Electronic Supplementary Material (ESI) for Lab on a Chip.

Supplementary Information

Enrichment of rare events using a multi-parameter high throughput microfluidic droplet sorter

Sheng-Ting Hung[a], Srijit Mukherjee[a,b], Ralph Jimenez[a,b]

[a]JILA, NIST and University of Colorado, Boulder, Colorado 80309, United States
[b]Department of Chemistry, University of Colorado, Boulder, Colorado 80309, United States

Table of contents:

Section 1: Optical layout and microfluidic device

Optical Layout

The optical layout of the instrument is depicted in Figure S1.

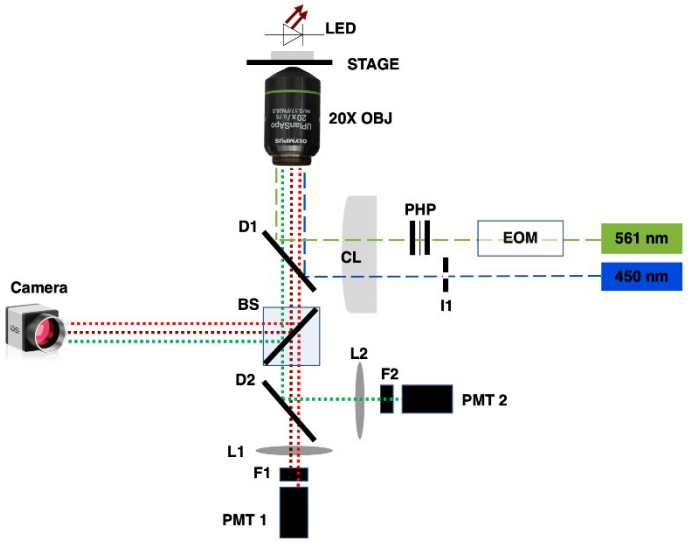


Figure S1: Optical diagram of the microfluidic sorter. The excitation light sources are two CW lasers with wavelengths 561 nm (green dashed line, Coherent Genesis MX) and 450 nm (Blue dashed line, ThorLabs L450P1600MM laser diode on a TCLDM9 temperature control mount shaped by a C230TMD-A f=4.51mm NA=0.55 aspheric lens). EOM is an electro-optic modulator (ThorLabs EO-AM-NR-C4) with a modulation frequency set to 29.5 MHz. PHP are a polarizer, a half-wave plate and another polarizer. I1 is an iris with an adjustable pinhole size. CL is a cylindrical lens with a focal length 150 mm. D1 (Semrock FF493/574-Di01-25x36) and D2 (Semrock FF573-Di01- 25x36) are dichroic mirrors. The laser beams are directed into an inverted objective lens (Olympus UPLSAPO 20X NA=0.75) under a 3-dimensional translation stage where the microfluidic device is located. L1 and L2 are convex lenses, and F1 (Semrock FF01- 629/56) and F2 (Semrock FF01-520/35) are band pass filters. PMT1 and PMT2 are photomultiplier tubes (Hamamatsu R9880-U-20). The far-red LED (dark red dotted line, Thorlabs M730D2) with a long pass filter (Thorlabs FGL715) allows one to visualize the flow on the Camera (Edmund EO-23122M). BS is a 10:90 beam splitter (ThorLabs BSN10R). The green and red fluorescence signals are shown with dotted green and red lines, respectively.

Both 561 nm and 450 nm continuous wave (CW) laser beams are the primary excitation beams to induce fluorescence in the cells encapsulated in droplets. The 561 nm beam is focused into an electro-optic modulator that can amplitude modulate the CW beam to a

sinusoidal profile. The half-wave plate and polarizers can be used to control the laser power and the polarization of the sinusoidally modulated beam. The 450 nm beam is spatially filtered using an adjustable iris (I1). Both beams are directed onto a cylindrical lens CL that reshapes the circular beams to elliptical ones. The beams are directed into a 20X inverted objective lens using a dichroic mirror D1. The epifluorescence from the sample staged on the objective passes through D1 into dichroic mirror D2 that reflects the green fluorescence towards a band pass filter F2 and focused onto a photomultiplier tube (PMT) PMT2, which is referred as the green channel. The red fluorescence that passes through D2 and band pass filter F1 is focused onto PMT1 referred to as the red channel. A far-red LED and a long pass filter are mounted above the objective lens, and the LED light is partially reflected off a beam splitter and collected onto a monochrome USB camera to provide a real-time display of the system in flow.

Microfluidic device

The microfluidic devices were fabricated in a clean room environment using standard soft lithography techniques. The photoresist (MicroChem SU-8 3050) was spin-coated evenly on a clean silicon wafer according to the protocol provided with the photoresist SU-8 3050 aiming for depth 50 μm. The coated wafer was illuminated with a uniform UV light source under an acetate mask printed with an inverted image of the device design that was created using CAD software as shown in Figure S2. The exposed wafers were baked shortly, rinsed with the developer (MicroChem SU-8 developer), and subsequently hard baked overnight according to the protocol provided with the photoresist SU-8 3050. These wafers were exposed to (1H,1H,2H,2H-perfluorooctyl) trichlorosilane (TCS) for more than 3 hours under vacuum for surface passivation. Polydimethyl-siloxane (PDMS) pre-polymer and bonding agent (Dow Corning Sylgard 184) were mixed in a 10:1 ratio and around 20-25 g of the mix was poured to the master wafer that was laid flat on a petri dish to achieve devices that were ~4 mm thick. These were baked in an oven at 70 ºC overnight to harden the PDMS. The PDMS slabs with the device design were cut out from the master and holes for the tubings to be connected to the devices were punched under a clean room hood. No. 1.5 glass coverslips and these slabs were washed with

isopropanol and dried on a hot plate at 75 ºC, and subsequently plasma cleaned their surfaces in a Reactive Ion Etcher (AXIC PlasmaSTAR) with O_2 plasma (50 W power, 50 SCCM and 5-10 sec exposure). The PDMS slabs were bonded to the coverslips immediately after plasma cleaning and heated on a hot plate at 75 ºC for a couple of hours, then were stored at room temperature for ~24 hours before filling the electrode channels with In-Sn solder (In 52% and Sn 48%, melting point 118 ºC). Terminal pins (Male, GT-150) were used to connect the solder filled channels with wire strips that can be clipped to high voltage power sources that are necessary for DEP sorting. The pins are fixed to the devices using epoxy glue to ensure a secure connection. Pneumatic pressure controllers (Parker Electronics) were used to control pressure from cell and oil reservoirs that are connected to the input channel of the devices.

