Electronic Supplementary Material (ESI) for Analyst.

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.

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

RESULTS AND DISCUSSION

Practical library size

As discussed in the main text, the sorting frequency in our drop experiments is ~0.8 kHz. Hence, within 1 h, a total of ~2 million drops are passing through the detection window and sorted out. On the other hand, the λ value is maintained at 0.3 for a marked level of single-cell occupancy. Hence, an hour of sorting interrogates ~500,000 cell-containing drops, which would contain ~500 "+" cell drops for a library of 1:1000. Assume we can harvest 100% of the "+" cell drops (which is not possible in practical experiments because of the continuous distribution curve in drop fluorescence), the total number of ~500 "+" drops equals to a total aqueous volume of ~3 nL. The minute amount of the collected sample poses a big challenge in sample retrieval and preparation for downstream analysis. We failed to retrieve the sorted sample from libraries larger than 1:1000 using the setup and procedures developed in this study.

PCR bias correction

Multi-template PCR is known to introduce bias between the templates due to different amplification efficiencies on different templates [1-3]. The PCR bias can result in an amplification difference of over 3.5 fold between the templates [4]. To evaluate the bias in the colony PCR between the "+" and "-" cells, we extract the intensity curves from the gel images, fit the curves with the 2-term Gaussian distribution model, and approximate the quantity of the PCR product from the AUC of the fitting curve (between the region of μ-σ and μ+σ). The quantification of the PCR products from the unsorted samples for the libraries of 1:2 and 1:10 reveals no substantial difference in amplification efficiency between the "+" and "-" cells. But we do notice a slight preference of amplification towards the "+" cells, suggested by the slightly higher intensity from the "+" cells. We attribute the PCR bias of the "+" cells over the "-" cells to the higher accessibility of the "+" cell plasmids in colony PCR, because after RE expression and sorting, "+" cells are less healthy with more compromised integrity.

To correct for the detected PCR bias, we adopt the log-ratio linear model developed by Suzuki and Giovannoni, which is simple and works well the two-template PCR bias [5]. In the log-ratio linear model, the ratio of the PCR products from the two DNA templates is expected to be $a_1/a_2 \times (b_1/b_2)^n$, where a_1/a_2 is the initial ratio of the DNA templates before PCR amplification, b_1/b_2 is the ratio of the amplification efficiency at each PCR cycle, and n is the cycle number of the PCR. Since we keep the same PCR condition throughout different libraries, the bias remains constant across different libraries. We calculate the correction factor $(b_1/b_2)^n$ from the results with the libraries of 1:2 and 1:10, respectively, and use the average value (1.18) to obtain the ratio before sorting for enrichment calculation.

