SUPPLEMENTARY INFORMATION

Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution

Derek Vallejo[#1], Ali Nikoomanzar[#1], Brian M. Paegel[4], and John C. Chaput*[1,2,3]

[1]Department of Pharmaceutical Sciences, [2]Department of Chemistry, [3]Department of Molecular Biology and Biochemistry. University of California, Irvine, CA 92697-3958.

[4]Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458

This file contains:

Materials and Reagents

DNA oligonucleotides (Table S1) were purchased from Integrated DNA Technologies (Coralville, Iowa). TNA triphosphates were chemically synthesized as previously described.[1] Oligonucleotides were quantified by UV absorbance with a NanoDrop 2000 (Thermofisher Scientific, Waltham, Massachusetts). All commercial buffers, enzymes, and the Q5 site-directed mutagenesis kit were purchased from New England Biolabs (Ipswich, MA), unless otherwise noted. Chemical reagents, including dNTPs, were purchased from Sigma Aldrich (St. Louis, Missouri). Ethylenediaminetetraacetic acid (EDTA), urea, acrylamide, and bis-acrylamide were purchased from Thermofisher Scientific. Tetramethylethylenediamine (TEMED) and Mini-PROTEAN precast gels were purchased from Bio-Rad (Hercules, California). Poly(dimethyl) siloxane (PDMS) base and curing agent was purchased from Dow Corning (Midland, MI). SU-8 2010 and 2025 photoresists (Microchem, Westborough, MA) were purchased from Fisher Scientific (Hampton, NH). 3-inch silicon wafers were purchased from the Polishing Corp. of America (Santa Clara, CA). (Tridecafluoro-1,1,2,2-tetrahydrooctyl) trichlorosilane (CAS: 78560-45-9**,** SIT8174.0) was purchased from Gelest Inc. (Morrisville, PA). Fluorinated oil HFE-7500 was purchased from 3M Novec (St Paul, MN), and Pico-Surf™ 1 surfactant, Pico-Glide™ 1, and Pico-Break™ 1 were all purchased from Dolomite Microfluidics (UK). All dichroic mirrors and optical filters were purchased from Semrock (Rochester, New York). Photon counting photomultiplier tubes (PMTs) were purchased from Hamamatsu Photonics (Model# H7828, Japan).

Optical Setup and Image Acquisition

An inverted epi-fluorescence microscope (Ti-S, Nikon, Japan) was used to monitor droplet production and fluorescent microscopy analysis. Fluorescence-activated droplet sorting (FADS) was monitored and controlled using a custom laser-induced fluorescent (LIF) microscope constructed from Thorlab (Newton, NJ) components. A high-speed camera (VEO-410S Phantom, Vision Research) was used to collect video and image data and ImageJ (NIH) was used for analysis. PAGE gels were visualized using a LI-COR Oddyssey CLx Imager (LI-COR Biosciences, Lincoln, NE).

Microfluidic Device Design and Fabrication

All microfluidic devices were fabricated in PDMS (Sylgard 184) using standard soft lithography techniques (Figure S1).[2] Channel geometries were designed in AutoCAD (Autodesk), with colors and layers added after importing into Adobe Illustrator (Adobe). Designs were printed on a transparency mask at 20,000 dpi by CAD/Art Services (Bandon, OR). All master molds were fabricated in a class 1000 cleanroom (Figure S1a). All wafers were pretreated for 1 min in either 2% hydrofluoric acid, or 6:1 Buffered Oxide Etch (BOE) to remove the surface oxide layer, washed in Di H_2O, dried with pressured air, and dehydrated in a 120 °C oven for 15 mins or on a 120 °C hot plate for 5 mins. For the single layer droplet generator, the 3-inch silicon wafer was spin-coated with a 20 μm thick layer of SU-8 2025 photoresist, and patterned with a photomask through exposure to UV light (350nm-450nm). For the 2-layer droplet sorter, SU-8 2010 was used to create both layers (10 μm and 20 μm) forming a maximum channel height of 30 μm. Briefly, after the first layer was patterned and baked, the second layer was spin-coated on top of the first layer and soft baked. The alignment marks on the first layer were revealed using acetone, and the 2nd layer photo mask was aligned to the 1st layer channel features on the wafer using a Karl Suss MA56 Mask Aligner (SUSS MicrotTec, Germany). After the final exposure, wafers were baked at 95 °C, and then submerged in SU-8 developer to remove the unexposed photoresist. The remaining SU-8 formed the positive channel features. Wafers were then hard-baked at 150 °C for 5 min to harden, smooth, and improve SU-8 adhesion.

Silicon wafers were then treated with tridecafluoro-1,1,2,2-tetrahydrooctyl)trichlorosilane in a vacuum chamber for 1 hr to reduce PDMS adhesion. PDMS was mixed at a ratio of 10:1 (w/w) pre-polymer base to curing agent, degassed for at least 30 mins, and then poured onto the patterned silicon wafer. The polymer mixture was cured at 65°C overnight. After curing, devices were cut out and peeled off of the wafer, and inlet/outlet tubing holes were punctured into the device using disposable biopsy punches (1.5 mm, Integra™ Miltex®). PDMS chips were then cleaned with compressed N_2 and tape and concurrently air plasma treated (Harrick Scientific, NY) with pre-cleaned (100% IPA) glass slides (Corning Inc., Corning, NY) at 300 mTorr (30 - 60 sec per device) to allow for irreversible bonding (Figure S1b). Following plasma treatment, devices

were surface treated to create fluorophilic channel walls. Channels were filled with <1 μL Pico-Glide and incubated for 1 hr at room temperature. Channels were then washed with Fluorinert FC-40 (Sigma-Aldrich), followed by a stream of nitrogen gas to dry. Chips were then incubated overnight at 120°C.

Microfluidic Reagent Delivery and Droplet Collection

All aqueous and oil solutions were sealed in 1.5 mL screw-cap plastic micro-centrifuge tubes (Sigma-Aldrich) and delivery was controlled by pressure driven flow with custom LabVIEW software (National Instruments, Austin, TX). Reagent vials consisted of two lengths of Tygon tubing (OD: 1.52 mm, ID: 0.51 mm, EW-06419-01, Cole-Parmer, Vernon Hills, IL) inserted through two holes drilled into the caps of the micro-centrifuge tubes and glued into place to create an airtight seal. One length of tubing remained in the pressure headspace above the reagent and was connected at the other end to a SMC ITV0011-2UMS digital pressure regulator (Automation Distribution, Hatfield, PA). Another length of tubing was submerged in the reagent solution with the other end connected to the appropriate inlet of the microfluidic device.[3] Applying a positive pressure to the reagent vial by the SMC digital regulator caused fluid to be driven into the channels of the microfluidic device. A length of Tygon tubing was also inserted in the outlet and placed in a micro-centrifuge tube for droplet collection. During production, droplets were collected under a layer of light mineral oil (~200 - 300 μL) in 1.5mL plastic micro-centrifuge tubes. The FADS system generated considerable amounts of oil from both the collection and waste outlets, necessitating the use of 5 mL screw cap vials (USA Scientific, Ocala, Florida) for droplet collection. In the FADS system, droplets tended to remain at the highest point of the collection tubing until forced out with air.

Formation of Single Emulsion Droplets

Emulsions were produced using custom PDMS chips (Figure S3a) utilizing a flow focusing geometry.[4] The aqueous phase contained the *E. coli* (expressing enzyme) or commercial enzyme and any associated activity assay reagents. A second design (Figure S3a, right) allowed cells to be co-encapsulated with a lysis agent to promote chemical or enzymatic lysis only following encapsulation. Co-encapsulation was accomplished by co-flowing two solutions, one containing the cells, and the other containing the lysis agent, for a short time before encapsulation. The aqueous phase was sheared by a continuous phase consisting of a low-viscosity fluorinated oil (HFE-7500, 3M Novec) containing 1-2% (w/w) Pico-Surf surfactant (Dolomite Microfluidics, UK). Pressures were maintained to achieve droplet diameters of 20 μm and production rates between 30-35 kHz. Cells undergoing heat lysis were incubated for 5 mins at 95°C, followed by incubation at 55°C for 1 hr – 18hrs.

***E. coli* Preparation for Droplet Encapsulation**

1 mL of *E. coli* from a 50 mL overnight growth or expression was transferred to a 14 mL round-bottom Falcon tube (#352059, Corning), centrifuged for 5 min at 1,811 r.c.f., and the supernatant discarded. The cells were washed three times with 1 mL of 1x ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% TritonX-100, pH 8.8) with the supernatant removed each time after centrifugation. The rinsed bacterial pellet was re-suspended in 2 mL of 1x ThermoPol buffer and the absorbance was measured at 600 nm. Cells were diluted in 1x Thermopol to an OD_{600} of 0.05 (0.1 for co-encapsulation with lysis agents) to enable encapsulation at occupancies of 0.1 cells per droplet according to a Poisson distribution (Figure 2S).

Encapsulation of GFP Expressing *E. coli* in Droplets

Cell populations were grown to express green fluorescent protein (GFP) by first transforming the pTrc99a-GFP plasmid DNA into XL1-blue cells following the manufacturer's recommended instructions. Recovered cells were then used to inoculate 50 mL of LB-ampicillin (100 μg/mL) liquid medium in a 500 mL baffled flask. The culture was grown to confluency overnight at 37 °C with shaking at 225 rpm. Cells were washed as described earlier, and suspended in 1x Thermopol. Either 1x lysozyme or 1x BugBuster in 1x Thermopol was co-encapsulated with *E. coli* expressing GFP. The lysis agents were introduced in a separate stream of the microfluidic chip at 2x final concentration and co-flowed with the *E. coli* containing stream for a short distance before encapsulation (Figure S3a, right). Given that the device operates at low Reynolds number, material transfer between the co-flowing streams is diffusion limited, ensuring the *E. coli* are not lysed until

following encapsulation. Drops were then heat incubated at 37 °C or 55 °C for up to 1 hr to promote lysis (Figure S3c-d).

Construction of Null Mutant KOD-D542G DNA Polymerase

KOD polymerase mutant KOD-D542G was constructed using the Q5 site-directed mutagenesis kit (New England Biolabs). Briefly, the forward primer containing the D542G mutation was designed using the NEBaseChanger software package (New England Biolabs) to ensure effective annealing during PCR amplification while the reverse primer aligned perfectly with the parent template. This forward/reverse primer pair (Table S1) was first used to conduct whole-plasmid amplification of the custom pGDR11 polymerase expression plasmid harboring the KOD exo$^-$ polymerase gene (Initial denaturation: 95°C-2 min followed by 25 cycles: 95 °C - 30 sec, 60 °C - 45 sec, 72 °C at 8 min followed by polishing step of 72 °C for 5 min.) This was followed by a kinase-ligase-DpnI (KLD) treatment (20 min at room temperature) to phosphorylate and ligate the blunt-ended linear PCR product. DpnI was used to digest the parent template background. This mixture (1 µL) was then transformed into DH5-alpha supercompetent cells (New England Biolabs), recovered for 1 hr in 250 uL of SOC media with shaking at 225 RPM, plated onto LB-ampicillin (100 µg/mL) agar plates and grown overnight with shaking at 225 RPM at 37 °C. Single colonies were picked and used to inoculate separate 4 mL aliquots of LB-ampicillin (100 µg/mL) liquid medium in 14 mL round-bottom Falcon-tubes (Corning) with shaking at 225 RPM and 37 °C overnight. Overnight cultures were spun for 10 min at 4,000 RPM and 4°C with the supernatant discarded. Cell pellets were purified using the Express Plasmid Miniprep Kit (Biomiga) following the manufacturer's recommended instructions. DNA constructs were sequence validated (Retrogen, San Diego, CA) and analyzed using the CLC Main Workbench (Qiagen) software package.