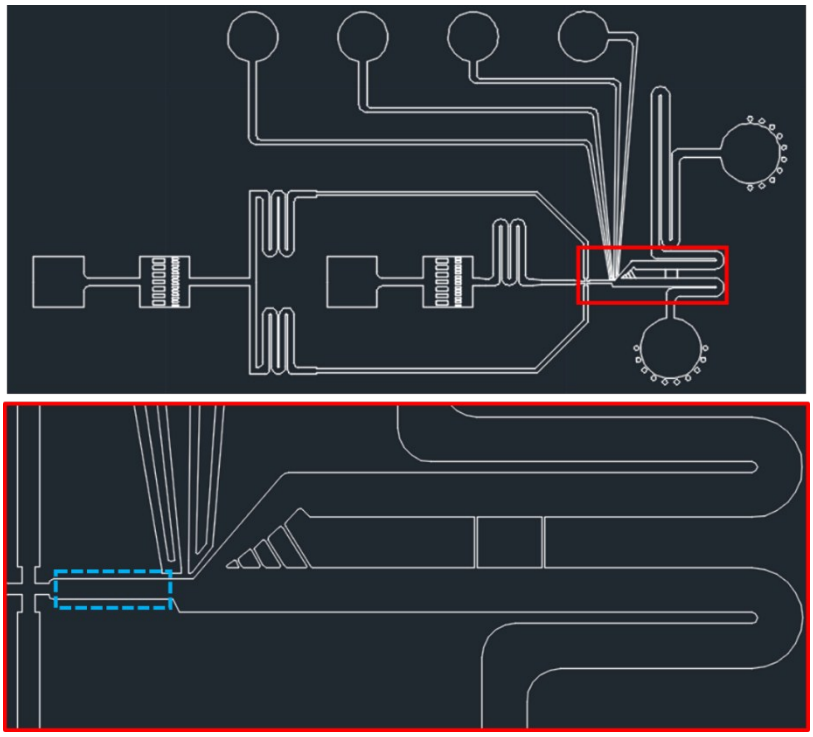


Figure S2: Microfluidic chip design. The droplet formation and selection junctions are indicated within the solid red box and zoomed in. The laser beams are placed in the dashed blue box to excite the fluorophores in droplets, where the dimension of the channel is $530 \times 90 \times 50$ μm^3 (length × width × depth).

Section 2: Instrument Operation Parameters

Excitation intensities for optical transitions are usually about 1-10 kW/cm^2 for both beams. The illumination is a Gaussian beam in the z- profile (out of imaging plane) and is elliptical in the x-y plane. The full width at half maximum (FWHM) of the elliptical beams have the dimension of ~2.5 μm and ~4.5 μm in the in-flow direction, and ~45 μm and ~35 μm in the perpendicular direction for 561 nm and 450 nm beams, respectively. The PMT voltages are usually set to be -1025 V. Electronically controlled pneumatic pressure controllers (Parker Inc.) are operated in the range of 0.5-5 psi for cell channel and 5-10 psi for oil channel in current chip configuration, which allows uniform droplet generation at the frequency tested up to 4 kHz. All sorting modes are tested for droplet frequencies up to 2.5 kHz in the current microfluidic chip configuration, and the droplet volume for such operation frequencies is measured to be around 50 pL. To generate effective force to sort droplets flowing at 1-2.5 kHz, 1.5 kV peak to peak pulses at 5 kHz frequency are applied on the electrodes of the chip. The high voltage pulses applied on the electrodes are set to 2-5 μs with variable delay time (>1 μs) allowing the droplet to travel from the last beam to the collection-waste junction depending on the position of the last beam in the dashed blue box in Figure S2.

Section 3: Cell culture and sample preparation

For this study, FPs were expressed in the unicellular bacterial system, Escherichia coli (*E. coli*). Bacterial systems offer high expression efficiencies and fast doubling times (~20 minutes) and are robust systems for screening and selecting fluorescent proteins with improved photo-physical properties. Bacterial cells (Top10) are transformed with the plasmid DNA of a fluorescent protein of interest and grown on agar plates with ampicillin resistance and arabinose for expression in an incubator at 37 ºC overnight, then a single bacterial colony is selected and passed into a 5 mL LB media with ampicillin resistance (referred as LB-amp) culture tube and grown overnight in a shaker at 37 ºC, 230 rpm. 100 ul of this culture is then passed into a 10 mL of LB-amp flask and grown in the shaker at 37 ºC, 230 rpm for 3 hours till optical density (OD) ~0.6 at 600 nm, post which arabinose is added to the culture for expression for 16 hours in the shaker at 28 ºC, 230 rpm. After 16 hours expression, 1~2 ml of cells are pelleted (8000 rpm for 2 minutes) and washed with aqueous blank buffer (composed of 0.17% Nitrogen Base and 0.5% Ammonium Sulfate both in wt%), then diluted to desired concentrations determined by the absorbance of the re-suspended cell solution in blank buffer at 600 nm. These cell solutions are filtered with sterile cell strainers (40 µm) before being used for screening or sorting. Sorted cells are usually collected on a suspension in oil and added to a flask with 25 mL LB-amp for overnight growth in a shaker at 37 ºC, 230 rpm, and can be expressed for further analysis and sorting depending on the needs.

Section 4: Data acquisition and signal processing

The laser beam focused on the microfluidic chip using a cylindrical lens has a Gaussian intensity profile I(x, y) with FWHM intensity spreads σ_x and σ_y as

$$I = I_0 exp(-\frac{2x^2}{{\sigma_x}^2} - \frac{2y^2}{{\sigma_y}^2})$$, [1]

where I_0 is the peak intensity at the center of the Gaussian beam.

Droplets flowing through this beam result in scattering signals. Despite the presence of emission filters some scattered signals of excitation light leak into the PMT. The scattered light is used to our advantage, for tracking droplet flow speeds and stability, counting number of droplets, and pair-matching two signals in green and red channels from the same droplet. Depending on the concentration of cells used in an experiment – droplets can be filled or empty. Only scattering signals are observed when the droplet is empty, but when the droplet contains a fluorophore – both scattering and fluorescence signals are recorded on the PMT. These cases may be represented by Gaussian profiles that follow the intensity profile in time: the scattering signal Sc(t) from an empty droplet, fluorescence signal Fl(t) and the total signal $C_0(t)$ from a filled droplet (assuming a 10:1 fluorescence: scattering) as presented in Eqs. [2] ~ [4]. In an instance when a droplet is filled with a single fluorophore, the signals are a convolution of the two functions. The ratio of scattering and fluorescence signals can be varied by adjusting the pump laser intensity, flow speeds, and the PMT gains for an experiment.

$$Sc(t) = Sc_0 \exp\left(-\frac{t^2}{2{\sigma_{st}}^2}\right), [2]$$

$$Fl(t) = Fl_0 \exp\left(-\frac{t^2}{2{\sigma_{ft}}^2}\right), [3]$$

$$C_0(t) = 0.1Sc_0 + Fl_0, [4]$$

The distributions in time for the beams σ_{st} and σ_{ft} are dependent on the flow speed v_x, which in turn is related to the droplet generation frequency. An example of simulated and experimentally observed signals is shown in Figure S3.

$$\sigma_{st} = \frac{\sqrt{{\sigma_x}^2 + radius\_droplet^2}}{2\, v_x}, \text{[5]}$$

$$\sigma_{ft} = \frac{\sqrt{{\sigma_x}^2 + radius\_cell^2}}{2\, v_x}, \text{[6]}$$

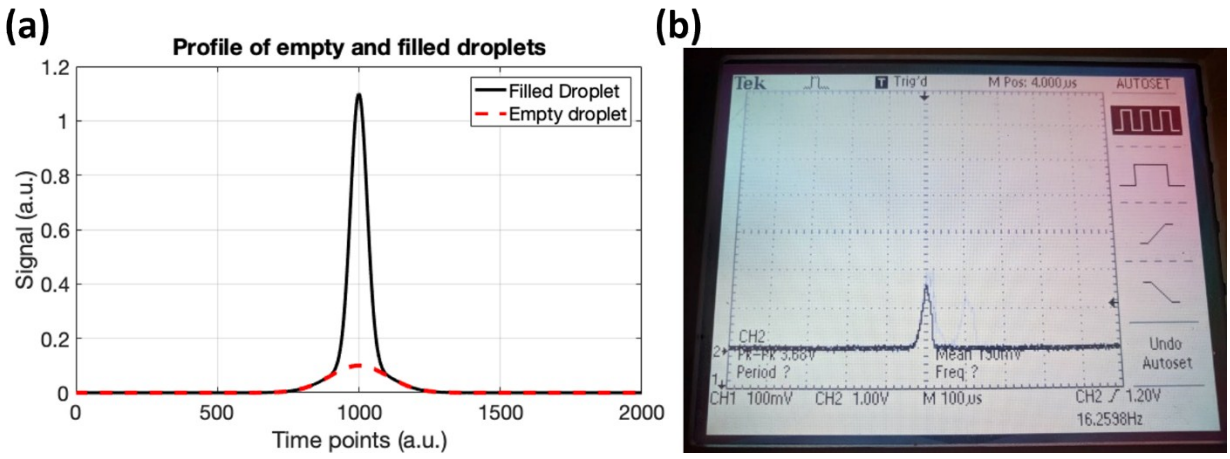


Figure S3: Simulated and observed signals of a filled droplet.

