Analyst

High-throughput direct screening of restriction endonuclease using microfluidic fluorescence-activated drop sorter based on SOS response in *Escherichia coli*

Journal:	*Analyst*
Manuscript ID	AN-ART-01-2024-000106.R1
Article Type:	Paper
Date Submitted by the Author:	13-Apr-2024
Complete List of Authors:	Zhang, Yizhe; Harvard University, Agresti, Jeremy; Harvard University, Physics/SEAS Zheng, Yu; New England Biolabs Inc Weitz, David; Harvard University, Department of Physics

SCHOLARONE™
Manuscripts

High-throughput direct screening of restriction endonuclease using microfluidic fluorescence-activated drop sorter based on SOS response in *Escherichia coli*

Yizhe Zhang[a,b,*], Jeremy J. Agresti[b], Yu Zheng[c], David A. Weitz[b,d,*]

[a] Department of Chemistry and Chemical Biology,
[b] John A. Paulson School of Engineering and Applied Sciences,
[d] Department of Physics, Harvard University, Cambridge, MA 02138, USA
[c] New England BioLabs, Inc., Ipswich, MA 01938, USA

ABSTRACT

A restriction endonuclease (RE) is an enzyme that can recognize a specific DNA sequence and cleave that DNA into fragments with double-stranded breaks. This sequence-specific cleaving ability and its ease of use have made REs commonly used tools in molecular biology since their first isolation and characterization in 1970s. While artificial REs still face many challenges in large-scale synthesis and precise activity control for practical use, searching for new REs in natural samples remains a viable route to expanding the RE pool for fundamental research and industrial applications. In this paper, we propose a new strategy to search for REs in an efficient manner. We construct a host bacterial cell to link the genotype of REs to the phenotype of β-galactosidase expression based on the bacterial SOS response, and use a high-throughput microfluidic platform to isolate, detect and sort the REs in microfluidic drops at a frequency of ~800 drops per second. We employ this strategy to screen for the XbaI gene from constructed libraries of varied sizes. In single round of sorting, a 90-fold target enrichment was obtained within 1 h. Compared to conventional RE-screening methods, the direct screening approach we propose excels at efficient search of desirable REs in natural samples - especially the unculturable samples, and can be tailored to high-throughput screening of a wide range of genotoxic targets.

INTRODUCTION

A restriction endonuclease (RE) is an enzyme that can recognize a specific DNA sequence and cleave that DNA into fragments with double-stranded breaks. This sequence-specific cleaving ability and its ease of use have made REs commonly used tools in molecular biology since their first isolation and characterization in the 1970s [1,2]. With continuous improvement in flexibility and specificity [3-9], REs are gradually gaining traction as promising agents for targeted gene disruption in gene therapy against various virus infectious diseases, such as HIV and HPV [10-13]. New REs with desirable features can be obtained through artificial synthesis or wild-type search. While artificial REs still face challenges in large-scale synthesis and activity control for practical use [5,14], searching for new REs in natural samples remains an indispensable route to expanding RE candidates, for growing needs in fundamental research and industrial applications.

* Corresponding authors.
Email addresses: yizhe_zhang@post.harvard.edu (Y. Zhang), weitz@seas.harvard.edu (D. A. Weitz)

The conventional RE-searching method utilizes modification methylases to indirectly screen for their companion REs. It predicts the potential RE genes by scanning through the sequence database for the companion modification methylase based on its conserved motif element, followed by the evaluation digestion tests to identify the new RE genes [15,16]. The major drawback of this indirect screening method is its dependence on the modification methylase, which precludes potential REs that do not have the companion methylases, for example, *Pac*I [17]. Moreover, the scanning step requires the sample to be pre-sequenced, and the evaluation step requires intense labor, making it a challenge to search for REs in natural samples with potentially large library sizes [18]. A few *in vitro* translation-based selection strategies have been proposed to directly search for new REs; the searching range of those methods is restricted by the stringent requirements on the recognition sequence [19,20]. Thus, more efficient methods to search for new REs would be valuable.

In this paper, we propose a new strategy based on the bacterial SOS response for RE-screening in high throughput [21]: the presence of an RE gene is indicated by the over-expression of β-galactosidase (β-gal) through a specially designed host bacterial cell, which can be detected and sorted by a fluorescence-activated microfluidic drop sorter at an interrogation rate of 0.8 kHz [22]. We apply this strategy to screen for the XbaI gene from a constructed library, and reach a 90-fold target enrichment for a 0.1% library within one hour.

MATERIALS AND METHODS

Host cell preparation

ER2745, an *Escherichia coli* (*E. coli*) derivative strain that contains a fusion of a DNA damage-induced SOS gene *dinD* [21] and an indicator gene *lacZ*, is constructed as the host cell. The strain is deficient in all known endogenous restriction systems and expresses T7 RNA polymerase under *lac* control from a chromosomal location. To suppress the basal level expression of T7 RNA polymerase, ER2745 is transformed with pLysY [23]. ER 2745/pLysY is then transformed with pTXB1_XbaI (*E. coli*^{XbaI} cell) or pTXB1_ΔXbaI (*E. coli*$^{\Delta XbaI}$ cell, with an internal deletion inside the XbaI gene) to compose the model library.

Primer design

For easy identification, primers for the post-sorting PCR are designed to amplify a fragment of the insert such that the amplicons from the target and the control are short enough to run at a detectable distance on the gel without losing the base-pair difference information. Specifically, instead of amplifying the intact target insertion of 628 bp and the control insertion of 572 bp, the primers p_{fwd} (5'-TAGGGGAATTGTGAGCGGATAAC-3') and p_{rev} (5'- GGAATCGGCCCTTGTTTTGATAG-3') target for a 263-bp fragment in XbaI and a 207-bp fragment in ΔXbaI, so the shorter amplicons from the target and the control become well separated in the gel allowing for quantification.

Cell culture

Host *E. coli* cells are cultured overnight in the standard LB medium with antibiotics Ampicillin (Amp) and Chloramphenicol (Cmp) (Sigma) until the OD_{600} reaches about 0.2. The harvested cells are spun down to remove the basal β-galactosidase (β-gal), and resuspended in LB medium and stored on ice for sample preparation.

Sample preparation

For signal generation and detection in bulk, the harvested *E. coli*^{XbaI} cell and *E. coli*$^{\Delta XbaI}$ cell cultures are both induced with 0.5 mM (final concentration) of Isopropyl β-D-1-thiogalactopyranoside (IPTG; Invitrogen), and incubated at 37 °C in the dark for a given period of time before being disrupted,

whereupon 1 mL of the cell suspension is spun down, resuspended in 0.7 mL sonication buffer (100 mM NaCl (Sigma), 25 mM Tris-HCl (Sigma), 1 mM β-mercaptoethanol (β-ME; Sigma); pH 8.0), then sonicated and spun down again to extract β-gal. 5 μL of the sonication extracts from *E. coli*$^{\text{XbaI}}$ cell and *E. coli*$^{\Delta\text{XbaI}}$ cell are added to individual wells of a 96-well plate, where 50 μL of 0.2 mM Fluorescein-Di-β-D-Galactopyranoside (FDG; Life Technologies) is added as the fluorogenic substrate. The mixture is incubated at 37 °C in the dark for 20 min and then a fluorescence measurement is performed with a microplate reader (EM: 490 nm/AB: 514 nm; ThermoFisher Scientific).

