Chapter 12

Homing Endonuclease Target Site Specificity Defined by Sequential Enrichment and Next-Generation Sequencing of Highly Complex Target Site Libraries

Hui Li and Raymond J. Monnat Jr.

Abstract

Homing endonucleases (HEs) are DNA sequence-specific enzymes that recognize and cleave long target sites (14–40 bp) to generate double-strand breaks (DSBs). Their high site recognition specificity and tight coupling of binding and cleavage make HEs attractive reagents for targeted genome manipulation. In order to delineate the target site specificity of HEs and facilitate HE engineering, we have developed a method for comprehensive target site profiling of HEs cleavage specificity using partially randomized target site libraries and high-throughput DNA sequencing.

Key words Homing endonuclease, Cleavage specificity, Sequential enrichment, Next-generation sequencing, Targeted genome engineering

1 Introduction

Highly site specific endonucleases are powerful tools for genome engineering that can enable targeted gene manipulation by generating DSBs in specific target genomic loci [1]. These DSBs stimulate cellular DNA repair to generate different outcomes depending on repair pathway activity and the presence or absence of a repair template. For example, DBSs in the absence of a repair template can promote non-homologous end joining (NHEJ) of DSBs to generate small deletions or insertions at or near the DSB site. These targeted modifications may alter or disrupt gene function depending on their size and the target gene reading frame [2]. Alternatively, homology-dependent recombinational repair (HDR/HR) of DSBs in the presence of a homologous repair template can be used to introduce specific genetic modifications [3].

Three different endonuclease protein scaffolds are now being used to generate highly site specific DSBs in vivo to facilitate genome engineering: zinc-finger nucleases (ZFNs), TAL effector nucleases (TALENs, TALNs, or TALs), and LAGLIDADG

David R. Edgell (ed.), *Homing Endonucleases: Methods and Protocols*, Methods in Molecular Biology, vol. 1123, DOI 10.1007/978-1-62703-968-0_12,

homing endonucleases (LHEs, also known as "meganucleases"). ZFNs and TALENs are novel endonuclease scaffolds that are generated by combining modular DNA binding motifs with a non-sequence-specific nuclease domain derived from the Type II restriction endonuclease FokI [4, 5]. The LHE nucleases, in contrast, are small, naturally occurring proteins in which DNA recognition and catalytic motifs are tightly integrated. Native LHE proteins are found throughout all kingdoms of life, most often as open reading frames in mobile introns or inteins [6, 7]. Despite their small size (<50 kDa), the LHEs recognize long DNA sequences (~20 base pairs) and cleave these sites in vivo to promote lateral gene transfer or "homing" of their respective mobile intron or intein. LHE target sites are cleaved with high though not absolute specificity: we and others have shown that LHEs can tolerate some target site base pair changes without losing their site binding or cleavage activities [8]. This modest degree of LHE site degeneracy is practically useful, as it can enable the engineering of new DNA recognition specificities [9].

In order to better delineate the cleavage specificity of LHEs, we have previously developed a sequential enrichment protocol to identify cleavage-sensitive HE target sites contained in highly complex target site libraries [10]. This method is very robust and has been used successfully to profile the target site sequence degeneracy of several other LHEs [11, 12]. In this chapter we describe the use of this sequential enrichment protocol together with next-generation sequencing (NGS) to identify and characterize cleavage-sensitive target sites contained in highly complex target site libraries. These results have begun to provide insight into practically important questions such as the cleavage co-dependence of target site base pair positions and base pair combinations, and the cleavage sensitivity of different "central 4" base pair combinations. The "central 4" are the four contiguous base pairs that reside at the center of LHE target sites between the scissile phosphates on the target site top and bottom strands. These four base pairs are known to strongly influence target site cleavage, despite the paucity of DNA–protein contacts to these base pairs. We provide an example of how sequential enrichment together with NGS was used to rank-order the cleavage sensitivity of all 256 central 4 base pair combinations for the canonical LHE I-CreI. Our results using enrichment and sequencing are in close agreement with prior results from our lab and independently reported in vivo cleavage assay data [13].

Many additional LHE proteins with different target site specificities are being identified by genome sequencing [14, 15]. The protocol outlined below can be used to rapidly define the target site specificity of these LHEs, or of other highly site-specific nucleases to facilitate targeted genome engineering applications.

