

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Droplet Microfluidics for Single-cell Genomics

**Permalink**

<https://escholarship.org/uc/item/6cn3t8js>

**Author**

Lim, Shaun Wei Sheng

**Publication Date**

2017

Peer reviewed|Thesis/dissertation

Droplet Microfluidics for Single-cell Genomics

by

Shaun Wei Sheng Lim

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Copyright 2017

by

Shaun W. Lim

## **Acknowledgements**

Elements of this dissertation have been published elsewhere or are in the process of being submitted to a peer-reviewed journal. Chapters 2-4 have been published in *Lab on A Chip*, *PLoS One* and the *Journal of Virological Methods* respectively.

When I first embarked on this journey through graduate school, I was focused on the science and the excitement of working with cutting-edge technologies and saw this as an opportunity to stretch and challenge my intellectual side, and did not fully foresee how the experience would shape me as a person. I think that this period of time has provided me with a significant paradigm shift in terms of thinking about how I should be living my life and the principles on which future endeavors can be built. The exploratory and mercurial nature of research has led me to not only question the status quo of scientific thinking, but also other aspects of my life, and I am appreciative of this opportunity for personal growth.

It was a struggle for me not only to identify problems worth working on, but also to locate the discrete elements necessary for a project to reach satisfactory progress. These elements came not only in terms of suitable materials and funding resources, but also colleagues and collaborators who had a common vision moving forward. I was lucky to have numerous factors ensuring that I was making steady progress towards the completion of a once-seemingly long and arduous journey through graduate school. The first of these would be the guidance of my advisor Professor Adam Abate, who has been my steadfast guide through the long years of research—this work would not have been possible without your amazing positivity to inspire me after countless failed experiments and negative results, and for providing me with opportunities within and out of the laboratory. I would not be anywhere near to completing this period of study

without your tireless mentorship. Secondly, I would like to acknowledge the members of my laboratory, especially the graduate students in the same cohort as I— Freeman , Shea, Tuan and John, it was a pleasure working alongside all of you, and I hope our friendship continues after our paths have diverged beyond this period of life. Thirdly, I would like to mention the members of my committees, Professors Xin Chen, Dorian Liepmann, Richard Mathies and Nicole King, for their steadfast support of my cause, and for providing invaluable insight into my research problems. Lastly, but most importantly, I would like to thank my parents and girlfriend, Huiyun, for their unconditional support and belief in me, and ever-ready willingness to lend a listening ear.

## **Abstract**

Microbial systems often exhibit staggering diversity, making the study of rare, interesting species challenging. Metagenomic analyses of mixed-cell populations are often dominated by the sequences of the most abundant organisms, while those of rare microbes are detected only at low levels, if at all. To overcome this, selective cultivation or fluorescence-activated cell sorting can be used to enrich for the target species prior to sequence analysis; however, since most microbes cannot be grown in the lab, cultivation strategies often fail, and while cell sorting requires techniques to uniquely label the cell type of interest, this not possible with uncultivable microbes. In this thesis, we introduce a strategy for the high-throughput targeted enrichment of microbial genomes, which we term PCR-activated cell sorting (PACS), and show that it enables the sequence-specific detection, sorting and recovery of microbial genomes. We then show that PACS can be applied to identify phage-host relationships by sorting bacteria based on infection by specific phages. PACS has many potential applications in microbiology, foremost among them the possibility to uncover the genetic makeup of microbial “dark matter”, in order for mankind to gain a greater understanding of these enigmatic creatures and the roles that they play in this world.

## Table of Contents

Chapter 1: Droplet microfluidics for single-cell genomic analysis .....	1
1.1 Background and introduction to droplet microfluidics.....	1
1.2 Motivation for single-cell analysis with droplets.....	2
1.3 Droplet microfluidic toolbox for biological reactions .....	4
1.5 Scope of the thesis .....	4
1.6 References.....	6
Chapter 2: Ultrahigh-throughput sorting of microfluidic drops with flow cytometry .....	8
2.1 Abstract.....	8
2.2 Introduction.....	8
2.3 Materials and methods .....	11
2.4 Results and discussion .....	14
2.5 Conclusion .....	34
2.6 Supplemental Information .....	35
2.7 References .....	37
Chapter 3: PCR-activated cell sorting for cultivation-free enrichment and sequencing of rare microbes .....	41
3.1 Abstract.....	41

3.2 Introduction.....	42
3.3 Materials and methods .....	45
3.4 Results and discussion .....	50
3.5 Supplemental Information .....	67
3.6 Conclusion .....	68
3.7 References.....	69
Chapter 4: PCR-activated cell sorting as a general, cultivation-free method for high-throughput identification and enrichment of virus hosts.....	
	72
4.1 Abstract.....	72
4.2 Introduction.....	73
4.3 Materials and methods .....	76
4.4 Results and discussion .....	79
4.5 Conclusion .....	92
4.6 Supplemental Information .....	95
4.7 References.....	101



## List of Figures

Figure 2.1 Schematic of microfluidic double emulsion FACS analysis.....	16
Figure 2.2 Non-planar device for emulsification of reinjected single emulsions.....	19
Figure 2.3 Characterization of double emulsions produced in bulk and with microfluidics.....	23
Figure 2.4 Device-made double emulsion fluorescence titration and spike-in analysis.....	26
Figure 2.5 Double emulsion TaqMan digital PCR FC analysis.....	30
Figure 2.6 Sorting with double emulsion FC.....	33
Figure 2.7 Secondary population analysis .....	35
Figure 2.8 Monitoring fluorescein dye leakage from double emulsions .....	36
Figure 3.1 The PACS workflow applied to a model microbial system .....	53
Figure 3.2 Taqman PCR detection of TolA gene on E. coli bacteria .....	59
Figure 3.3 DEP droplet sorting device.....	62
Figure 3.4 Droplet detection and sorted drops.....	63
Figure 3.5 Sequencing verification and genome enrichment.....	66
Figure 3.6 Droplet size distribution .....	67
Figure 4.1 Microfluidic workflow for PACS-based viral detection and host sorting.....	81
Figure 4.2 Digital detection of phage particles after droplet PCR.....	84

Figure 4.3 Digital detection of lambda particles and pre- and post-sorted drops containing lambda and its E. coli host .....	87
Figure 4.4 qPCR detection of host genomes before after droplet sorting based on the presence of lambda DNA .....	91
Figure 4.5 Device schematic for flow focus drop maker and sorter .....	96
Figure 4.6 LabVIEW interface .....	97
Figure 4.7 Optical setup of sorter .....	98
Figure 4.8 $\Phi$ X174 virus digital droplet assay .....	99
Figure 4.9 T4 virus digital droplet assay .....	100

**List of Tables**

Table 4.1 Primers used for qPCR and TaqMan droplet PCR .....94

## Chapter 1: Droplet microfluidics for single-cell genomic analysis

### 1.1 Background and introduction to droplet microfluidics

First envisioned as the liquid analog of silicon microelectronic devices, microfluidics is the technology of the small-scale manipulation of liquids in micrometer channels. There has been much interest in this area of research because of the potential that it holds for many diverse fields that span chemistry and biology, from drug delivery to point-of-care devices to single-cell genome analysis. The introduction of Polydimethylsiloxane (PDMS) as a medium for crafting microfluidic devices by Whitesides<sup>1</sup> was an important development in the history of the field as it brought about a new age of quick and cheap prototyping for those in academia. From the month or so of time that could have been required to fashion a silicon example, the new standard is now a few days or less for one to bring an idea from concept to testbed<sup>2</sup>.

The vision for microfluidic technologies is that any set of laboratory operations can be miniaturized, automated and integrated into a single device, providing significant savings in terms of cost, space and time. Besides capitalizing on smaller device and reagent requirements, microfluidic systems in general take advantage of fluid behavior at the microscale. These low Reynolds number flows mean that fluids operate within conditions that are laminar, leading to predictable streamlines and precise manipulation of the movement and gradients of molecules, cells or particles. Smaller volumes used for reactions mean that the kinetics of reactions can be accelerated due to a concomitant increase in the local concentration of a particular chemical compounds.

One subfield of microfluidics is droplet-based systems, which are characterized by generating small volumes of liquids, anything in the range of femto- and pico- to nano- liters, in an outer immiscible phase. Also referred to as emulsions, these droplets have been traditionally used in the food and chemical industry and only recently have been utilized for biological applications. Although work has been conducted where it is necessary to generate special cases such as air-in-water or air-in-oil emulsions, most biological work has been conducted in water-in-oil systems, due to water acting as a common solvent for many chemical reactions. Each droplet acts as a discrete compartment and can be used to confine cells or biomolecules, like nucleic acids or proteins so that there is no cross-talk between separate reactions. Surfactant is crucial for maintaining the stability of the droplets so that coalescence over time is attenuated.

These droplets can be generated at rates of thousands per second, with speeds of over 10kHz<sup>3</sup>, and are highly monodisperse, with deviations from the mean of 1-5%<sup>3</sup>. The main advantage of such systems over their chamber-based cousins is that the number of reactions that can be performed is many decades-fold in magnitude, and is only limited to the droplet generation rate. This advantage in throughput means that droplets are often put to use for applications that require high-throughput reactions that can be paired with a screen, for example dissecting out genetic variability in a population of heterogeneous cells<sup>4</sup> or the directed evolution of proteins<sup>5</sup>.

## 1.2 Motivation for single-cell analysis with droplets

The paradigm for much of biological research has been moving from bulk-type assays to a burgeoning array of single-cell profiling methods. To understand the appeal of studying individual components of a population, it is necessary to look at how research over time has

revealed new truths about cell populations once believed to be homogeneous. It is now apparent that heterogeneity exists in almost all types of cell populations, whether in the form of genetic variability and protein level differences, and these intrinsic differences can affect the phenotype of the single cell and of the population in general. Examining all cells and treating them as a whole often occludes subpopulations that may have defined and distinct functions that determine the biology of the whole population. Having greater resolution when studying biological populations holds unprecedented promise for many fields such as clinical science, where cancer tissues are dissected, to microbiology, where environmental microbes are analyzed to reveal their constituent populations.

Ever since Nossal and Lederberg generated polydisperse water-in-oil droplets containing bacteria to produce antibiotics in 1958<sup>6</sup>, droplets have been recognized and investigated for their capability to sequester reactions involving individual cells or molecules. Droplets are remarkably suited for studying cell-associated products such as secretory metabolites or for single-cell culture as the products or progeny will be confined within the aqueous compartments, and the target analyte can rapidly reach detectable concentrations. Since the work of Nossal and Lederberg, much work has been done to develop single-cell applications with droplet microfluidics, such as polymerase chain reaction (PCR)<sup>7</sup>, antibody secretion assays<sup>8</sup>, transcriptomic analysis<sup>9</sup>, and cell cytotoxicity studies<sup>10</sup>. However, one drawback of most droplet-based systems is the need to optimize assay conditions to minimize or eliminate the movement of analytes between droplets, as most systems use fluorosurfactants that can lead to micelle-assisted interdrop transport of small, hydrophobic molecules<sup>11</sup>.

### 1.3 Droplet Microfluidic toolbox for biological reactions

One of the largest advantages of using droplets for biological assays is the ability of the experimenter to execute multi-step iterative droplet manipulation. These range of operations run the gamut from cell/molecule encapsulation, droplet splitting, merger, picoinjection, and sorting. Depending on the protocol being followed, it is possible to take the droplets off-chip for incubation periods in a thermocycler, and then subject them to reinjection in another device for further manipulation and analysis. This is useful should one require temperature specific reactions, one example being PCR where temperature fluctuations for nucleic acid binding and dissociation are needed. Mixing and matching the aforementioned droplet handling techniques mean that many complex biochemical workflows can be executed.

### 1.4 Scope of the thesis

This dissertation introduces several platforms for single-cell genetic analysis with ultrahigh-throughput microbial genetic screening in mind. Droplets are used to stochastically encapsulate single microbial genomes and subject them to digital amplification, followed by droplet sorting to retrieve the genomes of interest. Most of the following chapters involve a PCR-based assay in droplets on bacteria followed by either an off-chip sorting process, or have an on-chip fluorescence-based sort, and these proof-of-concepts are intended to pave the way for future droplet microfluidic-facilitated applications that involve searching for rare species of bacteria or microbes with certain genetic constitution. These technologies could be used to interrogate any given microbial community, from searching through the human gut microbiome for certain metabolite-secreting bacteria or looking through hot spring microbes to retrieve specific phage-infected hosts.

Chapter 2 details a method to enable the off-chip screening of microfluidically-generated emulsions. Our contribution to the field is that we show how microfluidic drops can be converted to double emulsions for analysis and sorting off-chip, eliminating the need for bulky and technically complicated droplet sorting equipment. We show that commercially-available fluorescence activated cell sorters can detect and sort microemulsions with high-throughput and accuracy, and also introduce an application of our method for digital droplet quantification, comparing it favorably to existing nucleic quantitation technologies.

Chapter 3 applies PCR detection to microbes and presents a droplet workflow for genetic screening of microbes. We use a model system as a proof-of-concept to show that our system is a viable platform for screening large amounts of microbes for genetic differences and selecting for rare variants.

Lastly, Chapter 4 extends PCR-based detection and on-chip sorting and applies this workflow towards bacteria with phage sequences, demonstrating the selection of a phage host from a large background of bacteria and quantify the degree of enrichment. We also show that phage particles can be detected with our fluorescence detection platform with the same accuracy as the traditional phage-based plaque assay.



## 1.5 References

1. J. R. Anderson, D. T. Chiu, H. Wu, O. J. Schueller and G. M. Whitesides, *Electrophoresis*, 2000, **21**, 27–40.
2. G. M. Whitesides, *Nature*, 2006, **442**, 368–373.
3. T. P. Lagus and J. F. Edd, *J. Phys. Appl. Phys.*, 2013, **46**, 114005.
4. M. Najah, R. Calbrix, I. P. Mahendra-Wijaya, T. Beneyton, A. D. Griffiths and A. Drevelle, *Chem. Biol.*, 2014, **21**, 1722–1732.
5. B. Kintsjes, C. Hein, M. F. Mohamed, M. Fischlechner, F. Courtois, C. Lainé and F. Hollfelder, *Chem. Biol.*, 2012, **19**, 1001–1009.
6. G. J. N. Joshua Lederberg, *Nature*, 1958, **17**, 1419–1420.
7. B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, T. K. Kitano, M. R. Hodel, J. F. Petersen, P. W. Wyatt, E. R. Steenblock, P. H. Shah, L. J. Bousse, C. B. Troup, J. C. Mellen, D. K. Wittmann, N. G. Erndt, T. H. Cauley, R. T. Koehler, A. P. So, S. Dube, K. A. Rose, L. Montesclaros, S. Wang, D. P. Stumbo, S. P. Hodges, S. Romine, F. P. Milanovich, H. E. White, J. F. Regan, G. A. Karlin-Neumann, C. M. Hindson, S. Saxonov and B. W. Colston, *Anal. Chem.*, 2011, **83**, 8604–8610.
8. B. El Debs, R. Utharala, I. V. Balyasnikova, A. D. Griffiths and C. A. Merten, *Proc. Natl. Acad. Sci.*, 2012, **109**, 11570–11575.
9. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner, *Cell*, 2015, **161**, 1187–1201.

10. E. Brouzes, M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J. B. Hutchison, J. M. Rothberg, D. R. Link, N. Perrimon and M. L. Samuels, *Proc. Natl. Acad. Sci.*, 2009, **106**, 14195–14200.
11. M. Pan, F. Lyu and S. K. Y. Tang, *Anal. Chem.*, 2015, **87**, 7938–7943.

## **Chapter 2: Ultrahigh-throughput sorting of microfluidic drops with flow cytometry**

The following section is reprinted from “Ultrahigh-throughput sorting of microfluidic drops with flow cytometry” by Shaun W. Lim and Adam R. Abate. This article was published as a paper in Volume 13 of *Lab on a Chip* on 8 Oct 2013. Shaun W. Lim and Adam R. Abate planned the experiments. Shaun W. Lim performed the experiments and wrote the paper with Adam R. Abate.

### 2.1 Abstract

The detection and sorting of aqueous drops is central to microfluidic workflows for high-throughput biology applications, including directed evolution, digital PCR, and antibody screening. However, high-throughput detection and sorting of drops require optical systems and microfluidic components that are complex, difficult to build, and often yield inadequate sensitivity and throughput. Here, we demonstrate a general method to harness flow cytometry, with its unmatched speed and sensitivity, for droplet-based microfluidic sorting.

### 2.2 Introduction

Droplet-based microfluidic technologies are uniquely suited for applications requiring high-throughput processing of large numbers of reactions, such as in directed evolution of proteins<sup>1</sup>, targeted genomic sequencing<sup>2</sup>, rare mutation detection<sup>3</sup>, antibody screening<sup>4</sup>, and digital PCR<sup>5</sup>. Fluorescence assays are often used as readouts for the reactions because they provide bright signals that can be measured rapidly. However, measuring the fluorescence of microfluidic drops

requires sensitive and high-throughput optical systems incorporating lasers, photomultiplier tubes, beam splitters, and filters<sup>6</sup>. In addition, while several methods have been demonstrated for sorting drops with dielectrophoresis<sup>7,8</sup>, membrane valves<sup>9</sup>, and surface acoustic waves<sup>10</sup>, they require substantial time and prior experience to optimize. Moreover, sorting at high speeds requires integrated electronic and computer hardware to detect drops in real time and actuate the sorting mechanism<sup>7</sup>. While examples of this have been demonstrated, their complexity has limited widespread adoption.

Flow cytometry (FC) presents a viable alternative to on-chip droplet sorting. FC systems are unmatched in sensitivity, capable of simultaneously measuring 8-12 fluorescence colors<sup>11</sup>, and can achieve sorting rates of up to 40 KHz<sup>12</sup>; in contrast, the best microfluidic systems have only measured 3 colors and reached sorting rates of 2 KHz<sup>7,13</sup>. In addition, flow cytometers are far more common than microfluidic sorters, making them more accessible. These advantages have inspired efforts to sort drops with FC using a double emulsification strategy<sup>14</sup>, which has been successfully applied to protein evolution studies<sup>15-17</sup>. A challenge that remains, however, is that the double emulsions screened are formed using bulk emulsification methods like homogenization<sup>18</sup>, vortexing<sup>17</sup>, or extrusion through porous membranes<sup>19</sup>.

While bulk methods produce copious quantities of emulsion, they are of limited usefulness for high-throughput biology applications due to droplet non-uniformity. The drops are formed in an uncontrolled process, resulting in high polydispersity<sup>18</sup>. The large variation in drop size and shape leads to inconsistent conditions in the reactors, biasing assays and skewing sorting results. To reduce bias, drops can be “gated” during FC analysis to analyze only the ones falling within a tight size and shape range; however, these drops usually comprise a minute fraction of the emulsion, so that a majority of the reagent and screening time is wasted. Finally, and most

importantly, bulk emulsification is incompatible with droplet-based microfluidic workflows, which take great care to maintain drops as distinct reactors through the multiple steps of processing. If subjected to bulk emulsification, the fragile drops can merge or be broken into smaller drops, negating the benefits of the microfluidic workflow. To enable more effective screening and sorting of microfluidic drops for high-throughput biology applications, a new approach is needed that allows FC to be used with droplet-based microfluidic workflows.