For signal generation and detection in the droplet, a co-flow drop-maker microfluidic chip is used for better control over the onset of the enzymatic reactions [24]. For the inner flow, the cells are mixed with Cmp (34 μg/mL), Amp (100 μg/mL) and Pluronic F127 (0.001%, to prevent cells adhering to the PDMS surface [25]) (Sigma) in LB to the density of 10^8 cells/mL. For the middle flow, FDG is mixed in LB to 0.2 mM with IPTG (1 mM, final concentration), sodium *N*-lauroyl sarcosine solution (0.1%, cell lysate buffer to allow FDG in) (Sigma), Pluronic F127 (0.001%), Cmp (34 μg/mL) and Amp (100 μg/mL). The inner flow and the middle flow are infused at equal flow rates to form droplets that contain a mixture of LB with Cmp (34 μg/mL), Amp (100 μg/mL), Pluronic F127 (0.001%), FDG (0.1 mM), IPTG (0.5 mM) and sodium *N*-lauroyl sarcosine (0.05%). The average number of cells per drop is around 0.3, and the drop size is 23 μm in diameter.

Microfluidic device fabrication

The microfluidic devices are fabricated by patterning channels in polydimethylsiloxane (PDMS) using conventional soft lithography methods [26]. Briefly, for a 10-μm drop-maker in our experiments, SU8-3010 photoresist (MicroChem Corp.) is spin-coated onto the 3" silicon wafer and patterned by UV exposure through a photolithography mask. SU8-3025 photoresist is used for a 25-μm sorter and the dropspots devices. After baking and developing with SU-8 developer (propylene glycol methyl ether acetate; MicroChem Corp.), the 10-μm tall positive master of the drop-maker and the 25-μm tall positive master of the sorter and dropspots are formed on the silicon wafers. Then a 10:1 (w/w) mixture of Sylgard 184 silicone elastomer and curing agent (Dow Corning Corp), degassed under vacuum, is poured onto the master and cured at 65 °C for 2 h. Afterwards, the structured PDMS replica is peeled from the master and inlet and outlet ports are punched in the PDMS with a 0.75-mm diameter biopsy punch (Harris Unicore). The PDMS replica is then washed with isopropanol, dried with pressurized air, and bonded to a 50 × 75 mm glass slide (VWR) through oxygen plasma treatment to form the device.

To fabricate the electrodes in the sorter, a 0.1-M solution of 3-mercaptotrimethoxysilane (Gelest) in acetonitrile (99.8%; Sigma) is flushed through the electrode channels and blown dry with pressurized air. A low-melting point solder (Indalloy 19 (52 In, 32.5 Bi, 16.5 Sn) 0.020" diameter wire; Indium Corp.) is infused into the electrode channels at 80 °C. An eight-pin terminal block with male pins (DigiKey) is then glued with Loctite 352 (Henkel) to the surface of the device. The solid electrodes in the shape of the channels are formed when the device is cooled down to room temperature. Electrical contacts are made with alligator clips and connected to a high-voltage amplifier (Trek) that is then connected to the function generator on an FPGA (field-programmable gate array) card (National Instruments).

To form the aqueous-in-oil emulsion, the microfluidic channels are rendered hydrophobic by flushing Aquapel (PPG Industries) through the channels and drying with pressurized air. To produce biocompatible stable drops, we dissolve 1.8% (w/w) EA surfactant (RainDance Technologies) in the fluorinated oil Novec HFE-7500 (3M).

Microfluidic drop-making and sorting

To make the 23-μm-diameter drops, the cell suspension and substrate solution are infused into the inner and middle channels respectively of the 10 μm drop-maker at equal flow rates of 19 μL/h, while the oil is infused into the outer channel at 20 μL/h. The emulsion is collected in a 1-mL plastic

syringe and incubated at 37 ˚C in the dark for 3 h to allow sufficient enzymatic reactions to take place, whereupon the drops are re-injected into a 25-μm sorter. The closely-packed drops are injected at 20 μL/h and are separated by the carrier oil, injected at 200 μL/h, producing drops which flow by the laser-situated detection window at a frequency of ~1.5 kHz, where their fluorescence intensity is interrogated using a photomultiplier tube and a custom LabView program. Drops with a fluorescence intensity above the defined threshold trigger the sorter to which we apply a single-ended electric square wave of 0.8~1.2 kV_{pp} using a frequency of 20 kHz, generated by the function generator on the FPGA. Approximately 5 cycles of the square wave are applied to deflect the selected drop into the collection channel using the dielectrophoretic force [27]. The asymmetric design of the sorting junction defines the default flow to the waste channel when the electric field is not triggered [22,24].

Drop fluorescence imaging

To enable fluorescence imaging of the cell-containing drops over the signal-inducing period, we use the microfluidic dropspots device [28] to hold and trace individual drops longitudinally on the chip. Briefly, upon collecting the drops from the dropmaker device, we inject the emulsion to the dropspots device, so that each droplet occupies a drop-keep chamber (chambers are 40 μm in diameter, 25 μm in height, interconnected by a constriction of 20 μm in width). To immobilize the drops on the chip for timecourse imaging, we insert the pipette tips (filled with the continuous phase) at the inlets and outlet of the device. The device is then situated on a microscope stage with a heater set at 37 ˚C for incubation. To prevent water loss from the drops at elevated temperature over the incubation, we place a piece of damp tissue paper on the edge of the device. Microscopic images are acquired on an inverted epifluorescent microscope (Nikon Eclipse TE2000-E) equipped with a 10/0.30 Plan Fluor objective (Nikon), X-cite series 120 lamp (EXFO, Mississauga, Canada), and a CCD camera (Coolsnap HQ2). Over a 12-h incubation, the brightfield and fluorescence images are acquired at 3-min intervals with exposure times of 22 ms for brightfield and 15 ms for fluorescence.

Colony PCR

Samples from the collection channel are washed in Perfluorooctane solution (20% (v/v) in HFE 7500; Sigma) to break the emulsion. Cells are retrieved in 50 μL nuclease-free water (Life Technologies). For a 25-μL PCR reaction, 2.5 μL of the cell suspension is mixed up with 1.25 μL of p_{fwd} (10 μM), 1.25 μL of p_{rev} (10 μM), 0.5 μL of dNTPs (10 mM; Life Technologies), 0.25 μL of Phusion Hot Start DNA polymerase (2 unit/μL; NEB), 5 μL of 5x Phusion buffer (NEB) and 14.25 μL of nuclease-free water. The inserted DNA fragments on the vector pTXB1 are fully amplified through a 60-cycle PCR process under the fine-tuned condition (initial denaturing: 98 ˚C, 3 min; denaturing: 98 ˚C, 10 s; annealing: 59.2 ˚C, 30 s; extension: 72 ˚C, 1 min).