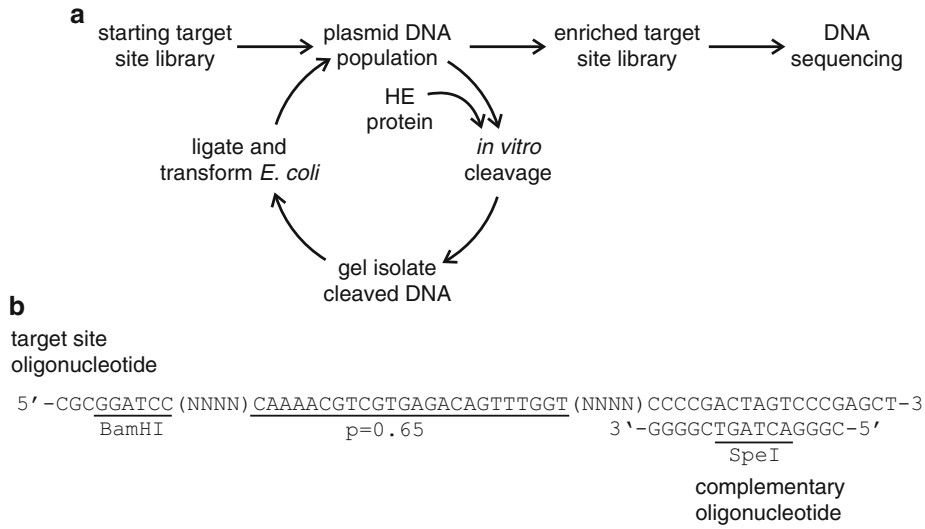

Fig. 1 Protocol overview and target site design. (**a**) Schematic overview of sequential enrichment protocol beginning (*upper left*) with construction of a complex target site library in a plasmid vector backbone. (**b**) Design of target site insert for library construction, using the I-CreI/mCreI target site as an example. The target site sequence (*underlined*) is partially randomized during synthesis, whereas flanking ("N") nucleotides are fully randomized to serve as simple barcodes to allow different versions of the same target site sequence to be distinguished. The complementary primer oligonucleotide shown right anneals to a constant region of the target site oligonucleotide, and primes the DNA synthesis needed to convert single-stranded, partially randomized target site oligonucleotides into dsDNA for plasmid capture (see text for additional detail). This oligonucleotide also contains a unique, target site-specific restriction site to allow the easy assessment of the fraction of a starting plasmid library that contains a target site insert

2 Materials

Figure 1 provides an overview of our sequential enrichment protocol, together with an example of target site oligonucleotide design to generate a highly complex target site library for sequential enrichment and NGS characterization. We have used the canonical homodimeric LHE I-CreI and variants such as monomerized I-CreI (referred to as mCreI; [16]) as an example in the following protocols.

2.1 Generation of Partially Degenerate Target Site Library

1. A single-stranded target site DNA oligonucleotide harboring the partially degenerate I-CreI LHE target site and constant flaking DNA sequences:

   5′-CGCGGATCCNNNNCAAAACGTCGTGAGACAGTTTGGTNNNNC CCCGACTAGTCCCGAGCT-3′ in which the LHE target site is underlined. This template oligonucleotide is synthesized using defined ratios of the four DNA bases to ensure the resulting target site library will be both complex and diverse. The example shown was synthesized with 65 % wild type base,

and 11.6 % of each of the three other bases at each position. The four "N" bases flanking the target site are fully randomized during synthesis to provide barcodes that help identify individual target sites (*see* Fig. 1 and **Note 1**).

2. Complementary primer oligonucleotide: 5′-CGGGACTAGT CGGGG-3′.
3. Klenow DNA polymerase, T4 DNA ligase, and restriction enzymes.
4. 25 mM dNTP mix: generate this by mixing equal volumes of 100 mM dNTP stocks. Store frozen at −20 °C until use.
5. Phenol–chloroform–isoamyl alcohol mix (25:24:1): made by mixing equal volumes of water-saturated phenol and a 24:1 mix of chloroform–isoamyl alcohol. Store at −20 °C until use.
6. DNA gel extraction kit.
7. 20 mg/mL glycogen.
8. pBluescript SK(+) plasmid (New England Biolabs).
9. DH5α-E electroporation-competent *E. coli* host cells.
10. SOC medium: 0.5 % yeast extract, 2 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
11. Plasmid midiprep kit.