Pair-matching

Our instrument has two beams in the current configuration and since we are interested in spectroscopic signatures from the same analyte from both beams, these two events need to be "pair matched" in time. The pair matching event depends upon the flow speed. Either using a distribution feature on the custom-built LabView program or through real time tracking on an oscilloscope we can calculate the flow speed of droplets. A distribution tab on the program allows us to estimate the pair matching times, and a narrow distribution of pair matching times indicates a steady and stable flow. The distance between the two beams can be calculated through a camera image by a calibrated pixel to length conversion ratio. These allow us to calculate the flow speed of the droplets in the system.

$$v\text{x} = \frac{Distance\ between\ the\ beams}{Pair\ matching\ time}, \text{[7]}$$

Pair-matching events are schematically illustrated in Figure S4. For example, when the separation distance of two beams d~250 µm and the pair matching time ~500 µs, the flow

speed of the droplets is v_x ~0.5 m/s. The averaged droplet counts per second from a single beam can be used to check whether the pair matching time is reasonable.

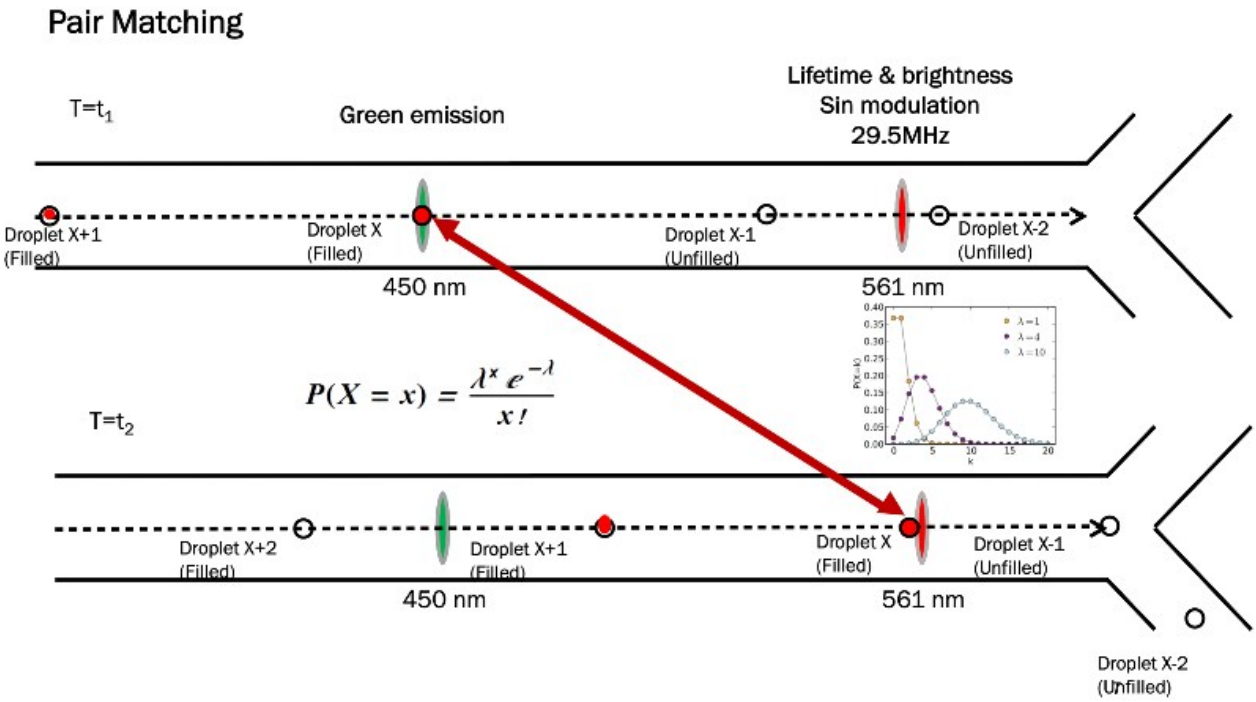


Figure S4: Pair matching events to estimate the transit time and the flow speed of the system.

Repeated measurements indicate that our droplets are around ~50pL, which gives the estimated droplet sizes to be ~46 μm in diameter. Our analyte is usually an *E. coli* cell, which is roughly assumed to be of the size of 5 μm in diameter. The FWHM of our beams on the device are usually ~3 μm on the x-(in-flow) dimensions. Using eqns. [5] and [6], σ_{st}~34 μs and σ_{ft}~ 25 μs. Thus, a fluorescent signal usually has an FWHM of ~25 μs whenever it passes a beam when we have a steady flow velocity ~0.5 m/s.

Raw Fluorescence Brightness

To operate the instrument in each mode, one needs to screen the distribution of empty droplets signals to separate from droplets that contain the fluorophore. Then by subtracting the background of empty droplets, one can estimate the distribution of the

fluorophores in the respective photophysical parameters that can be analyzed. The operation of the instrument to get the raw peak brightness of an event is carried out in the following fashion.

- The ADCs digitize the analog signals from the PMT.
- The data is then handled by the FPGA board.
- FPGA starts a search algorithm if the data point is higher than threshold.
- FPGA searches for the next 8 data points to see if they are above threshold.
- **YES:** Peak detection – 8 continuous points above threshold. Selects the maximum as the brightness reading from continuous data points above the threshold.
- **NO:** Abandons peak search and looks for next point above the threshold.

Figure S5 exhibits an example of screening fluorescence brightness of *E. coli* cells expressed with mScarlet, simulated signals and experimental signals output on an oscilloscope.