In this paper, we demonstrate a method that allows FC to be used in conjunction with any droplet-based microfluidic workflow. Our method is based on a novel microfluidic device that acts as a “bridge” between the microfluidic workflow and the flow cytometer. The droplets produced by the microfluidic workflow are introduced into the device, which rapidly disperses them into monodisperse double emulsions. The double emulsions, dispersed in an aqueous carrier phase, can then be screened with FC. Because we form the double emulsions with microfluidics in a controlled, coaxial flow-focusing process, the drops are monodisperse and fall within a narrow size and shape range; this ensures that the majority of drops screened are usable, significantly increasing throughput compared to bulk-prepared emulsions, and ensuring more consistent conditions in the drops. To demonstrate the utility of our method, we use it to quantitate drops for a digital PCR assay. In the largest digital PCR assays ever demonstrated, a million reactions were performed in a megapixel microfluidic chamber array<sup>20</sup>. By combining ultrahigh-throughput droplet-based microfluidics with the screening capacity of flow cytometry, one million reactions can be screened in ~100 s and nearly a billion in one day.

## 2.3 Materials and methods

### *Device design and fabrication*

The microfluidic device was fabricated in poly(dimethylsiloxane) (PDMS) using soft lithography techniques<sup>34</sup>. The construction of the non-planar device necessitated the use of two complementary molds that formed opposing sections of the double emulsion-forming junction<sup>35</sup>. Both molds were fabricated on the same silicon wafer using the following steps: (1) spinning on a 50  $\mu\text{m}$ -thick layer of SU-8 3050 photoresist; (2) soft baking at 95°C for 15 minutes; (3) patterned UV exposure; (4) post-exposure bake at 95°C for 4 minutes; (5) spinning on a second 125  $\mu\text{m}$ -thick layer of 2150 photoresist; (6) steps (2) – (4) repeated; (7) resist development in propylene glycol methyl ether acetate (PGMEA). PDMS monomer was then mixed with crosslinker at an 11:1 ratio, poured over the master molds, and baked at 65°C for 3 hr. The two halves of the device were then treated with 1 mbar oxygen plasma for 20 seconds in a 300 W plasma cleaner, followed by bonding with alignment by fiducial marks in the two halves of the design.

### *Production of fluorescent-BSA microfluidic double emulsions*

We first prepare monodisperse single emulsion consisting of FITC-BSA at 200  $\mu\text{g ml}^{-1}$ . This solution is introduced into a planar flow-focusing device with a 20  $\mu\text{m}$  nozzle at 400  $\mu\text{L hr}^{-1}$ . The carrier oil phase, consisting HFE- 7500 fluorinated oil to which is added a biocompatible fluorinated surfactant at 2% by weight, is introduced at 800  $\mu\text{L hr}^{-1}$ . The fluorosurfactant is a polyethylene glycol-perfluoropolyether based triblock surfactant and was synthesized in-house, following the method as detailed in Holtze et al<sup>36</sup>. For these flow rates and device dimensions, we generate monodisperse single emulsions  $\sim 25 \mu\text{m}$  in diameter. The single emulsion drops are

collected into a 1 mL polycarbonate syringe and allowed to cream for 2 min. To generate the double emulsion drops, the creamed single emulsion is introduced into the coaxial flow focusing device at a flow rate of 200  $\mu\text{L hr}^{-1}$ . Simultaneously, the carrier aqueous phase is introduced at 30,000  $\mu\text{L hr}^{-1}$ . The high carrier phase flow rate ensures that the oil of the single emulsion is lifted from the channel walls, enabling double emulsification. The carrier phase is thickened with 10% polyethylene glycol (molecular weight 35K), which allows us to achieve higher shear rates for these flow rates, enabling better double emulsification. It also contains Pluronic F-68<sup>37</sup> at 1% by weight, to stabilize the double emulsions generated.

#### *Production of fluorescent-BSA shaken double emulsions*

Water-in-oil shaken emulsions were first produced via vortexing a 1:2 mix of 200  $\mu\text{g ml}^{-1}$  FITC-BSA with HFE-7500 fluorinated oil and fluorosurfactant at 2% by weight. The vortexing step was carried out for 5 min in a 1.5 mL microcentrifuge Eppendorf tube. The emulsion was then transferred to a 15 mL Falcon tube and vortexed for another 5 min with a detergent solution (10% PEG 35K, 1% Pluronic F-68). All reactions were carried out at room temperature.

#### *FACS analysis of double emulsions*

Double emulsions were diluted with PBS in a 1:1 ratio and transferred to a 12 x 75 mm round bottom tube for analysis with a FACSAria IIu (BD Biosciences). PBS was used as sheath fluid and events were run with a flow rate that corresponded to 200 events  $\text{s}^{-1}$  unless otherwise specified. The cytometer was maintained at a temperature of 4°C and the tube was rotated at a speed of 300 rpm during event recording. FITC-BSA and the TaqMan probe were excited with a

488 nm laser and their emission passed through a  $530\pm 30$  nm bandpass filter with a 505 nm low-pass filter in series. Alexa-Fluor 647-containing droplets were excited with a 633 nm laser and had emissions passing through a  $710\pm 50$  nm bandpass filter with a 685 nm low-pass filter in series. Sorted drops were collected in water containing 1% Pluronic F-68 and 10% PEG 35K before imaging under a microscope (Nikon Eclipse Ti inverted) on a cover glass.

#### *Single emulsion digital PCR and qPCR conditions*

The TaqMan reaction mixture consisted of a 2X ddPCR Mastermix (Bio-Rad), template (1  $\mu$ L of various concentrations), primers and probes (final working concentrations of 1  $\mu$ M and 250 nM, respectively) in a reaction volume of 50  $\mu$ L. This mix was loaded into a syringe back-filled with 100  $\mu$ L of HFE-7500 oil which was connected to a coaxial flow-focus device. The oil used for the carrier phase was the droplet generation oil for probes (Bio-rad). The emulsion was collected into PCR tubes and thermocycled on a T100 thermocycler (Bio-Rad), using the following conditions: 10 min at 95°C, 35 cycles of 10 s at 95°C, 15 s at 72°C and 30 s at 55°C.

The sequence of the Taqman probe and primers for droplet PCR were as follows: BRAF Forward 5'-TCTTCATGAAGACCTCACAGT-3'; BRAF Reverse 5'-CCAGACAAGTGTTCAAACTG (Integrated DNA Technologies); BRAF probe 5'-/6-FAM/CTAGCTACAGTGAAATC/MGB/-3' (ABD Serotec). The BRAF probe amplification was validated by breaking the emulsion with chloroform and running the product on an agarose gel. No non-specific product was observed after imaging the PCR products with gel electrophoresis.

qPCR was performed using a 2X DyNAmo Flash probe Master Mix (Thermo Scientific) and the same concentration of primer and Taqman probe as in the droplet PCR reaction. 25  $\mu$ L



reactions were run on the Stratagene Mx3005P qPCR machine (Agilent Technologies) and analyzed on the MxPro qPCR software (Agilent Technologies).

## 2.4 Results and discussion

Advances in droplet-microfluidic technologies have yielded numerous “modules” for droplet manipulation, including merger<sup>8</sup>, picoinjection<sup>21</sup>, splitting<sup>22</sup>, and on-chip incubation<sup>23</sup> and these are now being used for a diversity of high-throughput studies, Fig. 2.1A. Independent of the arrangement of the steps in the microfluidic workflow, the final step is often to screen and sort the drops. Due to the number of drops created in most workflows, detection and sorting cannot be accomplished manually, which is why automated, high-throughput systems are needed. These systems consist of technologies from multiple fields including optics, electronics, microfluidics, and computer science, making them especially challenging to build and integrate seamlessly. Often, the resultant system is sub-optimal, lacking the needed sensitivity and throughput.

Flow cytometry, in contrast, is a mature technology with unmatched speed and sensitivity, and flow cytometers are widely available for purchase. Sorting water-in-oil drops with FC, however, is non-trivial, because the instruments are designed to screen suspensions of particles or cells in a conductive aqueous phase: To sort cells, flow cytometers encapsulate the cells in charged aqueous drops; the drops then pass through an electric field, deflecting in proportion to their charge. By modulating drop charge, the cells can thus be sorted into different reservoirs<sup>11</sup>. Charging the carrier fluid is therefore central to FC sorting. However, in essentially all droplet-based microfluidic workflows used in high-throughput biology, aqueous drops are encapsulated in

an insulating oil; this oil is both immiscible with FC sheath fluids and cannot be charged, making sorting drops in a flow cytometer impossible without significant modification to the instrument.

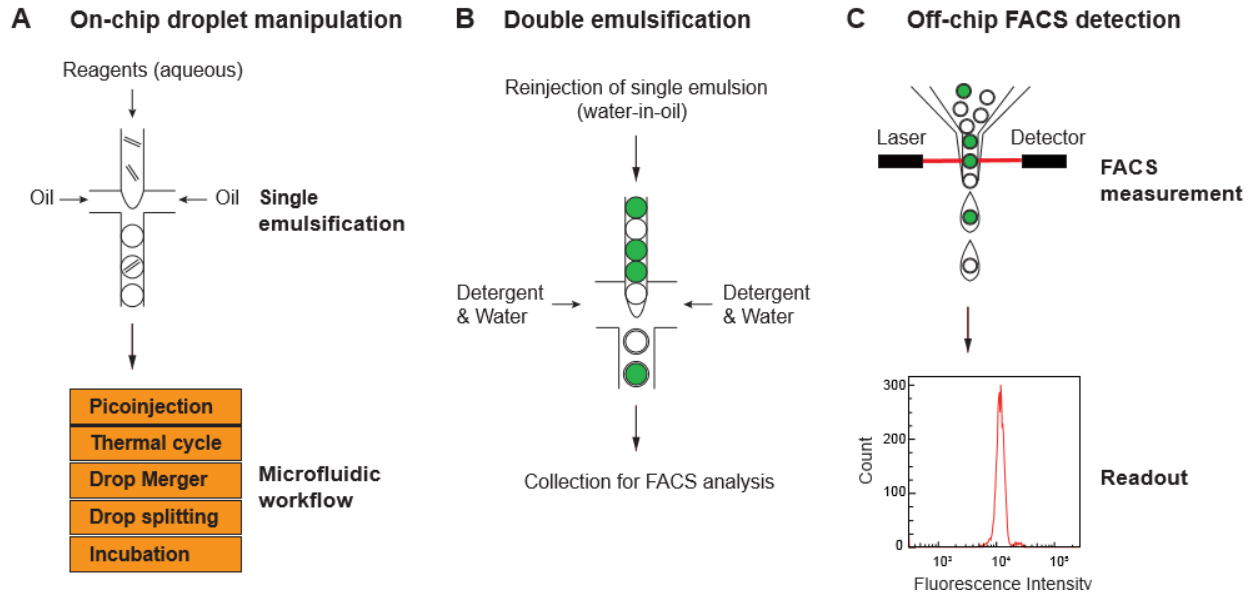


Figure 2.1. Schematic of microfluidic double emulsion FACS analysis. The non-planar emulsification device bridges FC analysis with microfluidic droplet manipulations. A) Example of microfluidic workflows in which drops are formed and manipulated using techniques like splitting and picoinjection, the results of which are single emulsions that must be sorted. B) The single emulsions are double emulsified using the non-planar device to C) enable analysis with FC.

To enable FC sorting of microfluidic drops, our strategy is to re-disperse the aqueous-in-oil drops in an aqueous carrier phase compatible with FC. To accomplish this, we use a novel coaxial flow-focusing device that encapsulates the drops in double emulsions. The double emulsion consists of a single emulsion drop sheathed in a thin shell of oil and surfactant, dispersed in an aqueous carrier phase, as shown in Fig. 2.1B. The oil shell keeps the drop contents encapsulated so that the drops retain their integrity as separate reactors and can be subject to individual FC analysis. The aqueous carrier phase, on the other hand, allows the drops to be introduced into an unmodified flow cytometer and sorted, as illustrated in Fig. 2.1C. Our double emulsification device thus acts as a “bridge” that allows microfluidic emulsions to be adapted for FC sorting.

#### *Dispersion of microfluidic single emulsions into double emulsions*

To encapsulate microfluidic single emulsions into double emulsions, we use a novel coaxial flow-focusing device that can be fabricated lithographically. The device consists of a channel 50  $\mu\text{m}$  tall into which the single emulsion drops are introduced as a close pack; close packing minimizes interstitial oil, allowing the formation of thin-shelled double emulsions, as in shown in Fig. 2.2A. The double emulsification junction consists of a channel taller and wider than the single emulsion channel; aqueous carrier fluid is introduced into the Y-shaped channel, as shown in Fig. 2.2A. The single emulsion channel is centered horizontally and vertically in the carrier phase channel; when the aqueous carrier phase is introduced at a sufficient velocity, this geometry ensures that the oil encapsulating the single emulsion lifts from the walls, forming a “cone” suspended in the flowing aqueous phase, as shown in Fig. 2.2B. This non-planar geometry allows us to form double emulsions in a device that is uniformly hydrophobic; other double

emulsification geometries require chemical patterning of channel wettability, adding cumbersome steps to the fabrication<sup>24,25</sup>.

Downstream of the cone is a constriction centered vertically and horizontally in the channel, as shown in the schematic of Fig. 2.2A. This novel feature is the reason we form thin-shelled double emulsions with just one core: As the cone extends into the constriction, it is hydrodynamically focused by the rushing carrier phase; this generates sufficient shear to rip individual drops from the tip of the cone, as illustrated in Fig. 2.2A and shown in Movie S1. A time series showing the budding off of double emulsions is provided in Fig. 2.2B, and two events are visible at  $t = 385 \mu\text{s}$  and  $t = 693 \mu\text{s}$ . Without the constriction, the double emulsions would contain multiple cores.

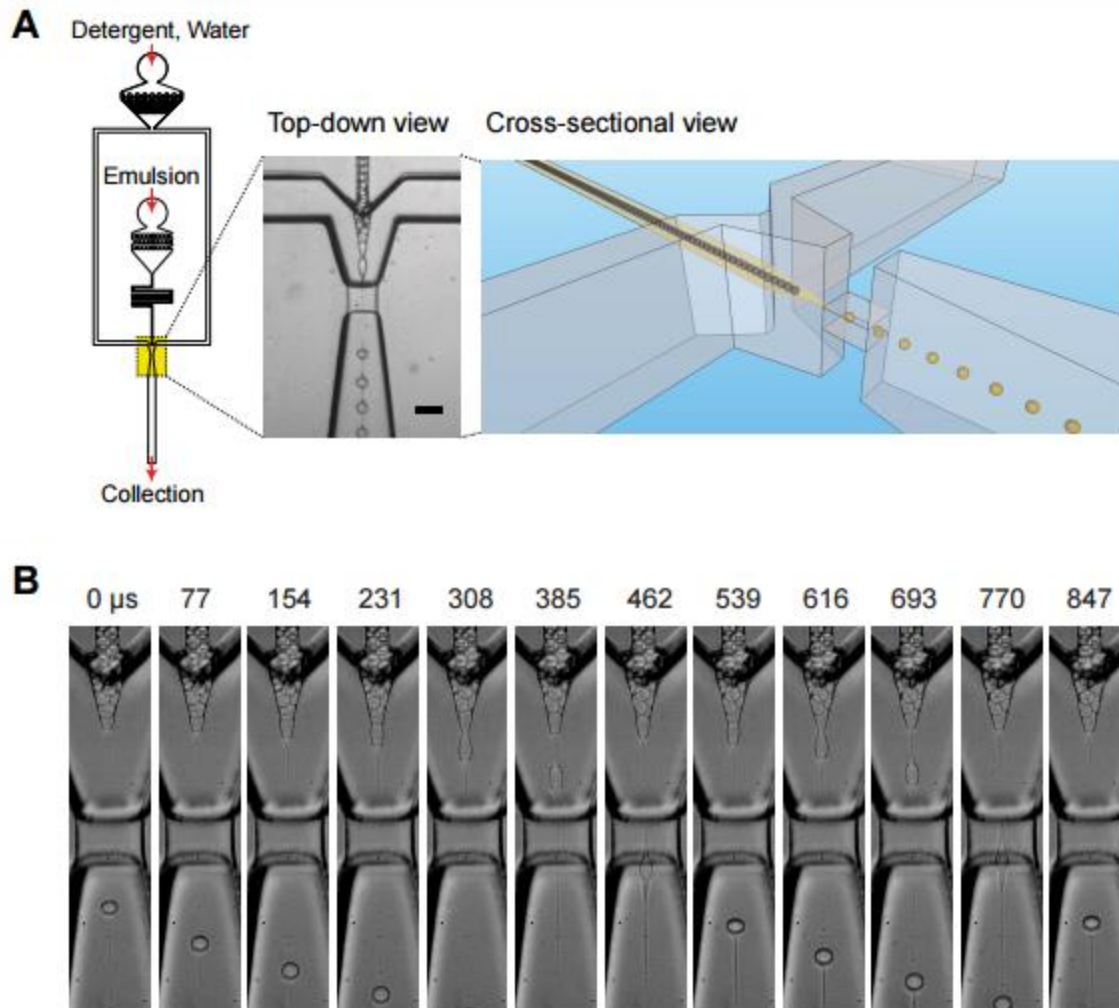


Figure 2.2. Non-planar device for emulsification of reinjected single emulsions. A) Device junction top and isometric views. The non-planar device has two streams converging at the junction, consisting of reinjected emulsion drops (inner phase) and a 10% PEG 35K and 1% Pluronic F-68 aqueous solution (outer phase). The inner phase channel is 50  $\mu$ m high and the outer phase channel 300  $\mu$ m high. B) Image sequence of double emulsion formation. Double emulsions are ripped from the tip of the single emulsion “cone” as they flow through the coaxial flow -focusing constriction. Scale bars are 50  $\mu$ m.

### *Comparison of bulk and device-made double emulsions*

The motivation for our device is that by using microfluidics to prepare double emulsions for FC analysis, we ensure that most of the emulsion is usable and the drops are uniform in size and shape, increasing throughput and sorting accuracy substantially. To directly compare our method with bulk methods, we produce emulsions containing fluorescent-BSA using mechanical homogenization in bulk and with our microfluidic device, as shown in Fig. 2.3A. As anticipated, the bulk emulsions are highly polydisperse, containing drops of many sizes and containing varying numbers of cores. Many drops are non-fluorescent, indicating they consist of carrier phase inadvertently encapsulated in an oil shell during the emulsification process, as shown in Fig. 2.3A; these drops waste reagent and screening time. Another important difference is that while the shells of the device-made emulsions are uniform in thickness, those of the bulk emulsions vary significantly, leading to lowered sorting accuracy.

To quantify the morphologies of the drops of both emulsions, we image 3000 drops of each type using fluorescence and bright-field microscopy and analyze the images with ImageJ to measure cross-sectional areas, as shown in Fig. 2.3C. The device-made drops are far more monodisperse than their bulk counterparts, having an average area of  $842 \mu\text{m}^2$  and standard deviation of  $248 \mu\text{m}^2$ , compared to  $331 \mu\text{m}^2$  and  $564 \mu\text{m}^2$ , respectively. The coefficient of variation (CV) of the device-made droplet diameters is 22.0%, which is higher than previous on-chip double emulsion production methods which cite CVs of around 1.0-6.0%<sup>26,27</sup>. This difference in polydispersity is almost entirely from the polydispersity of the single emulsion that was used for reinjection; examination of fast-camera recordings (Movie S1) reveal extremely stable emulsification of the reinjected emulsion. The single emulsion used was extremely stable and no coalescence events were observed upon reinjection.