Encapsulation of Polymerase Expressing *E. coli* in Droplets

Cell populations were grown to express the KOD-wt, KOD-D542G, TGO-wt, DV-QGLK, or KOD-RS polymerases by first transforming plasmid DNA into XL1-blue cells following the manufacturer's recommended instructions. Recovered cells were then used to inoculate 50 mL of LB-ampicillin (100 µg/mL) liquid medium in a 500 mL baffled flask. The starter culture was grown to confluency overnight at 37°C with shaking at 225 rpm and then used to inoculate (1:100 v/v) 50 mL of LB-ampicillin (100 µg/mL) liquid medium in a separate 500 mL baffled flask and grown at 37 °C with shaking at 225 rpm. At OD_{600} = 0.6, the expression culture was removed and cooled to 25°C. The culture was then induced with IPTG at a final concentration of 1 mM and incubated overnight at 25°C with shaking at 225 rpm. Cells were washed in 1x Thermopol as described earlier, and diluted to OD_{600} = 0.05. The appropriate volume of cells was transferred to a micro-centrifuge tube, centrifuged at 1,811 r.c.f. for 5 mins, the supernatant discarded, and then mixed with the appropriate volume of a premixed fluorescence polymerase activity assay (PAA). The PAA consisted of 1 µM of a self-priming hairpin template labeled with Cy3 at the 5' end (ST.1G.HP.44.Cy3, Table S1), 2 µM of a 3' end labeled Iowa Black FQ quencher sequence (QP08.Iowa, Table S1), and 100 µM mixture of the appropriate nucleoside triphosphates in 1x Thermopol buffer: dNTPs were used for DNA synthesis (KOD-wt, KOD-D542G, and TGO-wt), NTPs for RNA synthesis (DV-QGLK), fNTPS for FANA synthesis (TGO-wt), and tNTPS for TNA synthesis (KOD-RS).

Fluorescent Activated Droplet Sorting

Following incubation, droplets were injected into a second chip for sorting (Figure 2). To prevent droplet shearing and minimize the formation of satellite droplets, the spacing and bias oil were composed of 0.25% (w/w) of Pico-Surf surfactant in HFE 7500. The FADS system consists of a disposable sorting chip, a custom built LIF microscope, and associated electronics. The working principle of the FADS system is illustrated in Fig 1A. Designed for Cy3 excitation/emission, incident light from a Coherent OBIS LS 552 nm laser (Santa Clara, CA) is focused through a 20x plan apochromatic objective (Motic, Hong Kong) where droplets pass single file. Emitted light is led through a 405/488/543/635 nm Quad Band Dichroic into an optical train through a series of long-pass dichroics that can lead to a high-speed camera or 1 of 2 PMTs. Use of the quad-band dichroic allows for the use of a two-channel detection system (Cy3 and AlexaFlour 660), where a secondary dye can be used to decouple droplet size information from activity, if desired. The sample was

illuminated with blue light (450nm, 18nm bandwidth), which does not overlap with the spectral properties of Cy3 (or AlexaFlour 660) and was imaged with a high-speed camera at 35,000 frames per second (fps). The digital signals generated by the PMT were analyzed by a field-gated programmable array (FPGA, USB-7856R, National Instruments) that is controlled with custom LabView software.[5] Droplets falling within a user-defined threshold triggered the FPGA to send a square-wave pulse (50 kHz, 50% duty cycle, 60 μs), amplified to 600V by a high-voltage amplifier (2210, Trek, Lockport New York), to the salt-water electrode (4M NaCl) of the sorting chip.[6] The resulting non-uniform electric field generated a dielectrophoretic (DEP) force that polarized and deflected the droplet into a collection channel (Figure 2c).

DNA Polymerase Activity Assay

Droplet Scale Reactions: Droplet scale reactions consisted of 2 μM of the quencher probe QP08.Iowa, and 1 μM of the self-priming hairpin-template ST.1G.HP.44.Cy3 in 1x Thermopol buffer. The max standard (+) droplet control utilized 2 μM of a non-complimentary quencher probe (QP13.Iowa.BQ, Table S1). After the reaction mixtures were heat annealed (2 min at 95°C followed by 5 min at 4°C), commercial Q5 (1.6U), Taq (4U), Bst (6.4U), and nuclease free water as a (-) polymerase control were then added to their respective reaction mixtures, followed by the addition of 100 μM dNTPs. Following encapsulation, droplets were incubated at 55°C for 1 h (Q5 and Taq) or 5 mins (Bst, and all control reactions). For each population, 10 μL of droplets were diluted in 50 μL of HFE-7500 containing 1-2% (w/w) of Pico-Surf surfactant, and 10 μL of this mixture was pipetted into a Countess™ Cell Counting Chamber Slide (Thermofisher) for analysis by fluorescence microscopy. Each droplet population was then injected into a FADS chip for flow cytometric analysis of ~$1x10^6$ droplets.

Bulk Solution Reactions: 20 μL scale reactions used 1 μM of a primer-template complex (ST.1G.44 and ST.1G.44.Primer.IR800, Table S1) and 2 μM of the quencher probe QP08.Iowa in 1x Thermopol. After heat annealing (2 min at 95°C followed by 5 min at 4°C), commercial Q5 (1.6U), Taq (4U), or Bst (6.4U) were added to their respective reaction mixtures, followed by the addition of 100 μM dNTPs or nuclease free water as a (-) dNTP control. Reactions were incubated at 55°C for 1 h (Q5 and Taq) or 5 mins (Bst, and all control reactions). For analysis with denaturing PAGE, a 2 μL aliquot from each reaction was transferred to 38 μL of Stop Buffer [1x TBE, 25 mM EDTA, 40% formamide, pH 8.0] prior to gel loading. A 10 μL aliquot of each dilution was analyzed by denaturing PAGE.

Strand Displacement Sensor

Droplet Scale Reactions: Droplet scale reactions consisted of 2 μM of a 33mer blocking oligo (Drops.SDP.317.Iowa, Table S1) that also acted as the quencher probe, and 1 μM of the self-priming, 5' Cy3 labeled, hairpin-template 3.17.HP.Cy3 (Table S1) in 1x Thermopol buffer. The max standard (+) droplet control contained 2 μM of a non-complimentary quencher probe (QP13.Iowa.BQ, Table S1). After heat annealing (2 min at 95°C followed by 5 min at 4°C), commercial Q5 (1.6U), Bst (6.4U), and nuclease free water as a (-) polymerase control were then added to their respective reaction mixtures, followed by the addition of 100 μM dNTPs. Following encapsulation, droplets were incubated at 37°C for 2 hrs (Q5) or 5 mins (Bst, and all control reactions). For each population, 10 μL of droplets were diluted in 50 μL of HFE-7500 containing 1-2% (w/w) of Pico-Surf surfactant, and 10 μL of this mixture was pipetted into a Countess™ Cell Counting Chamber Slide (Thermofisher) for analysis by fluorescence microscopy. Each droplet population was then injected into a FADS chip for flow cytometric analysis of ~$1x10^6$ droplets.

Bulk Solution Reactions: 20 μL scale reactions used 1 μM of a primer-template complex (3.17 and IR800.PBS2, Table S1), and 2 μM of a non-Iowa labeled blocking oligo (Drops.SDP.317, Table S1) or an equivalent volume of nuclease free water as a negative control. After heat annealing (2 min at 95°C followed by 5 min at 4°C), commercial Q5 (1.6U) and Bst (6.4U) were then added to their respective reaction mixtures, followed by the addition of 100 μM dNTPs. Reactions were incubated at 37°C for 2 hrs (Q5) or 5 mins (Bst, and all control reactions), and 2 μL aliquots were collected at 30, 60, and 120 min time points for Q5 and 1, 2, and 5 min time points for Bst. For analysis with denaturing PAGE, aliquots were transferred to 38 μL of Stop

Buffer [1x TBE, 25 mM EDTA, 40% formamide, pH 8.0] points prior to gel loading. A 10 μL aliquot of each dilution was analyzed by denaturing PAGE.

Restriction Digestion Activity Assay

Droplet Scale Reactions: Droplet scale reactions consisted of 1 μM of a double stranded DNA fluorophore-quencher complex containing a Pst1 cut-site (PstI-Template, PstI-F-ON.Cy3, and Q-ON.Iowa, Table S1) in 1x Cutsmart buffer. The max standard (+) droplet control consisted solely of a mixture of oligos (1 μM each) that simulated the composition of a reaction after complete digestion of the double stranded probe (PstI-F-ON.Cy3.Cut.11, PstI-F-ON.Cut.15, PstI-Template.Cut.19, PstI-Template.Cut.26, and Q-ON.Iowa, Table S1). After heat annealing (2 min at 95°C followed by 5 min at 4°C), commercial NotI (5U), PstI (5U) and nuclease free water as a (-) enzyme control were added to their respective reaction mixtures. Following encapsulation, droplets were incubated at 37°C for 18 h. Restriction enzymes were then heat-inactivated by incubating the droplets at 80°C for 10 min. For each population, 10 μL of droplets were diluted in 50 μL of HFE-7500 containing 1-2% (w/w) of Pico-Surf surfactant, and 10 μL of this mixture was pipetted into a Countess™ Cell Counting Chamber Slide (Thermofisher) for analysis by fluorescence microscopy. Each droplet population was then injected into a FADS chip for flow cytometric analysis of ~$1x10^6$ droplets.

Bulk Solution Reactions: 25 μL scale reactions were identically composed as the droplet scale reactions, but the PstI-Template was IR labeled (PstI-Template.IR800, Table S1). After heat annealing (2 min at 95°C followed by 5 min at 4°C), commercial NotI (5U), PstI (5U) and nuclease free water as a (-) enzyme control were added to their respective reaction mixtures. Reactions were incubated at 37°C for 18 h. Restriction enzymes were then heat-inactivated by incubating at 80°C for 10 min. For analysis with denaturing PAGE, aliquots were transferred to 38 μL of Stop Buffer [1x TBE, 25 mM EDTA, 40% formamide, pH 8.0] points prior to gel loading. A 10 μL aliquot of each dilution was analyzed by denaturing PAGE.

Ligation Activity Assay

Droplet Scale Reactions: Droplet scale reactions consisted of 1 μM of a 5' Cy3 - 3' Iowa Black FQ labeled molecular beacon and an acceptor-donor sequence complex (MB.Ligation.Drops, Ligation.N1.6, and Ligation.N2.6.Phos, Table S1) in 1x T4 DNA ligase buffer. The acceptor-donor sequence pair consisted of a left hand piece (Ligation.N1.6) and right hand piece (Ligation.N2.6.Phos) that annealed to the inner ring of the molecular beacon, and promoted linearization of the beacon upon ligation of the two pieces, leading to a fluorescent signal. Two negative control groups were assayed alongside the experimental group: one contained the non-phosphorylated version of the donor sequence (Ligation.N2.6, Table S1) in place of Ligation.N2.6.Phos, and the other was prepared using the standard reaction mixture without enzyme. The droplet scale reactions also included a max standard (+) control that used 1 μM of a 12mer oligo (Ligation.N1+N2.12, Table S1) in place of Ligation.N1.6 and Ligation.N2.6.Phos that mimicked the effect of a ligated product. After heat annealing (2 min at 95°C followed by 5 min at 4°C), 400U of T4 DNA ligase, or nuclease free water for the (-) ligase control, was added to the reaction mixture. Following encapsulation, droplets were incubated at 37 °C for 1 hr. Ligase was then heat-inactivated by incubating the droplets at 80°C for 10 min. For each population, 10 μL of droplets were diluted in 50 μL of HFE-7500 containing 1-2% (w/w) of Pico-Surf surfactant, and 10 μL of this mixture was pipetted into a Countess™ Cell Counting Chamber Slide (Thermofisher) for analysis by fluorescence microscopy. Each droplet population was then injected into a FADS chip for flow cytometric analysis of ~$1x10^6$ droplets.

Bulk Solution Reactions: 20 μL scale reactions used an unlabeled molecular beacon (MB.Ligation, Table S1) in place of MB.Ligation.Drops, and an IR800 labeled acceptor sequence (Ligation.N1.6.IR800, Table S1) in place of Ligation.N1.6, but were otherwise identically composed to the droplet scale reactions. After heat annealing (2 min at 95°C followed by 5 min at 4°C), 400U of T4 DNA ligase, was added to the reaction mixture. Reactions were incubated at 16 °C, 25 °C, and 37 °C for 1 hr. Ligase was then heat-inactivated by incubating at 80°C for 10 min. For analysis with denaturing PAGE, a 10 μL aliquot of each reaction was transferred to 10 μL of Stop Buffer [2x TBE, 50 mM EDTA, 80% formamide, pH 8.0] prior to gel loading. A 10 μL aliquot of each dilution was analyzed by denaturing PAGE.