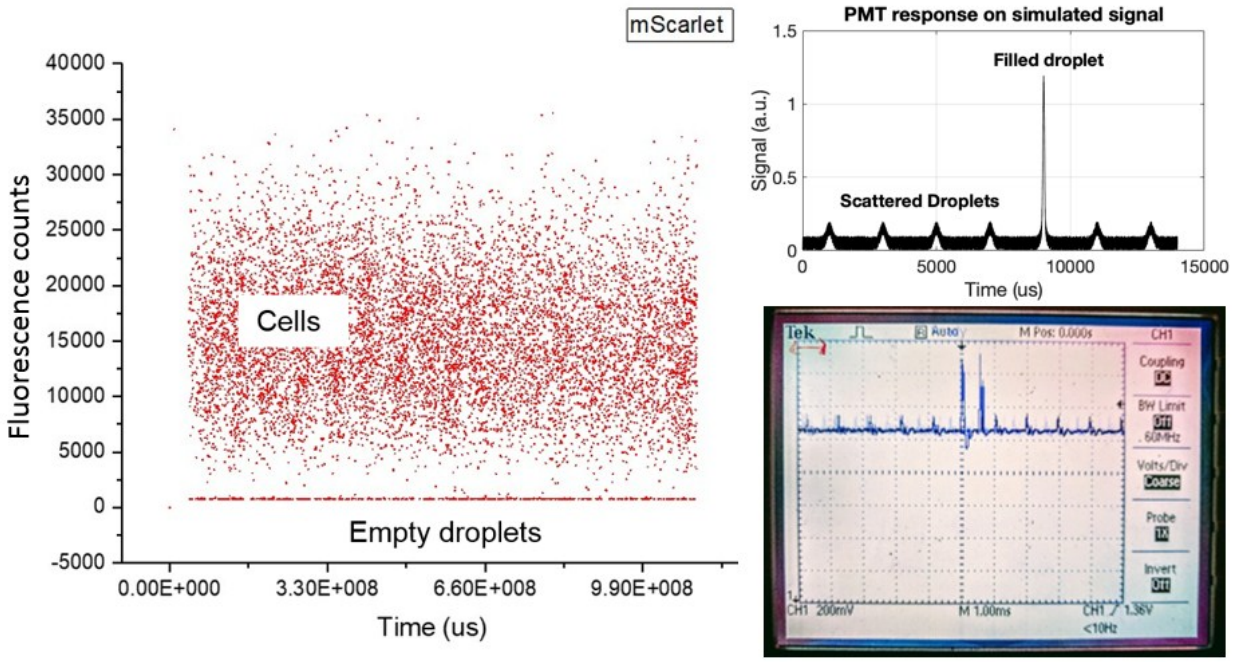


Figure S5: (Left) Fluorescence counts obtained in Red channel shows the signals from empty droplets and mScarlet. (Upper right) Signals simulated on Matlab. (Lower right)

Similar responses in flow are also noted real time on an oscilloscope during an experiment with average number of cells encapsulated per droplet λ=0.1 (9.5% of droplets filled).

Fluorescence Lifetime

For phase fluorimetry our laser is amplitude modulated in sine by the EOM. The modulation frequency of the EOM is set to *v*=29.5 MHz, (ω = 2π*v*) which resonates with the fluorescence lifetime of fluorophores in the order of ns. The fluorescence signal is therefore also amplitude modulated in time but is recorded with a phase delay and a lower modulation depth in comparison to the reference signal. The phase delay corresponds to the average time a fluorophore spends in the excited state before it emits a photon back to the ground state (eqn. [8]).

$$Fl(t) \propto A\ m\ sin(\omega t - \phi - \delta) \exp\left(-\frac{t^2}{2{\sigma_{ft}}^2}\right), \text{[8]}$$

The phase shift ϕ corresponds to the phase delay in the signal caused by the intrinsic fluorescence lifetime of the analyte whereas the constant phase shift δ can be attributed to the delay of the signal due optics, electronics etc.

$$\tau = \frac{1}{\omega} tan(\phi), \text{[9]}$$

$$m = \frac{1}{\sqrt{1 + \omega^2 \tau^2}}, \text{[10]}$$

In the experimental setup of the system, a biased-Tee filters the high frequency component (>83 kHz) of the signal from the larger gaussian envelope. The gaussian envelope is sent to a pre-amplifier (logarithmic/linear). The signal from the pre-amplifier is then digitized by the ADCs and a peak of the signal is determined as discussed previously.

The high frequency component is fed into one of the input ports of the lock-in amplifier. A reference signal is sent to the lock-in amplifier from a function generator that drives the EOM. The lock-in amplifier calculates the in-phase (cosine component of the signals) and the out-of-phase (sine component of the signal) with respect to the reference signal, and the phase shift is calculated from the inverse tangent of these two parameters.

It is critical to note that the phase value of empty droplets is usually a distribution that is close to the mean of the total phase shift observed from delay due to the optical and electronic lines and the offset value that sets the reference phase shift of a known RFP to be around 45 degree. Whenever a fluorescence event takes place, i.e. the analyte is in a droplet that passes the beam, an intrinsic phase shift is observed that corresponds to the average fluorescence lifetime of the analyte. Once the FPGA determines the peak fluorescence value, it picks up the corresponding phase value that is reported by the lock-in amplifier through an auxiliary output on the lock-in amplifier. However, the accuracy and stability of phase values are critical to many parameters including

1. The low-bandpass filter for the signal inputs
2. The time constant (TC) of the corresponding filter
3. The spread $2\sigma_{ft}$ which is critical in determining the order of the low-pass filter and the time constant of the UHF-Lock-in amplifier
4. The PMT gain and the excitation intensity: the scattered signals are not negligible, if fluorescent events are not very prominent, one can observe a skewing of the phase shift towards the mean value of the empty droplets (see Example 2 below).

It is critical to note that the Equivalent Noise Bandwidth (ENBW) is dependent on the time constant and the order of the filter in use. Lower order filters result in greater noise in the lifetime measurements, but the wait times for the phase to stabilize to the 99% of the accurate values are much lower (as an example, see SRS-830 Lock-in-amplifier Manual).

Slope	ENBW	Wait Time
6 dB/oct	1/(4T)	5T
12 dB/oct	1/(8T)	7T
18 dB/oct	3/(32T)	9T
24 dB/oct	5/(64T)	10T

(SRS-830 Lock-in-amplifier Manual)

Hence it is critical that we select a filter and a time constant that provides us enough resolution in lifetime but does not require a long wait time. The noise reduces when the

time constant increases, but the wait time also increases. As flow speed is increased the transient time of the fluorescence signal keeps dropping. Thus, it is critical to note that for a certain PMT setting and flow speed the appropriate time constant and the filters are selected for a reasonable phase resolution. Some examples are shown below to illustrate the effects of these settings.

Example 1: A mixture of different fluorescent proteins (mCherry, Kriek[1] and FusionRed-M) are mixed and are screened in a chip that operates at a speed of 2 kHz droplet generation frequency. The PMT was set at -650 V and the laser irradiance was at 75 kW/cm^2 to obtain strong scattering signals. The FWHM of the signal is around 25 µs, and the selection of a TC of 50 µs with an order 6 filter results in a noisy phase signal that is centered around the mean of the empty droplets. This means that the FPGA does not pick up the right phase and the values are centered around the distribution of the empty droplet. The results are shown in Figure S6(a).

Example 2: The same flow settings were used with a time constant of 2.5 µs and an order 4 filter. While the main populations of each RFP are distinguishable in the mixture, the heterogeneity in lifetime and brightness makes it difficult to distinguish each RFP clearly. Therefore, each RFP was screened individually and plotted on the same scatter plot in Figure S6(b). The lifetimes could be resolved but the scattering signals are also high due to the strong irradiance, resulting in fluorescence signals that are skewed in the lifetime axis. Here mCherry was selected as a reference and its mean lifetime was set at 1.6ns.