SUPPLEMENTARY FIGURES

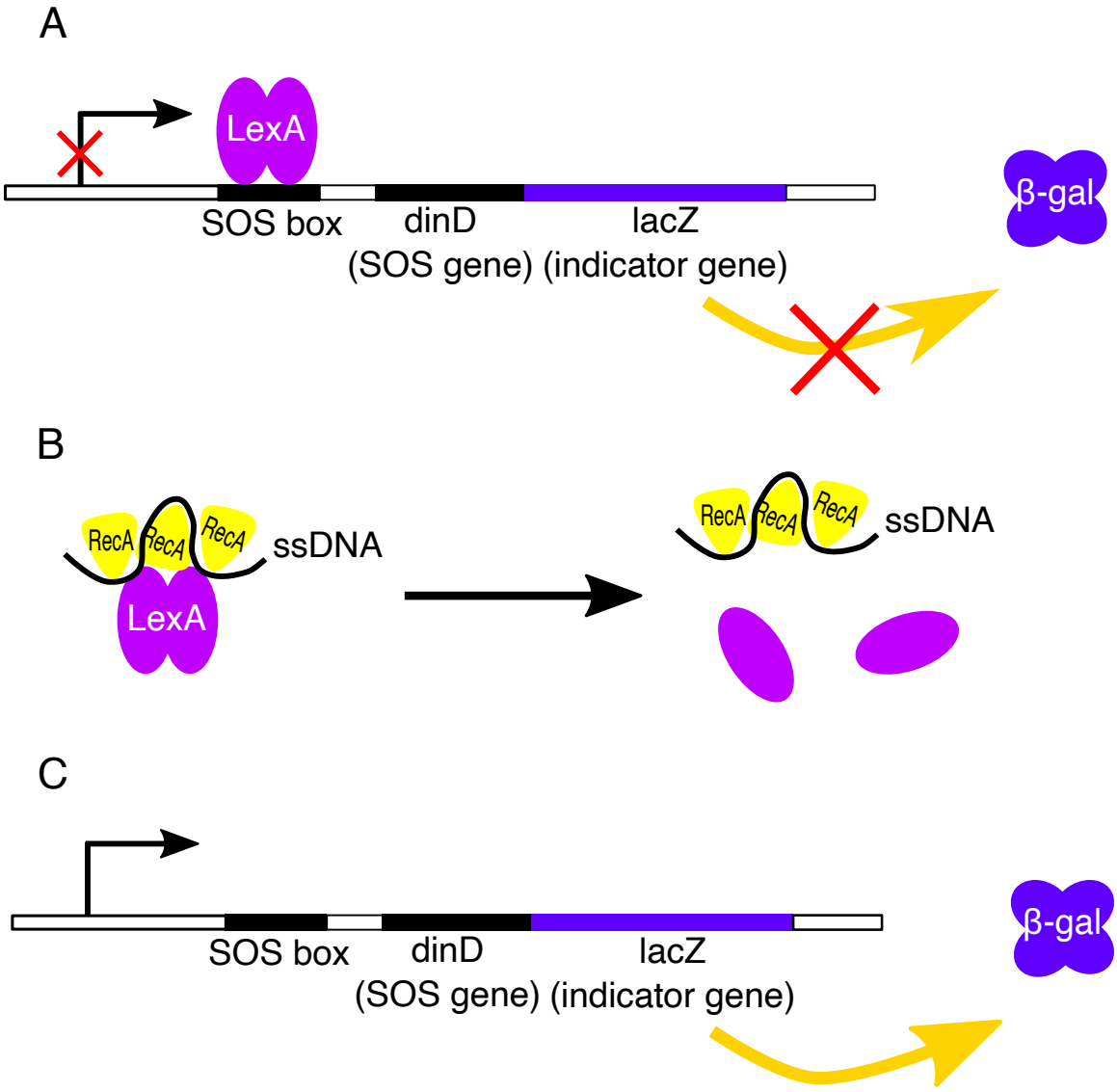

Figure S1 Restriction Enzyme (RE) detection mechanism. Presence of RE in the host *E. coli* triggers the over-expression of β-galactosidase (β-gal) on the DNA-damage-induced pathway. (A-C) SOS-regulated expression of indicator gene. (A) In the normal state, SOS genes are repressed by the repressor binding protein LexA. (B) When SOS is activated by the functional RE, SOS proteins such as RecA accelerate the hydrolysis of LexA, resulting in (C) the derepression of the *dinD :: lacZ* expression that leads to the production of the indicator protein β-gal. For clarity, the symbols in the legend are not to scale.