To determine how differences in polydispersity between device-made and shaken double emulsions impact flow cytometric characterization, we screen samples of both emulsions with our flow cytometer (BD Biosciences, FACSAria IIu), screening ~10,000 drops of each. The device-made double emulsions generate a distinct cluster on the forward and side scatter (FSC and SSC) plot, whereas the bulk emulsions display only a diffuse cluster, as seen in Fig. 2.3C. These results agree with previous studies of FC screened double emulsions<sup>14,19,18</sup>. In addition to the clusters that represent the double emulsions, we also observe dense clusters at low FSC and SSC for both emulsions. We suspect these events to be small oil drops devoid of an aqueous core. To confirm this, we plot the fluorescence of these drops. As expected, they have minimal fluorescence, ~3 orders of magnitude below the large drops, confirming that they are empty (Fig 2.7). These empty oil drops are a by-product of the double emulsification methods used, as the excess oil from the single reinjected emulsions that do not become part of the double emulsions' thin oil shells form oil-in-water emulsions. To reduce the number of these oil-in-water emulsions, the reinjected emulsion should be left to cream so that the drop-to-oil ratio is the largest possible.

Importantly, of all events scanned, only 7.41% for the bulk emulsion have fluorescence values corresponding with useful double emulsions, compared to 43.9% for the device-made emulsion. As expected, device-made double emulsions thus yield many more usable events than bulk-made double emulsions, even when the gating thresholds for the bulk emulsions are set to include drops of a wide size range. A higher percentage of usable events increases the net throughput of the method, allowing larger numbers of drops to be screened per experiment.

The device-made double emulsions are also more uniform in size and have thinner, more regular shells, allowing more precise measurement of their fluorescence. To confirm this, we plot the fluorescence intensity distributions for all drops falling within the depicted gating regions, Fig.



2.3C, *right*. As expected, the intensity distribution for the device-made emulsion is significantly narrower than for the bulk-made emulsion – half a decade compared to 2 decades – demonstrating that the readings are more accurate. Accurate measurement of drop fluorescence is key for high-throughput biology applications that yield limited signal, especially in directed evolution in which the difference in signal between the best and worst mutants may be small.

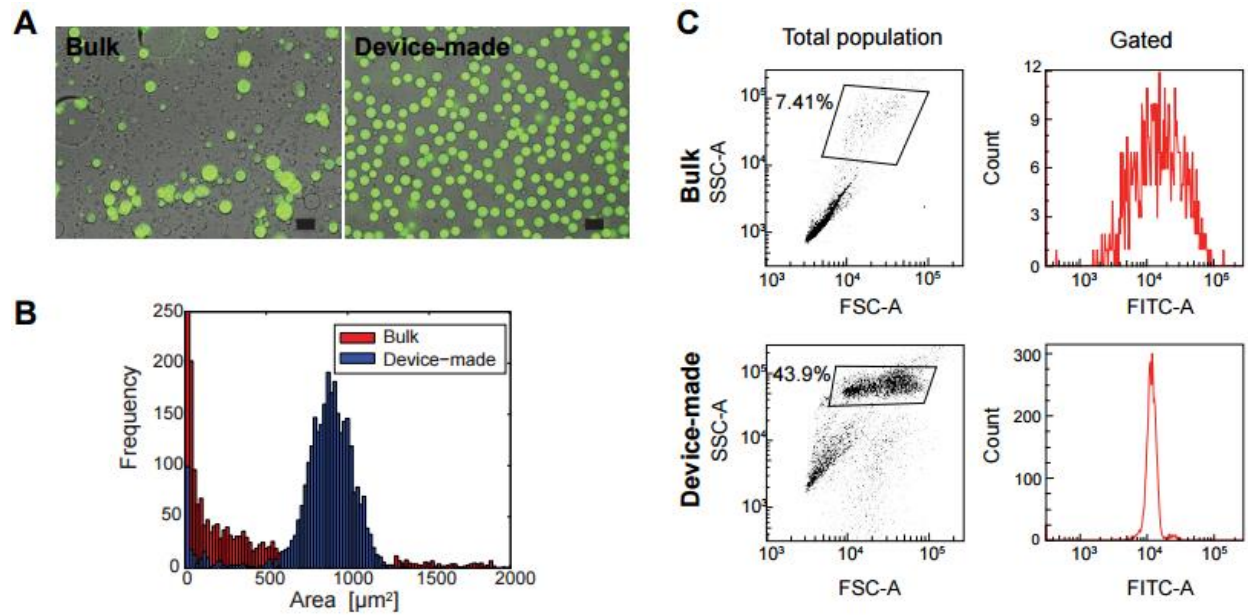


Figure 2.3. Characterization of double emulsions produced in bulk and with microfluidics. A) Double emulsions formed with vortexing in bulk (left) and with microfluidic coaxial flow focusing (right). Scale bars are 50  $\mu\text{m}$ . B) Area distributions of bulk and device-made double emulsions. C) FSC-A (Front Scatter-Area) against SSC-A (Side Scatter-Area) log-log plots and FITC channel fluorescence frequency histograms for 10,000 events. Device-made and shaken double emulsions contain 20  $\mu\text{g ml}^{-1}$  FITC-BSA. Fluorescence plots are derived by gating events in defined areas on the FSC-A/ SSC-A plots.

### *Monodispersity enables better fluorescence discrimination between double emulsions*

The ability to precisely discriminate between double emulsions based on their fluorescence is crucial for accurate screening. Monodisperse double emulsions enable more accurate discrimination, because they are more uniform in size and morphology, as shown in Fig. 2.3. To demonstrate that these accurate readings enable better discrimination between drops, we create single emulsions with four concentrations of FITC-BSA, 0.2, 2, 20 and 200  $\mu\text{g ml}^{-1}$ . The drops are double emulsified and analyzed with the flow cytometer. When we plot the intensity histograms of the double emulsions, we observe four separated peaks, each corresponding to one dye concentration, as shown in Fig. 2.4A. This shows that the double emulsions prepared with our system can be accurately discriminated by fluorescence. In contrast, if the fluorescence distributions were as wide as the bulk distribution of Fig. 2.3C, such discrimination would not be possible.

In many applications in which our approach will be used, such as in digital PCR quantitation or directed evolution, the events to be detected will be rare. This, however, plays directly into the strength of FC: Flow cytometers are capable of extremely high-throughput screening, allowing for the quantitation and sorting of very rare events. To confirm this capability, we perform spike-in experiments in which we add single emulsion drops dyed with FITC-BSA at 200  $\mu\text{g ml}^{-1}$  (emission  $\sim 520$  nm) to drops dyed with Alexa-Fluor 647-BSA at 200  $\mu\text{g ml}^{-1}$  (emission 647 nm) in volumetric proportions of 0.001, 0.01, 0.1, 0.5 and 1. The drops are double emulsified and analyzed with the flow cytometer. As expected, there is a strong correlation of the known spiked-in volume ratio to the ratio measured with the flow cytometer ( $R^2 > 0.99$ , forced fit to  $x = y$ ), as shown in Fig. 2.4B. This demonstrates we can detect droplets that are rare and, also, that

their frequency can be quantified accurately. As we show in the next section, this ability to enumerate rare events in a large population is crucial for accurate digital PCR.

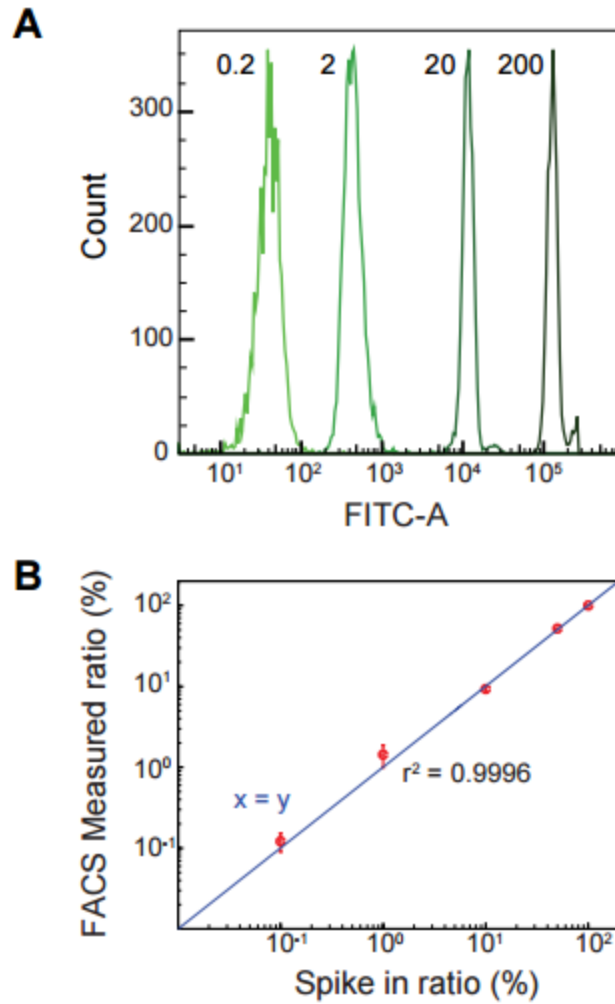


Figure 2.4. Device-made double emulsion fluorescence titration and spike-in analysis. A) Intensity histograms of device-made double emulsions containing 0.2, 2, 20 and 200  $\mu\text{g ml}^{-1}$  FITC-BSA analyzed by FC. B) Known spike-in ratios versus FC-determined ratios.

### *Megadroplet digital PCR*

In digital PCR, single molecule counting is used to estimate the concentration of a target molecule in solution. Digital PCR is thus an alternative to quantitative PCR (qPCR) that is less sensitive to amplification bias and provides absolute measurements of concentration rather than relative measurements compared to a control<sup>5</sup>. In addition, the dynamical range of digital PCR is often greater than conventional quantitative methods, making it especially useful for quantifying the concentration of rare target molecules<sup>28</sup>.

In digital PCR, a suspension of target molecules is dispersed into a large number of reactors such that, on average, most reactors are empty and a small fraction contain a single molecule. Under such limiting dilution, encapsulation of the molecules is governed by Poisson statistics, allowing the original concentration of the DNA molecules to be estimated by measuring the proportion of containers with a single molecule. The ultimate limit to the precision with which the target molecule concentration can be measured depends on the number of reactors that can be screened. Recent studies have demonstrated one million parallel digital PCR reactions using advanced chamber-based devices<sup>20</sup>. Based on the advertised throughput of flow cytometers, enumerating a million microdroplet digital PCR reactors should take ~2 min.

To investigate this capability, we use our approach to perform digital droplet PCR and rapidly enumerate the results with FC. In our experiment, we use TaqMan PCR for the single molecule amplification and counting because this reaction uses specific amplification primers and probes, adding two layers of specificity that ensure accurate quantification. To perform the reaction, we use planar flow focusing with a 25  $\mu\text{m}$  wide and 50  $\mu\text{m}$  tall nozzle to encapsulate gene-fragments on a plasmid together with TaqMan PCR reagents. The single emulsion is thermocycled and double-emulsified for FC quantitation. The double emulsions formed from

Taqman single emulsions are around twice as large as the single emulsions that they are formed from (Fig 2.5A). This may be due to osmotic differences causing water to diffuse across the double emulsion oil layer<sup>29</sup>.

Our TaqMan probes consist of DNA oligomers labeled with 6-carboxyfluorescein (6-FAM), an isomer derivative of fluorescein that has been observed to transfer between aqueous drops after cleavage during the TaqMan reaction<sup>30</sup>. To mitigate transfer, we include a high concentration of BSA (5%) in the outer phase when we make the double emulsions. BSA has been shown to attenuate inter-drop diffusion of dye for single emulsions, as it has excellent adherence to the water-oil interface<sup>31,32</sup>. With BSA outside the double emulsions, transfer of dye between the internal compartments is negligible after 2 hr, as compared to control drops that have no BSA added to the outer continuous phase. The dye diffuses so rapidly that minimal signal remains 20 min after drop making, as shown in Fig. 2.8.

As expected, we observe a strongly “digital” signal post-thermocycling, in which drops lacking a target molecule are dark and those with target molecules are bright, as illustrated in Figs. 2.5A-B. The loading of template molecules into drops follows a Poisson distribution,

$$P(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

Where  $x$  is the number of templates in a drop and  $\lambda$  is the average number of templates per drop. When  $x = 0$ , we obtain the number of drops without a template (a dark drop). The proportion  $p$  of fluorescent to non-fluorescent drops therefore depends on the average number of target molecules per drop  $\lambda$  via the relation<sup>5</sup>,

$$p = 1 - e^{-\lambda}. \quad (2)$$

The original target molecule concentration can thus be calculated based on the measured proportion of positive drops. To investigate this, we perform successive 2-fold dilutions of the gene-containing solution and process the dilutions with our digital droplet PCR workflow. As a basis of comparison with FC characterization, we also image samples of each concentration with fluorescence microscopy. Using image analysis, we estimate the proportion of positive drops by analyzing ~2000 drops for each sample. The remainder is then analyzed with FC, results for which are plotted to the right in Fig. 2.5B. We repeat this analysis for an additional 6 sets of 2-fold dilutions to estimate measurement error. If our FC approach is accurate, we expect a strong correlation between the proportions measured with imaging and double emulsion FC, which is what we find, as shown in Fig. 2.5C. This demonstrates that FC characterization of double emulsion digital droplet PCR agrees with results obtained by imaging single emulsion drops directly.

To further validate the accuracy of our method, we also analyze the 7 dilution sets using conventional qPCR, Fig. 2.5D. As described previously, qPCR is an alternative method for quantifying the concentration of DNA molecules in solution. The qPCR analysis provides concentration measurements for diluted samples relative to the most concentrated sample. Using Eq. (1), we calculate the expected corresponding number of positive drops for each dilution, provided in Fig. 2.5D. While the proportions estimated with qPCR are in good agreement with the known concentrations, the FC-digital data more closely approximates the known values, suggesting that it yields more accurate results, as shown in the figure.



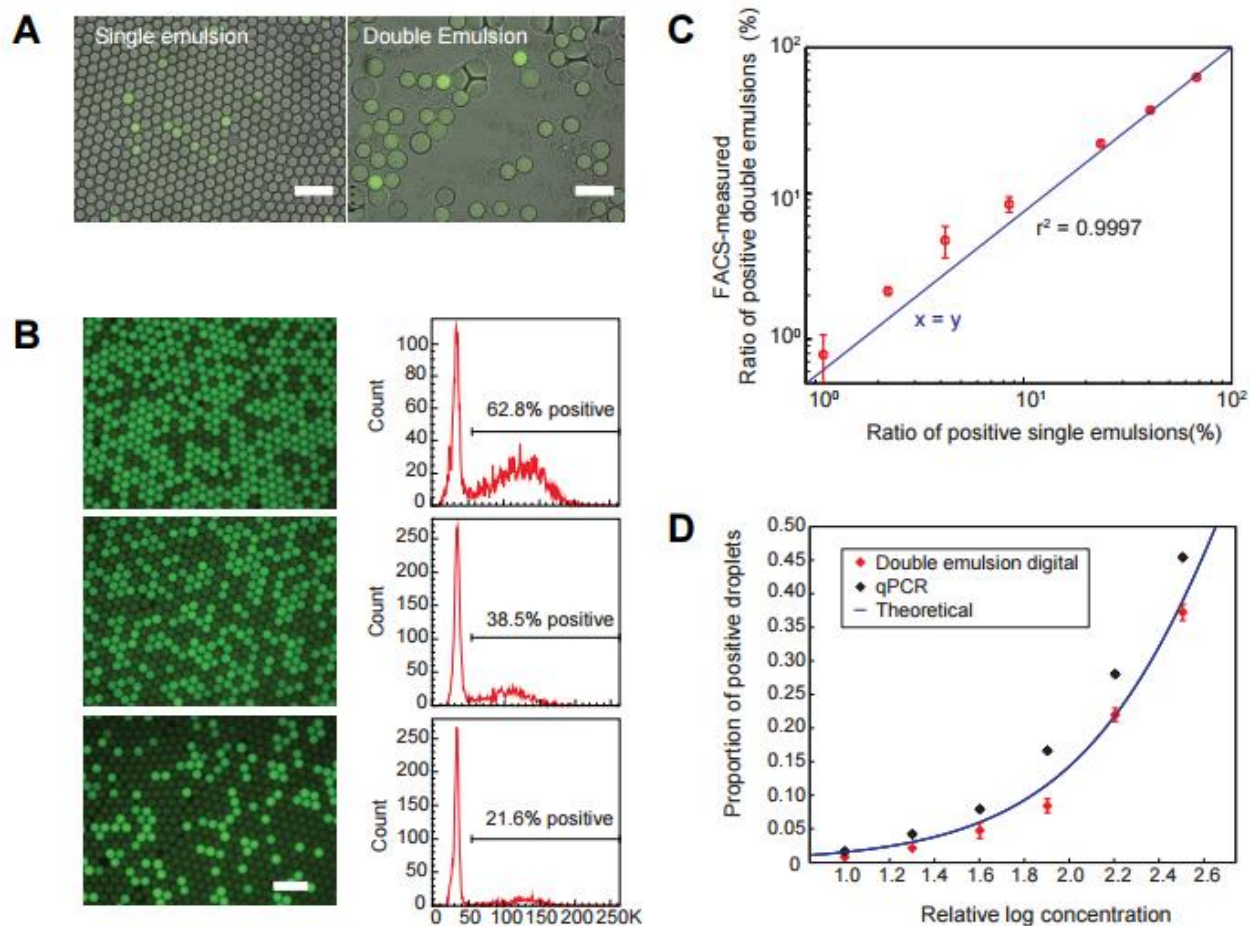


Figure 2.5. Double emulsion TaqMan digital PCR FC analysis. A) Single and double emulsions containing thermocycled cleaved TaqMan probe. B) Single emulsion digital PCR images and their associated double emulsion FC scatter plots. C) Ratio of positive double emulsions measured with double emulsification and FC against ratio of positive single emulsions measured with direct fluorescence imaging. D) Proportion of positive drops measured with double emulsion FC and estimated with qPCR analysis of serially-diluted plasmid samples. The blue line is the proportion expected based on the known dilution. Scale bars are 100  $\mu\text{m}$ .

### *Ultra-high-throughput droplet sorting with flow cytometry*

In addition to enabling ultrahigh-throughput optical characterization of individual drops, FC can also be used to sort drops. Sorting is important for high-throughput biology applications in which specific drops must be recovered from a heterogeneous sample, such as in directed evolution. To illustrate the use of our method for droplet sorting, we prepare a mixed population of drops containing FITC-BSA and Alexa-Fluor 647-BSA. The two droplet types are mixed at a ratio of 1:9, respectively, double emulsified, and FC screened. The fastest theoretical sorting possible with our flow cytometer is 40 KHz, although accuracy decreases above 10 KHz. This is because at higher sorting rates the probability that two drops enter at the same time and are sorted together increases<sup>12</sup>.

To determine the maximum speed with which microfluidic double emulsions can be sorted with our flow cytometer, we prepare mixed emulsions at different densities. In the first emulsion, the drops are diluted so that at the maximum flow rates of our flow cytometer the sorting speed is ~200 Hz. At this sorting rate we observe no red Alexa-Fluor 647-BSA drops in the green FITC-BSA drop container, as shown in Fig. 2.6. Despite analyzing multiple fields of view, we detect no red drops amongst the sorted green drops, demonstrating that at this rate sorting errors are rare. To push the sorting speed, we concentrate the emulsion by centrifuging and re-dispersing at a 50X higher concentration. At this concentration, the sorting rate is ~10 KHz and we begin to observe red drops in the collection container, as shown in Fig. 2.6. As the speed of FACS sorting increases, the degree of enrichment from an emulsion sort should decrease. This is mainly because multiple droplets can be charged and sorted at the same time on FACS equipment<sup>33</sup>, which leads to the presence of unwanted background signal in the sort bin. To maximize the sort purity, users can

utilize algorithms on FACS operating software that only sort drops with a predetermined number of adjacent empty drops. However, this will come at a cost of sorting yield as fewer drops will be sorted under the more stringent criteria.