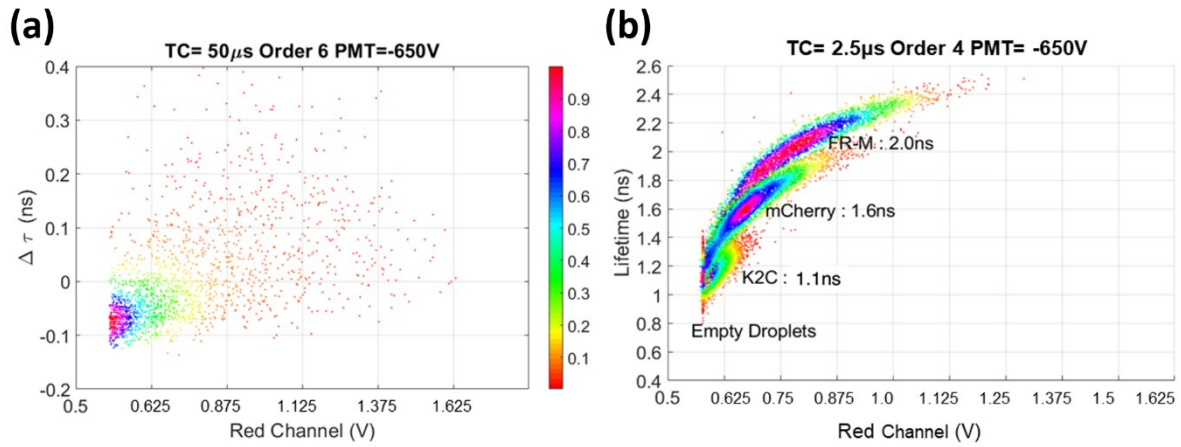


Figure S6: Screening results of a mix of mCherry, FusionRed-M and Kriek (K2C) from (a) Example 1 and (b) Example 2. The fluorescence lifetime is not resolvable in the settings of Example 1, whereas the main population of each RFP is resolvable in Example 2 but the lifetime data is skewed.

The major reason for the data skewing on the lifetime axis in Example 2 can be attributed to the fact that the scattered signals are comparable with the fluorescence signals. The scattered light is added to the fluorescence signal and both signals have the same modulation frequency but different phase shift values, so the lock-in amplifier extracts an averaged phase value from the combined signals. Therefore, the averaged lifetime becomes closer to the actual fluorescence lifetime as the fluorescence brightness increases, but the influence from the scattered light becomes noticeable as the fluorescence brightness decreases. A simulation of the reported phase values was carried out by assuming the phase values being the convolution of the offset phase shift of scattered light (set to zero in this simulation) and different intrinsic phase shifts from fluorescence lifetime as presented in Figure S7. As the fraction of scattered signal increases, i.e. the fluorescence is weaker, the measured phase shift of the total signals is skewed towards the phase of scattered light (zero phase shift).

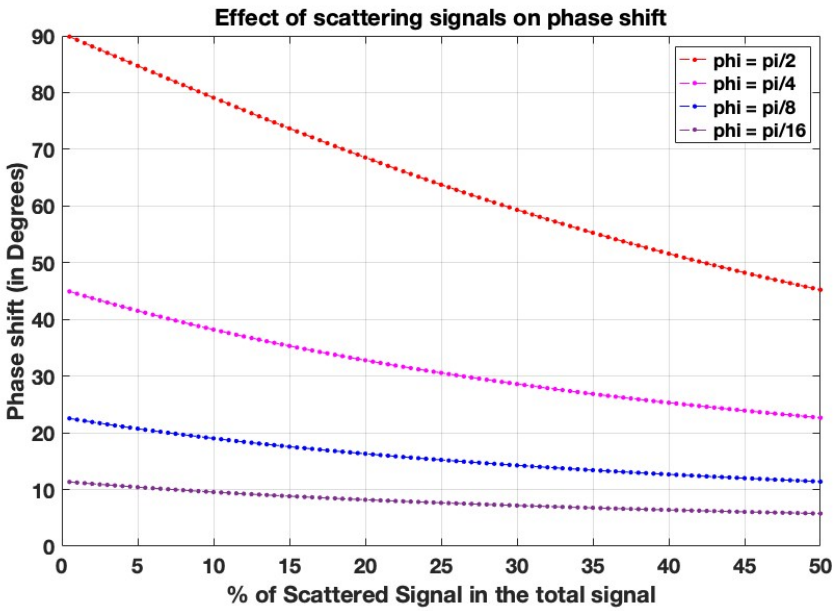


Figure S7: Simulation of phase shifts obtained from a convolution of scattered and fluorescence signals qualitatively indicates the brightness dependent lifetime data observed in Figure S6. It clearly demonstrates the effect is more dominant for greater phase shifts and operations at lower fluorescence:scattering.

Calculation of lifetimes from phase shifts

The lifetime mode of the instrument works on the concept of phase fluorimetry based on relative lifetime shifts. Therefore, lifetime characterization in the system needs a reference value for calibration. In most cases either RhodamineB containing fluorescent beads or a fluorescent protein of known lifetime is used as a reference lifetime value. We have used mCherry as the reference in most of our measurements. mCherry is set to 1.7 ns as a reference.

The phase shifts are reported to the FPGA in terms of voltages that are digitized using an ADC board. The least significant bit (LSB) value for the boards is 76 μV. The lock-in amplifier reports every degree of a phase shift as +10 mV increments.

Given our resonant frequency is 29.5 MHz. Each degree of phase shift is expressed as

$$\Delta\phi = 1^{\circ} = \frac{1}{360x29.5} \times 10^{-6}s,$$

and phase shift is expressed in LSB values,

$\Delta\phi\ (V) = \Delta(LSB) * 76 * 10^{-6}\ V.$

Equating the above two equations,

$\Delta\tau = \Delta(LSB) \times 7.15 \times 10^{-4} ns.$

Or the fluorescence lifetime of an analyte can be expressed as

$\tau = \tau(\text{reference}) + \Delta(LSB) \times 7.15 \times 10^{-4} ns.$

Section 5: Performance of the sorter

Two-color brightness sorting

The samples were prepared as described in Section 3 to obtain *E. coli* expressed with mScarlet and EGFP. *E. coli* cells expressing mScarlet or EGFP were mixed in the proportion of ~1:1 before sorting. The mixture was screened to obtain the mean brightness of mScarlet, then we sorted ~2,000 droplets from ~10^5 droplets with the selection gate greater than the mean brightness of mScarlet. The sorting process was repeated 2 times to collect totally 3 sorted samples. The sorted cells were subsequently grown overnight and screened 16 hours after induction of expression. The screening results are tabulated in Table S1.

Table S1:Screening results of sorting efficiency. The sorted cells were re-grown, expressed, and screened for totally ~1,000 cells.

# of trials	# of EGFP	# of mScarlet	Sorting efficiency
1	153	1016	87%
2	192	1057	85%
3	143	965	86%

Lifetime screening with purified proteins

Three purified proteins, mCherry, mApple, and mScarlet, and an organic dye, Rhodamine B, were prepared with concentrations about 10s μM to obtain brightness in the range comparable with cellular screening in the sorter. Pure proteins were prepared in Tris-HCl buffer (pH=7.4), and Rhodamine B was dissolved in deionized water (pH~7). Fluorescence lifetime of each sample was measured using Time-correlated Single Photon Counting (TCSPC), then transferred to the microfluidic droplet sorter to screen for the fluorescence lifetime. The results are tabulated in Table S2. The deviation between TCSPC and sorter measurements is due to the sorter averaging the phase shifts from the scattered excitation light and the fluorescence signals as described in Section 4. As the

fluorescence phase shift (lifetime) being further away from the scattering phase offset, the averaged phase shift deviates from the actual fluorescence lifetime more, in agreement with the simulation shown in Section 4.

Table S2: Fluorescence lifetime measured in TCSPC and the microfluidic droplet sorter. The lifetime of Rhodamine B was set as the reference for the sorter measurement, and the uncertainties are the standard deviations from ~25,000 screened droplets.