2.2 Sequential Enrichment of Cleavage-Sensitive Target Sites

1. 5 μM purified I-CreI or mCreI homing endonuclease or other enzyme (*see* **Note 2**).
2. 10× reaction buffer: 100 mM $MgCl_2$, 200 mM Tris–HCl pH 8.0 (*see* **Note 3**).
3. 6× stop buffer: 300 mM EDTA, 0.3 % SDS (w/v), 3.9 % Ficoll 400 (w/v).
4. 1× TAE buffer: 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA.
5. NuSieve GTG Agarose.
6. DNA gel extraction kit.
7. T4 DNA ligase.
8. DH5α-E electroporation-competent *E. coli* host cells.
9. SOC medium: 0.5 % yeast extract, 2 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
10. Plasmid midiprep kit.
11. Gel image analysis software (e.g., ImageQuant, ImageJ, etc.).

2.3 Sample Preparation for Illumina Sequencing

1. Oligonucleotides for target site PCR amplification: forward primer: 5′-AATGATACGGCGACCACCGAGTAAAACGACGGCC AGTG-3′ and reverse primer: 5′-CAAGCAGAAGACGGCAT ACGAGGAAAC AGCTATGACCATG-3′.
2. Oligonucleotide primer for target site sequencing: 5′-GAAT TCCTGCAGCCCGGGGGATCC-3′.
3. A high fidelity DNA polymerase (e.g., Phusion polymerase from New England Biolabs).
4. DNA gel extraction kit.
5. 1× TAE buffer: 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA.
6. NuSieve GTG Agarose.

3 Methods

3.1 Generation of Partially Degenerate Target Site Library

1. Dissolve the degenerate target site oligonucleotide and the complementary primer oligonucleotide in H_2O to generate 100 μM working stocks of each. Mix equal molar amount of both oligonucleotides to a final concentration of 5 μM in 1× Klenow polymerase buffer, and incubate at 95 °C for 10 min, then slowly cool to room temperature.
2. Add 6 units of Klenow DNA polymerase and 2 μL of the 25 mM dNTP mix, then incubate at 37 °C for 10 min to generate a double-stranded version of the target site oligonucleotide (dsDNA) (*see* **Note 4**).
3. Extract the polymerase extension reaction with an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1) by vortexing for 1 min followed by centrifugation at maximum speed in a tabletop centrifuge.
4. Carefully extract the aqueous phase with a pipettor and stand on ice of 5 min. If the aqueous phase is still hazy or milky, respin and extract residual phenol from the bottom of the tube with a pipettor.
5. Add 2 volumes of ethanol and 1 μL of 20 mg/mL glycogen, mix well, then precipitate overnight at −20 °C.
6. Spin down the precipitated dsDNA target site for 5 min at top speed in a tabletop centrifuge, and carefully wash once with 70 % ethanol. Air-dry or SpeedVac-dry, then dissolve in 20 μL H_2O.
7. Digest the dsDNA with 5 units of restriction enzyme/μg DNA, followed by a second round of extraction and precipitation. BamHI is shown in the example in Fig. 1b (**steps 3–6** above; *see* **Note 4**).