These sorts show that as sorting rate increases errors will increase, although even at 10 KHz high enrichment ratios are possible, as demonstrated by comparing emulsions before and after sorting. This sorting rate is five times faster than the fastest microfluidic sorters demonstrated<sup>7,13</sup>.

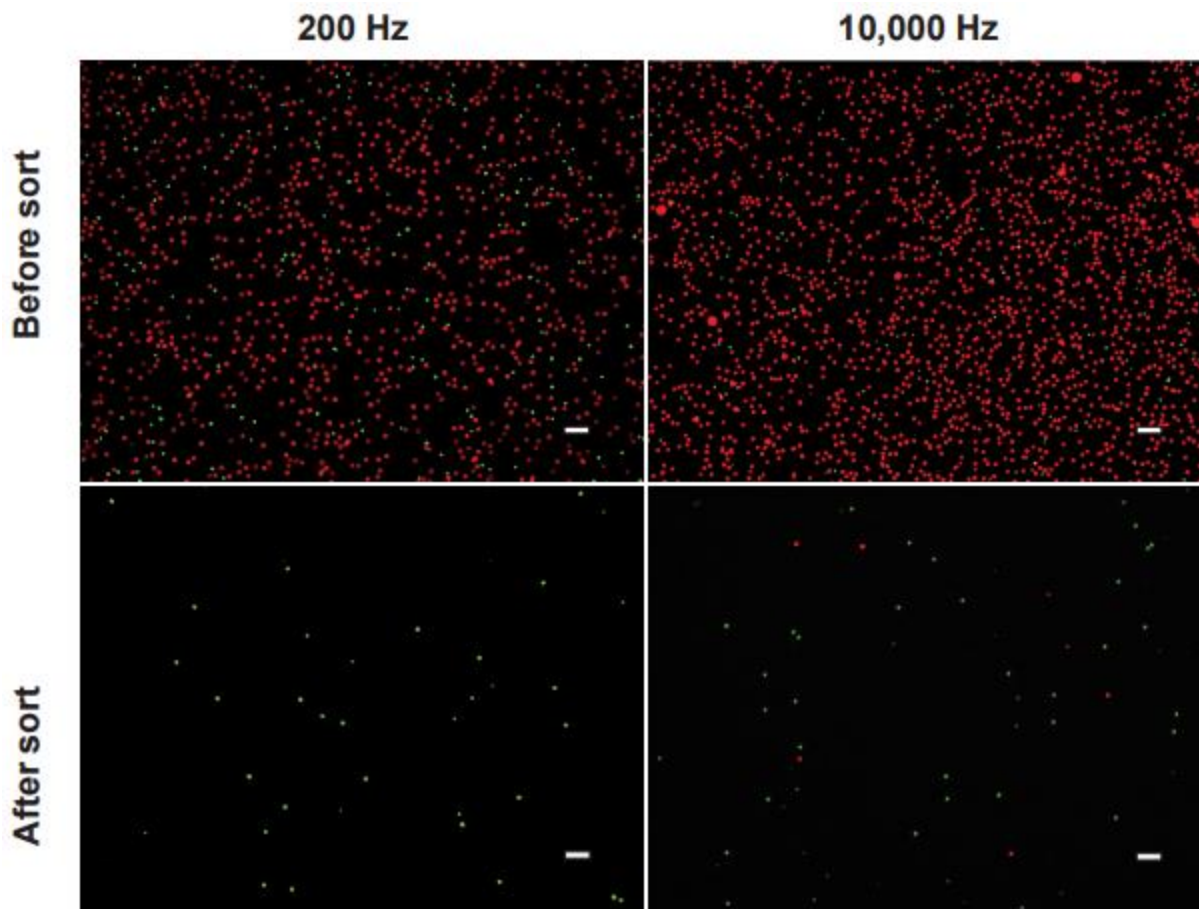


Figure 2.6. Sorting with double emulsion FC. Single emulsions containing  $200 \mu\text{g ml}^{-1}$  FITC-BSA (green) or Alexa-Fluor 647-BSA (red) were mixed at a 1:9 ratio, double emulsified and sorted at 200Hz and 10 KHz on the FC. Scale bars are  $100 \mu\text{m}$ .

## 2.5 Conclusion

We have developed a double emulsification strategy that allows droplets produced in nearly any microfluidic workflow to be sorted with flow cytometry. The high sensitivity of flow cytometers should enable the use of assays that, previously, yielded too little signal to be measured with custom detectors, while the ability to simultaneously characterize 8-12 fluorescence channels should enable greater reaction multiplexing. The higher throughput obtainable with flow cytometry should also allow larger samples to be screened, for more accurate digital PCR characterization and the screening of larger libraries in directed evolution.

## 2.6 Supplementary information

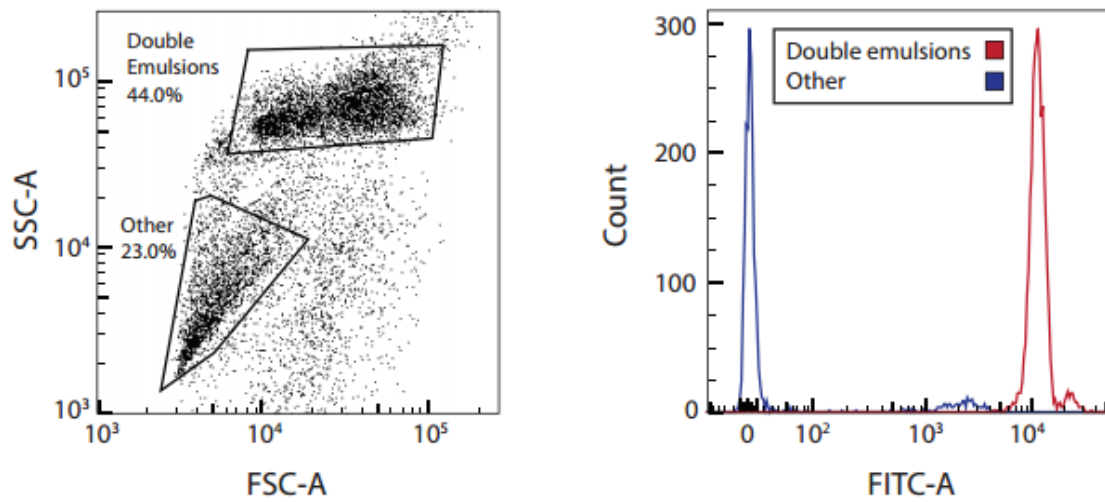
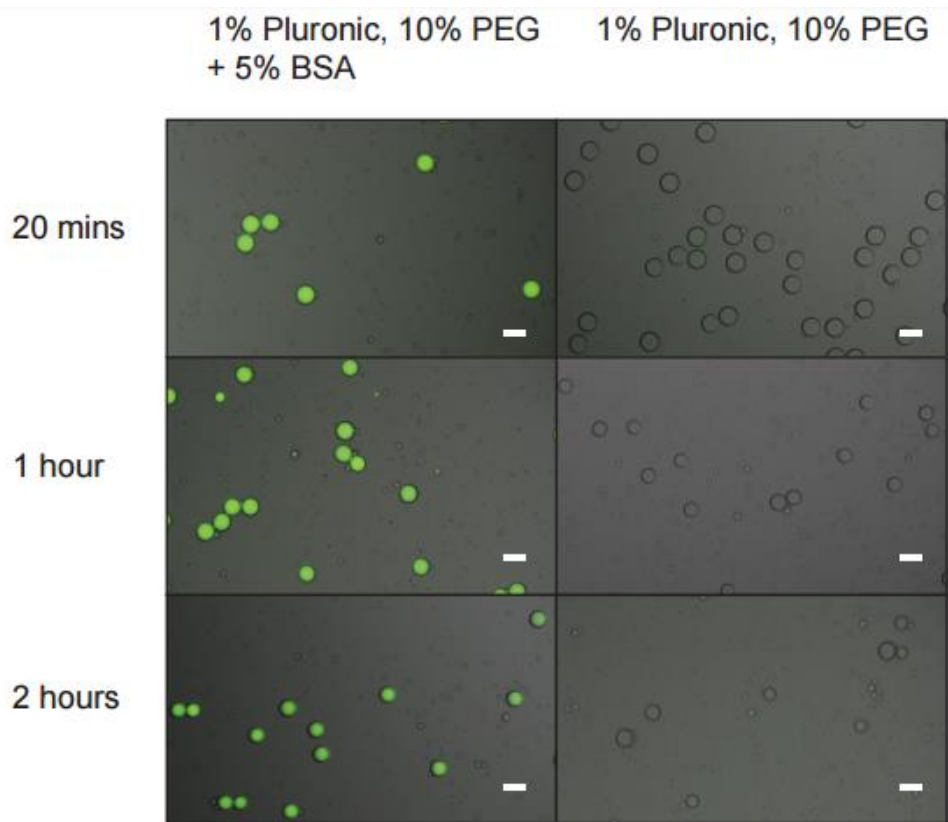


Figure 2.7. Secondary population analysis. Gating out a secondary population that comprise 23.0% of the total events reveal that the fluorescence levels of the main population is around 3 decades higher, confirming that the former are empty oil droplets that can be excluded from our analyses.



Note: Double emulsions for both sets of experiments were formed with the same set of single emulsion containing fluorescein at 2.5  $\mu\text{M}$

Figure 2.8. Monitoring fluorescein dye leakage from double emulsions. The left column shows double emulsions made with 5% BSA added to the outer phase and 200  $\mu\text{g ml}^{-1}$  Fluorescein in the inner phase, while the right column shows double emulsions omitting BSA in the outer phase. Fluorescein dye leaks so quickly out of the double emulsions that 20 min after generating them minimal fluorescence is observed when no BSA is added to the outer phase. Scale bars are 50  $\mu\text{m}$ .

## 2.7 References

1. J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, and D. A. Weitz, *Proceedings of the National Academy of Sciences*, 2010, **107**, 4004–4009.
2. R. Tewhey, J. B. Warner, M. Nakano, B. Libby, M. Medkova, P. H. David, S. K. Kotsopoulos, M. L. Samuels, J. B. Hutchison, J. W. Larson, E. J. Topol, M. P. Weiner, O. Harismendy, J. Olson, D. R. Link, and K. A. Frazer, *Nature Biotechnology*, 2009, **27**, 1025–1031.
3. D. Pekin, Y. Skhiri, J.-C. Baret, D. Le Corre, L. Mazutis, C. Ben Salem, F. Millot, A. El Harrak, J. B. Hutchison, J. W. Larson, D. R. Link, P. Laurent-Puig, A. D. Griffiths, and V. Taly, *Lab on a Chip*, 2011, **11**, 2156.
4. B. El Debs, R. Utharala, I. V. Balyasnikova, A. D. Griffiths, and C. A. Merten, *Proceedings of the National Academy of Sciences*, 2012, **109**, 11570–11575.
5. B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, T. K. Kitano, M. R. Hodel, J. F. Petersen, P. W. Wyatt, E. R. Steenblock, P. H. Shah, L. J. Bousse, C. B. Troup, J. C. Mellen, D. K. Wittmann, N. G. Erndt, T. H. Cauley, R. T. Koehler, A. P. So, S. Dube, K. A. Rose, L. Montesclaros, S. Wang, D. P. Stumbo, S. P. Hodges, S. Romine, F. P. Milanovich, H. E. White, J. F. Regan, G. A. Karlin-Neumann, C. M. Hindson, S. Saxonov, and B. W. Colston, *Analytical Chemistry*, 2011, **83**, 8604–8610.
6. G. Ryu, J. Huang, O. Hofmann, C. A. Walshe, J. Y. Y. Sze, G. D. McClean, A. Mosley, S. J. Rattle, J. C. deMello, A. J. deMello, and D. D. C. Bradley, *Lab on a Chip*, 2011, **11**, 1664.



7. J.-C. Baret, O. J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M. L. Samuels, J. B. Hutchison, J. J. Agresti, D. R. Link, D. A. Weitz, and A. D. Griffiths, *Lab on a Chip*, 2009, **9**, 1850.
8. K. Ahn, J. Agresti, H. Chong, M. Marquez, and D. A. Weitz, *Applied Physics Letters*, 2006, **88**, 264105.
9. A. R. Abate, J. J. Agresti, and D. A. Weitz, *Applied Physics Letters*, 2010, **96**, 203509.
10. T. Franke, S. Braunmüller, L. Schmid, A. Wixforth, and D. A. Weitz, *Lab on a Chip*, 2010, **10**, 789.
11. L. A. Herzenberg, D. Parks, B. Sahaf, O. Perez, M. Roederer, and L. A. Herzenberg, *Clinical chemistry*, 2002, **48**, 1819–1827.
12. S. F. Ibrahim and G. van den Engh, *Current Opinion in Biotechnology*, 2003, **14**, 5–12.
13. A. Fallah-Araghi, J.-C. Baret, M. Ryckelynck, and A. D. Griffiths, *Lab on a Chip*, 2012, **12**, 882.
14. K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, and D. S. Tawfik, *Analytical Biochemistry*, 2004, **325**, 151–157.
15. E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, and A. D. Griffiths, *Chemistry & Biology*, 2005, **12**, 1291–1300.
16. R. Ostafe, R. Prodanovic, U. Commandeur, and R. Fischer, *Analytical Biochemistry*, 2013, **435**, 93–98.
17. J. A. Stapleton and J. R. Swartz, *PLoS ONE*, 2010, **5**, e15275.
18. W.-C. Lu and A. D. Ellington, *Methods*, 2013, **60**, 75–80.
19. O. J. Miller, K. Bernath, J. J. Agresti, G. Amitai, B. T. Kelly, E. Mastrobattista, V. Taly, S. Magdassi, D. S. Tawfik, and A. D. Griffiths, *Nature Methods*, 2006, **3**, 561–570.

20. K. A. Heyries, C. Tropini, M. VanInsberghe, C. Doolin, O. I. Petriv, A. Singhal, K. Leung, C. B. Hughesman, and C. L. Hansen, *Nature Methods*, 2011, **8**, 649–651.
21. A. R. Abate, T. Hung, P. Mary, J. J. Agresti, and D. A. Weitz, *Proceedings of the National Academy of Sciences*, 2010, **107**, 19163–19166.
22. A. R. Abate and D. A. Weitz, *Lab on a Chip*, 2011, **11**, 1911.
23. T. M. Tran, F. Lan, C. S. Thompson, and A. R. Abate, *Journal of Physics D: Applied Physics*, 2013, **46**, 114004.
24. A. R. Abate, J. Thiele, and D. A. Weitz, *Lab on a Chip*, 2011, **11**, 253.
25. W.-A. C. Bauer, M. Fischlechner, C. Abell, and W. T. S. Huck, *Lab on a Chip*, 2010, **10**, 1814.
26. M. B. Romanowsky, A. R. Abate, A. Rotem, C. Holtze, and D. A. Weitz, *Lab on a Chip*, 2012, **12**, 802.
27. S.-H. Kim, J. W. Kim, J.-C. Cho, and D. A. Weitz, *Lab on a Chip*, 2011, **11**, 3162.
28. A. S. Whale, J. F. Huggett, S. Cowen, V. Speirs, J. Shaw, S. Ellison, C. A. Foy, and D. J. Scott, *Nucleic Acids Research*, 2012, **40**, e82–e82.
29. R. Mezzenga, B. M. Folmer, and E. Hughes, *Langmuir*, 2004, **20**, 3574–3582.
30. Y. Chen, A. Wijaya Gani, and S. K. Y. Tang, *Lab on a Chip*, 2012, **12**, 5093.
31. F. Courtois, L. F. Olguin, G. Whyte, A. B. Theberge, W. T. S. Huck, F. Hollfelder, and C. Abell, *Analytical Chemistry*, 2009, **81**, 3008–3016.
32. N. Gaiti, A. Aserin, and Y. Cohen, *Journal of controlled release*, 1994, **29**, 41–51.
33. J. F. Leary, *Methods Cell Biol*, 1994, **42**, 331–358.
34. Y. Xia and G. M. Whitesides, *Annual review of materials science*, 1998, **28**, 153–184.

35. A. Rotem, A. R. Abate, A. S. Utada, V. Van Steijn, and D. A. Weitz, *Lab on a Chip*, 2012, **12**, 4263.
36. C. Holtze, A. C. Rowat, J. J. Agresti, J. B. Hutchison, F. E. Angilè, C. H. J. Schmitz, S. Köster, H. Duan, K. J. Humphry, R. A. Scanga, J. S. Johnson, D. Pisignano, and D. A. Weitz, *Lab on a Chip*, 2008, **8**, 1632.
37. N. Garti, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 1997, **123**, 233–246.

## **Chapter 3: PCR-activated cell sorting for cultivation-free enrichment and sequencing of rare microbes**

The following section is reprinted from “PCR-activated cell sorting for cultivation-free enrichment and sequencing of rare microbes” by Shaun W. Lim, Tuan M. Tran, and Adam R. Abate. The article was published as a research article in *PLoS One* on January 28 2015. Shaun W. Lim and Adam R. Abate planned the experiments and wrote the manuscript. Shaun W. Lim and Tuan M. Tran performed the experiments.

### 3.1 Abstract

Microbial systems often exhibit staggering diversity, making the study of rare, interesting species challenging. For example, metagenomic analyses of mixed-cell populations are often dominated by the sequences of the most abundant organisms, while those of rare microbes are detected only at low levels, if at all. To overcome this, selective cultivation or fluorescence-activated cell sorting (FACS) can be used to enrich for the target species prior to sequence analysis; however, since most microbes cannot be grown in the lab, cultivation strategies often fail, while cell sorting requires techniques to uniquely label the cell type of interest, which is often not possible with uncultivable microbes. Here, we introduce a culture-independent strategy for sorting microbial cells based on genomic content, which we term PCR-activated cell sorting (PACS). This technology, which utilizes the power of droplet-based microfluidics, is similar to FACS in that it uses a fluorescent signal to uniquely identify and sort target species. However, PACS differs importantly from FACS in that the signal is generated by performing PCR assays on the cells in microfluidic droplets, allowing target cells to be identified with high specificity

with suitable design of PCR primers and TaqMan probes. The PACS assay is general, requires minimal optimization and, unlike antibody methods, can be developed without access to microbial antigens. Compared to non-specific methods in which cells are sorted based on size, granularity, or the ability to take up dye, PACS enables genetic sequence-specific sorting and recovery of the cell genomes. In addition to sorting microbes, PACS can be applied to eukaryotic cells, viruses, and naked nucleic acids.

### 3.2 Introduction

Microbial communities play crucial roles in geochemistry and exist in diverse ecosystems<sup>1,2</sup>. Understanding the genetics of the cells that inhabit an ecosystem is critical to understanding how microbes function individually and in the complex networks that make up the natural environment<sup>1,2</sup>. Studying the genetics of individual microbes, however, is difficult because most cannot be cultured in the laboratory and comprise uncultivable “microbial dark matter”<sup>3,4</sup>. To study uncultivable microbes, cultivation-free methods like shotgun sequencing are necessary. In this approach, nucleic acids are purified out of a heterogeneous sample via chemical means, sheared into short fragments, and sequenced. To assemble the resulting compilation of short sequences into a larger coherent dataset, computational algorithms are utilized, but this process is often hampered by the lack of sequencing depth and the complexity of the diverse set of sequences obtained<sup>5-7</sup>. As a result, next-generation sequencing of diverse communities commonly yields information about the genes present in a system but is unable to tell how those genes are bundled into genomes and packaged into individual cells. The inability to correlate sequences present within a single microbe prevents the association of distinct biosynthetic pathways that interact to form important phenotypes that can impact the global

ecology of the system. Moreover, in such analyses, the genes of rare microbes are difficult to detect since they tend to be swamped out by the sequences of the off-target microbes that greatly outnumber them <sup>7,8</sup>. This makes studying low-abundance microbes with interesting phenotypes particularly difficult.