Sample	**TCSPC Intensity weighted lifetime in cuvette (ns)**	**Lifetime from frequency domain measurement in flow (ns)**
Rhodamine B	1.63	1.63±0.05 (ref)
mCherry	1.54	1.58±0.09
mApple	3.07	2.78±0.09
mScarlet	3.78	3.45±0.08

Lifetime-based sorting

The samples were prepared as described in Section 3 to obtain *E. coli* expressed with mScarlet and mCherry. In each batch of experiment, both cell solutions were grown from their single colony, thus there were two biological duplicates for each RFP in two batches of experiments. In the first batch of experiment, *E. coli* cells expressing mScarlet or mCherry were mixed in the proportion of ~1:1 before sorting. With one round of sorting on mScarlet, an 85% sorting efficiency was obtained as described in the main text. In the second batch of experiment, *E. coli* cells expressing mScarlet or mCherry were mixed in the proportion of ~40:60 before sorting. The mixture was screened to obtain the mean brightness and fluorescence lifetime of mScarlet and mCherry, then sorted 3,000 droplets from ~250,000 droplets with the selection gate greater than the mean brightness and fluorescence lifetime of mScarlet. The sorting process was repeated 2 times to collect totally 3 sorted samples. Another set of sorting was performed similarly by gating at brightness and fluorescence lifetime lower than the mean of mCherry. The sorted cells were subsequently grown overnight and screened 16 hours after induction of expression. The screening results are tabulated in Table S3.

Table S3: Screening results of sorting efficiency. The sorted cells were re-grown, expressed, and screened for totally ~50,000 cells

	# of trials	# of mCherry	# of mScarlet	Sorting efficiency
Sort for mScarlet	1	9688	40030	81%
	2	10745	39288	79%
	3	9780	38522	80%
Sort for mCherry	1	46799	1208	97%
	2	50053	1833	96%
	3	47740	1341	97%

Section 6: Enriching rare events

Poisson distribution

The Poisson distribution for λ = 1 to λ = 10 are plotted in Figure S8. The probability of encapsulated cells per droplet decreases quickly with the increasing number of encapsulated cells, so the p_F can be numerically calculated using n≤50 for λ≤10.

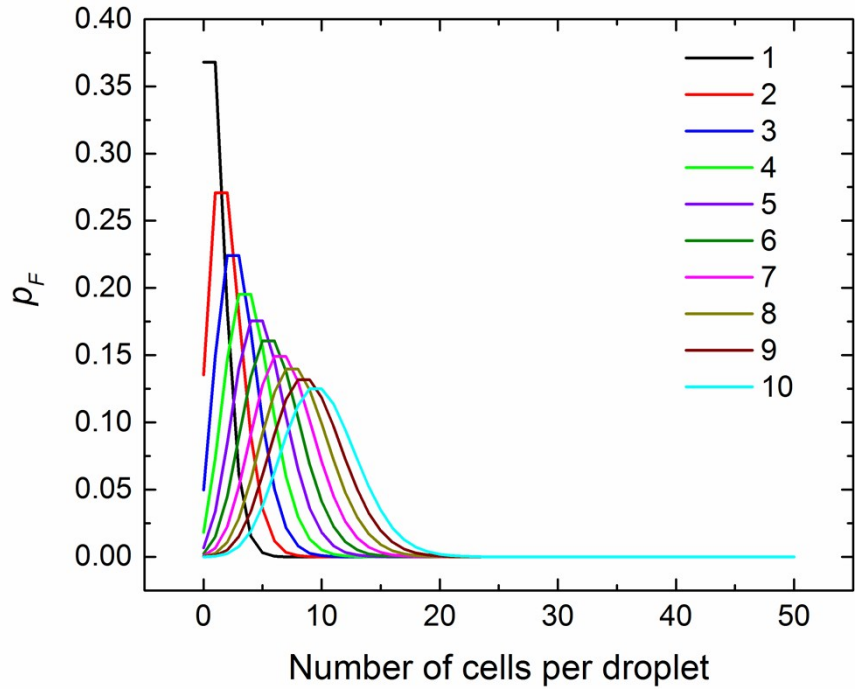


Figure S8: Poisson distribution for λ = 1 to λ = 10.

Enrichment efficiency

The sorter with current microfluidic chip configuration is capable of enriching rare events from 10^8 population within 3 hours without losing cell viability, when choosing $\lambda \geq 4$. As the targeted events become rarer, the enrichment efficiency varies depending on how many unwanted cells are encapsulated with the targeted cells in one droplet. We estimate the convergence of the enrichment efficiency below.

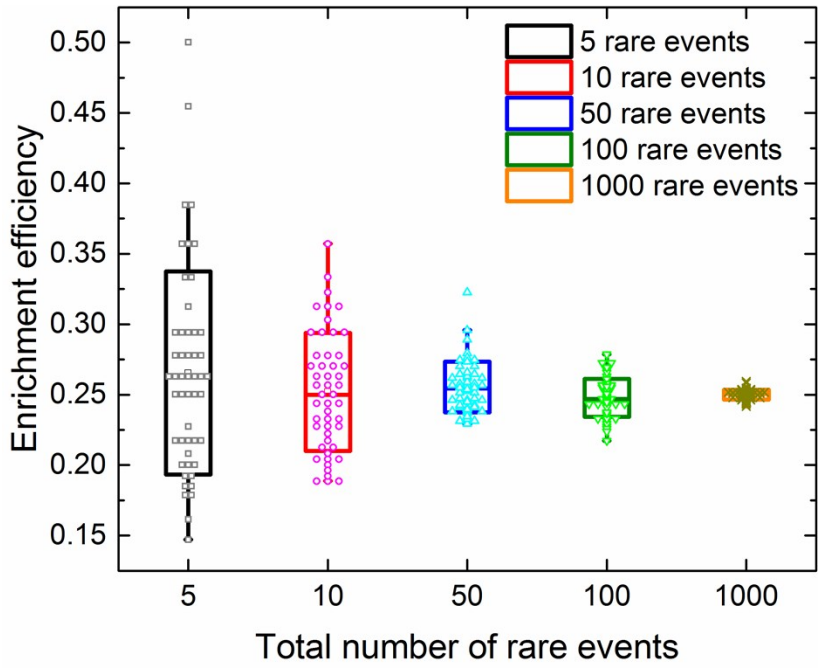


Figure S9: Enrichment efficiency converges as the number of rare events increases.

Assume that there are totally n targeted cells in 10^8 population for enrichment. The droplet encapsulation obeys Poisson distribution, we assume that there is only one targeted cell found in each sorted droplet since they are rare events. This is only an approximation, but it does not lose the generality in the convergence of enrichment efficiency. We simulate sorted n (n = 5, 10, 50, 100, and 1000) rare cells (i.e. droplets) by generating random numbers based on Poisson distribution with λ = 4, as the numbers of cells encapsulated in droplets. Repeating the process 50 times for each case to attain a distribution of enrichment efficiency and the mean and standard deviation, as plotted in Figure S9. The average enrichment efficiency is lower than the predicted values shown at Figure 6 in the main text, because this is only an approximation considering only 1 targeted cell per droplet. The standard deviation is ~7% of the mean for enriching 50 cells from 10^8 population, and ~5% of the mean for enriching 100 cells from 10^8 population. The results indicate that the calculated enrichment efficiency can be expected when sorting 0.5~1 ppm of rare events from 10^8 population. Please note that it does not limit the enrichment efficiency for sorting smaller fraction of rare events. With smaller fraction, the enrichment efficiency may deviate from the expected value, but it still provides approximately same order of magnitude of enrichment. For example, sorting 5 rare cells

from 10^8 population, the enrichment efficiency varies between approximately 15-50% in Figure S9, i.e. enriching rare events from fraction 10^{-8} to ~10^{-1} resulting in 10^7 times enrichment.