Mock Enrichment

To evaluate the performance of our FADS system, we performed a mock enrichment for DNA synthesis activity by spiking *E. coli* expressing a wild-type polymerase (KOD-wt) into a population expressing an inactive null mutant (KOD-D542G) at cellular ratios of 1:1,000 and 1:10,000 (active to inactive polymerase) and co-encapsulated them in drops with the PAA. *E. coli* were heat lysed at 95 °C for 5 mins, incubated for 1 hour at 55 °C to allow DNA extension on the ST.1G.HP.44.Cy3 template (Table S1), and then sorted. Enrichment was measured by comparing the number of positive droplets detected in the naïve and enriched populations (see Supplemental). To increase the accuracy of the analysis, we measured the hit rates of clonal KOD-wt and clonal KOD-D524G and compared them to the expected values in order to calculate the rate of false positives and false negatives. Using the false positive rate (FPR) and false negative rate (FNR), the true composition of a mixed population could be estimated (see Supplemental, Figure S6). A model was also derived (see Supplemental) to calculate the maximum theoretical value of enrichment for a perfectly performing instrument, as a reference point to assess our results. The maximum theoretical enrichment (η_{max}) was calculated with the following expression:

$$\boldsymbol{\eta_{max} = \frac{1}{1 - e^{-\varepsilon_0 \lambda}(1 - \varepsilon_0)}}$$

where λ is the cell occupancy, and ε_0 is the fraction of KOD-wt cells in the naïve population.

Recovery of Sorted DNA for Mock Enrichment

Plasmid DNA was recovered from sorted droplet emulsions by extraction with Pico-Break (Dolomite Microfluidics, UK) following the vendor protocol. After sorting, a 23-gauge needle was inserted into one end of the collection tubing, with the other end left in the 5 mL collection vial. A 1mL syringe was screwed onto the needle, and used to push air through the tubing and force the droplets into the collection vial. The tubing was then washed a few times by pulling the contents of the vial back into the tubing ~ 3/4 of the total length, followed by flushing with air. To create adequate aqueous volume to facilitate DNA extraction, at least 125 µL of pre-formed, well packed droplets made from 1x Thermopol buffer were added to the 5 mL collection tube. An equivalent volume of nuclease free water could also be added and vortexed to create a layer of droplets. The contents were then mixed to ensure the sorted droplets were randomly distributed within the added droplets. 1 mL of this mixture was added to a 1.5 mL microcentrifuge tube and spun at 1000 r.c.f. for 1 min, the bottom oil layer discarded, followed by addition of another 800 µL – 1 mL of the initial mixture. This process was repeated until all the droplets were transferred into the 1.5 mL microcentrifuge tube. After removing as much of the bottom oil layer as possible to create a compact droplet layer, the droplets were transferred to a 0.5 mL low-adhesion microcentrifuge tube (USA Scientific), followed by the addition of 2 equivalent volumes of Pico-Break 1 (Dolomite). The contents were then vortexed for 15 sec and centrifuged (2 min, 1000 r.c.f.) to attain phase separation. The top, aqueous layer containing the plasmid DNA was recovered and then concentrated using a spin column (DNA Clean & Concentrator-5, Zymo Research) and eluted with molecular biology grade water (10 µL).

Cloning of Enriched DNA Polymerase Domains

Custom DNA primers (Table S1, see Tile 1-Fwd and Tile 9-Rvs) were used to PCR amplify the polymerase domain from the extracted FADS-sorted plasmid DNA (Initial denaturation: 95°C-30 sec followed by 30 cycles: 95°C-30 sec, 56°C-45 sec, 72°C 1 min followed by polishing step of 72°C-2 min). The amplicon was then purified using a spin column (DNA Clean & Concentrator-5, Zymo Research) and size-validated by 1% agarose gel prior to cloning using the Gibson assembly. In a separate reaction, custom DNA primers (Table S1, see KOD-poldom_381-422-Fwd and KOD-poldom_21-65-Rvs) were used to PCR amplify the KOD-wt pGDR11 plasmid region flanking the polymerase domain PCR amplicon and DpnI treated to digest the parent template. Gibson assembly was performed in a 20 µL reaction volume containing a final concentration of 100 ng FADS-sorted polymerase domain amplicon, 100 ng linear vector, and 1x Gibson assembly Mastermix (NEB). The reaction was incubated at 50°C for 60 minutes. This mixture (5 µL) was then transformed into

DH5-alpha supercompetent cells (New England Biolabs), recovered for 1 hr in 250 uL of SOC media with shaking at 225 RPM, plated onto LB-ampicillin (100 μg/mL) agar plates and grown overnight at 37°C. LB-ampicillin (4 mL, 100 μg/mL) was pipetted onto the overnight plate and colonies were scraped in a clockwise motion with a sterile L-shaped cell spreader (Fisher Scientific) until a uniform slurry was created by sloughing the colony forming units (CFUs) from the surface of the plate. The slurry was pipetted into a 14 mL round-bottom Falcon-tube (Thermofisher Scientific) and spun for 10 min at 4,000 RPM and 4 °C with the supernatant discarded. Cells were then minipreped using the Express Plasmid Miniprep Kit (Biomiga) following the manufacturer's recommended instructions. Purified plasmid DNA was quantified using a Nanodrop spectrophotometer. Purified plasmid DNA (300 ng) was transformed into XL1 blue supercompetent cells following the manufacturer's recommended instructions. The recovered transformation (50 μL cells + 200 μL SOC media) was pipetted directly into 50 mL LB-ampicillin (100 μg/mL) media in a 500 mL baffled flask and grown at 37 °C with shaking at 225 rpm overnight. This ensures that the aggregate composition of the enriched plasmid DNA (KOD-wt : KOD-D542G) is evenly represented in the overnight culture used for inoculating a fresh round of cell growth and polymerase expression. An aliquot from the overnight transformation culture (500 μL) was used to inoculate a fresh cell culture (50 mL LB-ampicillin (100 μg/mL) media in a 500 mL baffled flask) for polymerase expression of the enriched mixture of KOD-wt and KOD-D542G enzyme variants. This culture was grown and expressed as previously described. Cells were harvested and encapsulated with the PAA as previously described.

Table S1. DNA Primer, Template, and Sensor Sequences. Oligos are grouped according to assay/workflow. Modifications are written using IDT nomenclature.

Polyerase Extension	Sequence (5' -> 3')	Description	Format
ST.1G.HP.44.Cy3	/5Cy3/ACA ACC ATA CTC TTC CAC ATC TGC ATA CGC GAA AGC GTA TGC AG	Cy3 Labeled Self-Priming Hairpin Template	Drops/Mock Enrichment
ST.1G.44	ACA ACC ATA CTC TTC CAC ATC TGC ATA CGC	Template	Gel
QP08.Iowa	ATG GTT GT/3IABkFQ/	Quencher Probe	Drops/Mock Enrichment
ST.1G.44.Primer.IR800	/5IRD800/ GCG TAT GCA G	IR800 Labeled Primer	Gel
QP13.Iowa.BQ	ACA ACC ATA CTC T/3IABkFQ/	Incorrect Quencher Sequence for Max Standard	Droops

Strand Displacement	Sequence (5' -> 3')	Description	Format
3.17.HP.Cy3	/5Cy3/ ACA ACC ATT TAT GTA GCA TTT ATG AAA TTT TTA AAT CAA TTT ACT ATT GGC TAC TGC ATA CGC GAA AGC GTA TGC AGT AGC C	Cy3 Labeled Self-Priming Hairpin Template	Drops
3.17	ACA ACC ATT TAT GTA GCA TTT ATG AAA TTT TTA AAT CAA TTT ACT ATT GGC TAC TGC ATA CGA GTG TC	Template	Gel
Drops.SDP.317.Iowa	TAA AAA TTT CAT AAA TGC TAC ATA AAT GGT TGT/3IABkFQ/	Blocking Oligo/Quencher	Drops
Drops.SDP.317	TAA AAA TTT CAT AAA TGC TAC ATA AAT GGT TGT	Blocking Oligo	Gel
IR800.PBS2	/5IRD800/GAC ACT CGT ATG CAG TAG CC	IR800 Labeled Primer	Gel
QP13.Iowa.BQ	ACA ACC ATA CTC T/3IABkFQ/	Incorrect Quencher Sequence for Max Standard	Drops

Restriction Digestion	Sequence (5' -> 3')	Description	Format
PstI-Template	ATA CGC ATA CCT GTC TGC AGT GGC TAA AAG CAC ACG CAC GGA GAC	Template w/ Pst1 Cut Site	Drops
PstI-Template.IR800	/5IRD800/ATA CGC ATA CCT GTC TGC AGT GGC TAA AAG CAC ACG CAC GGA GAC	IR800 Labeled Template w/ Pst1 Cut Site	Gel
PstI-F-ON.Cy3	/5Cy3/TAG CCA CTG CAG ACA GGT ATG CGT AT	Cy3 Labeled Probe Compliment to Pst1 Cut Site	Drops/Gel
PstI-F-ON.Cy3.Cut.11	/5Cy3/TAG CCA CTG CA	Cy3 Labeled Probe/Digestion Product Piece/Max Standard	Drops
PstI-Template.Cut.19	ATA CGC ATA CCT GTC TGC A	Digestion Product Piece/Max Standard	Drops
PstI-Template.Cut.26	GTG GCT AAA AGC ACA CGC ACG GAG AC	Digestion Product Piece/Max Standard	Drops
PstI-F-ON.Cut.15	GAC AGG TAT GCG TAT	Digestion Product Piece/Max Standard	Drops
Q-ON.Iowa	GTC TCC GTG CGT GTG CT/3IABkFQ/	Quencher Probe	Drops/Gel

Ligation	Sequence (5' -> 3')	Description	Format
MB.Ligation.Drops	/5Cy3/CG TTG ATG GTT CCA CTT CTC GTG CGT TCA ACG/3IABkFQ/	Molecular Beacon (5' Cy3, 3' Iowa Black FQ)	Drops
MB.Ligation	CGT TGA TGG TTC CAC TTC TCG TGC GTT CAA CG	Non-labeled Molecular Beacon for Gel Analysis	Gel
Ligation.N1.6.IR800	/5IRD800/ACG AGA	IR800 Labeled Acceptor Sequence	Gel
Ligation.N1.6	ACG AGA	Acceptor Sequence	Drops
Ligation.N2.6	AGT GGA	Donor Sequence Fragment/Negative Control	Drops/Gel
Ligation.N2.6.Phos	/5Phos/AGT GGA	5' Phosphorylated Donor Sequence	Drops/Gel
Ligation.N1+N2.12	ACG AGA AGT GGA	Full Length Product/Max Standard	Drops

KOD-D542G Mutation	Sequence (5' -> 3')	Description
Tile 1 Fwd	GAA TGA GCT GGC CCC GAA CAA G	Polymerase Domain Forward PCR Primer
Tile 9 Rvs	CTA AT TAA GCT TGG CTG CAG GTC GAC TC	Polymerase Domain Reverse PCR Primer
KOD-poldom_381-422 Fwd	GAG TCG ACC TGCA GCC AAG CTT AAT TAG	pGDR-11 Vector PCR Forward Primer
KOD-poldom_21-65 Rvs	CTT GTT CGG GGC CAG CTC ATT C	pGDR-11 Vector PCR Reverse Primer
AN-SK-01-Fwd	CTA CAG CGA CAC CGG AGG ATT TTT TGC	Q5 Site-Directed Mutagenesis Forward PCR Primer
AN-SK-02-Rvs	ATT ACC TTA AAG CCG TAC TTT TCC TCT ATC TCC TT	Q5 Site-Directed Mutagenesis Reverse PCR Primer