One strategy for obtaining the genomic sequences of rare microbes in a diverse population is target enrichment. In this approach, fragments of the genomes of the target microbes are recovered by hybridization to capture probes. Sequence complementarity between the probes and targets allows the molecules to anneal, so that the target fragments can be recovered via probe enrichment <sup>9,10</sup>. A limitation of probe capture, however, is that recovering whole microbial genomes requires hundreds or thousands of overlapping capture probes <sup>11</sup>, necessitating substantial knowledge of the target sequence, which may not be available. Moreover, even when capture probes can be designed, the fragments captured are limited to those near the sequences targeted by the probes, biasing what can be detected by what is already known; this precludes recovery of whole genomes in many instances and, thus, prevents complete genetic characterization of the species of interest. This is particularly problematic when horizontal transfer of genetic elements occurs because, in such unpredictable instances, these sequences are not known to exist in the species of interest and, thus, it is not possible to construct probes with which to capture them. Horizontal gene transfer is an important method by which microbes transfer genetic information and generate phenotypic diversity <sup>12</sup>, which is why detecting such events is essential for increasing our understanding of microbe evolution.

To overcome the limitations of probe hybridization capture, a superior method would be to label the target microbes with a specific reporter; the labeled cells could then be recovered, together with their whole genome, using ultrahigh-throughput fluorescence-activated cell sorting

(FACS). One method for accomplishing this is to chemically fix and permeabilize microbes and then bind to their nucleic acids probes labeled with fluorescent dyes; the then fluorescent cells can be sorted with FACS, a method known as fluorescence in-situ hybridization, fluorescence-activated cell sorting (FISH-FACS)<sup>13</sup>. FISH-FACS has enormous benefits over probe hybridization capture because it allows cultivation-free enrichment of the whole genome of the microbe of interest. However, FISH-FACS also has drawbacks that significantly limit its applicability for sequencing microbes. For example, fixation can chemically modify DNA and introduce sequencing bias and errors into the genomes recovered, yielding poor sequence data<sup>14,15</sup>. More importantly, achieving bright, specific labeling of the cell type of interest requires substantial trial-and-error optimization of the fixation and permeabilization procedure, something that is not possible when seeking to recover the genome of a cell that cannot be cultured in the lab. This is challenging when screening natural samples containing large numbers of different microbes with distinct cell wall and membrane properties<sup>16</sup>. Consequently, while FISH-FACS holds enormous utility for the *in situ* identification of nucleic acid sequences in uncultivable microbes, it has drawbacks which limit its routine use for sequencing purposes. To enable the robust whole-genome sequencing of rare, uncultivable microbes, a new method for enriching intact microbial genomes out of a diverse ecosystem is needed.

In this paper, we introduce PCR-Activated Cell Sorting (PACS) for the cultivation-free enrichment of rare microbial genomes. In PACS, microbes from a diverse ecosystem are individually encapsulated in picoliter-volume aqueous droplets and subjected to TaqMan PCR, interrogating them for the presence of specific nucleic acid sequences. If the sequences are present, TaqMan amplification yields a bright fluorescent signal that fills the droplet encapsulating the cell, allowing us to recover the cell's whole genome by sorting the droplet.

PACS has a number of advantages for recovering rare microbial genomes. Because it utilizes PCR to identify the microbes of interest, harsh lysis procedures that include temperatures near the boiling point of water can be used, reducing erroneous identification compared to methods that rely on room-temperature probe hybridization. In addition, because these reactions are performed on single cells using ultrahigh-throughput microfluidics, millions of cells can be screened in hours, making the approach well adapted for recovering rare microbial genomes in a large and diverse sample. PACS thus realizes the potential of FISH-FACS for targeted metagenomics while being more robust and simpler to implement, since it does not require fixation of the cells and, rather than relying on room-temperature probe hybridization, relies on robust, easy to optimize, and straightforward to target TaqMan PCR.

### 3.3 Materials and methods

#### *Microfabrication of devices*

Fluidic chips are fabricated using standard photolithography techniques in poly(dimethylsiloxane) (PDMS)<sup>34</sup>. To produce a master, we first spin a layer of SU-8 photoresist (Microchem) onto a silicon wafer, and then expose the photoresist to UV light from a Blakray device under a mylar mask (Fineline Imaging). The wafer is then baked at 95°C on a hotplate for 1 min and then developed in Propylene glycol monomethyl ether acetate (PGMEA). We then pour PDMS polymer and crosslinker mixed in a 11:1 ratio over the master and then bake it at 75°C for 4 hours. The device is then peeled from the master and holes are punched using a 0.75mm biopsy coring needle. After that, the device is bonded to a glass slide following oxygen plasma treatment. To make the device channels hydrophobic, Aquapel is flushed into the



channels, after which the device is baked in an oven for 20 mins at 65°C. For all the devices (DEP and flow focusing) used in this paper, the thickness of the photoresist was maintained at 25µm while the channel widths at the flow-focusing junctions were 20µm.

### *Bacterial strain construction and growth*

The parental wild type strain is BW25113<sup>35</sup>. The entire lpoA ORF was deleted and replaced with a sacB-cat cassette using lambda Red recombinase-mediated allelic exchange (33). The Red recombinase was expressed using plasmid pKD46. The sacB-cat cassette was generated by PCR using plasmid pDS132<sup>36</sup> as template and primers ANG188 and ANG189. Transformants were selected on LB Cam10 and verified by diagnostic PCR.

Next, the mutant lpoA allele was generated by two-step overlap-extension PCR. The first-round PCR products were generated using primers ANG065 and AG4 together with AG3 and ANG066, with BW25113 genomic DNA as template. The PCR products were treated with DpnI and gel-purified to get rid of the initial template DNA. The final PCR product was generated using primers ANG065 and ANG066, with the first-round PCR products as template (present in equimolar amounts). This PCR product was used to replace the sacB-cat cassette as above, with selection on LB 0% NaCl 7% (w/v) sucrose. The sacB gene confers sucrose sensitivity, allowing counterselection. Transformants were screened for chloramphenicol sensitivity (indicating loss of cassette) and verified by diagnostic PCR and sequencing. The strain produced is the LpoA K168A E. coli.

A  $\Delta$ tolA::kan insertion was introduced into wild type strain BW25113 by sequential P1 transductions, with selection on LB with 10 mM sodium citrate and ampicillin at 50 µgml<sup>-1</sup>, and LB with 10 mM Na citrate and kanamycin at 30 µgml<sup>-1</sup>, respectively. The  $\Delta$ tolA::kan allele is

from the Keio collection E. coli gene knock-out library<sup>35</sup>. The strain produced has TolA knocked out but with a wild-type copy of the BW25113 LpoA gene.

The bacteria are grown in 2% Luria-Bertani (LB) broth at 37°C for around 10 hours. The bacterial cultures are then assayed for their optical density (OD) via spectrophotometrical measurement of absorption at 600nm. The correlation between OD and bacterial number is taken to be that 1 OD is equivalent to  $5 \times 10^8$  bacteria.

#### *Primer sequences for the construction of mutant bacterial strains*

The primer sequences used for the construction of the bacterial strains are as follows:

ANG188 5'-

TGCCGATTTAATATTGAGCATTGCGTAAAAAAAATATCACTGGATACATTGCCCGTA  
GTCTGCAAATCC-3' (50 bp upstream of lpoA and forward sacB-cat cassette primer; the 50 bp

upstream of lpoA allows homologous recombination to replace the gene), ANG189 5'-

CAGCCAGCGACGCGCTTGTGCTTCCCACGCATCGCCGGTCTGTTTGGTGGCCATGAC  
CCGGGAATTACG-3' (50 bp downstream of lpoA (reverse-complement) and rev sacB-cat

cassette primer), ANG065 5'-CGCAAACAACCGGGCATTAATC-3' (forward upstream lpoA  
primer, anneals 256 bp upstream of lpoA),

ANG066 5'-TTTGCTGCGGGTCACACTG-3' (reverse downstream lpoA primer, anneals 209

bp downstream of lpoA), AG3 5'-gctgcttgcgcgGCagaaaaacagcag-3' (forward lpoA(K168A)

mutagenesis primer; upper-case letters represent changes for lpoA(K168A) point mutation),

AG4 5'-ctgctgttttctGCcgcgccaagcagc-3' (reverse lpoA(K168A) mutagenesis primer).

### *Encapsulation of bacteria in monodisperse droplets*

Before mixing bacteria together with the other components of the reaction, the bacterial suspension is washed 3 times by centrifugation at 3000 rpm (Eppendorf) followed by resuspension of the pellet in distilled water. The bacteria is mixed together with primers, Taqman probe and PCR mix (2X ddPCR MasterMix, Biorad). The primers and Taqman probe are used at a working concentration of 1 $\mu$ M and 250nM respectively. This mix is loaded into a 1 ml syringe back-filled with HFE-7500 oil, which was connected to a coaxial flow-focus device. The oil used for the carrier phase was the droplet generation oil for probes (Bio-rad). The oil flow rate is run at 400  $\mu$ lhr<sup>-1</sup> while the aqueous flow rate is set at 200  $\mu$ lhr<sup>-1</sup>. The total handling time for bacteria is ~20 minutes. The emulsion is collected into PCR tubes and thermocycled on a T100 thermocycler (Bio-Rad), with the following conditions: 10 min at 95°C, 35 cycles of 10 s at 95°C, 15 s at 55°C and 30 s at 70°C. To verify that the PCR reactions were specific, both bacterial samples were electrophoresed on a 2% agarose gel. No non-specific product was observed after imaging.

### *Primer and probe sequences for Taqman PCR, LpoA amplification and sequencing*

The primers for the detection of TolA are: TolA Forward 5'- GTTGATTCAGGTGCGGTAGTT-3', TolA Reverse 5'- GCCTGCTGTTTCCTTCATCTT-3'. The TolA probe sequence is 5'- /6-FAM/ATCAAACCT/ZEN/GAAGGTGGCGATCCC /3IABkFQ/-3'. The primers for LpoA amplification are: LpoA Forward 5'- TTTACTGCGCGCGTTAATTG-3', LpoA Reverse 5'- TTGCGGCTGAGGTTGTT-3'. The

primer for TOPO Vector sequencing is : M13 Forward (-20) 5'-GTAAAACGACGGCCAG -3'.

All probes and primers are from IDTDNA Technologies.

### *DEP sorting*

Thermocycled drops are collected into a syringe filled with HFE-7500 (3M) fluorinated oil, and are left to cream for 10 minutes before starting the syringe pump. The drops are then reinjected into the DEP device at a flow rate of  $50 \mu\text{hr}^{-1}$ , with the spacer oil flow rate set at  $1000 \mu\text{hr}^{-1}$ . The flow rate for the 2<sup>nd</sup> oil spacer at the sorting junction is set at  $100 \mu\text{hr}^{-1}$ . All the oil used for spacing droplets is HFE-7500. The moat is filled with 2M NaCl salt solution, as are the salt electrodes. The PMTs are connected to a computer with LABVIEW software and a FPGA data acquisition card (National Instruments) for droplet fluorescence intensity recording and electrode activation. Custom LABVIEW software is written to enable dynamic adjustments of PMT gain, droplet fluorescence intensity thresholds for sorting, electrode AC voltage pulse frequency and magnitude. The data acquisition rate for this system is 200 kHz.

### *LpoA sequencing verification*

Droplets from the positive DEP sort are collected into 1.5ml Eppendorf tubes. Chloroform (Sigma-Aldrich) and distilled water are pipetted over the oil, with  $20\mu\text{L}$  of water used for every  $200\mu\text{L}$  of chloroform and  $200\mu\text{L}$  of oil. The droplets are then vortexed for 10 minutes on a shaker, and then centrifuged at 14,000rpm. The top layer of immiscible water is then extracted, of which  $9\mu\text{l}$  is used for PCR amplification. The PCR amplification mixture consists of  $1\mu\text{M}$  forward and reverse LpoA sequencing primers, 1X Toptaq PCR master mix (Qiagen), and template from the broken drops in a total volume of  $20\mu\text{L}$ . The mixture is then

thermocycled with the following conditions: 10 min at 95°C, 35 cycles of 10 s at 95°C, 15 s at 55°C and 30 s at 72°C. The PCR product is then cloned into a pCR4-TOPO vector (Life Technologies) using a TOPO TA cloning kit for sequencing (Life Technologies), as following the manufacturer's instructions. This is transformed into electrocompetent *E. coli* TOP10 bacteria and streaked onto LB plates with 50µgml<sup>-1</sup> kanamycin for growth at 37°C overnight. Colonies are picked at random for overnight growth in LB with 50µgml<sup>-1</sup> kanamycin at 37°C, DNA extracted using a Qiagen miniprep kit, and then sent for Sanger sequencing (Quintara Biosciences). The primer used for sequencing is the M13 Forward (-20) primer.

### 3.4 Results and discussion

#### *The strategy of PACS*

The motivating concept of PACS is to enable ultrahigh-throughput sorting of cells based on the sequences contained within their genome. We accomplish this by utilizing TaqMan PCR, which has known advantages of sensitivity and specificity. Upon exponential amplification of target DNA from single-copy genomic material, a bright signal is generated in the droplets containing the target cells, allowing robust discrimination and recovery of these cells within a large population of off-target cells. Additionally, TaqMan PCR enables specific identification of microbial genotypes by allowing for multiplexing, which enables simultaneous interrogation of several genomic regions in the same microbe.

The benefits of PCR have been recognized previously and implemented into microchamber devices for single cell analysis. It has enabled the amplification of two distinct sequences from individual microbes for sequencing<sup>17</sup>, and the association of specific viruses

with their bacterial hosts <sup>18</sup>. A limitation of microchamber methods, however, is that the number of single cell reactions that can be performed is limited by the number of microchambers that can be fabricated. In addition, in microchamber methods, recovering the genomes of target microbes requires individually accessing the positive chambers using tedious micromanipulation techniques <sup>17,18</sup>, limiting the ability of the method for screening and recovering large numbers of cells.

In PACS, we utilize microfluidic droplets in place of microchambers. Compared to microchambers, droplets are more scalable and enable the individual PCR analysis and sorting of many more cells. This is because, with microfluidic devices, droplets can be generated and sorted at kilohertz rates and each droplet utilizes just tens of picoliters of reagent, allowing millions of PCR reactions on single cells with microliters of total reagent <sup>19</sup>. Moreover, because the aqueous droplets are suspended in an inert liquid oil, they can be flowed through microfluidic devices, which allows multiple steps of processing, such as sampling fluid from, adding reagents to, and incubating and sorting the droplets; this allows multistep reactions in the droplets not possible in sealed microchambers <sup>19-21</sup>.

In PACS, small genomic regions hundreds of bases in length serve as “sequence biomarkers” to identify cells of interest. Based on the PCR signal produced when a cell containing the sequence biomarker is present, droplet sorting is used to recover the entire genome of the cell. This sorting can be accomplished microfluidically, as we demonstrate here, or using double emulsification and FACS <sup>22</sup>, which provides the benefit of allowing single genomes to be dispensed into individual wells.

In contrast to conventional antibody labeling methods and FACS sorting, PACS is easier to implement and quicker to optimize, because the detection of the microbe depends on the

efficiency of robust PCR reactions rather than on antibody-epitope binding. Moreover, PACS is more easily targeted at different identifying biomarkers, since this requires only the design of different probe sequences, compared to the development of a new antibody. Another benefit of PACS is that it can utilize either PCR or RT-PCR<sup>23,24</sup> to analyze cell DNA or RNA; this allows differentiation between cells merely containing a pathway in their genome and those that are actively expressing it. Similarly, because PACS derives its sensitivity from PCR, it is capable of detecting individual copies of the target sequence in the cell, making it applicable to sorting based on sequences in the genome and transcriptome, or even plasmid or viral sequences.

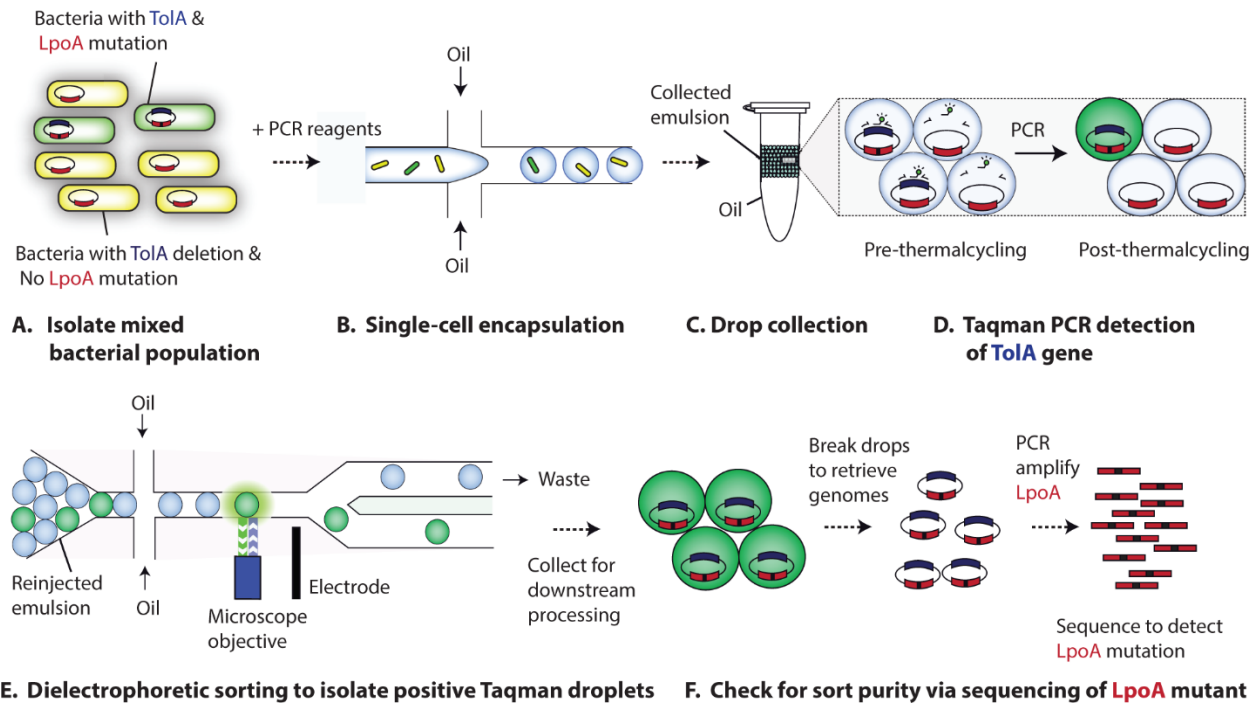


Fig 3.1. The PACS workflow applied to a model microbial system. A microbial sample consisting of K-12 E coli harboring wild type *TolA* and a spike-in variant ( $\Delta TolA$ ) is created from growth cultures (Fig 3.1A). This sample is then encapsulated together with PCR reagent to form a single emulsion (Fig 3.1B). This emulsion is then collected and thermocycled, with PCR-positive droplets experiencing an increase in Taqman fluorescence (Fig 3.1C). This emulsion is then DEP sorted for bright drops (Fig 3.1D), and these drops are ruptured to release genomic content which is sequenced to verify sorting efficacy (Fig 3.1E).