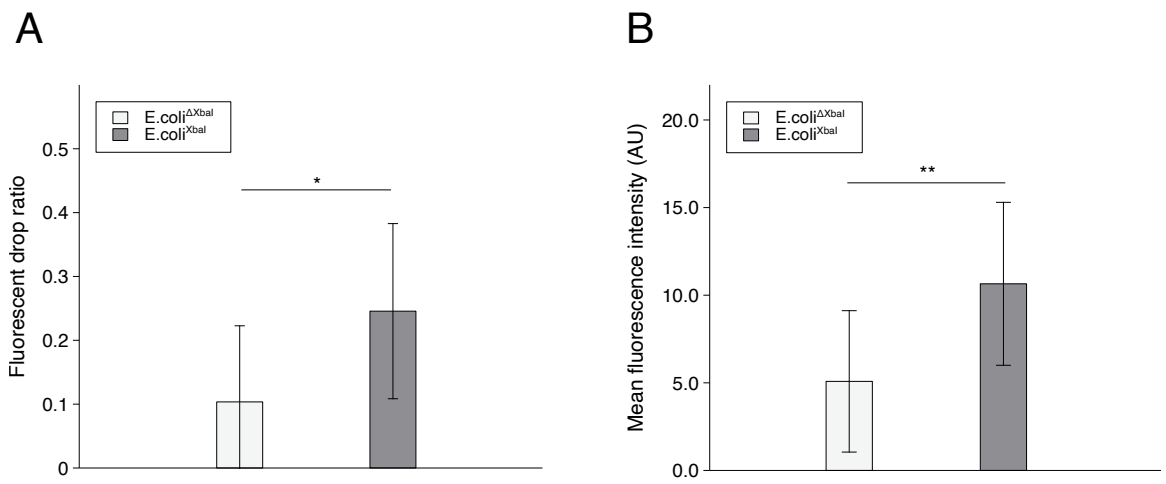

Figure S2 Quantification of the detected fluorescence signals in drop experiments. (A) Fraction of detectable fluorescent drops in total drops for *E. coli*^{XbaI} and *E. coli*$^{\Delta XbaI}$. (B) Average drop fluorescence intensity for *E. coli*^{XbaI} and *E. coli*$^{\Delta XbaI}$. The average cell number per drop was kept at 0.3 in all of the droplet experiments in this study. Mann-Whitney significance levels: *, p (0.0245) < 0.05; **, p (0.0058) < 0.01. n > 10 for each group.

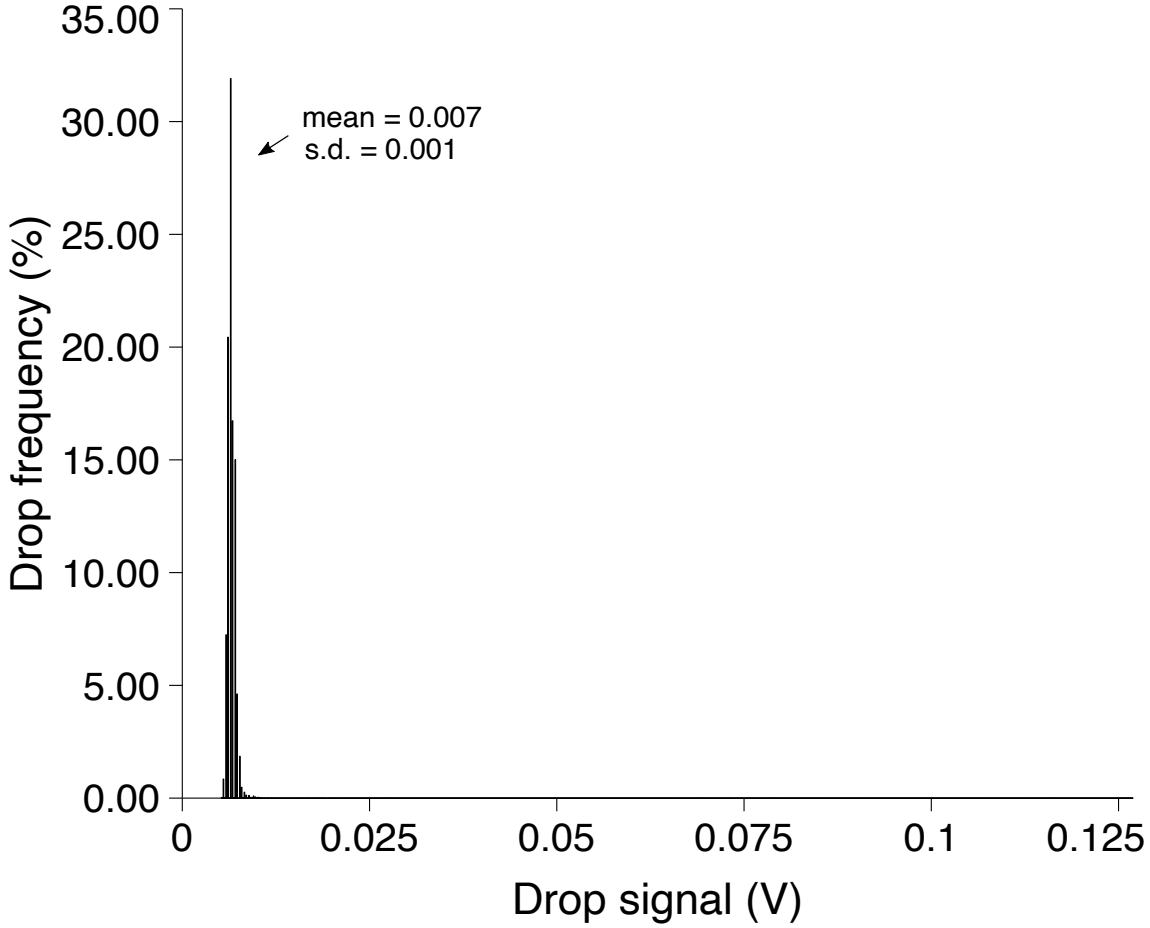

Figure S3 Distribution of the drop fluorescence signal detected from empty drops through the photomultiplier tube at the custom detection setup. The mean values and the standard distribution (s.d.) values were obtained from the Gaussian fitting to the population. The same incubation condition (3 h at 37 °C) was applied to the empty drops before performing fluorescence detection. We chose Gaussian distribution in fitting because the fluorescence intensity from each droplet was a result of a series of enzymatic reactions (through transcription and translation) and electrical (through PMT) amplification, which respectively generated normally distributed amplicons in drop population.