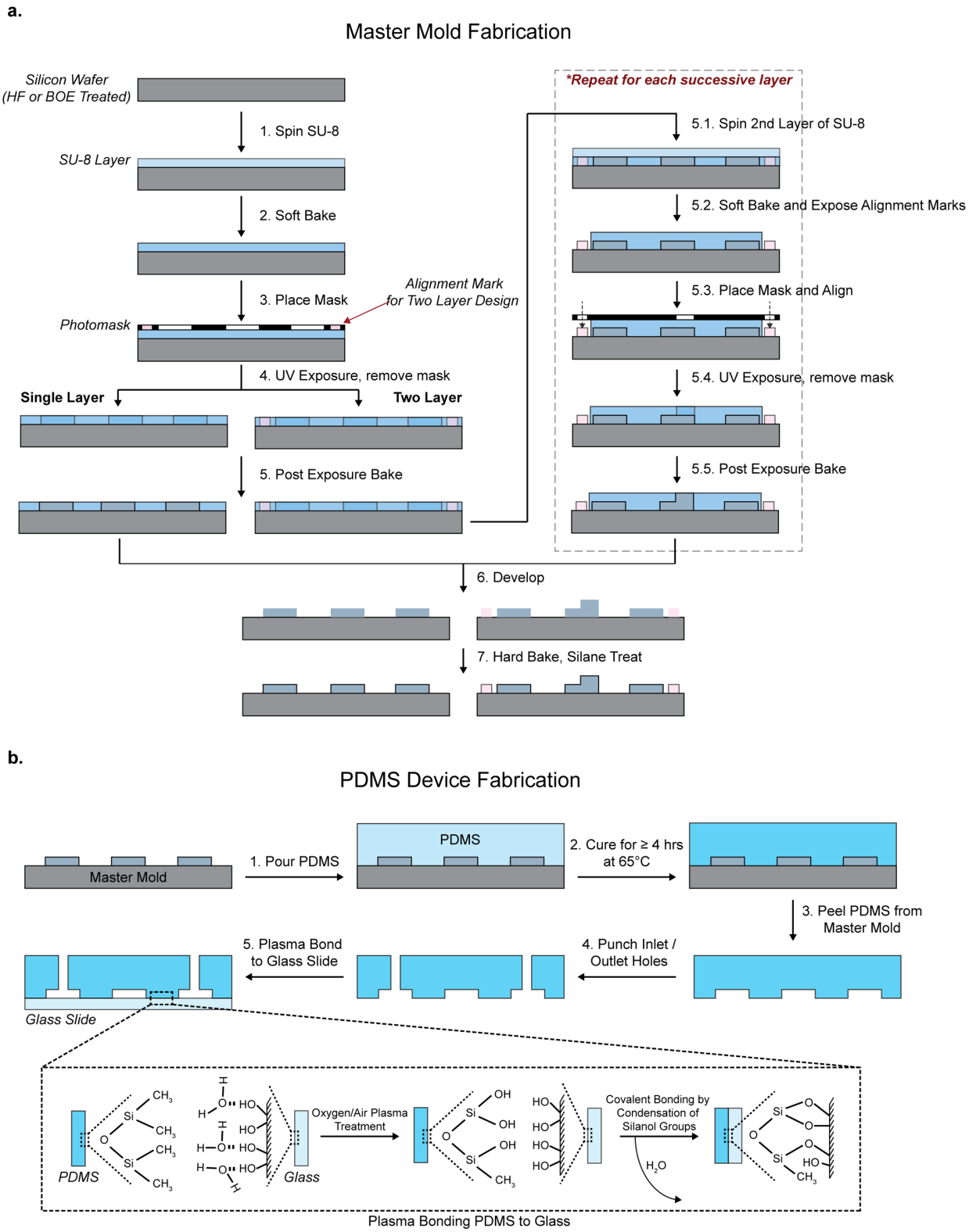


Figure S1. Fabrication Process for a PDMS Microfluidic Device. (a) Production of the SU-8 master mold that will form the channel reliefs in PDMS. Multiple layers can be added sequentially to create channel geometries with different heights. (b). Production of a PDMS chip from an SU-8 master mold. Channels are enclosed by plasma bonding to a glass slide (bottom).

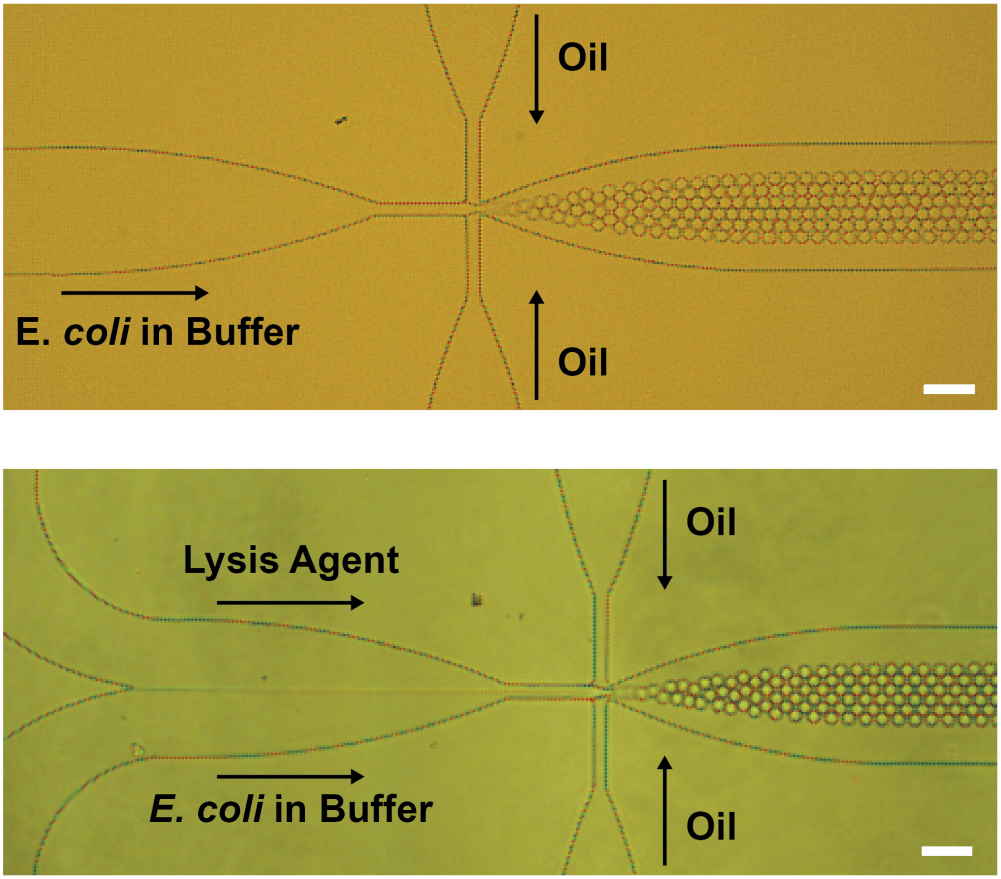


Figure S2. Droplet generation devices. Microscopic images of w/o droplet generation. Droplets can be generated from a single aqueous stream (top), which is suitable for heat lysis, or from co-flowing two different aqueous streams (bottom) prior to droplet production to co-encapsulate cells and lysis agents for enzymatic or chemical lysis following droplet production. Scale bars are 75 μm.

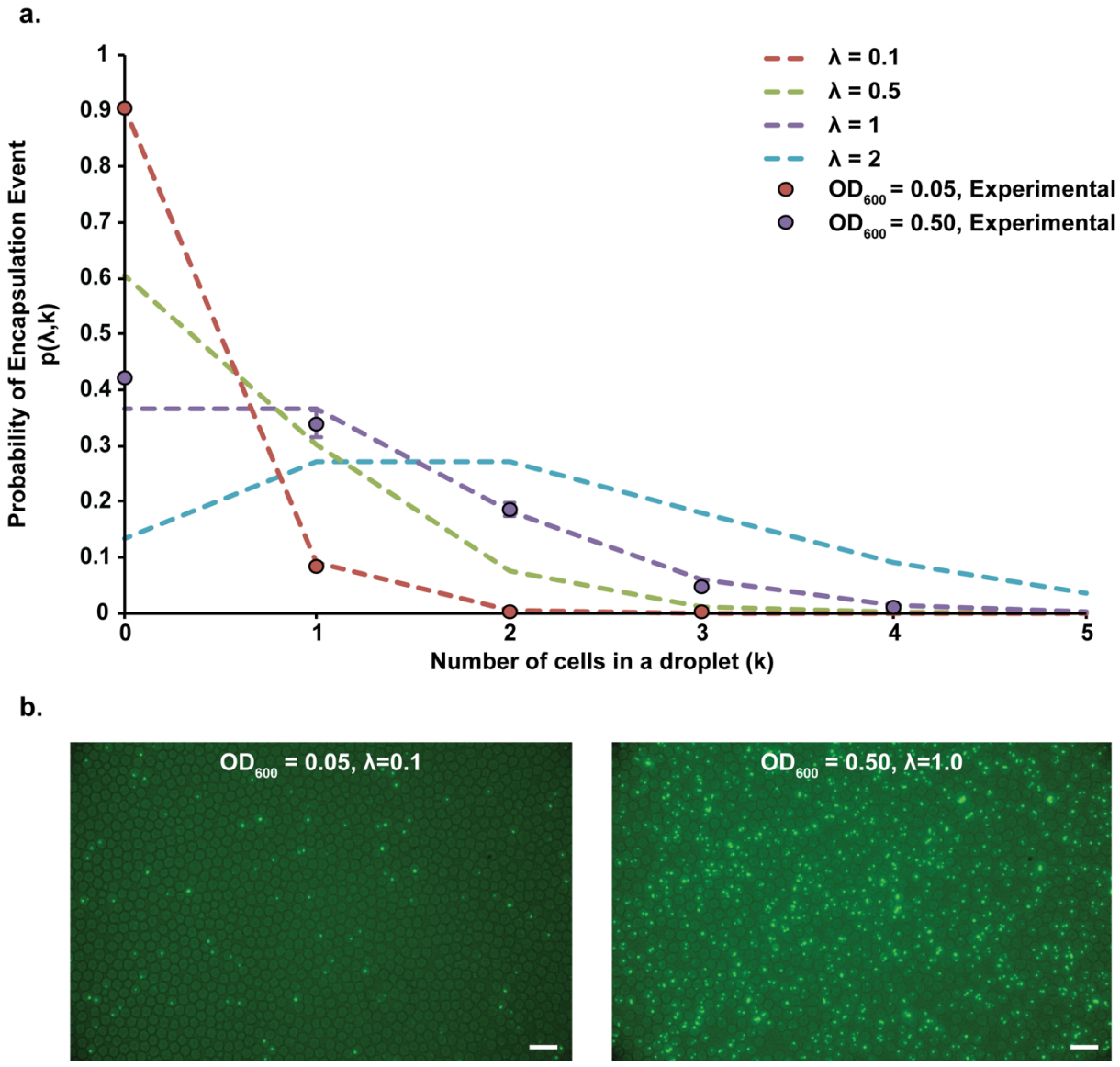


Figure S3. Theoretical Model and Empirical Validation of *E. coli* Occupancy. (a) Poisson probability distributions calculated for different *E. coli* occupancy levels in uniform droplets. A Poisson distribution was used to model the proportion of droplets *p(λ,k)* containing a given number of cells (*k*) at four different occupancy levels (λ), a parameter describing the average number of cells per droplet. The OD_{600} of the pre-encapsulated *E. coli* solution can be used to predict the value of λ (and vice-versa), which is critical for experiments that target a specific occupancy level. An OD_{600} of 0.05 and 0.5 were calculated to result in values of λ of 0.1, and 1.0, respectively, and were validated experimentally. (b) Predicting occupancy level from OD_{600}. Fluorescent images showing the difference in cellular droplet occupancy of GFP expressing *E. coli* at OD_{600} = 0.05 and OD_{600} = 0.5. Scale bars are 50 μm.