### *Targeted recovery of microbial genomes with PACS*

The first step in PACS is to encapsulate the microbes (Fig. 3.1A) in individual water-in-oil droplets using microfluidic emulsification (Fig. 3.1B); here, we include the PCR reagent in the microbial suspension, although the microbes and PCR reagent, which can include detergents to enhance lysis, can also be combined on-chip via laminar co-flow followed by droplet generation. The microbes are diluted so that there is on average less than one per droplet, loading the droplets randomly in accordance with a Poisson distribution. The droplets are ~35 pL in volume, although this can be varied over >5X up or down, and are collected into a PCR tube and thermocycled (Fig. 3.1C). We perform the thermocycling on a standard PCR machine, although on-chip thermocyclers could also be used for an unbroken workflow<sup>25</sup>. During PCR, the elevated temperature lyses the microbes and disrupts DNA-protein and DNA-DNA interactions, providing the PCR primers with access to the cell's DNA. Droplets containing the genetic sequences being assayed for will result in TaqMan PCR amplification, yielding a droplet that is bright with fluorescence at the emission wavelength of the TaqMan probe due to its degradation by the 5' exonuclease activity of Taq polymerase<sup>26</sup>. At this point in the process, we have millions of droplets, some of which are now fluorescent because they contain a microbe with the sequence targeted by our assay. The next step is to screen these droplets using ultrahigh-throughput dropometry and recover the positives with dielectrophoretic (DEP) sorting (Fig. 3.1D)<sup>27</sup>. The sorted droplets can be loaded into individual wells or pooled together and chemically ruptured to access their contents, providing us with the genomic DNA of the target microbes. Amplicons mixed together with the sorted genomes can be removed by a variety of methods, such as Uracil-DNA Glycosylase enzymatic digestion, if using a PCR mix with uracil, or with streptavidin beads that can extract biotin-labelled amplicons. Downstream transcriptomic

analysis might necessitate the development of lysis and isothermal PCR methods that do not require high heat, as RNA is known to be heat-sensitive.

### *Validation of PACS*

To validate that PACS can recover specific microbes out of a mixed population, we spike target cells into a background of non-target cells. We use two different *E. coli* strains that have two differences in their genome: The first strain has the genetic sequence for the membrane protein TolA knocked out ( $\Delta$ TolA), whereas the second has TolA intact but is a double mutant on the LpoA gene, which is an outer membrane lipoprotein (LpoA K168A)<sup>28</sup>. The mixed population is then run through our PACS workflow (Fig. 3.1) sorting based on the presence of TolA, which should only recover the LpoA double mutants. To characterize the efficiency of the PACS sorting, we recover the genomic DNA of the sorted microbes and PCR-amplify and sequence the portion of the LpoA gene containing the mutations (Fig. 3.1E). By comparing the number of sequences containing the double mutant and those absent of it, we are able to estimate the efficiency with which PACS can discriminate between these cell types based on TolA. That is, the objective of this experiment is to show that we can differentiate between cells based on the presence of a gene (TolA) at one location of the genome, and then confirm correct sorting by analyzing a different gene (LpoA) far away on the same genome. Importantly, the sequences analyzed post-sorting are not the product of the first PCR; they are present in the sorted mixture only because they existed in the same genome that contained TolA and, thus, were sorted with it.

### *Efficiency of single cell droplet PCR*

The ability to PACS sort the microbes is thus contingent on the ability to specifically discriminate between the two cell types with our TaqMan assay. To investigate the specificity of the TaqMan assay, we perform control experiments in which we emulsify clonal populations of the two cell types separately, and then analyze the cells using droplet single cell TaqMan PCR with primers and probes for the TolA gene, as shown in Fig. 3.2A. For the droplets containing the double mutants (LpoA K168A), in which TolA is present, we observe a “digital” fluorescence signal, in which a small fraction of the droplets are bright, and the remainder exhibit no fluorescence, as illustrated in Fig. 3.2A, *upper*; the fluorescent droplets contain individual K168A microbes, while the dim droplets are devoid of any cells and thus constitute what we expect to see when the target sequence (TolA) is not present within the droplet. To confirm this, we perform the same experiment with the knockout population ( $\Delta$ TolA), the results of which are shown in Fig. 3.2A, *lower*. As expected, even though the stoichiometry of the  $\Delta$ TolA cells is comparable to that of the K168A cells in the first experiment, so that we expect similar loading rates into the droplets, we observe no fluorescent droplets; this demonstrates that our TaqMan assay is specific to cells that have sequences targeted by our primers. This is consistent with control experiments performed in bulk on large numbers of the cells and also with the properties of TaqMan PCR. To validate that, indeed, the positive droplets in the K168A experiment correspond to “digital” amplification resulting from a TolA positive cell, we repeat the experiment for different concentrations of K168A cells. For Poisson loading of the cells in droplets, the probability that a given droplet has  $x$  cells is given by,

$$P(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad , \quad (1)$$

where  $\lambda$  is the average number of cells per 35 pL droplet (Fig. 3.6). Bright drops correspond to  $x \geq 1$ , whereas  $x = 0$  relates to dark drops. The proportion  $p$  of bright to dark drops depends on  $\lambda$  according to,

$$p = 1 - e^{-\lambda}. \quad (2)$$

This is a simple statement that as the concentration of cells in suspension increases, more of the droplets contain at least one cell. To relate the number of cells in the droplets to the number of fluorescent droplets observed at the conclusion of the assay, we must account for the fact that not all droplets containing single cells undergo amplification. That is, due to inefficiencies in the PCR, the probability that the reaction undergoes amplification is less than unity. We can account for this by rewriting the equation as

$$p = 1 - e^{-k\lambda}, \quad (3)$$

where  $k$  is the probability that a droplet containing a target cell yields a fluorescent signal. To measure,  $k$ , an important parameter that describes the sensitivity with which we detect positive cells, we repeat the experiment at different concentrations, Fig. 3.2B. For  $k = 1$ , the TaqMan reaction can be said to be perfectly efficient so that every drop containing a cell yields a fluorescent signal. For  $k < 1$ , the reaction is imperfect so that some droplets containing positive cells do not yield a fluorescent signal. Based on our data, we determine  $0.6 < k < 0.7$ , indicating

that we detect approximately 65% of the positive cells in the sample. This inefficiency may be a consequence of the natural stochasticity of PCR, particularly in picoliter volumes in which reagents may be limiting. Another explanation is that cell lysis is not perfectly efficient, as heat lysis might not be sufficient to lyse all the bacteria. This effect can be mitigated by including PCR-compatible detergents in the droplets, which aid cell lysis and solubilization of DNA targets and may improve single cell PCR efficiency. Using more sophisticated multistep microfluidic techniques, it is also possible to include PCR incompatible lysis reagents, such as alkaline buffers, lysozyme, or proteases, to enable lysis of particularly durable microbes. Further development of lysis cocktails compatible with our workflows will be essential towards the unbiased lysis of unknown microbes in environmental samples.

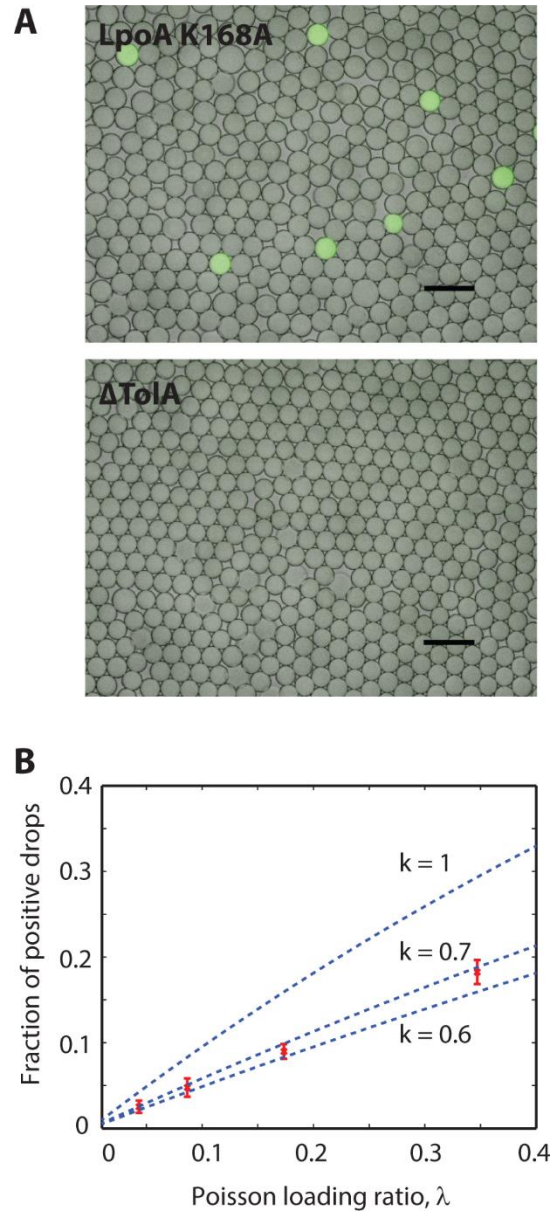


Fig 3.2. Taqman PCR detection of TolaA gene on E. coli bacteria. E coli bacteria are encapsulated with PCR reagents in droplets and are thermocycled. Fig 3.2A, *upper*, Drops containing bacteria with the TolaA gene are bright, whereas this is absent in Fig 3.2A, *lower*, which has E coli without this gene. Fig 3.2B shows the dependency of the fraction of loading drops which are bright versus the poisson loading ratio. The different curves represent different calculated curves if the E. coli lysis factor  $k$  was varied.

### *Recovery of whole bacterial genomes with droplet sorting*

At the conclusion of our single cell droplet PCR, we have a collection of millions of droplets, some of which contain target microbes and are fluorescent. To recover the positive droplets and the genomes of the cells they contain, we use ultrahigh-throughput dielectrophoretic (DEP) droplet sorting. DEP droplet sorting has previously been employed to robustly sort droplets at several kilohertz for applications like enzyme evolution<sup>29</sup> and the detection of high-affinity antibodies<sup>30</sup>. Our droplet sorter consists of a droplet reinjection inlet, a spacing inlet, and a sorting junction. The device is surrounded by conducting aqueous “moats” that shield the injected droplets from stray electric fields, which can unintentionally coalesce droplets. These “moats” are designed so that they surround the oil channels, as illustrated in Fig. 3.3.

Upon injection into the device, the thermocycled droplets are close packed and spaced by oil in the spacing junction, as shown in Fig. 3.3, *left*. Spacing ensures that the droplets pass the detection region (Fig. 3.3, *middle*) one at a time, so that the fluorescence of each droplet can be measured individually. It also ensures that the droplets do not crowd the sorting junction (Fig. 3.3A, *right*), which can result in droplet collisions that interfere with controlled sorting. After spacing, the droplets pass through the detection region (Fig. 3.3, *middle*) and pass through a focused laser beam; the laser excites the fluorescent dyes in the droplets, causing them to emit light in proportion to the amount of cleaved TaqMan probes they contain. Droplets that underwent successful TaqMan amplification emit bright fluorescent light, while those that did not appear dim. The fluorescent light is captured by the objective of a microscope, filtered through dichroic mirrors and bandpass filters, and focused onto the sensor of a photomultiplier tube (PMT). The PMT outputs a voltage proportional to the intensity of the fluorescent light. The oil surrounding each droplet is not fluorescent; hence, when a droplet passes through the

detection laser, the PMT records a peak as a function of time, as shown in Fig. 3.4A; each peak in the time trace corresponds to a distinct droplet. The amplitude of a given peak is proportional to the intensity of the droplet, allowing bright TaqMan positive droplets to be differentiated from dim TaqMan negative droplets, as illustrated by the bright droplet at  $t = 32.5$  ms.

To recover the bright droplets, we set a threshold voltage to 0.12V; this value varies between runs depending on the focusing optics and PMT gains and is selected in this experiment to cleanly distinguish between positive and negative droplets, as shown in Fig. 3.4A. Above this value, the computer is instructed to sort the droplet, which it does by outputting an alternating current (AC) pulse that is amplified to ~1500 V and applied to a conducting aqueous electrode in the sorting junction, as illustrated in Fig. 3.3, *middle*. Energizing the electrode generates an electric field that polarizes the droplet in the sorting junction; this produces a dielectrophoretic attraction that pulls the droplet towards the electrode<sup>31</sup>, deflecting it into streamlines that carry it into the collection channel. When the electrode is not energized, the geometry of the sorting junction is designed so that the droplet follows streamlines that carry it into the waste channel. Hence, by selectively energizing the electrodes based on the measured fluorescence of the droplets, we are able to recover the TaqMan positive droplets and discard the negative droplets, as can be seen by comparing the images in Fig. 3.4B. Other sorting methodologies such as valve-based sorting<sup>32</sup>, surface acoustic waves<sup>33</sup>, and double emulsion FACS<sup>22</sup> could also be used to recover the positive droplets.



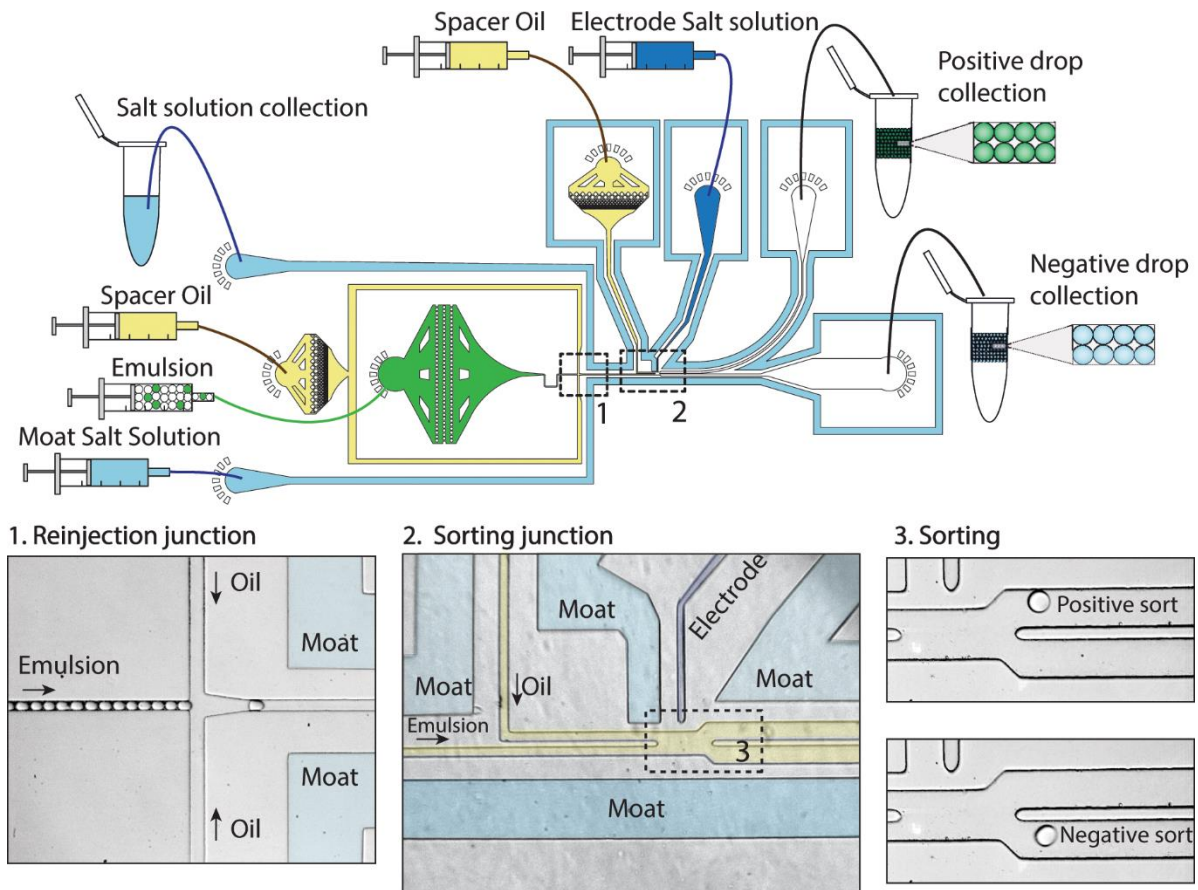


Fig 3.3. DEP droplet sorting device. Fig 3.3 , *upper*, shows the device layout, with the salt “moat” insulating the drops from any stray electric fields potentially originating from the salt electrode. This device consists of the reinjection junction, Fig 3.3, *left*, at which the reinjected emulsion is spaced out, as well as the sorting junction, Fig 3.3, *middle*, which is where detection and sorting occurs. Fig 3.3, *right*, shows positive and negative droplet sorting events.

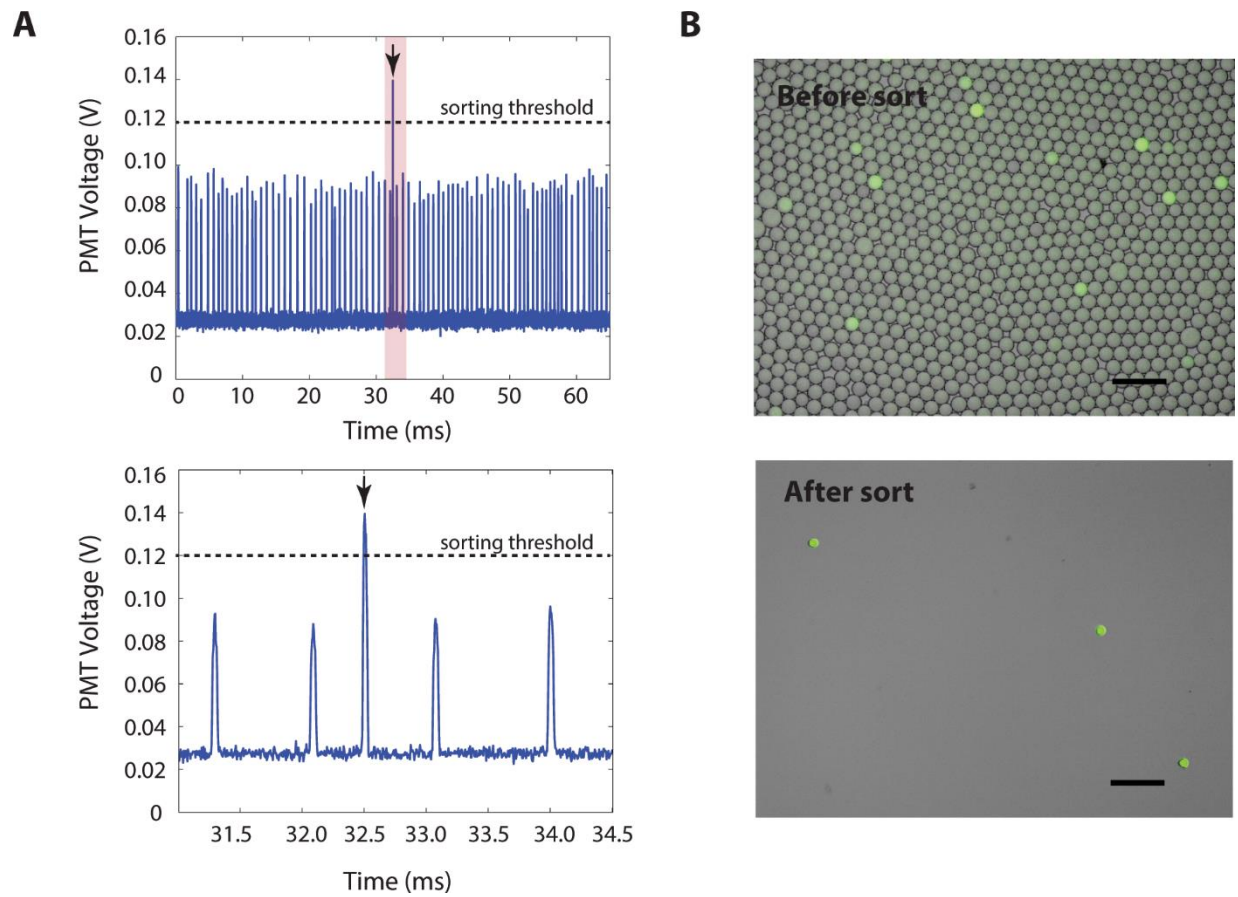


Fig 3.4. Droplet detection and sorted drops. Fig 3.4, *left*, is the PMT timetrace of recorded signals from the optical droplet detection setup. There is a clear peak at 32.5 ms, which corresponds to a bright drop that is sorted. Fig 3.4, *right*, are the fluorescence images of thermocycled drops before and after DEP sorting. Scale bars are 100  $\mu\text{m}$ .