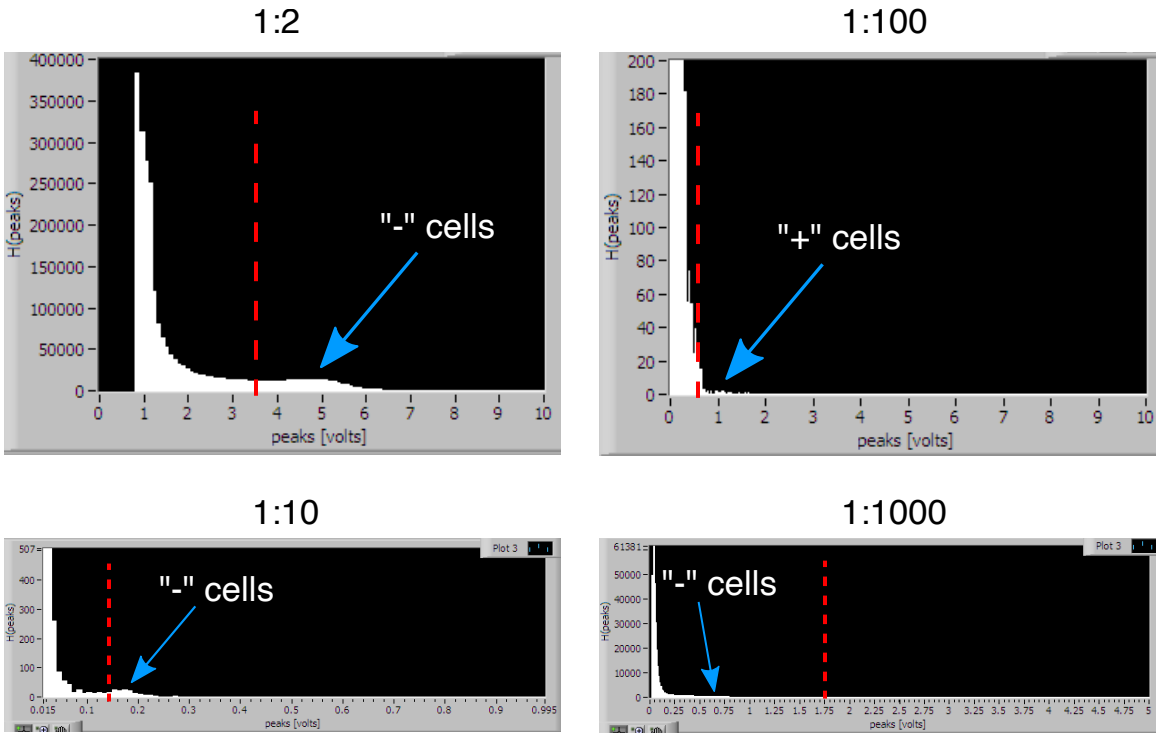

Figure S4 Screenshots of the FPGA control panels in the sorting experiments, showing the real-time distribution of the drop fluorescence signal, and the sorting threshold selected *in situ* for varied-sized libraries. The red line indicates the selected sorting threshold for each library.