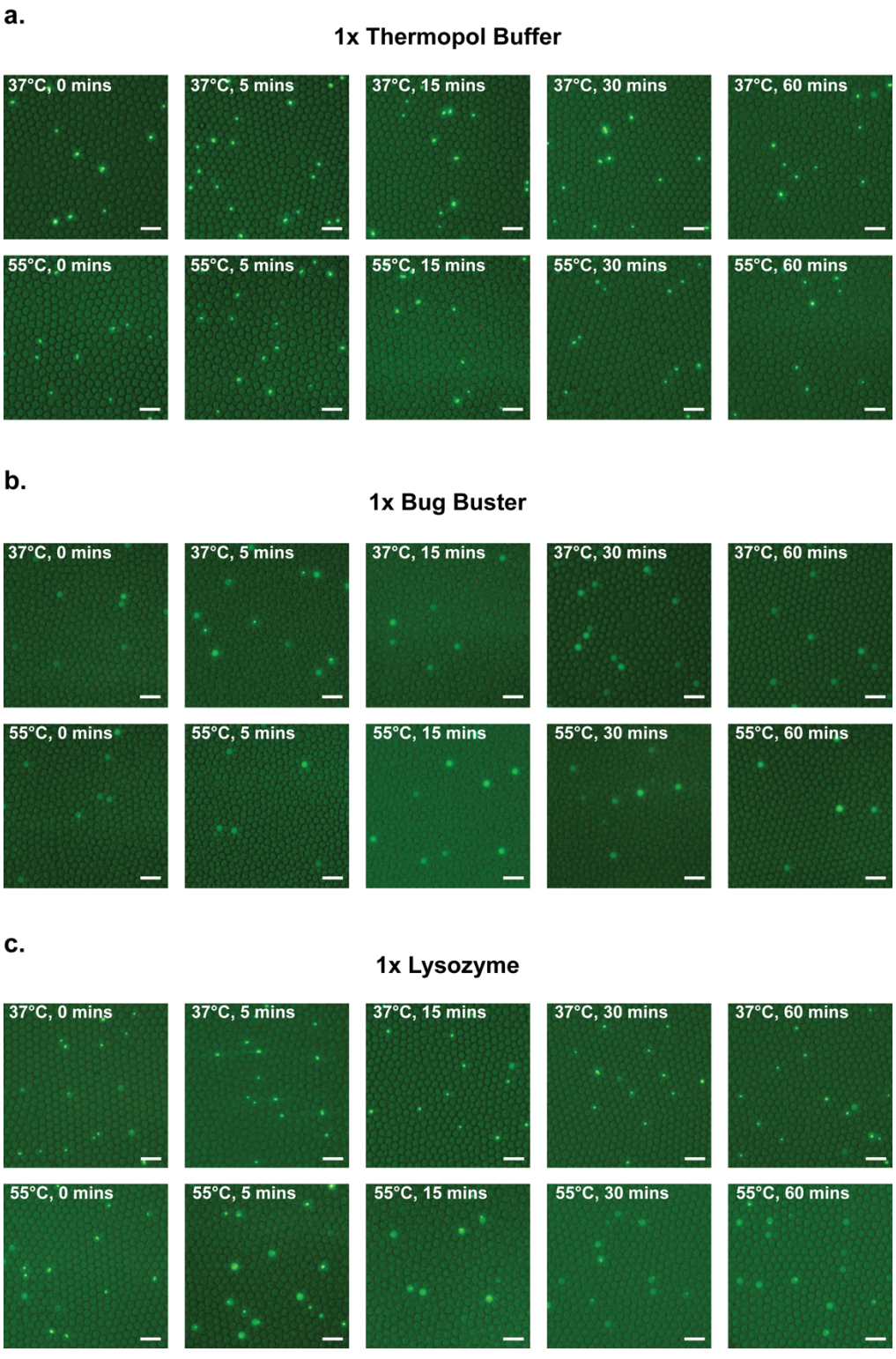


Figure S4. Time and Temperature Dependence of Chemical and Enzymatic Lysis. Fluorescent images of GFP expressing *E. coli* encapsulated in droplet microcompartments at different time points at 37 °C (top rows) and 55 °C (bottom rows) for (a) 1x Thermopol, (b) 1x Bug Buster, and (c) 1x Lysozyme. Scale bars are 50 μm.

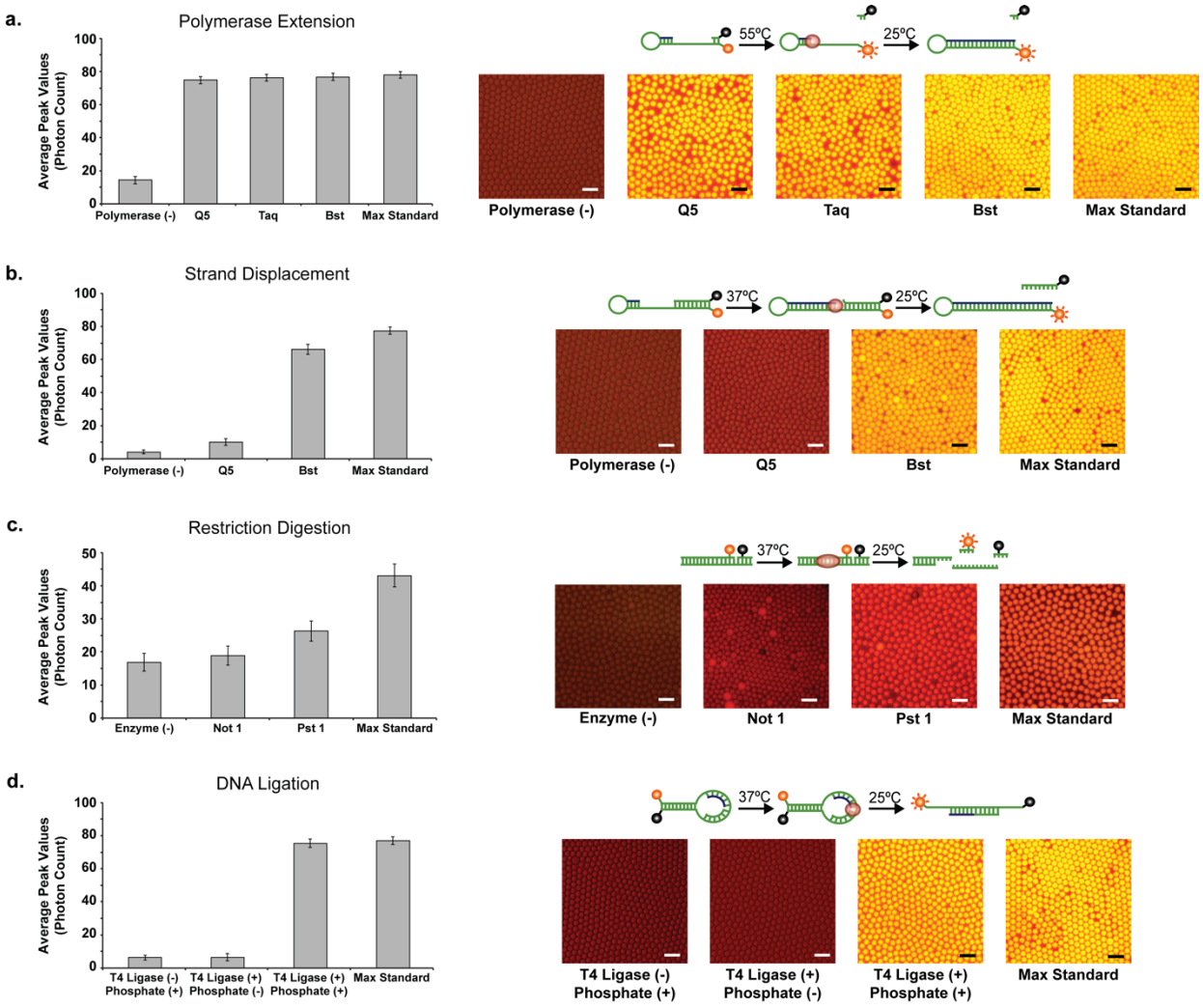


Figure S5. Optical Sensors for Detected Enzymatic Activity in Droplet Microcompartments. Recombinant enzymes were co-encapsulated in droplets with fluorescent sensors for detecting (a) polymerase extension, (b) strand displacement, (c) restriction digestion, and (d) DNA ligation along with the appropriate substrates. Bar charts summarize the average peak droplet fluorescence for each population (right). Error bars represent the distribution of peak fluorescence values within the population. Fluorescent microscopic images of each droplet population are shown to the left. Scale bars are 50 μm.

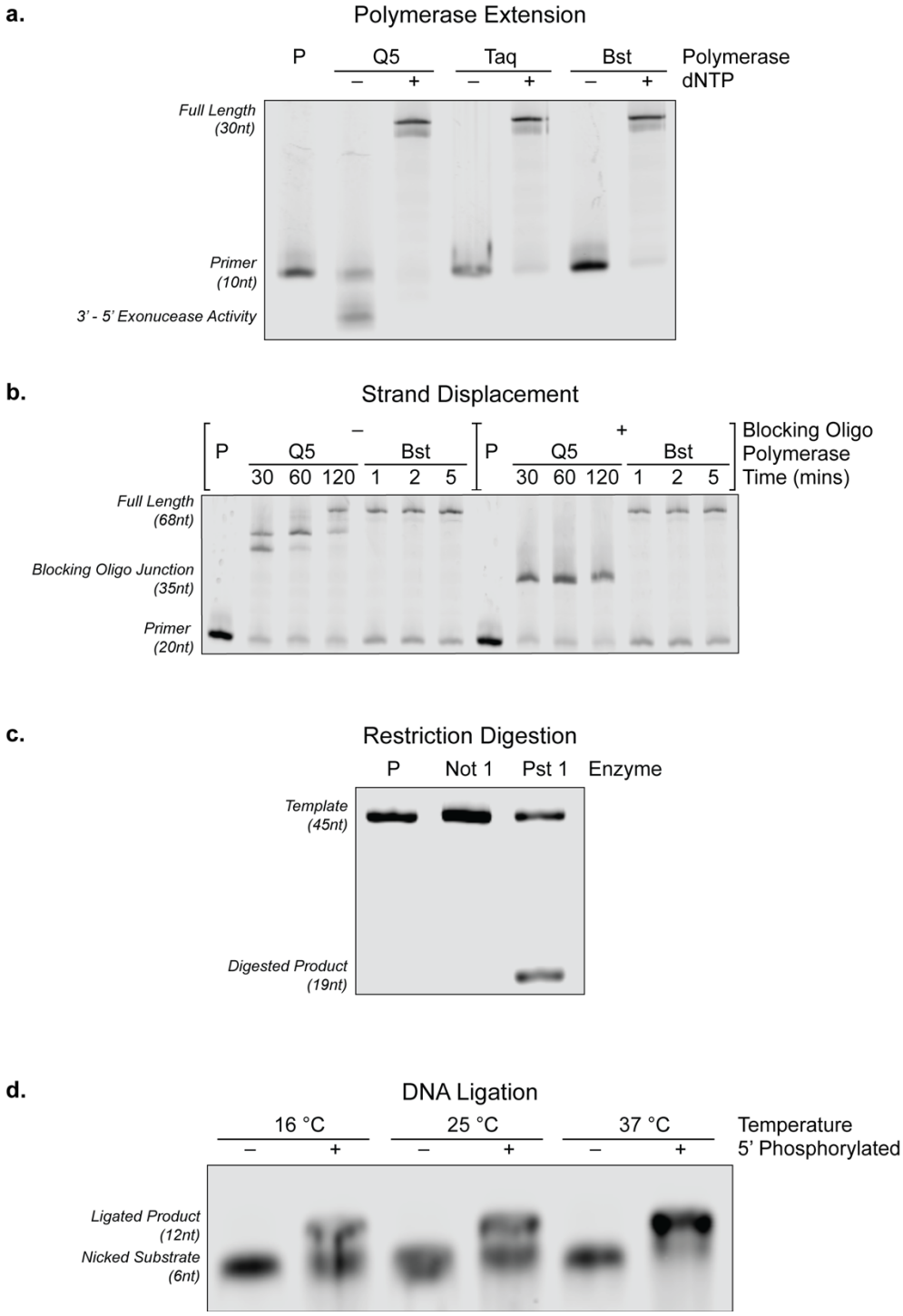


Figure S6. Validation of Optical Sensors in Bulk Solution. (a) Polymerase extension assay. Primer extension analysis on the polymerase extension sensor using Q5 (1 hour), Taq (1 hour), and Bst (5 mins) polymerases in the presence and absence of dNTP substrates at 55 °C. (b) Strand displacement assay. Time course at 37 °C on the strand displacement sensor with Q5 and Bst polymerases in the presence and absence of a 33nt blocking oligo. (c) Restriction digestion assay. Pst 1 restriction site sensor incubated at 37 °C for 18 hours with Pst 1, and a non-specific restriction enzyme, Not 1. (d) DNA ligation assay. Ligation of two 6nt oligos in the presence of the complementary molecular beacon sensor at three temperatures for 1 hour.