### *Sequence verification of sorted genomes*

Epifluorescence microscopy images like those of Fig. 3.4B demonstrate that the dielectrophoretic sorter is capable of accurately sorting the bright from the dim drops. To further validate that PACS enables accurate single cell sorting based on nucleic acid sequences, we recover the genomic DNA from the positively sorted droplets for Sanger sequencing. The sorted droplets are chemically ruptured with the addition of chloroform and application of mechanical shear, and the microbial genomes dispersed into aqueous buffer. The K168A cell line has a double amino acid mutation of “AAA” encoding lysine to “GCA” encoding alanine, as seen in Fig 3.5A. Even if there are errors in the PCR preparation or the Sanger sequencing, these are expected to be rare and, extremely so, for 2 consecutive nucleotides, providing a high confidence read out with which to validate the PACS sorting. We tested two mix ratios of the  $\Delta$ TolA and TolA bacteria, one where the mutant was present at 20% in the total population, and the other at 1%. For the 20% spike-in, 5 of 10 sequences before PACS were positive for the mutant, whereas 9 of 10 were positive after PACS, as shown in Fig. 3.5A. The high pre-PACS frequency of the mutant is likely the result of random sampling variation, since we sequenced only ten molecules. Similarly, the 1% spike-in yielded no pre-PACS positives in our ten molecule sample, while the post-PACS library was again 9 of 10. Thus, for both spike-in ratios, we observe a reasonable number of mutants pre-PACS and very few mutants post-PACS. For these two samples, the droplets were sorted at 1 kHz for 1 hour each, and at a cell loading ratio of around 0.1 cells per droplet. This translates to 3.6 million drops put through the sorter, with around 360,000 cells processed per sequencing run.

The deviation from 100% mutant sequence post-PACS may result from multiple sources. An unlikely source is library preparation or sequencing errors, since these should occur very

rarely. Another possibility is erroneous sorting of the droplets. This can be due to malfunction of the sorting device, which is also rare for the conditions used or, more likely, from droplets coalescing during the thermocycling. Unintended droplet coalescence allows the contents of multiple droplets to mix together, so that a droplet with a positive signal can have added to it, by way of merger, the genomic DNA of a negative cell, thereby resulting in contamination of the sorted material. While we do observe some merger during thermocycling, it is also rare. The most likely source of the error is multiple encapsulations while loading the cells into the droplets. Multiple encapsulations can occur for several reasons, such as the cells clumping together or two or more cells entering into the droplet generation junction at the same time. Multiple encapsulations are the primary events that limit our sorting accuracy, although they can be mitigated by handling cells carefully, filtering cells to remove clumps, and diluting the suspension to limit double encapsulations. In any case, even with the error, we are able to obtain an enrichment ratio of  $0.90/0.01=90X$  with one sort for the 1% spike in. If this enrichment is not sufficient, the system can be further diluted or additional rounds of PACS used to obtain high purity genomes.

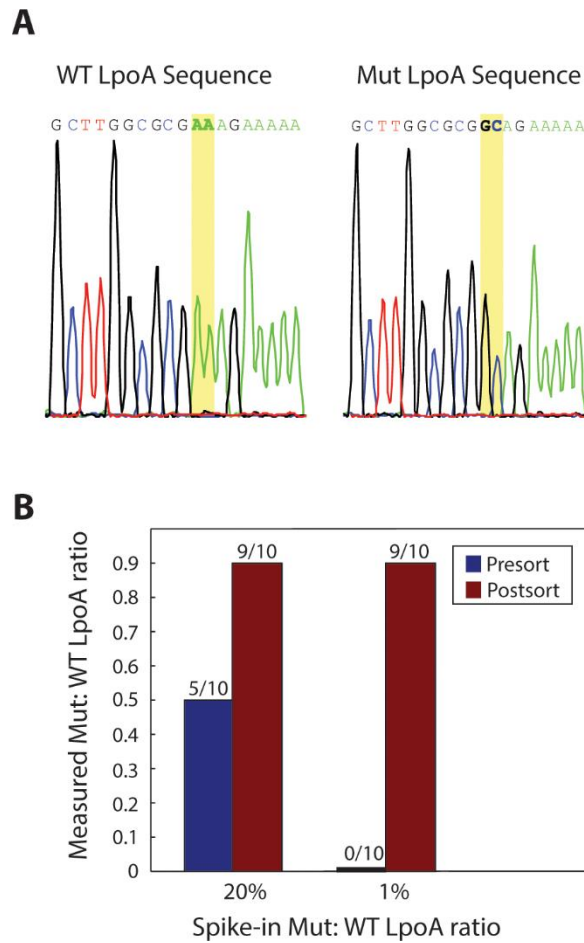


Fig 3.5. Sequencing verification and genome enrichment. Fig 3.5A is a representative electropherogram of the LpoA gene and its mutant counterpart after Sanger sequencing LpoA from sorted bacterial genomes. There are clear base calls on the double nucleotide mutation. Fig 3.5B shows the sequencing results, with clear enrichments for the TolA/ $\Delta$ LpoA bacterial strain for both spike-in ratios.

### 3.5 Supplemental Information

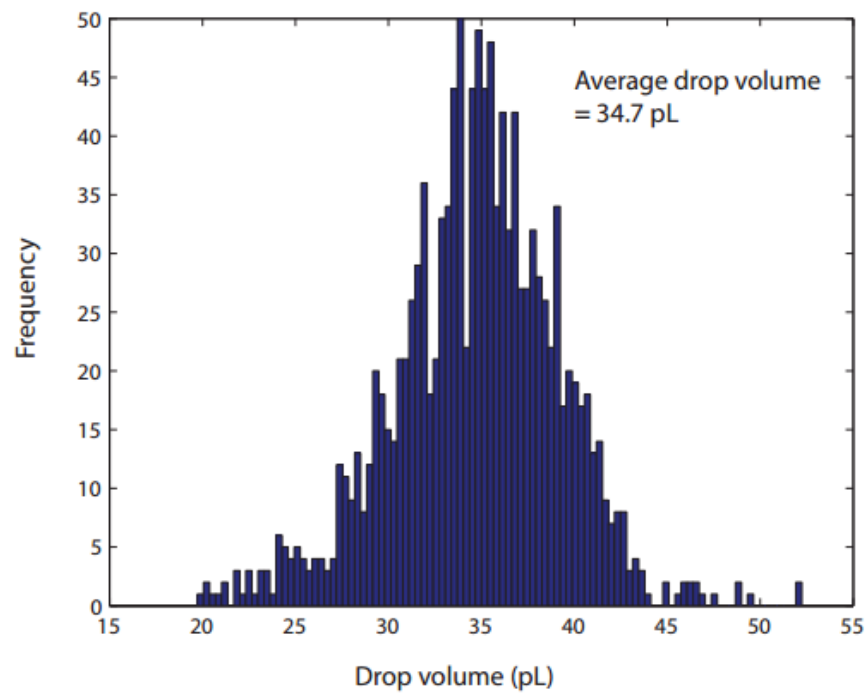


Fig 3.6. Droplet size distribution. The single emulsion droplet diameters was quantified using ImageJ, with a total of 1200 drops measured for all concentrations of bacteria. The average drop volume was calculated to be 34.7 pL.

### 3.6 Conclusion

PACS is a new cell sorting technology that has great utility for sorting microbes out of heterogeneous populations without the need for prior cultivation. PACS derives this power from its ability to directly sort cells off of sequence-specific nucleic acids; this has advantages over antibody labeling when the cell type is not known to express a unique surface marker. PACS preserves the genomic information from individual cells allowing them to be analyzed with downstream methods like next generation sequencing. This method can be used to search microbial “dark matter” for bacteria with certain postulated metabolisms, for example chemolithoautotrophy in the ocean. We have demonstrated PACS with bacterial cell genomes and believe that this method can also be applied to analysis of cellular transcriptomes in the future, which would open up opportunities for studying single-cell gene expression in complex ecosystems.

By utilizing droplet-based mammalian cell PCR methods, PACS can be applied to human health applications, such as, detecting and analyzing circulating tumor cells or characterizing heterogeneity in cultured stem cells and tumor cell populations. PACS should also enable direct sorting of viruses or naked DNA, capabilities that have never been available in the fields of virology and genomics. For example, this should enable populations of viruses and cells to be screened for metagenomic analysis or to detect the hosts of uncultivable viruses. We expect that PACS will impact genomics by allowing large fragment enrichment. By combining PACS with recently-developed methods for sorting microfluidic droplets using flow cytometry, PACS can also be used to load individual cells, viruses, and DNA molecules into microwells for whole-genome amplification and sequencing; this should aid single cell analysis and genetic haplotyping efforts.

### 3.7 References

1. C. S. Riesenfeld, P. D. Schloss and J. Handelsman, *Annu. Rev. Genet.*, 2004, **38**, 525–552.
2. P. C. Blainey, *FEMS Microbiol. Rev.*, 2013, **37**, 407–427.
3. C. Rinke, P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J.-F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W.-T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz and T. Woyke, *Nature*, 2013, **499**, 431–437.
4. S. R. Vartoukian, R. M. Palmer and W. G. Wade, *FEMS Microbiol. Lett.*, 2010, no-no.
5. V. Iverson, R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales and E. V. Armbrust, *Science*, 2012, **335**, 587–590.
6. M. B. Scholz, C.-C. Lo and P. S. Chain, *Curr. Opin. Biotechnol.*, 2012, **23**, 9–15.
7. M. Albertsen, P. Hugenholtz, A. Skarszewski, K. L. Nielsen, G. W. Tyson and P. H. Nielsen, *Nat. Biotechnol.*, 2013, **31**, 533–538.
8. Z. Taghavi, N. S. Movahedi, S. Draghici and H. Chitsaz, *Bioinformatics*, 2013, **29**, 2395–2401.
9. L. Mamanova, A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure and D. J. Turner, *Nat. Methods*, 2010, **7**, 111–118.
10. E. H. Turner, S. B. Ng, D. A. Nickerson and J. Shendure, *Annu. Rev. Genomics Hum. Genet.*, 2009, **10**, 263–284.
11. A. Gnirke, A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander and C. Nusbaum, *Nat. Biotechnol.*, 2009, **27**, 182–189.



12. H. Ochman, J. G. Lawrence and E. A. Groisman, *Nature*, 2000, **405**, 299–304.
13. M. G. Kalyuzhnaya, R. Zabinsky, S. Bowerman, D. R. Baker, M. E. Lidstrom and L. Chistoserdova, *Appl. Environ. Microbiol.*, 2006, **72**, 4293–4301.
14. S. Yilmaz, M. F. Haroon, B. A. Rabkin, G. W. Tyson and P. Hugenholtz, *ISME J.*, 2010, **4**, 1352–1356.
15. M. Srinivasan, D. Sedmak and S. Jewell, *Am. J. Pathol.*, 2002, **161**, 1961–1971.
16. A. Wendeborg, *Cold Spring Harb. Protoc.*, 2010, **2010**, pdb.prot5366-prot5366.
17. E. A. Ottesen, J. W. Hong, S. R. Quake and J. R. Leadbetter, *Science*, 2006, **314**, 1464–1467.
18. A. D. Tadmor, E. A. Ottesen, J. R. Leadbetter and R. Phillips, *Science*, 2011, **333**, 58–62.
19. T. M. Tran, F. Lan, C. S. Thompson and A. R. Abate, *J. Phys. Appl. Phys.*, 2013, **46**, 114004.
20. A. R. Abate, T. Hung, P. Mary, J. J. Agresti and D. A. Weitz, *Proc. Natl. Acad. Sci.*, 2010, **107**, 19163–19166.
21. S.-Y. Teh, R. Lin, L.-H. Hung and A. P. Lee, *Lab. Chip*, 2008, **8**, 198.
22. S. W. Lim and A. R. Abate, *Lab. Chip*, 2013, **13**, 4563.
23. D. J. Eastburn, A. Sciambi and A. R. Abate, *Nucleic Acids Res.*, 2014, **42**, e128–e128.
24. D. J. Eastburn, A. Sciambi and A. R. Abate, *Anal. Chem.*, 2013, **85**, 8016–8021.
25. N. R. Beer, B. J. Hindson, E. K. Wheeler, S. B. Hall, K. A. Rose, I. M. Kennedy and B. W. Colston, *Anal. Chem.*, 2007, **79**, 8471–8475.
26. P. M. Holland, R. D. Abramson, R. Watson and D. H. Gelfand, *Proc. Natl. Acad. Sci.*, 1991, **88**, 7276–7280.

27. K. Ahn, C. Kerbage, T. P. Hunt, R. M. Westervelt, D. R. Link and D. A. Weitz, *Appl. Phys. Lett.*, 2006, **88**, 24104.
28. A. Typas, M. Banzhaf, B. van den Berg van Saparoea, J. Verheul, J. Biboy, R. J. Nichols, M. Zietek, K. Beilharz, K. Kannenberg, M. von Rechenberg, E. Breukink, T. den Blaauwen, C. A. Gross and W. Vollmer, *Cell*, 2010, **143**, 1097–1109.
29. J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths and D. A. Weitz, *Proc. Natl. Acad. Sci.*, 2010, **107**, 4004–4009.
30. B. El Debs, R. Utharala, I. V. Balyasnikova, A. D. Griffiths and C. A. Merten, *Proc. Natl. Acad. Sci.*, 2012, **109**, 11570–11575.
31. A. Sciambi and A. R. Abate, *Lab. Chip*, 2014, **14**, 2605.
32. A. R. Abate, J. J. Agresti and D. A. Weitz, *Appl. Phys. Lett.*, 2010, **96**, 203509.
33. T. Franke, S. Braunmüller, L. Schmid, A. Wixforth and D. A. Weitz, *Lab. Chip*, 2010, **10**, 789.
34. Y. Xia and G. M. Whitesides, *Annu. Rev. Mater. Sci.*, 1998, **28**, 153–184.
35. K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci.*, 2000, **97**, 6640–6645.
36. N. Philippe, J.-P. Alcaraz, E. Coursange, J. Geiselmann and D. Schneider, *Plasmid*, 2004, **51**, 246–255.

## **Chapter 4: PCR-activated cell sorting as a general, cultivation-free method for high-throughput identification and enrichment of virus hosts**

The following section is reprinted from “PCR-activated cell sorting as a general, cultivation-free method for high-throughput identification and enrichment of virus hosts” by Shaun W. Lim, Shea T. Lance, Kenneth M. Stedman and Adam R. Abate. This article was published as a research paper in the *Journal of Virological Methods* on 29 Dec 2016. Shaun W. Lim, Kenneth M. Stedman and Adam R. Abate planned the experiments. Shaun W. Lim and Shea T. Lance did the experiments. Shaun W. Lim, Kenneth M. Stedman and Adam R. Abate wrote the manuscript.

### 4.1 Abstract

Characterizing virus-host relationships is critical for understanding the impact of a virus on an ecosystem, but is challenging with existing techniques, particularly for uncultivable species. We present a general, cultivation-free approach for identifying phage-associated bacterial cells. Using PCR-activated cell sorting, we interrogate millions of individual bacteria for the presence of specific phage nucleic acids. If the nucleic acids are present, the bacteria are recovered via sorting and their genomes analyzed. This allows targeted recovery of all possible host species in a diverse population associated with a specific phage, and can be easily targeted to identify the hosts of different phages by modifying the PCR primers used for detection. Moreover, this technique allows quantification of free phage particles, as benchmarked against the “gold standard” of virus enumeration, the plaque assay.

## 4.2 Introduction

Viruses substantially impact the health and dynamics of all communities, from causing disease in individuals to influencing global biogeochemical cycles<sup>1,2</sup>. Cyanophages, for example, play an important role in oceanic carbon fixation, since a core photosynthetic protein is of phage origin<sup>3,4</sup>. Moreover, there are an astronomical number of viruses in the biosphere<sup>2</sup>, but it is unclear what hosts the vast majority of these viruses infect. Studying the impact of environmental viruses necessitates methods for characterizing virus-host interactions. However, the challenges involved in growing heretofore “uncultivable” bacteria means that many of these potential hosts and their viruses cannot be studied in isolation<sup>5,6</sup>.

Next generation sequencing (NGS) is a powerful, culture-independent method for studying virus-host interactions in diverse populations<sup>7-10</sup>. In this approach, metagenomic sequences are interrogated for molecular traces left by the viruses co-occurring with host genes such as small subunit ribosomal DNA sequences. These viral traces are often resistance-conferring, such as CRISPR-associated sequences<sup>11,12</sup>. Identifying virus-host relationships using NGS is only possible, however, if such traces are identifiable. Moreover, even when they occur, these sequences are rare, making their identification challenging. Single cell genomics of uncultivated organisms is another powerful method for identifying virus-host pairs<sup>9,10</sup>. In one study, 127 single amplified genomes (SAGs) from a known clade of uncultivated organisms were screened for viruses and 69 new virus clades were found<sup>9</sup>. In another study, bioinformatic analysis of 58 SAGs allowed identification of 30 new virus genomes<sup>10</sup>. To isolate these organisms for SAG analysis, fluorescence-activated cell sorting (FACS) sorts individual cells into wells based on non-specific dyes or scattering properties; once in the wells, the cells can be identified with PCR or other sequence analysis. This limits throughput to identification of just hundreds of cells, making it

challenging to identify the hosts of rare viruses. A higher-throughput method for detecting infected cells is PhageFISH (Fluorescence *In Situ* Hybridization) which allows phage hosts to be separated based on virus infection by staining infected hosts with fluorescent probes complimentary to virus sequences<sup>13</sup>. PhageFISH is potentially high throughput since millions of hosts can be stained in parallel and sorted at >10,000 per second with FACS. However, probe hybridization is performed in the complex milieu of a fixed cell, necessitating substantial assay optimization extremely challenging with uncultivable organisms. Viral tagging, on the other hand, uses fluorescently labeled viruses to isolate cells that they are associated with by flow cytometry<sup>14</sup> or for identifying new viruses that infect specific hosts<sup>15</sup>.

A powerful, cultivation-free method for identifying specific phage hosts is single-cell PCR in individual microfluidic chambers<sup>16</sup>. In this approach, TaqMan PCR interrogates individual microbes for phage sequences such that, if the sequences are present, a fluorescent signal is produced. This allows identification and recovery of phage-associated microbial genomes by interrogating positive chambers. Because this method uses PCR, which yields exponential amplification when phage sequences are present, the signal difference between phage-associated and phage-free genomes is large, permitting unambiguous identification. Moreover, harsh PCR thermocycling facilitates bacterial lysis and enhances access of hybridization primers to their targets, reducing false-negatives. The principal limitation of this method is that it lacks scalability, enabling interrogation of just thousands of bacteria; this limits its utility for studying most environments, comprising billions of microbes per milliliter<sup>17</sup>. To allow comprehensive characterization of phage-host relationships, an optimal method would be applicable to uncultivable species, capable of analyzing millions of microbes, and enable the specific recovery of the genomes of all bacteria infected with the target phage.