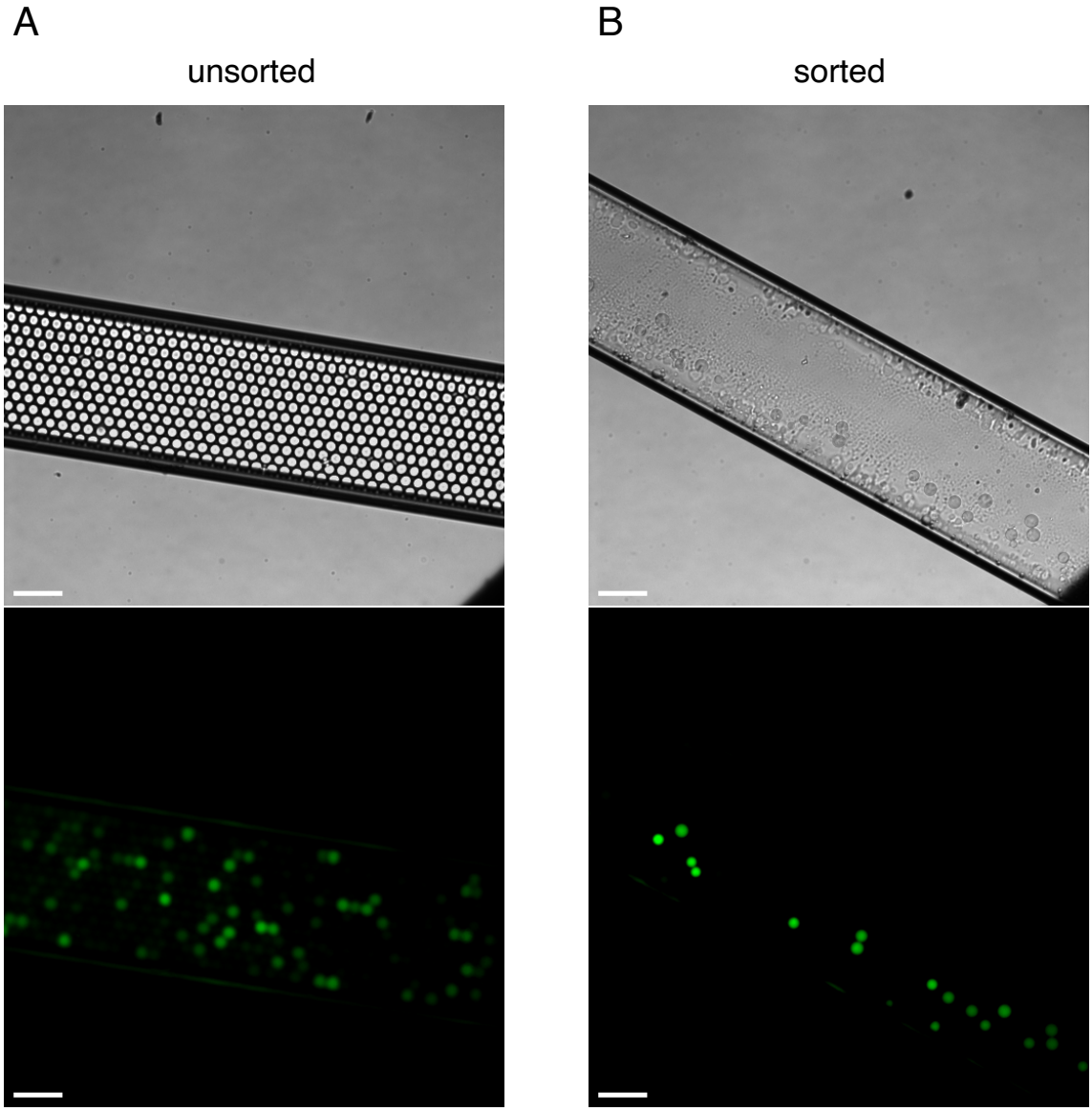

Figure S5 Representative bright-field and fluorescence images of the microfluidic drops before and after sorting experiments. The ratio of the "+" and "-" cells was 1:4 before sorting. Incubation: 2.5 h at 37 °C. Scale bar: 100 μm. The drops were captured in a quartz capillary tube with a cross section of 300(W) x 30(H) μm^2 for imaging.

REFERENCES

[1] S. G. Acinas, R. Sarma-Rupavtarm, V. Klepac-Ceraj, M. F. Polz, PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl. Environ. Microbiol. 71 (2005) 8966. https://doi.org/10.1128/AEM.71.12.8966-8969.2005.

[2] M. Eisenstein Microbiology: making the best of PCR bias. Nat. Methods 15 (2018) 317. https://doi.org/10.1038/nmeth.4683.

[3] M. I. Love, J. B. Hogenesch, R. A. Irizarry, Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat. Biotechnol. 34 (2016) 1287. https://doi.org/10.1038/nbt.3682.

[4] M. F. Polz, C. M. Cavanaugh, Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. 64 (1998) 3724. https://doi.org/10.1128/AEM.64.10.3724-3730.1998.

[5] M. T. Suzuki, S. J. Giovannoni, Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62 (1996) 625. https://doi.org/10.1128/aem.62.2. 625-630.1996.