8. Digest pBluescript SK(+) plasmid DNA using the same restriction enzyme(s) and digest conditions, and purify the digested linear plasmid DNA using a DNA gel extraction kit.
9. Mix cleaved target site dsDNA with digested pBluescript SK(+) plasmid vector DNA at 3:1 molar ratio of vector to insert, add 10 units of T4 DNA ligase, and incubate overnight at 16 °C.
10. Heat the ligation reaction at 65 °C for 10 min, then purify the DNA by repeating **steps 3–6** above and resuspend in 20 μL sterile H_2O.
11. Electroporate the purified and resuspended ligation products into 50 μL DH5α-E electroporation-competent cells, then add 1 mL SOC medium and incubate for 1 h at 37 °C with gentle shaking.
12. Dilute 1 μL of transformed cells in 1 mL SOC or sterile H_2O, and plate 50 μL of the diluted cells on a 10 cm LB Agar plate containing 100 μg/mL carbenicillin. Incubate overnight inverted at 37 °C. Add the rest of the transformed cells to 50 mL LB medium with 100 μg/mL carbenicillin and grow overnight at 37 °C in a shaking incubator.
13. Count colonies on the LB Agar plate to calculate the library size by multiplying the colony number by the dilution factor.
14. Purify target site library DNA from the 50 mL overnight culture using a plasmid midiprep kit.
15. Randomly pick at least 12 colonies from the LB Agar plate, and inoculate each into 5 mL fresh LB medium with 100 μg/mL carbenicillin. Grow cultures overnight in a 37 °C shaking incubator.
16. Extract plasmid DNA and sequence the target site region of each colony plasmid to estimate the likely complexity of the target site library.

3.2 Sequential Enrichment of HE Cleavage-Sensitive Target Sites

1. Mix reaction components in a 1.5 mL microcentrifuge tube as follows: 10 μL of 10× reaction buffer, 10 μg of pBluescript target site library DNA (final concentration: ~10 nM), 5 μL of 5 μm purified I-CreI or mCreI, and H_2O to 100 μL (*see* **Note 5**).
2. Incubate at 37 °C for 1 h, then add 20 μL of 6× stop buffer and incubate at room temperature for 30 min (*see* Fig. 2 and **Note 5**).
3. Electrophorese the cleavage reaction products through a 1 % TAE-buffered NuSieve GTG agarose at 1.4 V/cm overnight at room temperature.
4. Stain the resulting gel with 100 ng/mL ethidium bromide and visualize products on a UV light box, taking care to wear proper UV eye protection.