Gel electrophoresis

One μL of amplicons from colony PCR is mixed with 1 μL of Gel Loading Dye (6x; NEB) and 4 μL of nuclease-free water in each well on the 1.2% agarose gel (Sigma) stained with Ethidium Bromide (1 mg/μL; Life Technologies). 1 μL of 2-Log DNA Ladder (200 μg/mL; NEB) is mixed with 1 μL of Gel Loading Dye (6x) and 4 μL of nuclease-free water for the ladder well. The gel electrophoresis is running at 70 V in 0.5x TBE buffer (Life Technologies) containing 0.5 mg/μL Ethidium Bromide for 45 min.

Data analysis

Data analysis including data fitting and significant analysis in this study was performed with Matlab (MathWorks). All the other data quantitation was performed with Fiji (NIH). In statistical analysis, Mann-Whitney *U* test was used to evaluate the difference between the groups that do not follow the normal distribution, including the quantification of the drop fluorescence intensity between the

drop populations from the "-" cells and "+" cells. Each experimental group contains >10 drops for statistical validity.

RESULTS AND DISCUSSION

Restriction Enzyme Screening Strategy

We accomplished the direct RE screening in three steps: target isolation, signal generation, and target selection. Specifically, the host *Escherichia coli* (*E. coli*) cells that carry the library on the vectors are separated from each other into picoliter aqueous droplets after flowing through the microfluidic drop-maker from the inner channel; chemicals required by the enzymatic reactions for the signal-generation step such as Fluorescein-Di-β-D-Galactopyranoside (FDG) and Isopropyl β-D-1-thiogalactopyranoside (IPTG) are co-encapsulated with cells into the droplets through the middle channel on the co-flow device; fluorinated oil is flowed through the outer channel, as shown in Figure 1. This co-flow design of the device enables a better control over the onset of the enzymatic reactions. The collected emulsion is then incubated at 37 °C in the dark for a given period of time to allow the production of the fluorescence signal molecules through a series of enzymatic reactions in the target-carrying drops. The target-carrying fluorescent drops are then detected and sorted in the microfluidic sorting system.

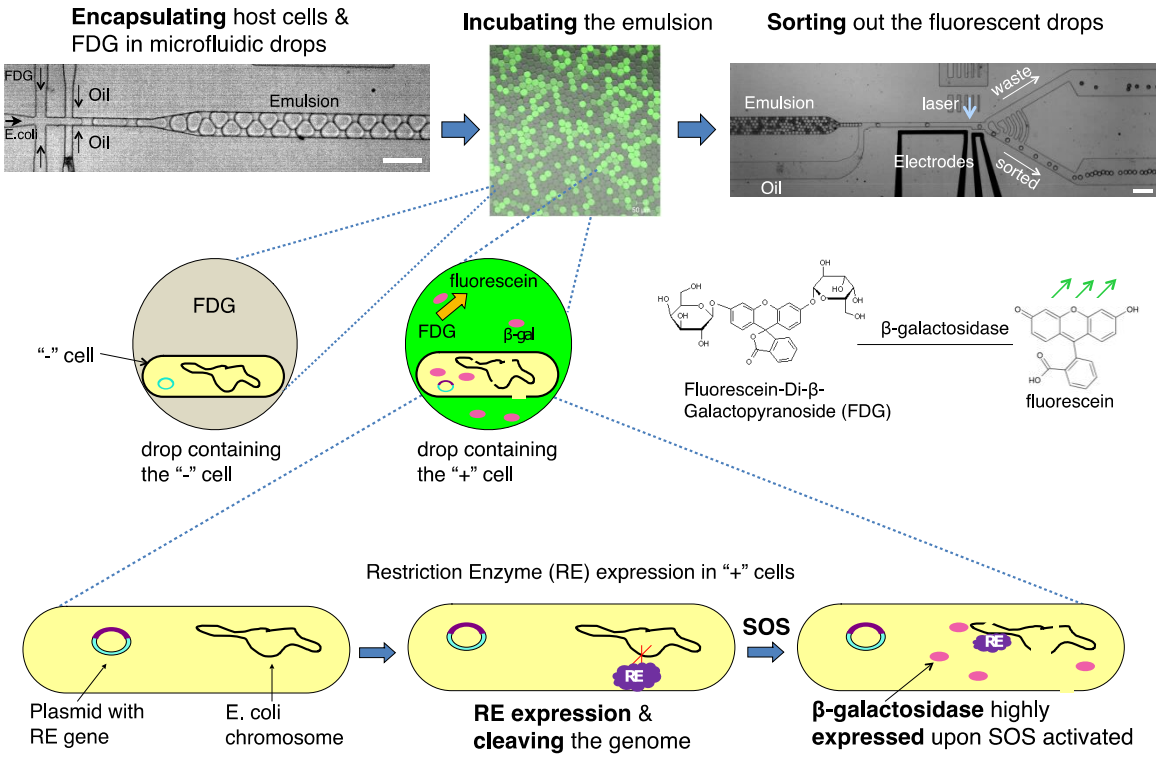

Figure 1 Restriction Enzyme (RE) screening strategy. Screening of RE is comprised of three steps: (1) Co-encapsulation of the host cell *E. coli* and the fluorogenic substrate fluorescein-Di-β-galactopyranoside (FDG). (2) Off-chip incubation to allow for enzymatic reactions. (3) Detecting and sorting on the re-injected emulsion for RE based on the following mechanism: Expressed RE in the host *E. coli* ("+" cell) cleaves the genome and triggers the over-expression of β-galactosidase (β-gal) on the DNA-damage-induced pathway; β-gal can be fluorescently detected through its catalytic activity in the hydrolysis of FDG. Scale bar: 50 μm in the drop-maker micrograph; 100 μm

in the sorter micrograph. "-" cell is the *E. coli* cell that contains the truncated fragment of the RE gene.

To enable signal generation from the target drops, we utilize the host cell's SOS response to activate the expression of the indicator protein. Specifically, we construct a host cell that lacks all known endogenous restriction systems and can express T7 RNA polymerase under *lac* control from a chromosomal location. We then fuse the indicator gene *lacZ* downstream to an SOS-inducible gene *dinD* (Supplementary Information, Figure S1A). If the cell contains the target, REs are expressed upon the induction of IPTG during incubation, as illustrated in Figure 1. Functional REs cause DNA damage to the host cells due to their DNA cleavage activities, and thus trigger the cell's SOS response in an attempt to repair the damage, leading to the over-expression of a myriad of SOS proteins [21], such as RecA. In the form of nucleoprotein filaments, RecA-ssDNA complex catalyzes the self-cleaving reaction of the repressor LexA (Supplementary Information, Figure S1B), thus derepressing the *dinD :: lacZ* expression that leads to the production of the indicator protein β-gal (Supplementary Information, Figure S1C). β-gal can be conveniently detected through the fluorogenic substrate FDG co-encapsulated in drops using our microfluidic fluorescence-activated drop sorter [29].

Signal-generation assay validation

To verify the SOS response-based signal-generation assay, we induce the enzymatic reactions in bulk and in drops respectively, and quantitatively examine the fluorescence signals through the detection experiments. We choose a typical RE XbaI (628 bp) as our target, and its truncated fragment ΔXbaI (572 bp) as the control.