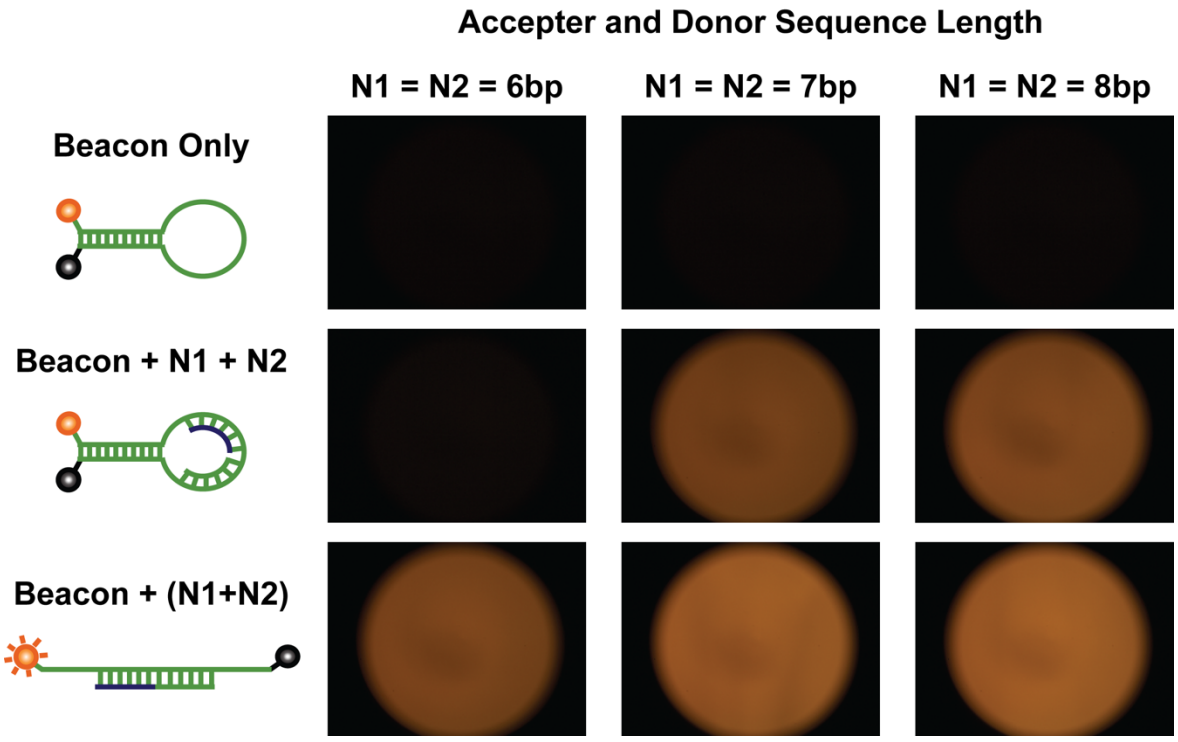


Figure S7. Identifying Optimal Substrate Lengths for the DNA Ligase Sensor. The SNR of the ligase sensor was evaluated for different donor and acceptor oligonucleotide lengths by fluorescence microscopy. Fluorescent images of the beacon only (top row), the beacon with different length donor/acceptor strands (middle row), and the beacon with chemically synthesized full-length product (bottom row). Donor and acceptor lengths of 6nts allow the molecular beacon to function with optimal fluorescence activity. Longer substrate lengths lead to unwanted fluorescence of the unligated complex.

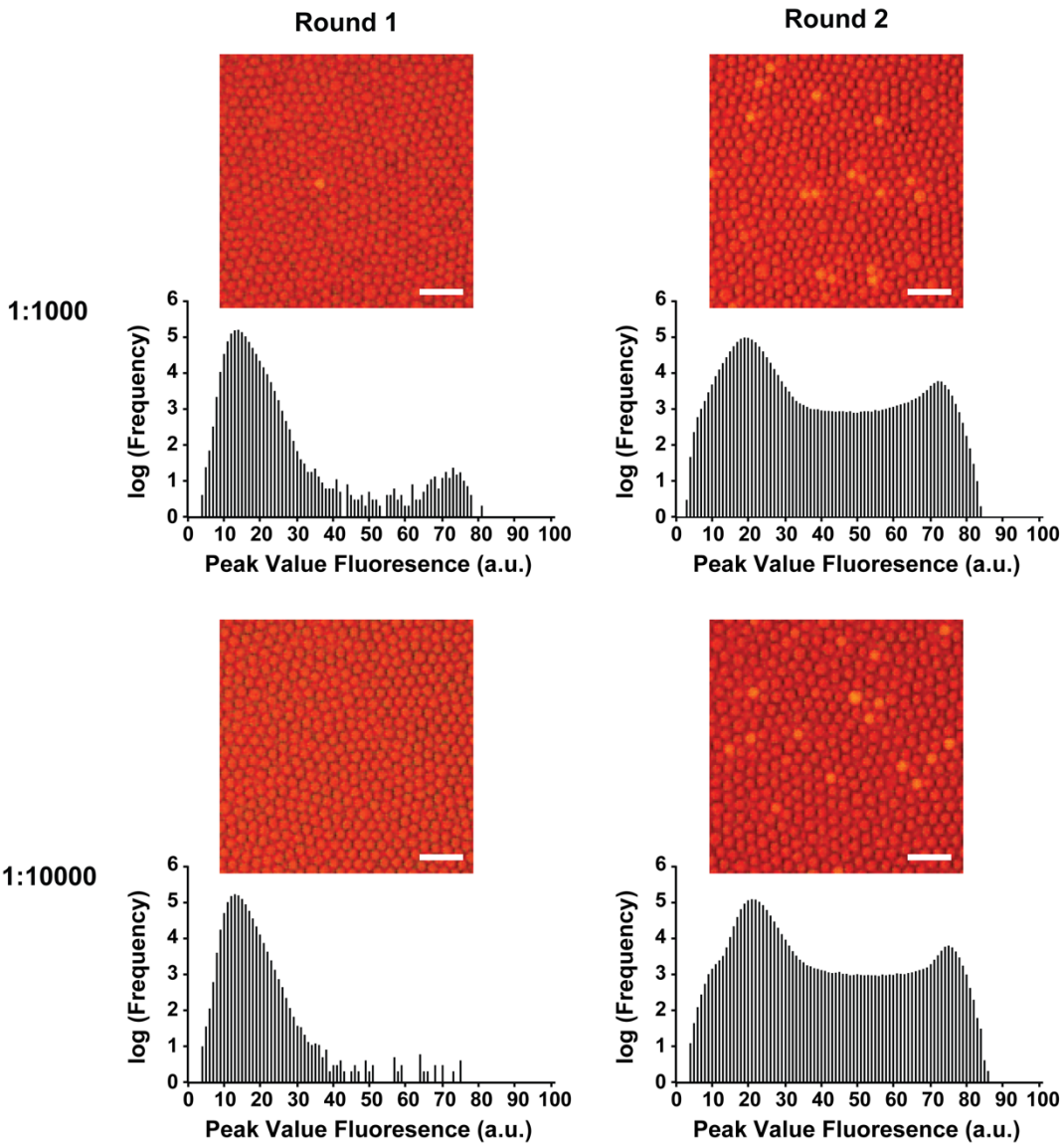
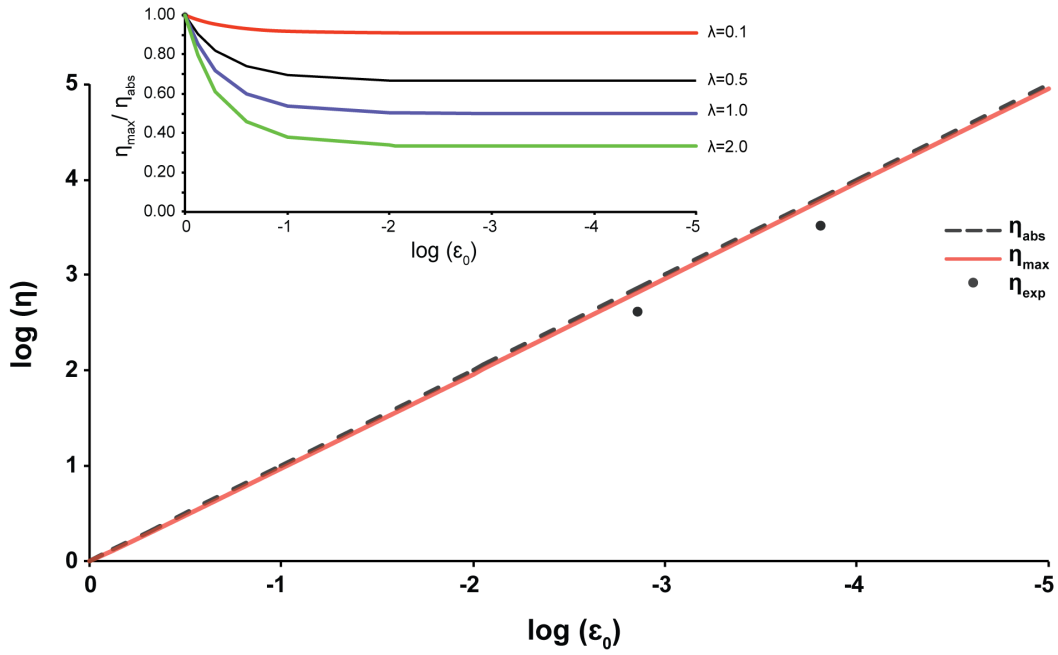


Figure S8. Raw Data from Polymerase Sensor Mock Enrichment. Images and peak fluorescence distributions for naïve (round 1, left) and enriched (round 2, right) droplet encapsulated populations of KOD-wt expressing *E. coli* spiked into populations of *E. coli* expressing KOD-D542G null mutant at ratios of 1:1000 (top) and 1:10000 (bottom). Scale bars are 75 μm.

a.



b.

		True Counts	
		N_+	N_-
		=	=
Experimental Counts	$N_{Hit,+}$ =	$N_+ \bullet (TPR)$ True Positives	+ $N_- \bullet (FPR)$ False Positives
		+	+
	$N_{Hit,-}$ =	$N_+ \bullet (FNR)$ False Negatives	+ $N_- \bullet (TNR)$ True Negatives

Figure S9. Theoretical Model for Enrichment. a) Plot of absolute maximum enrichment (η_{abs}) and theoretical maximum enrichment (η_{max}, λ=0.1) over a range of ε_0. Enrichment values measured with the FADs system are plotted for ε_0 = 1/1000, and 1/10000. Inset plots the ratio of the η_{max} to η_{abs} vs. ε_0 for λ=0.1, 0.5, 1, and 2, which can also be interpreted as the maximum ε_1 that can be achieved with a given ε_0. As ε_0 approaches 1, η_{max} approaches η_{abs}, and as ε_0 becomes increasingly small, η_{max} approaches a finite value that decreases with increasing λ, meaning the maximum degree to which a population can be enriched decreases for a given ε_0 as λ increases. b) Contingency table used to derive equations to calculate true hit and true non-hit values from measured hit and measured non-hit values given known TPR, FPR, FNR, and TNR.