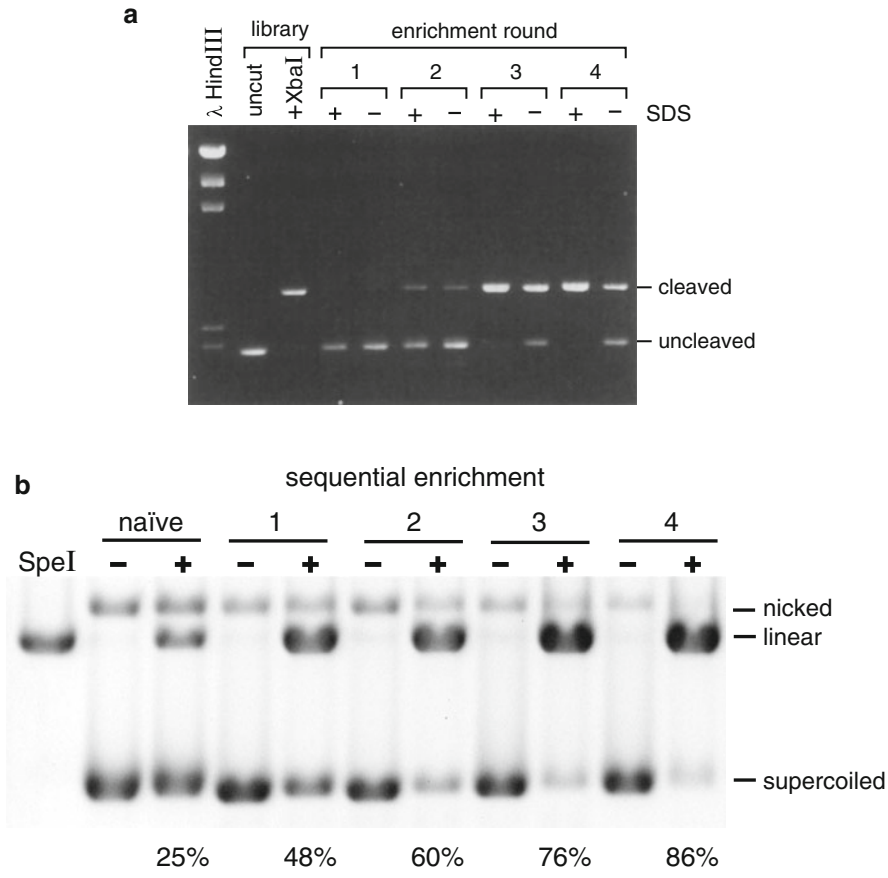

Fig. 2 Sequential enrichment examples. The two gels shown demonstrate sequential enrichment of cleavage-sensitive I-CreI/mCreI LHE target sites using low (**a**) and high (**b**) enzyme:substrate ratios of cleavage-sensitive sites present in the starting library over four successive rounds of cleavage-based recovery and amplification. Enrichment cycles in (**a**) were performed with or without the addition of SDS after completion of the cleavage reactions to facilitate product release [10]. *Numbers below* (**b**) represent the proportion of linearized, cleavage-sensitive plasmid DNA molecules after each round of enrichment. Panel (**a**) has previously been published as Fig. 3a in ref. 10

5. Excise the gel regions containing linear plasmid DNA (which contains cleavage-sensitive target sites) and supercoiled plasmid DNA (which contains cleavage-resistant target sites).
6. Extract DNA using a gel extraction kit and elute DNAs in sterile H_2O.
7. Mix the purified linear DNA plasmid with 800 units of T4 DNA ligase, 10 μL of 10× ligase buffer, and sterile water to 100 μL, then incubate at 16 °C overnight.
8. Repeat Subheading 3.1 **steps 11–14**.
9. Repeat **steps 1–6** for the supercoiled plasmid DNA fraction if you are interested in characterizing cleavage-resistant target sites in addition to cleavage-sensitive target sites.

10. Repeat **steps 1–8** above three times, or until the enrichment of cleavage-sensitive target sites reaches a plateau as assessed by the fraction of cleaved, linear DNA molecules (*see* Figs. 1 and 2).

3.3 Sample Preparation for NGS/ Illumina Sequencing

Note: In the following protocol we are using Illumina 36 base pair single end reads as a NGS sequencing platform for target site library analyses. The same general protocol can be easily adapted to different read lengths and Illumina sequencers, and to other NGS sequencing platforms.

1. Set up PCR reaction: mix 20 μL of 5× polymerase buffer, 0.8 μL of 25 mM dNTPs, 0.5 μL of 100 μM forward primer, 0.5 μL of 100 μM reverse primer, 200 ng plasmid DNA, and 1 μL of high fidelity DNA polymerase (at 2 units/μL) with sterile H_2O to 100 μL.
2. Run PCR using the following conditions: initial denaturation: 98 °C for 30 s; 30 elongation cycles: 98 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s; and final elongation step: 72 °C for 5 min.
3. Electrophorese PCR products through a 4 % TAE-buffered NuSieve GTG agarose at 6 V/cm for 2 h at room temperature. Be sure to include an appropriate size standard!
4. Stain the resulting gel with 100 ng/mL ethidium bromide and visualize products on a UV light box taking care to wear proper UV eye protection.
5. Excise the region containing PCR products of the desired size range, and extract the DNA using a gel extraction kit. Elute the resulting DNA in 20 μL of sterile H_2O.
6. Determine the quality and quantity of the purified PCR product by checking 1 μL of the isolated DNA on a 4 % TAE agarose gel (*see* **Note 6**).
7. Send three samples (naïve library, cleavage-sensitive plasmids and, if isolated, cleavage-resistant plasmids) for Illumina single end 36 bp sequencing. Each sample should consist of 10 μL of 200 nM DNA and 20 μL of 100 μM sequencing primer (*see* **Note 7**).

3.4 Data Analysis of Illumina Sequencing Results (*see* Note 8)

1. Search across all positions of each read, and eliminate any reads that have read positions with quality scores = "B".
2. Among these high quality reads identify the target read subset that has the expected end sequence for the target site oligonucleotide shown in Fig. 1b (CCCC, from target site base pair positions 33 to 36; Fig. 1b).