For validation experiments in bulk, the target cells (*E. coli*$^{\text{XbaI}}$) and the control cells (*E. coli*$^{\Delta\text{XbaI}}$) are both incubated for varied durations (0, 1, 2, 3, 4, 5 h) after IPTG induction, followed by the release of β-gal through sonication. The cell lysates from sonication are then added to the FDG-filled microtiter plate for fluorescence measurement. Without incubation, there is only a weak background fluorescence for both cell types; after incubation, *E. coli*$^{\text{XbaI}}$ gives higher fluorescence intensity than *E. coli*$^{\Delta\text{XbaI}}$; as the incubation time is prolonged, the fluorescence intensity measured from *E. coli*$^{\text{XbaI}}$ continues to increase whereas the measurement on *E. coli*$^{\Delta\text{XbaI}}$ shows no obvious changes, as shown by the black and grey bars in Figure 2A. The ratios of fluorescence intensity from *E. coli*$^{\text{XbaI}}$ to that from *E. coli*$^{\Delta\text{XbaI}}$ at different incubation times, defined as the signal-to-noise ratio (s/n), are calculated from the measured data and plotted as the curve in Figure 2B. The time-dependent strong fluorescence from *E. coli*$^{\text{XbaI}}$ indicates the SOS-inducing DNA-cleavage activity of the intact XbaI, and hence validates the proposed SOS response-based signal-generation assay in bulk. We assume the background fluorescence from *E. coli*$^{\Delta\text{XbaI}}$ is probably due to the baseline leak of regulation on the *dinD::lacZ* pathway.

For validation experiments in drops, we choose to generate drops with a diameter of 23 μm, a reliable size for single microbial isolation. To balance the ratio of single-cell drops to the number of empty drops, based on a Poisson distribution, we keep the ratio of the cell number to droplet number at 0.3 when preparing the cell suspension. To induce signal-generation in the drops, we incubate the emulsion at 37 °C for 3 h based on the results from the bulk detection experiments, at which point the signal is large enough (*s*/*n* = 7.5) for gating (Figure 2A). Longer incubation raises the chance of drop coalescence without providing a significant improvement in the *s*/*n* level. One thing distinct from signal generation in bulk is the addition of sodium *N*-lauroyl sarcosine (sarkosyl) to the droplet to facilitate the cell's uptake of FDG, since the mechanical rupture used in the bulk experiments is not used in the drops. To minimize the potential damage to cells, we carefully control the added sarkosyl to maintain a minimum concentration of 0.1%.

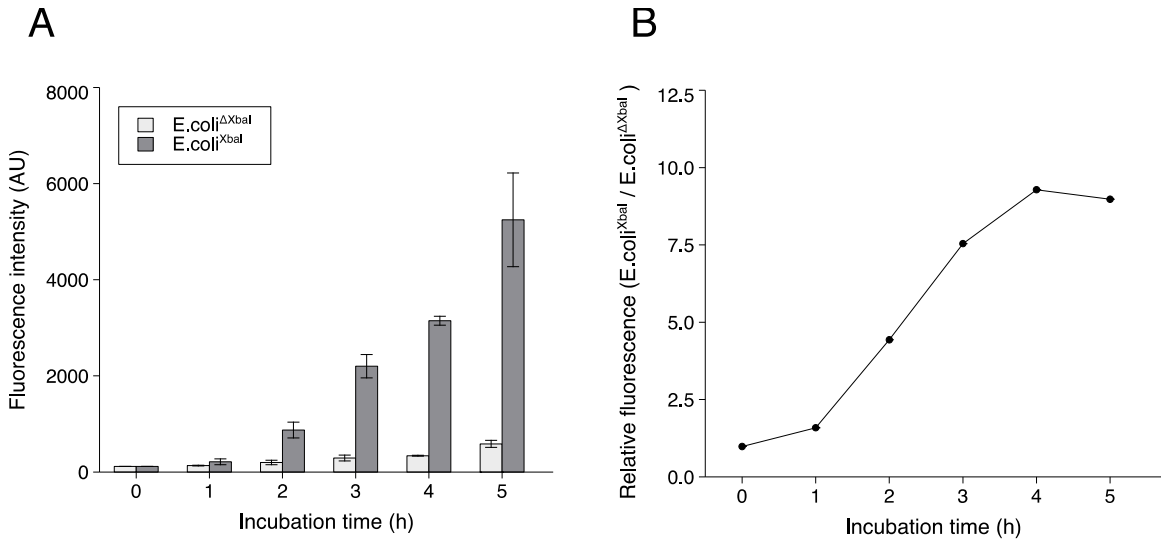

Figure 2 Validation of the signal-generation assay in bulk. (A) Fluorescence intensity from target cells (*E. coli*^{Xbal}) and negative control cells (*E. coli*$^{\Delta Xbal}$) as a function of cell-incubation time. (B) Comparison of the fluorescence intensity between *E. coli*^{Xbal} and *E. coli*$^{\Delta Xbal}$.

As a quick test, we incubate the drops at 37 °C in a dropspots device, and record the change of the fluorescence intensity over a period of 12 h. Figure 3A shows the representative drop fluorescence micrographs (overlaid with bright field images) captured after 3 h of incubation. For *E. coli*^{Xbal}, the detectable fluorescent drops compose about 25% of the total drops (Supplementary Information, Figure S2A), consistent with the Poisson distribution that predicts about 26% of the drops for a 0.3 λ-value. For *E. coli*$^{\Delta Xbal}$, as in the bulk experiments, we see some background fluorescence in the cell-carrying drops, but the fluorescence signal is weak in general: only 10% of the drops give detectable fluorescence (Supporting Information, Figure S2A) and the average drop fluorescence intensity is less than 50% of that measured in the *E. coli*^{Xbal} experiments (Supplementary Information, Figure S2B). There is a significant difference in the fraction of the detectable fluorescent drops, and in the average drop fluorescence intensity between the "+" cell and "-" cell drop populations (Supplementary Information, Figure S2), indicating the potential selectivity of the assay in drops under the incubation condition.

To assess the fluorescence intensity with higher sensitivity in a larger drop population, we re-inject the incubated drops in a sorter chip and interrogate the drop fluorescence individually using a photomultiplier tube and a custom LabView program, when the drops pass by the laser-situated detection window. We examine over 30,000 drops for "-" cells, "+" cells, and empty drops, respectively, and obtain the drop fluorescence distribution for each group (Figure 3B). In all the groups, the majority population is the empty drops, because the cell suspension is diluted below one cell per drop for single-cell encapsulation. In "+" cell drops, we observe a subpopulation with a higher level of fluorescence intensity (mean = 1.90 V), as shown in Figure 3B, corresponding to the drops that contain the "+" cells. In "-" cell drops, we also detect a subpopulation (mean = 0.34 V) apart from the empty drops, albeit having a lower fluorescence intensity level, compared to the subpopulation in the "+" cell drops. We attribute the above-baseline level of the fluorescence intensity from the non-empty drop subpopulation in the "-" cell drops to the baseline expression of β-gal from the "-" cells. Not surprisingly, we observe a single population (mean = 0.007 V) in the drop fluorescence distribution from the no-cell drop group (Supplementary Information, Figure S3).