Mathematical Model Relating OD_{600} to Droplet Cell Occupancy

When encapsulating cells within microfluidically generated droplets, the probability, p(k,λ), of a droplet containing 0, 1, or more cells within a population of droplets can be estimated using a Poisson distribution

$$\boldsymbol{P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}}$$

where k is the number of cells in a given droplet, and λ is the average number of cells per droplet volume. This model agrees well with experimental data if the cells are significantly smaller than the droplets, and they are sparsely distributed so as to not influence the position of neighboring cells [7]. If the cells are significantly smaller than the droplets that contain them, $\boldsymbol{\phi_d}$ <<< 1, where $\boldsymbol{\phi_d}$ is defined as the volume fraction of single cell in one drop, and $\boldsymbol{\emptyset_d = \bar{V}_c/\bar{V}_d}$, where $\boldsymbol{\bar{V}_c}$ is the average volume of a single *E. coli* cell and $\boldsymbol{\bar{V}_d}$ is the average droplet volume. Similarly, if the cells are sparsely distributed, the volume fraction of cells in the pre-encapsulated solution $\boldsymbol{\phi_S}$ <<< 1 where $\boldsymbol{\emptyset_s = V_C/V_{sol}}$. $\boldsymbol{V_C}$ is the total volume of cells in the pre-encapsulated solution, and $\boldsymbol{V_{sol}}$ is the total solution volume. Realizing that λ can be expressed as a ratio of $\boldsymbol{\emptyset_s}$ and $\boldsymbol{\emptyset_d}$[7]:

$$\boldsymbol{\lambda = \frac{\emptyset_s}{\emptyset_d}}$$

it is possible to derive an expression to relate the OD_{600} of the pre-encapsulated cell solution to the parameter λ. If we rearrange the equation and expand it:

$$\boldsymbol{\emptyset_d \lambda = \emptyset_s}$$

$$\boldsymbol{\frac{\bar{V}_c}{\bar{V}_d}\lambda = \frac{V_C}{V_{sol}} = \frac{n_{cells}\bar{V}_c}{V_{sol}} = \frac{(OD_{600} \cdot \sigma_{1.0} \cdot V_{sol}) \cdot \bar{V}_c}{V_{sol}} = (OD_{600} \cdot \sigma_{1.0}) \cdot \bar{V}_c}$$

we obtain an expression relating λ to the number of cells present in the pre-encapsulated solution (n_{cells}), the OD_{600}, and the *E. coli* concentration at OD_{600} = 1.0 ($\sigma_{1.0}$). Further simplifying and re-arranging, we derive the following expression showing that the target OD_{600} for a given λ can be calculated from the desired droplet volume, $\bar{V}_d$, and $\sigma_{1.0}$.

$$\boldsymbol{OD_{600} = \frac{\lambda}{\bar{V}_d \cdot \sigma_{1.0}}}$$

If we assume that $\sigma_{1.0}$ = 5 x 10^8 cells/mL [8] and desire a droplet diameter of 20 μm, we need an OD_{600} of ~0.05 to get λ = 0.1, and an OD_{600} of ~0.5 to get λ = 1, which was validated experimentally (Figure S2). At λ = 0.1, termed "single occupancy", 90.5% of the droplets will be empty, ~9% will contain a single cell, and ~0.5% will contain more than one cell. Of the cell containing droplets, approximately 95% will contain only a single cell. This is desired when a phenotype-genotype linkage is critical, such as in directed evolution and selection experiments. At λ = 1, we have what is considered "full occupancy", whereby we have the maximum number of single cell containing drops (36.8%) across all values of λ, but a larger proportion contain multiple cells (26.4%). This occupancy level can be beneficial to enrich a very dilute population by sacrificing the degree of the phenotype-genotype linkage for higher throughput.[9] Once enriched, further rounds of selection can be accomplished using single occupancy.

Theoretical Model for Cellular Enrichment

In sorting experiments, the efficiency of a sort is typically defined in terms of the degree of enrichment of a target cell from a population of cells that contains a mixture of both target and non-target cells. For the purposes of this exercise, target cells will be referred to as "positive" cells, while non-target cells will be referred to as "negative" cells. The term "enrichment" describes how the frequency of positive cells in the sorted pool of cells has increased in comparison to the naïve, non-sorted pool. This can be expressed mathematically as:

$$\eta = \frac{N_{+,1}}{N_{+,1} + N_{-,1}} \Big/ \frac{N_{+,0}}{N_{+,0} + N_{-,0}} = \frac{\varepsilon_1}{\varepsilon_0}$$

where η is enrichment, $N_{+,0}$ and $N_{+,1}$ are the number of positive cells before and after sorting (or in an enriched pool) respectively, $N_{-,0}$ and $N_{-,1}$ are the corresponding values for negative cells, and ε_0 and ε_1 represent the fraction of positive cells in the naïve and sorted populations, respectively. This definition of enrichment describes by what factor the fraction of positive cells in the population has increased after a round of sorting. For example, if the fraction of positive cells in a population increased by a factor of 90 after sorting, from 0.01 (1%) to 0.9 (90%), or from 0.001 (0.1%) to 0.09 (9%), then η = 90. Since the former example only contained a 0.01 fraction of positive cells, if it were enriched perfectly such that no negative cells we present after sorting, this would give η = 100 = $(0.01)^{-1}$. This puts an upper limit on the enrichment value as it would not make sense to enrich more than what it would take to make the sorted population 100% composed of positive cells. Consequently, the absolute maximum value and the dynamic range of enrichment increase with decreasing fractions of positive cells in the naïve population. Furthermore, given that a maximum value is present, the degree of success or efficiency of a particular enrichment exercise can be assessed based on the composition of the initial population.

An alternative approach is to measure enrichment based the ratio of positive to negative cells in the naïve and sorted populations[10]:

$$m = \frac{N_{+,1}}{N_{-,1}} \Big/ \frac{N_{+,0}}{N_{-,0}}$$

In this case, there is no theoretical upper limit on the enrichment (i.e., no real dynamic range). The more the positive cells out number the negative cells after sorting, the larger this value can climb regardless of the composition of the initial population, which can inflate enrichment values. Furthermore, when only positive cells are collected, this expression of enrichment is undefined, and has to be reported as "$m > h$", where h is a finite lower limit. This approach to calculating enrichment has merit in that a "x" fold increase in the ratio of positive to negative cells after sorting will give the same "x" fold increase in the enrichment value regardless of the initial ratio of positive to negative cells. The increase in the value of η (for a given population) will not reflect this trend though, as it will tend towards the absolute maximum value as positive cells take up a larger portion of the sorted population. However, η is a good representation of the sorting efficiency in regards to assessing how well an instrument performs in comparison to a perfect sorting system and it is a superior indicator of the composition of the sorted population. It also better reflects the increase in difficulty of enriching a target population when it makes up a lower proportion of the naïve population by not allowing enrichment values to become artificially large. Lastly, sorting efficiency can also be also quantified using η by comparing it to the absolute maximum value, which can allow one to assess how sorting efficiency trends with changes to the composition of the naïve population (i.e., sorting efficiency decrease with increasing proportion of negative cells).

Given that we wish to assess how well our instrument performs in comparison to a perfect system across a range of compositions for the naïve population, we will use η to characterize the

sorting efficiency of our FADS system. If we assume a perfect sorting system, we can see that the best scenario would be to collect only positive cells, and zero negative cells. In this situation, our absolute maximum enrichment would be defined as:

$$\boldsymbol{\eta_{abs}} = \frac{\boldsymbol{N_{+,1}}}{\boldsymbol{N_{+,1}}} \Big/ \frac{\boldsymbol{N_{+,0}}}{\boldsymbol{N_{+,0} + N_{-,0}}} = \frac{\boldsymbol{1}}{\boldsymbol{\varepsilon_0}} = \boldsymbol{\varepsilon_0^{-1}}$$

Thus, the absolute maximum obtainable enrichment can be easily calculated by taking the inverse of ε_0. This model would be sufficient if each cell-containing droplet only contained a single cell, but single cell encapsulation in microfluidic droplets follows a Poisson distribution, which means that inevitably, some negative cells will be sorted if they are co-encapsulated with positive cells. Consequently, even in a perfect system with no false positives or negatives, and 100% accurate sorting, it will never be possible to achieve $\boldsymbol{\eta_{abs}}$. To derive an expression for the theoretical maximum enrichment, $\boldsymbol{\eta_{max}}$, we must develop a model that includes the number of negative cells that are sorted with the positive cells, taking into consideration all possible combinations of co-encapsulation that can occur where at least one positive cell is present and n=0 to ∞ negative cells are also potentially present in a given sorted droplet.

To begin, only a single cell type will be considered for simplicity. Before encapsulation, a number of cells, n_{cells}, are randomly distributed in a volume of solution V_{sol}. After encapsulation, n_{cells} are distributed among N_0 droplet compartments, where the average number of cells per droplet can be defined as $\boldsymbol{\lambda} = n_{cells} / N_0$. Thus, each cell has a probability of $p = 1/N_0 = \lambda/n_{cells}$ of being encapsulated. To determine the probability of encapsulating k cells within a given compartment we can use the binomial distribution:

$$\boldsymbol{B(n_{cells}, k, p)} = \frac{\boldsymbol{n_{cells}!}}{\boldsymbol{k!\,(n_{cells} - k)!}} \cdot [\boldsymbol{p^k \cdot (1-p)^{n_{cells}-k}}]$$

Since we know the number of drops N_0 >>>*1* and cells n_{cells} >>>*1* and that the probability of encapsulation for each cell remains constant during the course of the experiment (as they are sparsely distributed and are being encapsulated at a constant rate), we can rewrite the binomial distribution in terms of the average rate of encapsulation $\boldsymbol{\lambda}$, and if we take the limit as $n_{cells} \to \infty$, then we can define a probability P(X=k) independent of the number of events:

$$\boldsymbol{p} = \frac{\boldsymbol{\lambda}}{\boldsymbol{n_{cells}}}$$

$$\boldsymbol{P(X=k)} = \lim_{\boldsymbol{n_{cells}\to\infty}} \left(\frac{\boldsymbol{n_{cells}!}}{\boldsymbol{k!\,(n_{cells}-k)!}} \cdot \left(\frac{\boldsymbol{\lambda}}{\boldsymbol{n_{cells}}}\right)^{\boldsymbol{k}} \cdot \left(\boldsymbol{1} - \frac{\boldsymbol{\lambda}}{\boldsymbol{n_{cells}}}\right)^{\boldsymbol{n_{cells}-k}} \right)$$

$$\boldsymbol{P(X=k)} = \frac{\boldsymbol{\lambda^k}}{\boldsymbol{k!}} \cdot \lim_{\boldsymbol{n_{cells}\to\infty}} \left(\frac{\boldsymbol{n_{cells}!}}{\boldsymbol{(n_{cells}-k)!}} \cdot \left(\frac{\boldsymbol{1}}{\boldsymbol{n_{cells}}}\right)^{\boldsymbol{k}} \cdot \frac{\left(\boldsymbol{1} - \frac{\boldsymbol{\lambda}}{\boldsymbol{n_{cells}}}\right)^{\boldsymbol{n_{cells}}}}{\left(\boldsymbol{1} - \frac{\boldsymbol{\lambda}}{\boldsymbol{n_{cells}}}\right)^{\boldsymbol{k}}} \right)$$

The terms inside the limit can be calculated individually. Expanding the product of the first two terms, we see that it reduces to 1 ($n=n_{cells}$):

$$\lim_{n\to\infty} \frac{n!}{(n-k)!} \cdot \left(\frac{1}{n}\right)^k = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k) \cdot (n-k-1) \cdots 1}{(n-k) \cdot (n-k-1) \cdots 1} \cdot \left(\frac{1}{n^k}\right)$$

$$\lim_{n\to\infty} \frac{n!}{(n-k)!} \cdot \left(\frac{1}{n}\right)^k = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{n^k} = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-k+1}{n} = 1$$

The term in the denominator also reduces to 1:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^k = 1^k = 1$$

We can evaluate the last term in the numerator on the right if we rewrite it into the form of a commonly known limit:

$$\lim_{n\to\infty} \left(1 + \frac{1}{x}\right)^{x \cdot a} = e^a \quad \rightarrow \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \lim_{n\to\infty} \left(1 + \frac{1}{\frac{-n}{\lambda}}\right)^{\frac{-n}{\lambda} - \lambda} = e^{-\lambda}$$

which leaves us with:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

As the number of events approaches ∞, p becomes increasingly small, the binomial distribution can be represented by the Poisson distribution, P(X=k), which we previously established is a good approximation for modeling the distribution of droplet encapsulation events for microfluidically generated droplets. To reiterate, P(X=k) gives the probability of finding k cells encapsulated in a droplet, independent of n_{cells} as long as the number of cells and drops is sufficiently large and p is sufficiently small.