Rare RFP enrichment

1. Enrichment of rare mScarlet from a mixture with EGFP

The estimated droplet size is approximately 50 pL, so *E. coli* cells expressing EGFP and mScarlet were individually prepared at a concentration less than 2 x 10^6 cells/ml to ensure an average of $\lambda \leq 0.1$ cell/droplet to screen and determine their mean brightness. To prepare the sample for enrichment experiment, the concentration of the mixed cells was estimated to be 5.9 x 10^7 cells/ml for an average of $\lambda = 3$ cells/droplet encapsulation. A portion of the mixed cells were diluted to less than 2 x 10^6 cells/ml to ensure an average of $\lambda \leq 0.1$, and was used to determine the fraction of mScarlet in the mixture being F~0.01 by screening ~10,000 mixed cells. The selection threshold was set to be greater than the mean brightness of mScarlet. After one round of enrichment by sorting ~2,000 droplets with $\lambda = 3$ encapsulation, the sorted cells were subsequently grown, expressed and screened with $\lambda \leq 0.1$ encapsulation. The same process was repeated twice in parallel to obtain totally 3 measurements. The results are tabulated in Table S4. The mScarlet population was enriched from F~0.01 to a weighted average 35±4%.

Table S4: The enrichment of rare mScarlet from a mixture with EGFP screened at a rate of 2 kHz.

# of trials	EGFP cells	mScarlet cells	%RFP in mix
1	1785	1218	41
2	5109	2722	35
3	1070	299	22

2. Enrichment of rare mScarlet from a mixture with mCherry

The sample preparation is similar as the mScarlet and EGFP mixture described above. The mean of fluorescence brightness and lifetime was determined by screening individual mScarlet and mCherry cells at $\lambda \leq 0.1$. The fraction of mScarlet in the mixture before

enrichment was estimated to be F~5x10^{-3} by screening ~10,000 mixed cells at $\lambda \leq 0.1$ for both batches of experiments. The sorting gates were set to be greater than the mean of fluorescence brightness and lifetime of mScarlet. After one round of enrichment by sorting ~2,000 droplets with $\lambda = 3$ encapsulation, the sorted cells were subsequently grown, expressed and screened with $\lambda \leq 0.1$ encapsulation. We attained an enrichment of the mScarlet population to 40% in the first batch of experiment. In the second batch of experiment, the same process was repeated 3 times in parallel to obtain totally 3 measurements. The results of the second batch of experiment are tabulated in Table S5. The mScarlet population was enriched from F~5x10^{-3} to a weighted average 30±5%.

Table S5: The enrichment of rare mScarlet from a mixture with mCherry screened at a rate of 2.5 kHz.

# of trials	mCherry cells	mScarlet cells	%RFP in mix
1	31846	17623	36
2	29775	15519	34
3	32251	7750	19

Section 7: Directed evolution of an RFP library

Site directed libraries and the sorting for bright mutants

We saturated mScarlet-I with the RRN codon that codes for R, D, E, K, G (2 copies), N and S amino acids (see Table S6). The library size with equal probability for each available amino acid in this case is 8^8 possible mutants (~$1.7x10^7$). To ensure the search covering 95% of a library with equal probability for each available amino acid, at least 3 times of the library size must be screened[2], which is >$5.1x10^7$ cells.

The library DNA was transformed to bacterial cells (Top10 strain), with a standard heat-shock protocol. These transformed cells were added to LB+ampicillin (~7 mL) and grown overnight for the cells to reach higher optical densities. 1.5 mL of this culture was then added to 100 mL 2XYT+ampicillin media, grown for ~2 hours to achieve an OD ~ 0.6. Arabinose was then added to this culture to initiate expression of RFPs in the system. The library selection/sorting was carried out ~16-20 hrs post induction to select mutants with fast RFP maturation.

Before each sorting round, the to-be sorted library was screened with low cellular concentrations λ=0.1 to ensure single cell encapsulation in droplets in order to estimate the percent fluorescent cells in the given population. The first two rounds of sorting namely 'S1' and 'S2' were carried out using multiple cells per droplet (λ=3 for S1 and λ=1 for S2), because each round had significant non-fluorescent mutants. S1 and S2 led to significant enrichment of the library with fluorescent mutants. Then onwards, the library was sorted on two independent pathways at λ=0.1 with higher thresholds in fluorescence brightness. The sorts were repeated to ensure library convergence. After every sort, the cells in oil were transferred to a culture with 25 mL LB+ampicillin to grow overnight, consequently, expressed in 100 mL of 2XYT+ampicillin for the next round of sorting. A few ml of the re-grown cells from each sort were made into glycerol stocks stored at -80 °C for future use.

Sorting processes and results

S1
Size~1.7 x10^7 mutants (~5% fluorescent from screening ~40,000 cells)
Sort Speed: 2000 droplets/s (~2.2x10^7 cells/hr)
Sorted at λ=3 cells/droplet in 2 batches
Brightness sorting gate : 0.38V (Blank Subtracted-just above empty droplet brightness)
Screened Cells: ~8 x10^7 (~4.5x coverage)
Sorted Cells: ~5.6x10^6 (~7% of screened cells)

S2
~30% of initial population are fluorescent from screening ~10,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^6 cells/hr)
Sorted at λ=1 cell/droplet (in 1 batch)
Brightness sorting gate : 0.38V (Blank Subtracted just above empty droplet brightness)
Screened Cells: 1.7 x10^7
Sorted Cells:2.8 x10^6

S3A
~90% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 0.71±0.6 V (Empty droplet subtracted)
Brightness sorting gate: 1.9V
Screened Cells: 2.3 x10^6
Sorted Cells: ~1 x10^6

S3B
~90% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 0.76±0.5 V (Empty droplet subtracted)
Brightness sorting gate : 1.9V
Screened Cells: 1.7 x10^6
Sorted Cells: 0.9 x10^5

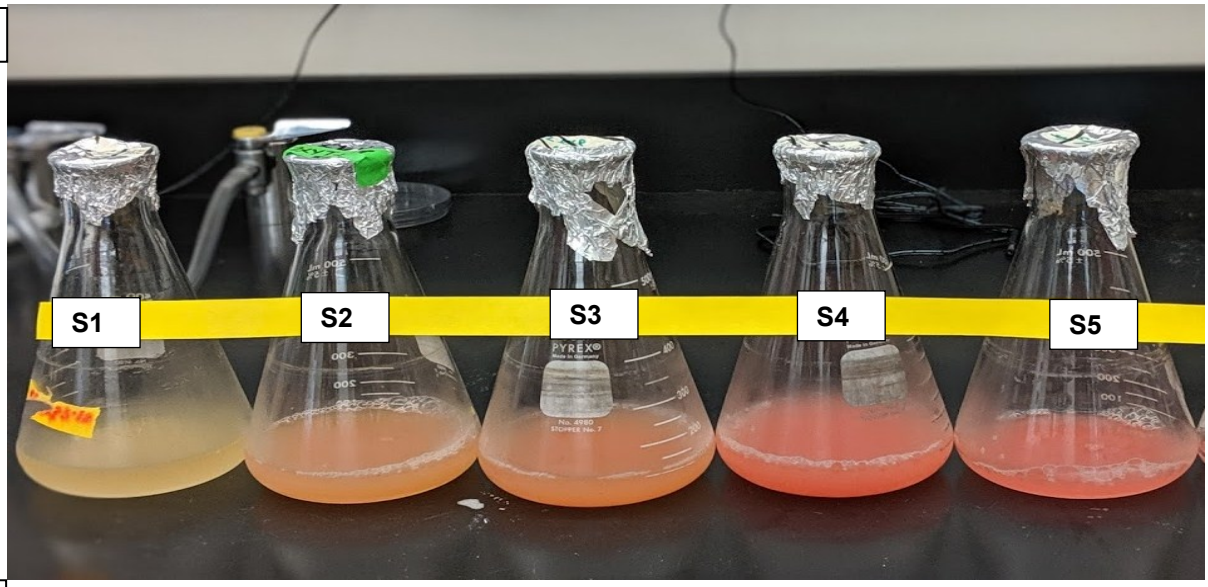
S4A
>90% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 1.25±0.8V
Brightness sorting gate : 2.3V
Screened Cells: 8x10^5
Sorted Cells: 4.5x10^4