Interestingly, we notice that there is an increase in the fluorescence intensity level of the empty drop population from the "-" cell drop experiments (mean = 0.02) to the "+" cell drop experiments (mean = 0.29), and the average empty drop fluorescence level from the "+" cell drop experiments and the "-" cell drop experiments are both higher than the average empty drop fluorescence level

from the no-cell drop experiments. We attribute the increase of the empty-drop fluorescence intensity level in the cell-containing drop experiments to the inter-drop diffusion of the fluorescent product fluorescein during the enzymatic incubation [30].

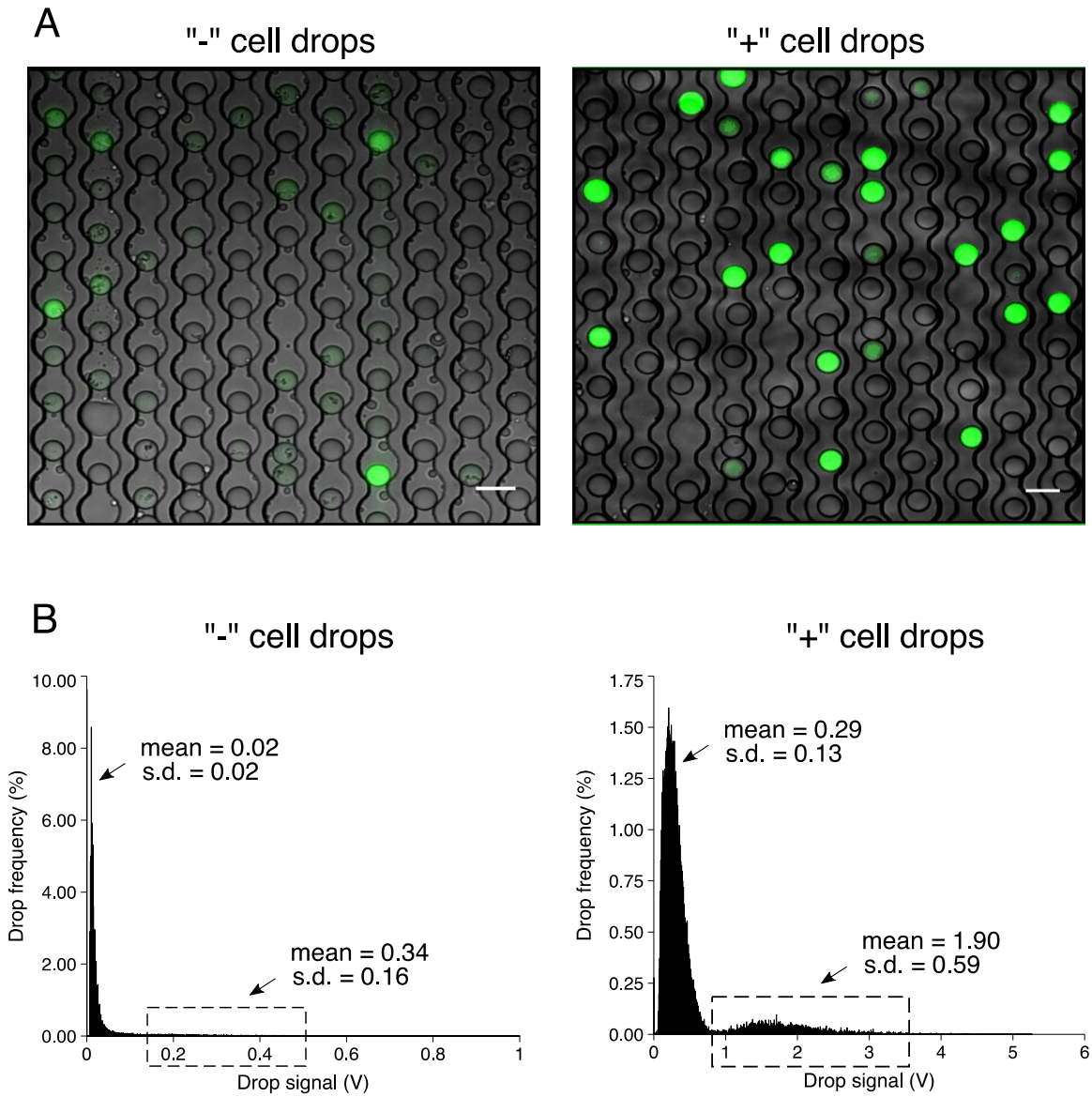

Figure 3 Validation of the signal-generation assay in drops. (A) Representative drop fluorescence micrographs (overlaid with bright field images) captured after signal-generation incubation in the signal detection experiments. The emulsion was allowed to flow into a dropspots device for imaging. Left: *E. coli*$^{\Delta Xbal}$ - encapsulated drops, or "-" cell drops for brevity. Right: *E. coli*^{Xbal} - encapsulated drops, or "+" cell drops for brevity. Scale bar: 40 μm. (B) Distribution of the drop fluorescence signal detected from "-" cell drops and "+" cell drops respectively, upon incubation through the photomultiplier tube at the custom detection setup. The mean values and the standard distribution (s.d.) values were obtained from the Gaussian fitting to individual subpopulations. Incubation condition: 3 h at 37 °C. The average cell number per drop was kept at 0.3 in all of the droplet experiments in this study.

Both in bulk and in drops, we observe strong fluorescence signals from the target cells after the enzymatic activities in the cell are induced, which validates the proposed signal-generating assay on the basis of SOS response in the host cell.

Sorting of XbaI

Upon validating the assay, we performed the sorting of the target RE, XbaI, from our model libraries. We construct our model libraries using *E. coli*^{Xbal} as target cells and *E. coli*$^{\Delta Xbal}$ as control cells. As a demonstration of the principle, four libraries with sizes of 1:2, 1:10, 1:100, and 1:1000 are prepared by mixing *E. coli*^{Xbal} and *E. coli*$^{\Delta Xbal}$ at ratios of 1:1, 1:9, 1:99, 1:999, respectively.

For each library, cells are encapsulated into aqueous drops along with IPTG, FDG and sarkosyl through the drop-making process. The collected emulsion is incubated off-chip in the dark at 37 °C for 3 h before being injected into the microfluidic sorter for detection and sorting. By monitoring the real-time distribution of the drop fluorescence intensity as calculated by LabView, we are able to instantly estimate the target-drop population and set an appropriate sorting threshold accordingly. Drops with fluorescence intensity above the threshold trigger the electric field and are deflected into the collection channel under the dielectrophoretic force [27]. The consecutive snapshots from a fast-camera movie show the process of a target drop being deflected (indicated by white arrows in Figure 4). Compared to the non-target drop flowing through the sorting junction (Figure 4, the right drop in t_1), the target drop (Figure 4, the right drop in t_3) shows a distortion in shape and a shift in transversal position as a result of dielectrophoresis.