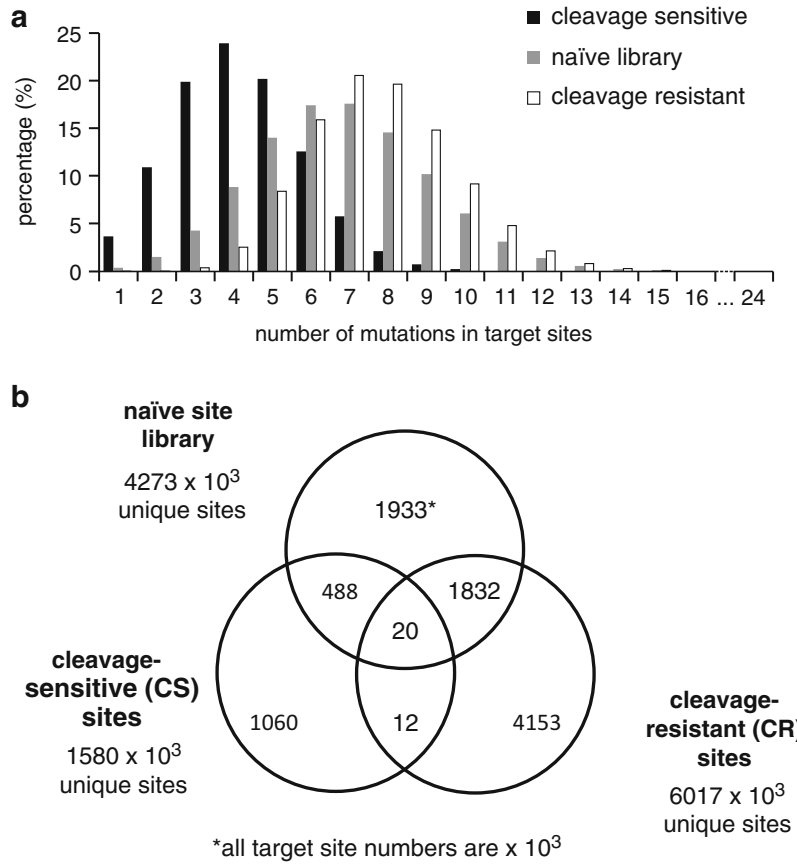

Fig. 3 Starting site library and enriched fractions. (**a**) Distribution of number of target site mutations across the starting/naïve library and cleavage-sensitive and cleavage-resistant library fractions. The distribution and type(s) of mutations in the starting library serve as a quality control measure for library synthesis. The most common oligonucleotide synthesis errors are single base deletions and insertions. The *left* shift to lower numbers of mutation in cleavage-sensitive sites, and the *right* shift in cleavage-resistant sites provide two additional indications of successful functional enrichment for these functional site classes. (**b**) Venn diagram showing the distribution of unique target sites and sites found in more than one site library (naïve, cleavage sensitive (CS) and/or cleavage resistant (CR) libraries) after four rounds of enrichment

3. Identify the subset of high quality target site sequences with the correct ends that are present in both the naïve and cleavage-sensitive (and/or -resistant) target site library fractions (*see* Fig. 3b).
4. Sum the number of reads for each unique target site in each library. Identical target sites with different barcodes are considered as different target sites, as they originated from different library templates (*see* Table 1).
5. The frequency of each target site is calculated by dividing the read number of each target site by the total read number of the sample.

Table 1
Target site enrichment sequencing results[a]

Samples	Total reads[b]	Total sites[c]	Unique sites[d]
Naïve library	13,089,102	10,925,984	6,303,279
Cleavage-sensitive	12,524,976	11,275,812	1,399,651
Cleavage-resistant	24,200,450	20,344,154	8,595,486

[a] This experiment followed closely the protocol and used the target site library design outlined in Fig. 1. The read numbers differ from the experimental analysis shown in Fig. 3b
[b] Number of sequencing reads with quality score above "B" at each position (**step 1** of Subheading 3.4)
[c] Number of sequencing reads with the expected target site oligonucleotide end sequence ("CCCC" from position 33 to 36) (Fig. 1b; **step 2** of Subheading 3.4)
[d] Number of unique sites in each library with different target site sequences or the same target site sequence but different barcodes (**step 3** of Subheading 3.4)

6. The "fold enrichment" of each target site is calculated by dividing the frequency of one target site in cleavage-sensitive sample by its frequency in naïve library sample (Fig. 4).