In a typical enrichment, a library of cells expressing many different mutant variants may be present in a given population. To simplify the analysis, we will only consider cells which express a positive/active protein and those that express a negative/non-active protein, identified by either the "+" or "-" subscript respectively in the following derivation. Since each encapsulation event is said to be random and independent, and the positive and negative cells will be present in different amounts in the pre-encapsulation solution, they can each be said to have their own associated probability of encapsulation, defined as $P_+(\lambda_+, k_+)$ for positive cells, and $P_-(\lambda_-, k_-)$ for negative cells [10]:

$$P_+(\lambda_+, k_+) = \frac{\lambda_+^{k_+}}{k_+!} e^{-\lambda_+}$$

$$P_-(\lambda_-, k_-) = \frac{\lambda_-^{k_-}}{k_-!} e^{-\lambda_-}$$

Where λ_+ and λ_- are the average number of encapsulated positive and negative cells, respectively. Assuming perfect sorting, whereby all sorted drops contain at least one or more positive cell co-encapsulated with 0 or more negative cells, we can express the theoretical number of collected

positive and negative cells in terms of the number of drops screened, N_0, and their associated probabilities. The number of positive cells collected during sorting can be expressed as:

$$\boldsymbol{N_{+,1} = N_0 \cdot \left(\sum_{k_+=1}^{\infty} k_+ \cdot P_+(\lambda_+, k_+) \right)}$$

which simplifies to

$$\boldsymbol{N_{+,1} = N_0 \cdot \left(\sum_{k_+=1}^{\infty} k_+ \cdot P_+(\lambda_+, k_+) \right) = N_0 \cdot \sum_{k_+=1}^{\infty} k_+ \cdot \frac{\lambda_+^{k_+}}{k_+!} e^{-\lambda_+}}$$

$$\boldsymbol{N_{+,1} = N_0 \cdot e^{-\lambda_+} \cdot \sum_{k_+=1}^{\infty} k_+ \cdot \frac{\lambda_+^{k_+}}{k_+!} = N_0 \cdot \lambda_+ \cdot e^{-\lambda_+} \cdot e^{\lambda_+} = N_0 \cdot \lambda_+}$$

$$\boldsymbol{N_{+,1} = N_0 \cdot \lambda_+}$$

Similarly, the number of negative cells sorted simplifies to:

$$\boldsymbol{N_{-,1} = N_0 \cdot \sum_{k_+=1}^{\infty} P_+(\lambda_+, k_+) \cdot \left(\sum_{k_-=1}^{\infty} k_- \cdot P_-(\lambda_-, k_-) \right)}$$

$$\boldsymbol{N_{-,1} = N_0 \cdot \lambda_- \cdot \sum_{k_+=1}^{\infty} P_+(\lambda_+, k_+)}$$

The fraction of positive cells in the pre-sorted (ε_0) and post-sorted (ε_1) populations can written as:

$$\boldsymbol{\varepsilon_0 = \frac{N_{+,0}}{N_{+,0} + N_{-,0}} = \frac{\lambda_+}{\lambda_+ + \lambda_-} = \frac{\lambda_+}{\lambda}}$$

$$\boldsymbol{\varepsilon_1 = \frac{N_{+,1}}{N_{+,1} + N_{-,1}} = \frac{N_{+,1}}{N_{+,1} + N_{-,1}} \cdot \frac{1/N_0}{1/N_0} = \frac{\lambda_+}{\lambda_+ + \lambda_- [\sum_{k_+=1}^{\infty} P_+(\lambda_+, k_+)]}}$$

Expanding and simplifying the series in the denominator of ε_1 reduces it to:

$$\boldsymbol{\sum_{k_+=1}^{\infty} P_+(\lambda_+, k_+) = \sum_{k_+=1}^{\infty} \frac{\lambda_+^{k_+}}{k_+!} e^{-\lambda_+} = e^{-\lambda_+} \cdot \sum_{k_+=1}^{\infty} \frac{\lambda_+^{k_+}}{k_+!} = e^{-\lambda_+} \cdot \left(\sum_{k_+=0}^{\infty} \frac{\lambda_+^{k_+}}{k_+!} - 1 \right)}$$
$$\boldsymbol{= e^{-\lambda_+}\left(e^{\lambda_+} - 1\right) = 1 - e^{-\lambda_+}}$$

$$\boldsymbol{\varepsilon_1 = \frac{N_{+,1}}{N_{+,1} + N_{-,1}} = \frac{N_{+,1}}{N_{+,1} + N_{-,1}} \cdot \frac{1/N_0}{1/N_0} = \frac{\lambda_+}{\lambda_+ + \lambda_-(1 - e^{-\lambda_+})}}$$

Using the expression for ε_0, λ_+ and λ_- can be written as:

$$\boldsymbol{\lambda_+ = \lambda \cdot \varepsilon_0 \qquad \lambda_- = \lambda(1 - \varepsilon_0)}$$

Allowing ε_1 to be written as:

$$\boldsymbol{\varepsilon_1 = \frac{\lambda_+}{\lambda_+ + \lambda_-(1 - e^{-\lambda_+})} = \frac{\lambda \cdot \varepsilon_0}{\lambda \cdot \varepsilon_0 + \lambda(1 - \varepsilon_0)(1 - e^{-\lambda \cdot \varepsilon_0})} = \frac{\varepsilon_0}{\varepsilon_0 + (1 - \varepsilon_0)(1 - e^{-\lambda \cdot \varepsilon_0})}}$$

$$\boldsymbol{= \frac{\varepsilon_0}{1 - e^{-\lambda \cdot \varepsilon_0} + \varepsilon_0 \cdot e^{-\lambda \cdot \varepsilon_0}} = \frac{\varepsilon_0}{1 - e^{-\varepsilon_0 \lambda}(1 - \varepsilon_0)}}$$

Dividing through by ε_0 to both gives the theoretical maximum enrichment, η_{max}:

$$\boldsymbol{\eta_{max} = \frac{\varepsilon_1}{\varepsilon_0} = \frac{1}{1 - e^{-\varepsilon_0 \lambda}(1 - \varepsilon_0)}}$$

Experimental Enrichment

To calculate ε_0 for an experimental enrichment, one can do so easily based on the number of positive drops detected, N_{0+}, which can also be expressed mathematically in relation to λ_+ and the number of droplets screened, N_0, as:

$$\boldsymbol{N_{0+} = N_0 \cdot \left(\sum_{k_+=1}^{\infty} P_+(\lambda_+, k_+) \right) = N_0 \cdot \left(1 - e^{-\lambda_+}\right)}$$

$$\boldsymbol{e^{-\lambda_+} = \left(1 - \frac{N_{0+}}{N_0}\right)}$$

Solving for λ_+ we get the following expression:

$$\boldsymbol{\lambda_+ = ln\left(\frac{N_0}{N_0 - N_{0+}}\right)}$$

Knowing the values of λ and λ_+, ε_0 can be calculated for any population, which allows for the calculation of the observed/experimental enrichment, η_{exp}:

$$\boldsymbol{\eta_{exp} = \frac{\varepsilon_{1,exp}}{\varepsilon_{0,exp}}}$$

If the value of λ remains constant, η_{exp} can simply be expressed as:

$$\boldsymbol{\eta_{exp} = \frac{\lambda_{1,+}}{\lambda_{0,+}}}$$

where $\lambda_{1,+}$ and $\lambda_{0,+}$ are the experimental values for the average number of positive cells in the enriched and naïve populations, respectively.

To evaluate the performance of our FADS system, we spiked *E. coli* expressing a wild-type polymerase (KOD-wt) into a population of *E. coli* expressing a null mutant polymerase (KOD-D542G) at spiking ratios of 1:1000 and 1:10000 (KOD-wt:(KOD-D542G+KOD-wt)) and performed a single round of enrichment. To measure enrichment, we compared the number of positive droplets detected in the naïve and enriched populations, and used those values to back calculate the number of positive and negative cells actually present in the pool. Since we would inherently have false positives and false negatives, we to took this into account to predict what our number of hits would be if the assay were perfect in order to get a more accurate estimation of positive and negative cell count in a real sample.

To calculate the true value of KOD-WT cells in a mixed sample, we took clonal populations of KOD-WT and KOD-D542G and compared the number of positive droplets to the theoretical number given λ = 0.1, in order to calculate the true positive rate (TPR) and false positive rate (FPR). For KOD-WT, we would expect 100% of the cell containing droplets to be counted as a hit (0 false positives from sensor), and conversely, for KOD-D542G, we would expect 0% of cell containing droplets to be counted as a hit (0 false negatives from sensor) in a perfect test, which is a valid assumption if the error rate of the sensor is almost 0. Thus the expected values of real KOD-WT containing droplets (N_+) in the clonal KOD-WT population and real KOD-D542G containing drops (N_-) in the clonal KOD-D542G population can be estimated as:

$$\begin{aligned} \boldsymbol{N_+ = N_0 \cdot \left(1 - e^{-\lambda}\right)} \qquad & \boldsymbol{\lambda_+ = \lambda = 0.1\,, \quad \lambda_- = 0} \\ \boldsymbol{N_- = N_0 \cdot \left(1 - e^{-\lambda}\right)} \qquad & \boldsymbol{\lambda_- = \lambda = 0.1\,, \quad \lambda_+ = 0} \end{aligned}$$

which was also derived earlier. Using this data, the TPR and FPR can be calculated using the following equations:

$$\boldsymbol{TPR = \frac{N_{Hit,+}}{N_+}, \;\; FPR = \frac{N_{Hit,-}}{N_-}}$$

where $N_{Hit,+}$ and $N_{Hit,-}$ are the number of positive droplets counted in the clonal KOD-WT and KOD-D542G populations, respectively. With the FPR, we can calculate the true negative rate (TNR) as TNR = 1 – FPR, and with the TPR, we can calculate the false negative rate (FNR) as FNR = 1 – TPR. Using these values, we can estimate the total number of actual KOD-WT (N_+) containing droplets in any mixed population, and thus calculate the true values for $\lambda_{0,+}$ and $\lambda_{1,+}$ (see below) to determine η_{exp}.

$$\boldsymbol{\lambda_{0or1,+} = ln\left(\frac{N_0}{N_0 - N_+}\right)}$$

where N_0 is the number of drops screened and N_+ is the predicted number of actual KOD-WT containing drops in a given population. In a given test of a mixed population, the values for TPR,

FPR, FNR, and TNR can be used to back calculate the expected values of N_+ and N_-, which can also be visualized by a contingency table (**Fig. S6b**).

$$\boldsymbol{N_+ \cdot (TPR) + N_- \cdot (FPR) = N_{Hit,+}}$$

$$\boldsymbol{N_+ \cdot (FNR) + N_- \cdot (TNR) = N_{Hit,-}}$$

We can rewrite this as matrix equation:

$$\begin{bmatrix} \boldsymbol{TPR} & \boldsymbol{FPR} \\ \boldsymbol{FNR} & \boldsymbol{TNR} \end{bmatrix} \begin{bmatrix} \boldsymbol{N_+} \\ \boldsymbol{N_-} \end{bmatrix} = \begin{bmatrix} \boldsymbol{N_{Hit,+}} \\ \boldsymbol{N_{Hit,-}} \end{bmatrix}$$

which gives the solution:

$$\begin{bmatrix} \boldsymbol{N_+} \\ \boldsymbol{N_-} \end{bmatrix} = \begin{bmatrix} \boldsymbol{TPR} & \boldsymbol{FPR} \\ \boldsymbol{FNR} & \boldsymbol{TNR} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{N_{Hit,+}} \\ \boldsymbol{N_{Hit,-}} \end{bmatrix}$$

References

1. Sau, S. P., and Chaput, J. C. (2017) A Gram-Scale HPLC-Free Synthesis of TNA Triphosphates Using an Iterative Phosphorylation Strategy, *Org. Lett.19*, 4379-4382.

2. Xia, Y. N., and Whitesides, G. M. (1998) Soft lithography, *Angew. Chem. Int. Edit 37*, 551-575.

3. Vallejo, D., Lee, S. H., and Lee, A. (2017) Functionalized Vesicles by Microfluidic Device, *Methods Mol. Biol. 1572*, 489-510.

4. Tan, Y. C., Cristini, V., and Lee, A. P. (2006) Monodispersed microfluidic droplet generation by shear focusing microfluidic device, *Sens. Actuators B Chem. 114*, 350-356.

5. Sciambi, A., and Abate, A. R. (2015) Accurate microfluidic sorting of droplets at 30 kHz, *Lab on a chip 15*, 47-51.

6. Sciambi, A., and Abate, A. R. (2014) Generating electric fields in PDMS microfluidic devices with salt water electrodes, *Lab on a chip 14*, 2605-2609.

7. Collins, D. J., Neild, A., deMello, A., Liu, A. Q., and Ai, Y. (2015) The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation, *Lab on a chip 15*, 3439-3459.

8. Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C., and Chaput, J. C. (2016) A general strategy for expanding polymerase function by droplet microfluidics, *Nat. Commun. 7*.

9. Zinchenko, A., Devenish, S. R. A., Kintses, B., Colin, P. Y., Fischlechner, M., and Hollfelder, F. (2014) One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution, *Anal. Chem. 86*, 2526-2533.

10. Baret, J. C., Miller, O. J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M. L., Hutchison, J. B., Agresti, J. J., Link, D. R., Weitz, D. A., and Griffiths, A. D. (2009) Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity, *Lab on a chip 9*, 1850-1858.