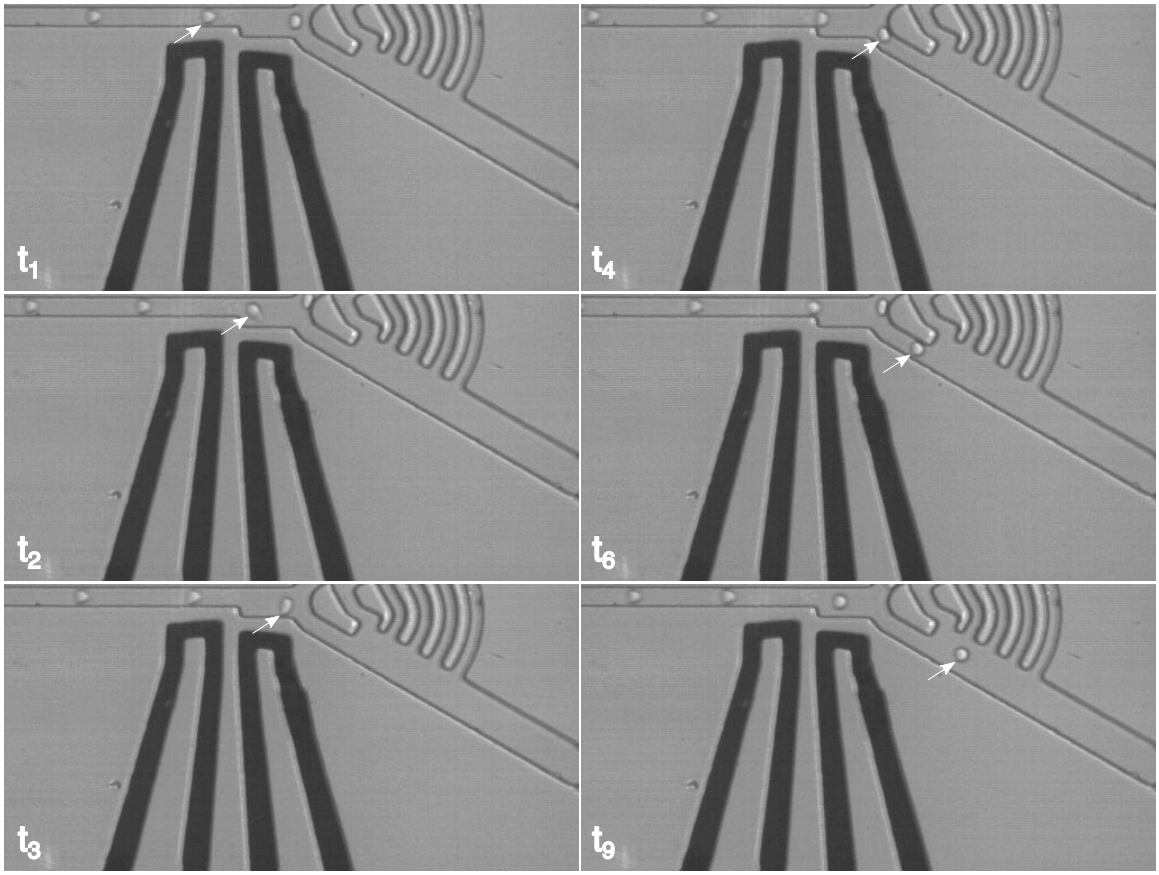

Figure 4 Microfluidic sorting process of XbaI. Snap shots of the sorting process on a target drop recorded by the fast camera (Photron Ultima512, 8000 fps). White arrows mark the target drop

being deflected by dielectrophoresis. t_i (i = 1, 2, 3, 4, 6, 9) indicates the time points when the snap shots were taken. t_{i+1} - t_i = 0.125 ms.

The duration from the fluorescence detection to the completion of the dielectrophoretic action on a droplet is ~1.25 ms, so the sorting frequency is ~0.8 kHz in our validation experiments. With this sorting rate, our system can complete screening for a library of 10^6 in size within 30 min. Further increase in sorting frequency is possible by tuning the parameters such as the drop size and flow rates [24].

We analyze the sorting result by evaluating the target enrichment in the sorted sample using gel electrophoresis. Briefly, the inserted DNA fragments XbaI or ΔXbaI on the vector pTXB1 from the sorted sample are recovered into distilled water after oil removal, and then amplified by colony PCR. The amplicons from the two templates are then separated by length on the agarose gel through electrophoresis. We notice that the separation resolution in the gel electrophoresis is not sufficient to resolve the similar-sized fragments of the amplicons from the templates that contain the complete XbaI or ΔXbaI (830 bp and 774 bp), as shown by the dual peaks in Figure 5A. Therefore, we choose the small segments of XbaI and ΔXbaI as the PCR templates for "+" and "-" cells (263 bp and 207 bp), respectively, which covers and enhances the fragment difference between the "+" and "-" cells for better separation in gel electrophoresis, as shown in Figure 5B-5D.