4 Notes

1. A 35 % randomization ratio of the target site was chosen so that a target site library for I-CreI/mCreI with a site length of 22 bp and a typical *E. coli* plasmid-based library size (10^6–10^7 independent colonies/transformants) would contain several independent copies of the native/wild type I-CreI/mCreI target site in the starting library.
2. Make sure your purified HE sample is free of contaminating DNases. This can be done by incubating serial dilutions of your purified HE, with supercoiled and linear fractions of a target site plasmid that *does not* contain a cognate target site, under optimal reaction conditions. Even trace amounts of nuclease contamination will lead to plasmid nicking or linearization (endonucleolytic activity/contamination) or linear DNA degradation (endo- and/or exonucleolytic activity/contamination). Either contaminating activity will reduce your recovery of desired target site-specific plasmid molecules.
3. Optimal cleavage reaction conditions vary between HEs and need to be determined in advance to optimize the recovery or cleavage-sensitive or -resistant target site fractions.
4. dsDNA product generation and digestion can be easily monitored by checking 1 μL of the reaction mix on a 4 % TAE agarose gel.
5. The cleavage conditions used here were chosen to favor near-complete target site cleavage. Cleavage conditions and sampling of the cleavage time course can be used to isolate target sites with differing cleavage sensitivities (*see* Fig. 2 for an example).

