735..1959</td>
<td>label=BGH-pA</td>
<td></td>
</tr>
<tr>
<td>rep_origin</td>
<td></td>
<td>2005..2433</td>
<td>label=f1 origin</td>
<td></td>
</tr>
<tr>
<td>rep_origin</td>
<td>complement</td>
<td>3338..4008</td>
<td>label=pUC origin</td>
<td></td>
</tr>
<tr>
<td>CDS</td>
<td>complement</td>
<td>4153..5013</td>
<td>label=bla</td>
<td></td>
</tr>
<tr>
<td>promoter</td>
<td>complement</td>
<td>5014..5112</td>
<td>label=p promoter</td>
<td></td>
</tr>
<tr>
<td>polyA_signal</td>
<td></td>
<td>2825..2955</td>
<td>label=SV40 pA</td>
<td></td>
</tr>
<tr>
<td>ORIGIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GACGGATCCGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>GAGATCTCCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>CAGAAGCGAAAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>TTAGGGTATAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>GATTATTTATAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>CGGCACTACA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>361</td>
<td>CCCGCCTATT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>421</td>
<td>ATTCGGTCTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>481</td>
<td>ATTCGGTCTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>541</td>
<td>ATTCGGTCTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>601</td>
<td>TCGGATGAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>661</td>
<td>ACTCAGGCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>721</td>
<td>AAAATCAACG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>781</td>
<td>GTTAGGCTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>841</td>
<td>CTGATCGAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>901</td>
<td>GATGTTGCTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>961</td>
<td>CAGGGGGGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1021</td>
<td>GAGGCGGGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4321 ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC
4381 CGCCTCCATC CAGTCTATTA ATTTGGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
4441 TAGTCTTGG CACATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG
4501 TATGCTTCCA TTCAGCTCCG GTTCCCAACG ACAAAGGCGA CTTCTGTCTC CCCCATGTT
4561 GTGCAAAAAA GCCGGTTAGCT CTTGGGTTCC TCCGATCGTT GTCAAGAAGT AGTTGGCCGC
4621 AGTTGATTCA TCATGTTGTT GGGCAGCAGT GTATAATTCT CTAAAACTGTCA TGCCATCCGT
4681 AAGATGCTTT TCTGTGACTG CTAGTACTTC ACAAAGATCA TTCTGAAAAAT AGTGTTATCC
4741 GCCGACGAGT TGCTCTTGGC CGGCGTTCAAT ACCGGATATA ATCGCGGCAAT ATACGAGAAC
4801 TTTAAAAATG CTCATCAATT GAAAAAGGTC TTCCGCGGCGA AAACCTCTCAAA GGATCTTACC
4861 GCTGTGAGAG TGGCTGGCTCA TCTGACCCCA ATCTGATCTT CAGCATCTTT
4921 TACCTTTCACC AGCGTTTCTG GGTGACAAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG
4981 AATAAGGGCG AAGGCAAGAT GTTTACAACT CATACTCTCT CTTCTTCCTT ATATTGGAAG
5041 CATTCTACG GGGTTATTGT CTCATGAGCGG ATACATATTGT GAATGTATTT AGAAAAATAT
5101 ACAATAGGG GTTCCCGCGCA CATTTCCCCG AAAAGTGCCC CCTGACGTC