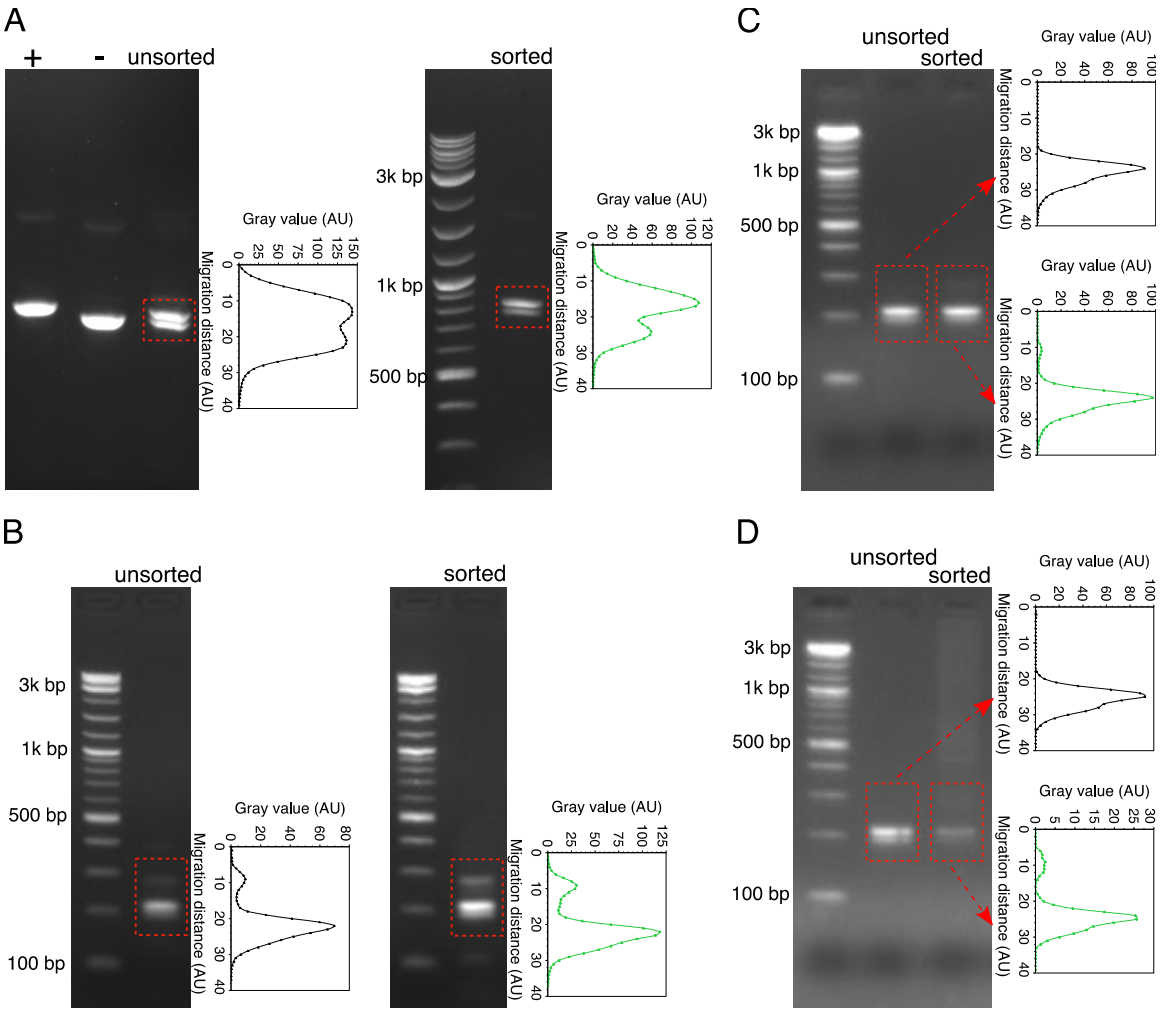

Figure 5 RE gene XbaI is enriched through drop-based microfluidic sorting. Agarose gel images and the corresponding intensity curves of the PCR amplicons from XbaI and ΔXbaI, in the unsorted and sorted samples from a library of (A) 1:2 (B) 1:10 (C) 1:100 and (D) 1:1000 (ratio of "+" cells to total cells). The sorting enrichment in the library of 1:2, 1:10, 1:100, and 1:1000 is 1.7, 1.8, 2.5, 90.0 at the selected sorting threshold. For the 1:2 library, the PCR templates are the entire XbaI and ΔXbaI; for all the other libraries, the PCR templates are the characteristic fragments of XbaI and ΔXbaI for a higher separation resolution in gel electrophoresis. The sorted results were obtained from the single-round of sorting for all the libraries.

For each library, through quantification of the intensity of the DNA bands and curve fitting, we can obtain the ratio of the concentrations of XbaI and ΔXbaI in the sorted sample. We define our sorting enrichment as:

$$E = \frac{N_{final+}/N_{final-}}{N_{initial+}/N_{initial-}}, \qquad (1)$$

where E is the enrichment, N_{final+} is the number of target DNA insertion (XbaI) in the sorted sample, N_{final-} is the number of truncated insertion (ΔXbaI) in the sorted sample, $N_{initial+}$ and $N_{initial-}$ represent the numbers of XbaI and ΔXbaI in the unsorted sample.

After a single round of sorting, "+" cells are enriched in all the libraries, as indicated by the enhanced relative intensity between the XbaI band and ΔXbaI band in the sorted samples compared to the unsorted samples. Further quantification on the gel intensity reveals that the enrichment is ~2 fold with a low sorting threshold (Supplementary Information, Figure S4), as shown in Figure 5A-5C for libraries of 1:2, 1:10, and 1:100. At a high sorting threshold for the library of 1:1000 (Supplementary Information, Figure S4), the sorting enrichment increases to ~90 fold, indicated by the detectable XbaI band in the sorted sample of Figure 5D.

As described above, the assessment of the sorting enrichment is accomplished through analytical gel electrophoresis of the colony PCR products from the collected samples. Colony PCR is a convenient assay in rapidly evaluating the sorting efficacy for the proof of concept. For the library of 1:2 and 1:10, dual peaks across 800bp (in 1:2 library) and between 200-300 bp (in 1:10 library) are detected in the unsorted sample, indicating the successful colony PCR from both "+" and "-" cells. The intensity ratio of the dual bands is 1:0.8 (in 1:2 library), and 1:8.2 (in 1:10 library) from the "+" and "-" cells, suggesting no substantial bias of the colony PCR from the "+" and "-" cells (See further discussion in Supplementary Information). For the large-sized library (<1:100), the simple agarose gel assay is no longer able to detect the minority component of the binary library even for its PCR products. Therefore, we use the corrected nominal values as the initial ratios ("+":"-" = 1:84 and 1:849 for 1:100 and 1:1000 libraries, respectively; See details in Supplementary Information) to calculate the sorting enrichment.

Ideally, if the fluorescence is only generated from the target, the enrichment would be determined by the distribution of the targets over the drops upon encapsulation, which, in our drop experiments, is the Poisson distribution. Therefore, merely by diluting the cell suspension to minimize the drops with more than one cell, the enrichment can be improved [22]. In fact, we observe in our experiments that control cells ("-" cells that are transformed with the truncated insertion ΔXbaI) also generate a background fluorescence (Figure 3B), which we assume is due to some leakiness of the *dinD::lacZ* pathway, so a higher sorting threshold can also effectively suppress the number of the false positive drops and lead to a higher enrichment, as demonstrated in Figure 5. However, improving the enrichment by diluting the cell suspension or increasing the sorting threshold requests longer collection time to obtain sufficient samples for post-sorting processing. Thus, the sorting threshold we chose in our experiments is a trade-off between enrichment and efficiency, which allows us to reach a reasonable enrichment within an hour of sorting for the large-sized model libraries. For screening of larger libraries, it would be more practical to run a multi-round

sorting than to optimize the enrichment with single-round sorting through modulating the sorting threshold or the cell density.

Despite the high throughput of the microfluidic sorting (~800 drops/s) in drop detection and interrogation, the practical library size is largely determined by the considerations of the time restraint and downstream processing requirement. In this study, we completed the sorting experiment within 1 h, to reduce the effect of progressive enzymatic activities in live cells over extended periods of time, such as generation and secretion of proteins that could cause drop coalescence and/or channel clogging. Therefore, we limit the library size to 1:1000 for the consideration of the detection capacity of the colony PCR and analytical gel electrophoresis in analyzing the sorted samples (See more discussion in Supplementary Information). For libraries larger than 1:1000, a multi-round sorting strategy can be adopted where the target cells are enriched gradually with a decreasing λ value for every sorting round; alternatively, a higher-resolution DNA analytical assay, for example, digital PCR or high-throughput sequencing, can be used to validate the sorting results with minute quantities of the samples.

CONCLUSIONS

We demonstrate the principle of an SOS response-based direct RE-screening strategy on the microfluidic platform through successful enrichment of the XbaI gene from model libraries with varied sizes. Compared to the conventional RE-screening method [16], our direct screening approach shows the potential for efficient search of desirable REs in natural samples. It does not rely on the modification methylase and therefore has a broader searching range; it does not involve any database-scanning and thus not requires pre-sequencing for the sample; and it runs on a microfluidic platform and therefore is highly cost- and time-efficient.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (DMR-1310266 to D.A.W.) and the Harvard Materials Research Science and Engineering Center (DMR-1420570 to D.A.W.). We thank Dr. Richard J. Roberts at New England BioLabs for critical reading of the manuscript.