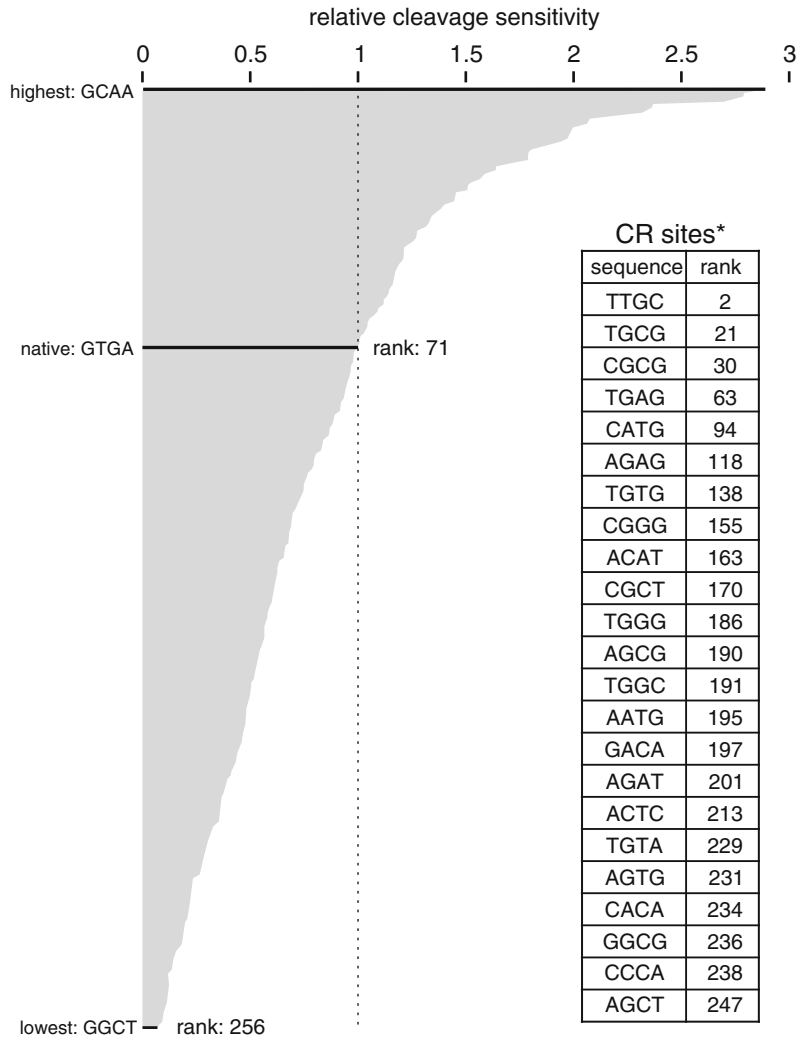

Fig. 4 Use of sequential enrichment and sequencing to define central 4 bp cleavage preferences. The main panel shows the rank ordering of relative cleavage sensitivity of 256 different, "central 4" base pair possibilities in the I-CreI/mCreI target site. Cleavage sensitivity is scaled to a native target site central 4 bp GTGA cleavage sensitivity of 1.0, versus the most-sensitive (GCAA) and least sensitive (GGCT) central 4 bp sequence in the context of different flanking sequences (Hui Li, Zafer Aydin, William S. Noble, and R.J.M., Jr., unpublished results). *Inset*: an independent determination of the central 4 bp sequences most resistant to cleavage by I-CreI using a reduced complexity target site library in which only the central 4 bp positions of the I-CreI/mCreI target site were fully randomized (Thai Akasaki, Megan S. Chadsey, and R.J.M., Jr., unpublished results)

6. A Bioanalyzer (Agilent) can be used to characterize DNAs going to NGS, and is a better alternative if available as it will give provide DNA quality and sizing data that make it easier to assess sample adequacy for NGS.

7. A typical Illumina sequencing run on a GAII should generate more than ten million reads with reasonable quality scores.
8. Data analysis of NGS sequencing results is a *very* rapidly evolving area of bioinformatics. The general outline we give can be used to implement a sequence analysis pipeline and generate and analyze results with any of a growing number of open source or commercial NGS analysis software. Some useful online resources include Galaxy tools, especially the NGS Toolbox and SAMTools (https://main.g2.bx.psu.edu/), the Phred, Phrap, and Consed suite developed by Phil Green (University of Washington Department of Genome Sciences: http://www.phrap.org/), and The Broad Institute's Software Tools resource (http://www.broadinstitute.org/scientific-community/software).

Acknowledgments

This work was supported by US National Institutes of Health Training Grant award to H.L. (5RL9HL092555); by a US National Institutes of Health U54 Interdisciplinary Research Roadmap award (1RL1 CA133831) to R.J.M. Jr; and by a grant from the Foundation for the National Institutes of Health funded by the Bill and Melinda Gates Foundation as a Grand Challenges in Global Health award to R.J.M. Jr.

References

1. Scharenberg AM, Duchateau P, Smith J (2013) Genome engineering with TAL-effector nucleases and alternative modular nuclease technologies. Curr Gene Ther 13:291–303
2. Certo MT, Gwiazda KS, Kuhar R, Sather B, Curinga G et al (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. Nat Methods 9:973–975
3. Yusa K, Rashid ST, Strick-Marchand H, Varela I, Liu P-Q et al (2011) Targeted gene correction of a1-antitrypsin deficiency in induced pluripotent stem cells. Nature 478:391–394
4. Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763
5. Miller JC, Tan S, Qiao G, Barlow KA, Wang J et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29: 143–148
6. Belfort M (2005) Back to basics: structure, function, evolution and application of homing endonucleases and inteins. In: Belfort M, Stoddard BL, Wood DW, Derbyshire V (eds) Homing endonucleases and inteins. Springer, Berlin, pp 1–10
7. Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure 19:7–15
8. Li H, Ulge UY, Hovde BT, Doyle LA, Monnat RJ (2012) Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. Nucleic Acids Res 40:2587–2598
9. Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL et al (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. Nucleic Acids Res 38:5601–5608
10. Argast GM, Stephens KM, Emond MJ, Monnat RJ Jr (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J Mol Biol 280:345–353
11. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate

specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128:2477–2484

12. Scalley-Kim M, Connell-Smith A, Stoddard BL (2007) Coevolution of a homing endonuclease and its host target sequence. J Mol Biol 372: 1305–1319
13. Molina R, Redondo P, Stella S, Marenchino M, D'Abramo M et al (2012) Non-specific protein–DNA interactions control I-CreI target binding and cleavage. Nucleic Acids Res 40:6936–6945
14. Takeuchi R, Lambert AR, Mak AN-S, Jacoby K, Dickson RJ et al (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. Proc Natl Acad Sci U S A 108:13077–13082
15. Szeto MD, Boissel SJS, Baker D, Thyme SB (2011) Mining endonuclease cleavage determinants in genomic sequence data. J Biol Chem 286:32617–32627
16. Li H, Pellenz S, Ulge U, Stoddard BL, Monnat RJ Jr (2009) Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. Nucleic Acids Res 37:1650–1662