S4B
>90% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 1.25±0.8V
Brightness sorting gate : 2.3V
Screened Cells: 7.8x10^5
Sorted Cells: 5.1x10^4

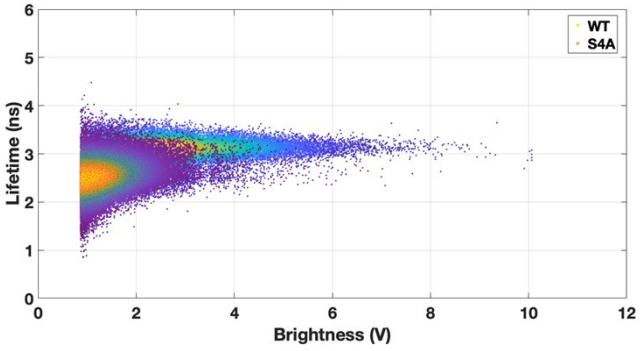
S5A
>95% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 1.8±1.1V
Brightness sorting gate : 2.7V
Screened Cells: 7.6x10^5
Sorted Cells: 7x10^4

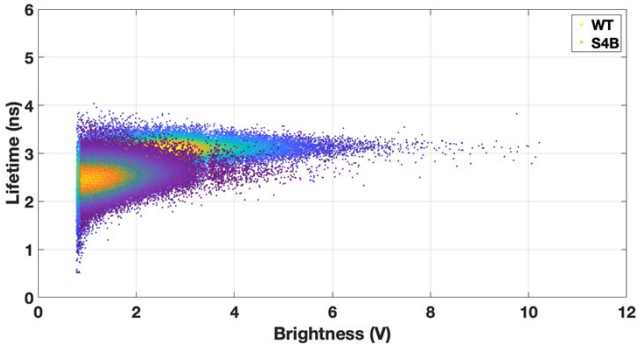
S5B
>95% of initial population are fluorescent from ~20,000 cells.
Sort Speed: 2000 droplets/s (~7.2x10^5 cells/hr)
Sorted at λ=0.1 cells/droplet
Mean Brightness= 1.9±1V
Brightness sorting gate : 2.8V
Screened Cells: 7.6x10^5
Sorted Cells: 2.5x10^4

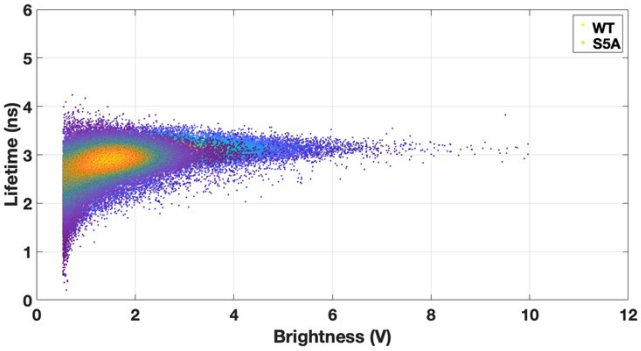
a)


S1
S2
S3
S4
S5

b)


WT
S4A
Lifetime (ns)
Brightness (V)


WT
S4B
Lifetime (ns)
Brightness (V)


WT
S5A
Lifetime (ns)
Brightness (V)

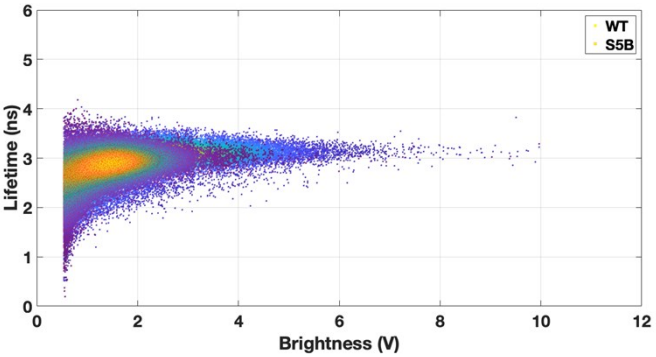

WT
S5B
Lifetime (ns)
Brightness (V)

Figure S10: (a) The progressive enrichment of the library across the rounds of sorting. All samples were re-grown from their pre-sort glycerol stocks for expression. (b) The lifetime and brightness screening plots of the post-sort S4A/S4B and S5A/S5B sorting with respect to the wild type m-Scarlet-I population.

Clones and photophysical characterization

We plated 5 plates of S5A and S5B respectively. Both S5A and S5B plates displayed mostly pink colonies. We picked 20 colonies of varying brightness with a check on lifetime using our lifetime beam. All picked colonies had phase shifts that were very close to the shifts observed in mScarlet-I. We had also plated one plate from the S4B and S4A. Interestingly one of the S4 plates showed a quick maturing bright colony. On sequencing it turned out to be unique and different from the wild type sequence, we named that clone S4-1 and purified it for further characterization. From the 20 clones we picked from the S5A and S5B plates, we had converged to 2 clones. Both plates had these clones 7:3 in one case and 8:2 in the other. This was sufficient evidence that our library had converged from both pathways. We named these clones S5-1 and S5-2 and selected them for further photophysical characterization. The mutated amino acids of these clones are listed in Table S6.

Table S6: The sequence of mScarlet-I mutants in the mutated positions.

Protein	114	119	121	147	164	174	196	198
mSc-I	E	I	K	T	A	L	D	K
S4-1	L	I	K	T	S	S	D	E
S5-1	E	I	K	T	K	K	N	N
S5-2	L	N	K	R	N	S	D	K

The protein purification and photophysical characterization were described in our previous work[3]. Fluorescence lifetime measurements were made with the PicoQuant TCSPC with a 560nm pulsed picosecond laser head. The photophysics of each RFP is listed in Table S7.

Table S7: Photophysical properties of mScarlet-I mutants.

Protein	**QY (3x trial / Avg)**				**Lifetime (ns) (3x trial / Avg)**			
mSc-I (ref)	54	54	54	54	3.22	3.29	3.27	3.26±0.03
S4-1	40	42	-	41±1.4	2.88	2.93	-	2.90±0.04
S5-1	49	53	49	50.3±2.3	3.00	3.04	3.01	3.02±0.02
S5-2	48	47	-	47.5±0.7	3.00	3.03	-	3.02±0.01

References:

1. K. M. Dean et al., *Integr. Biol.*, 2015, **7**, 263–273
2. Y. Nov, *Appl. Environ. Microbiol.*, 2012, **78**, 258–262
3. P. Manna et al., *Integr. Biol.*, 2018, **10**, 516–526