REFERENCES

[1] H. O. Smith, K. W. Wilcox, A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties, J. Mol. Biol. 51 (1970) 379. https://doi.org/10.1016/0022-2836(70)90149-x.
[2] P. H. Johnson, A. S. Lee, R. L. Sinsheimer, Production of specific fragments of φX174 replicative form DNA by a restriction enzyme from *Haemophilus parainfluenzae* endonuclease HP, J. Virol. 11 (1973) 596. https://doi.org/10.1128/jvi.11.4.596-599.1973.
[3] J. A. Townsend, D. A. Wright, R. J. Winfrey, F. Fu, M. L. Maeder, J. K. Joung, et al., High-frequency modification of plant genes using engineered zinc-finger nucleases, Nature 459 (2009) 442. https://doi.org/10.1038/nature07845.
[4] T. Li, S. Huang, W. Z. Jiang, D. Wright, M. H. Spalding, D. P. Weeks, et al., TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain, Nucleic Acids Res. 39 (2011) 359. https://doi.org/10.1093/nar/gkq704.
[5] M. M. Mahfouz, L. Li, M. Shamimuzzaman, A. Wibowo, X. Fang, J.-K. Zhu, De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 2623. https://doi.org/10.1073/pnas.1019533108.

[6] A. J. Wood, T.-W. Lo, B. Zeitler, C. S. Pickle, E. J. Ralston, A. H. Lee, et al., Targeted genome editing across species using ZFNs and TALENs, Science 333 (2011) 307. https://doi.org/10.1126/science.1207773.
[7] L. Cong, R. Zhou, Y.-C. Kuo, M. Cunniff, F. Zhang, Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains, Nat. Commun. 3 (2012) 968. https://doi.org/10.1038/ncomms1962.
[8] N. E. Sanjana, L. Cong, Y. Zhou, M. M. Cunniff, G. Feng, F. Zhang, A transcription activator-like effector toolbox for genome engineering, Nat. Protoc. 7 (2012) 171. https://doi.org/10.1038/nprot.2011.431.
[9] I. Fonfara, U. Curth, A. Pingoud, W. Wende, Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases, Nucleic Acids Res. 40 (2012) 847. https://doi.org/10.1093/nar/gkr788.
[10] M. Wayengera, H. Kajumbula, W. Byarugaba, Identification of restriction endonuclease with potential ability to cleave the HSV-2 genome: inherent potential for biosynthetic versus live recombinant microbicides, Theor. Biol. Medical Model. 5 (2008) 18. https://doi.org/10.1186/1742-4682-5-18.
[11] J. T. Schiffer, M. Aubert, N. D. Weber, E. Mintzer, D. Stone, K. R. Jerome, Targeted DNA mutagenesis for the cure of chronic viral infections, J. Virol. 86 (2012) 8920. https://doi.org/10.1128/JVI.00052-12.
[12] N. Manjunath, G. Yi, Y. Dang, P. Shankar, Newer gene editing technologies toward HIV gene therapy, Viruses 5 (2013) 2748. https://doi.org/10.3390/v5112748.
[13] P. Tebas, D. Stein, W. W. Tang, I. Frank, S. Q. Wang, G. Lee, et al., Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV, N. Engl. J. Med. 370 (2014) 901. https://doi.org/10.1056/NEJMoa1300662.
[14] V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection, Nat. Methods 8 (2011) 765. https://doi.org/10.1038/NMETH.1670.
[15] R. J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE: a database for DNA restriction and modification: enzymes, genes and genomes, Nucleic Acids Res. 51 (2023) D629. https://doi.org/10.1093/nar/gkac975.
[16] R. Schleif, Assaying of organisms for the presence of restriction endonucleases, Methods Enzymol. 65 (1980) 19. https://doi.org/10.1016/s0076-6879(80)65004-6.
[17] M. Nelson, M. McClelland, Site-specific methylation: effect on DNA modification methyltransferases and restriction endonucleases, Nucleic Acids Res. 19 (1991) 2045. https://doi.org/10.1093/nar/19.suppl.2045.
[18] M. L. Sogin, H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, et al., Microbial diversity in the deep sea and the underexplored ‘‘rare biosphere’’, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 12115. https://doi.org/10.1073/pnas.0605127103.
[19] N. Doi, S. Kumadaki, Y. Oishi, N. Matsumura, H. Yanagawa, *In vitro* selection of restriction endonucleases by *in vitro* compartmentalization, Nucleic Acids Res. 32 (2004) e95. https://doi.org/10.1093/nar/gnh096.
[20] Y. Zheng, R. J. Roberts, Selection of restriction endonucleases using artificial cells, Nucleic Acids Res. 35 (2007) e83. https://doi.org/10.1093/nar/gkm410.
[21] C. Janion, Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli, Int. J. Biol. Sci. 4 (2008) 338. https://doi.org/10.7150/ijbs.4.338.
[22] J.-C. Baret, O. J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, et al., Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity, Lab Chip 9 (2009) 1850. https://doi.org/10.1039/b902504a.
[23] F. W. Studier, Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system, J. Mol. Biol. 219 (1991) 37. https://doi.org/10.1016/0022-2836(91)90855-z.
[24] J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, et al., Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 4004. https://doi.org/10.1073/pnas.0910781107.
[25] V. N. Luk, G. C. H. Mo, A. R. Wheeler, Pluronic additives: a solution to sticky problems in digital microfluidics, Langmuir 24 (2008) 6382. https://doi.org/10.1021/la7039509.
[26] J. C. McDonald, D. C. Duffy, J. R. Anderson, D. T. Chiu, H. Wu, O. J. A. Schueller, et al.,

Fabrication of microfluidic systems in poly(dimethylsiloxane), Electrophoresis 21 (2000) 27. https://doi.org/10.1002/(SICI)1522-2683(20000101)21:1<27::AID-ELPS27>3.0.CO;2-C.
[27] K. Ahn, C. Kerbage, T. P. Hunt, R. M. Westervelt, D. R. Link, D. A. Weitz, Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices, Appl. Phys. Lett. 88 (2006) 024104. https://doi.org/10.1063/1.2164911.
[28] C. H. J. Schmitz, A. C. Rowat, S. Köster, D. A. Weitz, Dropspots: a picoliter array in a microfluidic device, Lab Chip 9 (2009) 44. https://doi.org/10.1039/B809670H.
[29] A. Plovins, A. M. Alvarez, M. Ibañez, M. Molina, C. Nombela, Use of fluorescein-di-β-D-galactopyranoside (FDG) and C12-FDG as substrates for β-galactosidase detection by flow cytometry in animal, bacterial, and yeast cells, Appl. Environ. Microbiol. 60 (1994) 4638. https://doi.org/10.1128/aem.60.12.4638-4641.1994.
[30] P. Gruner, B. Riechers, B. Semin, J. Lim, A. Johnston, K. Short, et al., Controlling molecular transport in minimal emulsions, Nat. Commun. 7 (2016) 10392. https://doi.org/10.1038/ncomms10392.