In this paper, we present a general method for identifying phage-host relationships applicable to uncultivable organisms; the method combines the scalability of FACS with the generality and precision of single-cell PCR. Using PCR-activated cell sorting (PACS) we sort bacteria based on infection by specific phages<sup>18,19</sup>. Besides applying PACS to phage sequences in infected bacteria, we also show in this paper that certain bacteriophage viral particles can be directly quantified via droplet-based PCR. This is accomplished by isolating individual bacteria from a heterogeneous population in water-in-oil droplets and performing PCR in each droplet with primers specific to the target phage genome (Fig. 4.1). If the phage is present in- or bound to- the host, the PCR produces a fluorescent signal, allowing the host bacterial and phage genomes to be recovered by sorting the fluorescent droplets. The sorted genomes can then be analyzed to identify the host species. Importantly, the PCR performed in the droplets is used only as a means of identifying whether the phage is present with the bacteria; the material that is recovered and sequenced is not amplified and, thus, does not need to be known in order to be recovered and sequenced.

PACS has a number of advantages over other methods for characterizing phage-host relationships: It does not require that the target bacteria or phage be cultivable. It is ultrahigh-throughput, allowing millions of individual microbial genomes to be sorted based on phage association, permitting detection of bacteria infected with rare phages. The TaqMan assay can be multiplexed using probes of different color allowing, for example, identification of specific microbes in a sample containing multiple, specific phage sequences, such as co-infection by distinct phages. The sorting can be easily re-targeted against different species by modifying the PCR primers, while maintaining highly specific sorting. This requires minimal optimization, and as a direct comparison to antibody-based methods of cell identification, does not require access to

microbial antigens. PACS is particularly applicable to the identification of putative hosts for phages whose sequences have been identified in metagenomic studies. In a recent study attempting to computationally connect viruses to their hosts, metagenomic viral sequences were associated with pathogenic species like *Fusobacterium* and *Leptotrichia* that had heretofore not been associated with any viruses<sup>20</sup>. Our method can build on similar known short assembled segments of phage sequences to design primers that target specific hosts.

#### 4.3 Material and methods

##### *Preparation of bacteriophages, plaque assays and bacterial hosts*

Bacteriophage T4 (T4), bacteriophage  $\Phi$ X174 ( $\Phi$ X174), and *E.coli* hosts were obtained from Carolina Biological Supply. T4 was propagated by infection of *E.coli* B (ATCC 11303)<sup>21</sup> and  $\Phi$ X174 by infection of *E.coli* C (ATCC 13706). Bacteriophage lambda (lambda cI857<sup>ts</sup>) was obtained from the lambda lysogen *E.coli* strain KL470 provided by R. Raghavan. Bacteriophage lambda was propagated by infection of *E.coli* C600 (Carolina) and plaques formed as previously described<sup>22</sup>. A lambda lysogen of lambda cI857<sup>ts</sup> in *E.coli* C was prepared<sup>22</sup> and purified by two rounds of single colony isolation on LB agar. We use a derivative of *E.coli* strain BW25113, containing mCherry as an uninfected control<sup>23</sup>. Titers of T4,  $\Phi$ X174 and lambda were determined with plaque assays as previously described<sup>24</sup>. Bacteriophage-containing lysates were separated from cellular debris by 5 minutes of centrifugation at 3000g, filtered through 0.2  $\mu$ m filters (Sartorius, Minisart) and preserved with a single drop of chloroform in the preparation.

### *Microfabrication of devices*

The microfluidic chips are fabricated using standard soft lithography techniques in poly(dimethylsiloxane) (PDMS)<sup>25</sup>. To fabricate a device master from which the PDMS replicates are molded, SU-8 photoresist (MicroChem) is spun onto a 3" silicon wafer at a thickness of 25  $\mu\text{m}$ , and exposed to UV light from a UV photodiode (ThorLabs) through a UV-absorbent Mylar mask containing an inverse-image of the microfluidic chip (Fineline Imaging). The wafer is baked at 95°C on a hotplate for 1 minute, and developed in propylene glycol monomethyl ether acetate (PGMEA) to remove uncrosslinked resist, followed by post-baking in accordance with the manufacturer's instructions. PDMS polymer and crosslinker is combined at a ratio of 11:1, poured over the master, degassed to remove trapped air bubbles, and baked at 75°C for 4 hours to crosslink the device. The device is peeled from the master and holes are punched using a 0.75 mm biopsy coring needle. The punched device is washed with isopropyl alcohol and patted with scotch tape to remove debris prior to plasma bonding to a glass slide. To render the channels hydrophobic for water-in-oil emulsification, Aquapel<sup>TM</sup> is flushed into the channels, after which the device is baked in an oven for 20 min at 65°C.

### *Sample encapsulation and droplet PCR*

Prior to encapsulation, bacterial suspensions are washed three times by centrifugation at 3000g and the pellets are re-suspended in distilled water. Phage suspensions are encapsulated without washing. The viral or bacterial suspensions are mixed with the appropriate primers, TaqMan probes, and PCR mix (2X ddPCR MasterMix, Bio-Rad). The primers and TaqMan probes are used at 1  $\mu\text{M}$  and 250 nM, respectively (Primers and probes are listed in Table 4.1). Droplet loading is assumed to follow Poisson statistics<sup>26</sup>. For the experiments where two strains of bacteria



are mixed, *E. coli* C and BW25113, the bacteria are first measured for their optical density at OD<sub>600</sub> and then mixed in appropriate volumes, so that their final concentration ratio was 1:9 or 1:999, corresponding to a spike-in ratio of *E. coli* C: *E. coli* BW25113 of 10% and 0.1% respectively. The mix is loaded into an upright 1 ml syringe pre-filled with 200  $\mu$ l HFE-7500 fluorinated oil (3M), connected to a PDMS flow-focus droplet generator (Supp. Data S3) through a 21 gauge needle and polyethylene tubing. Droplet generation oil for probes (Bio-Rad) is introduced into the carrier-phase inlet of the microfluidic device through another syringe and tube; the oil comes with a proprietary surfactant included to stabilize the generated droplets during the heating and cooling steps of PCR. Using computer-controlled syringe pumps, the aqueous phase is injected at 200  $\mu$ lhr<sup>-1</sup> and the oil at 400  $\mu$ lhr<sup>-1</sup> (New Era), generating 25  $\mu$ m diameter droplets at ~3.6 kHz in a droplet maker with nozzle 20  $\mu$ m wide and 25  $\mu$ m tall. The emulsion is collected into 200  $\mu$ l PCR tubes and thermocycled on a T100 thermocycler (Bio-Rad), using the following conditions: 10 min. at 95°C, 35 cycles of 10 s. at 95°C, 15 s. at 55°C and 30 s. at 70°C. To verify specificity of the PCR, the emulsions are chemically ruptured with chloroform and DI water and the aqueous fractions electrophoresed on a 2% agarose gel to confirm amplicon length.

#### *Detection and sorting of droplets*

After thermocycling the emulsions, the droplets must be sorted based on fluorescence. This is accomplished by loading the thermocycled emulsions into a syringe with 200  $\mu$ l HFE-7500, maintaining the syringe vertically so that the needle faces up, and allowing the emulsion to cream for ~10 min; this ensures the droplets are at the top of the emulsion before the syringe pump is started so that they flow into the device at a controlled rate and closely packed. The droplets are injected into the detection and sorting device (Fig. 4.5)<sup>18,19,27,28</sup> at a flow rate of 50  $\mu$ lhr<sup>-1</sup>, with

spacer oil flow rate  $1000 \mu\text{hr}^{-1}$ . The flow rate for the second oil spacer at the sorting junction is set to  $100 \mu\text{hr}^{-1}$ . All droplet spacing is performed with pure HFE-7500. All electrodes on the device, including the sorting electrode and moat shielding the droplets from stray field, are filled with 2M NaCl solution<sup>29,30</sup>. A 100 mW, 532 nm laser is focused upstream of the sorting junction to excite droplet fluorescence. Photomultiplier tubes (PMTs) focused on the same spot measure emitted light and output a voltage proportional to the light intensity to a computer outfitted with an FPGA card (National Instruments) programmed in LabVIEW. The card detects droplets as peaks in fluorescence over time and, when a droplet is to be sorted, outputs a 40 kHz, signal amplified to 1000 V (Trek) applied to the salt-water electrodes on the microfluidic chip. Custom LabVIEW software (available on request) allows adjustment of PMT gain, droplet fluorescence intensity thresholds for sorting, and electrode AC voltage pulse frequency and magnitude (Figure 4.6). An image of the optical setup can be found in figure 4.9.

#### 4.4 Results and discussion

##### *PACS workflow for identifying virus-host relationships*

To enable the ultrahigh-throughput sorting of microbes based on the presence of phage nucleic acids, we use PACS, a droplet-based microfluidic technology<sup>18,19</sup>. In PACS, picoliter-volume aqueous droplets are used as reactors in which to perform TaqMan PCR on single bacteria. Using flow-focus emulsification<sup>31–33</sup>, individual phage or bacteria from a heterogeneous sample are isolated in  $\sim 10^7$  droplets with PCR reagents and probes targeting specific phage genes (Fig. 4.1). At a loading rate of  $\sim 0.1$  bacterial cells per droplet, we interrogate  $10^6$  cells. At that loading ratio, we can use Poisson statistics to determine at over 95% of droplets that are filled contain a single cell. After all bacterial cells are encapsulated, the emulsion is thermocycled, performing  $10^6$

parallel single-cell PCR reactions, as illustrated in Fig. 4.1. During thermocycling, the bacteria lyse and their nucleic acids are subjected to TaqMan amplification. If a droplet contains the nucleic acids of a phage targeted by the probes, the nucleic acids are amplified, generating a fluorescent signal that fills the encapsulating droplet. This marks the droplet as containing a bacteria associated with the target virus, allowing the putative host cell nucleic acids to be recovered by sorting the droplet, as illustrated in Fig. 4.1.

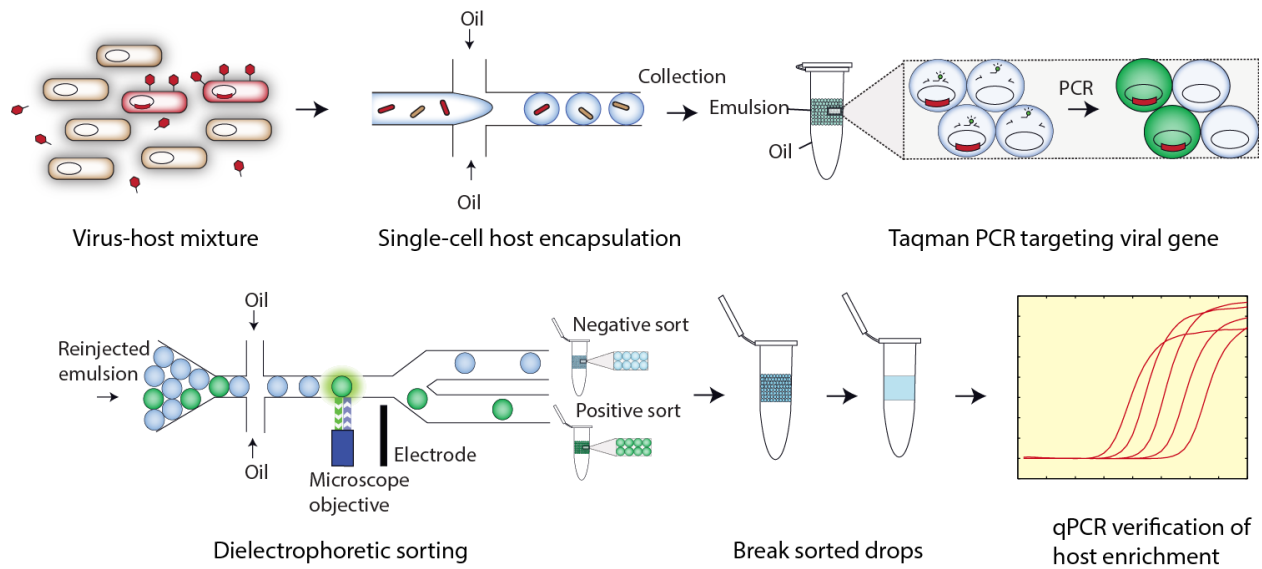


Figure 4.1. Microfluidic workflow for PACS-based viral detection and host sorting. Virus-infected hosts are first encapsulated with PCR reagent, primers and probe in picoliter-volume droplets, then thermocycled to yield fluorescent drops. These drops contain targeted genomes of viruses and their hosts, which are then sorted to yield a purified population of DNA. This material can be used for downstream sequencing analysis in this case qPCR, but any sequence analysis is possible.

*Specific detection and quantification of viral genomes from bacteriophage T4 and  $\Phi$ X174.*

Detecting cells associated with a specific virus depends on the ability to reliably and specifically amplify virus nucleic acids. To investigate the robustness of this step in the PACS process, we perform TaqMan PCR in microfluidic droplets on two distinct virus species, bacteriophage T4 (T4) and bacteriophage  $\Phi$ X174 ( $\Phi$ X174). Pure preparations of the viruses are combined with PCR and TaqMan reagents immediately prior to microfluidic emulsification and thermocycling. After thermocycling, we observe clearly fluorescent droplets in a population of non-fluorescent droplets, as shown in Fig. 4.2A, indicating successful amplification of the viruses. In negative controls, we swap the TaqMan probes by including T4 probes in  $\Phi$ X174 preparations and  $\Phi$ X174 probes in T4 preparations. Neither of these controls yield detectable fluorescent droplets, demonstrating that the probes are specific to their target virus.

An important factor when using PACS to enrich bacterial cells from a heterogeneous sample is the rate of false negatives since this limits the number of positive events detected. To characterize the sensitivity of the method, we scan the emulsions created in the previous experiment using flow dropometry<sup>18</sup>, recording fluorescence values for ~30,000 individual droplets (Figs. 4.9,4.10). Using the known droplet volumes and assuming phage encapsulation is governed by Poisson statistics<sup>34</sup>, we estimate phage concentrations in the starting samples and compare them to estimates from plaque assays, Fig. 4.2B. For bacteriophage T4, plaque assays on samples used for PACS yield  $3.0 \times 10^9$  pfu/ml compared to the  $3.1 \times 10^9$  particles/ml for PACS. For  $\Phi$ X174, the plaque assay yields  $3.0 \times 10^9$  pfu/ml versus  $3.5 \times 10^9$  particles/ml for PACS. The estimates using both methods are in excellent agreement indicating that the PCR conditions are sufficient to efficiently lyse the viral particles. The slightly higher values estimated by PACS may

reflect that some phage genomes are incorrectly packaged, mutated, or are in non-infectious particles. In addition to validating droplet PCR for PACS, this demonstrates that droplet digital PCR using this apparatus is an alternative approach for quantitating phage genomes in a sample. Our method of directly encapsulating phage particles for digital PCR circumvents the need to pre-lyse the virus before measurement which is needed for quantitative PCR protocols<sup>35,36</sup>.

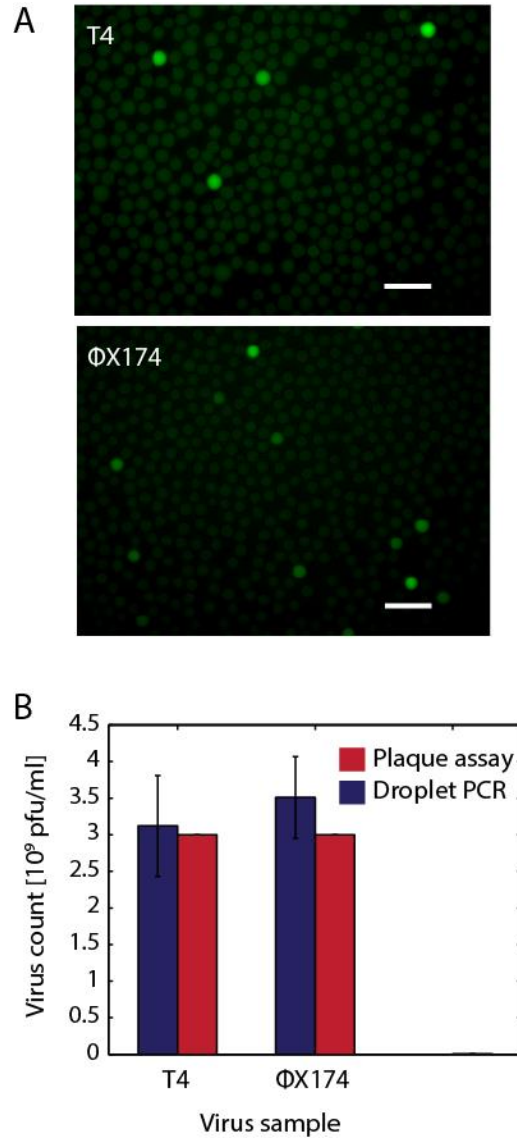


Figure 4.2. A) Digital detection of phage particles after droplet PCR. Bacteriophages T4 and  $\Phi$ X174 virions are partitioned into droplets for TaqMan PCR detection. Scale bars are 100 $\mu$ m. B) Plaque assay results closely mimic digital viral particle quantitation, suggesting that phage genomes are accurately measured with this new method. Error bars represent the standard deviation of 3 technical replicates.

### *Sorting E. coli infected with lambda bacteriophage.*

PACS enables the detection of bacterial cells associated with specific phages, including lysogens, and recovery of the host cell genomes. To illustrate this, we construct a test system comprising two *E. coli* strains: a lambda lysogenic C-strain, and an uninfected BW25113 strain. To validate the TaqMan probes, we analyze a sample of pure lambda virus in suspension and observe digital signals in droplets, indicating single-virion detection (Fig. 4.3A). We then spike *E. coli* C (lambda). into uninfected *E. coli* BW25113 in a ratio of 1:9, and wash the mixture to remove any free phage. We subject this sample to droplet PCR and again observe a digital signal corresponding to a small subpopulation of positive droplets which contain *E. coli* cells with the lambda genome, as shown in Fig. 4.3B. To verify that the digital fluorescence corresponds to droplets containing lambda genomes, we sort the emulsion to recover the positive droplets, which we accomplish using dielectrophoretic droplet sorting, Fig. 4.1<sup>18,19,27-29</sup>. Dielectrophoresis is a phenomenon in which a particle experiences polarization forces in a non-homogenous electric field<sup>37</sup>. The thermocycled emulsion is injected into the sorting device, which flows the droplets spaced by oil individually through the focused excitation laser. As a droplet passes through the laser, its fluorescence is excited and the resulting emitted light is measured with a photomultiplier tube (PMT), which outputs a voltage proportional to the fluorescence intensity analyzed by the computer and FPGA card. The droplets appear as peaks in voltage as a function of time, in which the amplitude of the peak is proportional to the droplet intensity, as shown in Fig. 4.3C. When a positive droplet passes through the laser, an abnormally tall peak is observed, as seen at  $t = 0.0185$  seconds (Fig. 4.3C). Upon detection of a positive droplet, the computer outputs an alternating voltage amplified to  $\sim 1000$  V applied to on-chip electrodes generating an attractive force that pulls the positive droplet into the collection channel<sup>18,19,28</sup>. When a negative droplet passes through the



laser the peak amplitude does not fall above the user-defined threshold; the electrode remains unenergized and the droplet flows passively into the waste channel. In this way, the droplet sorter separates the positive from the negative droplets. This method of sorting can separate drops at the rate of a couple of kilohertz, although modifications to the sorter design can further push sorting rates up to 30kHz<sup>38</sup>.

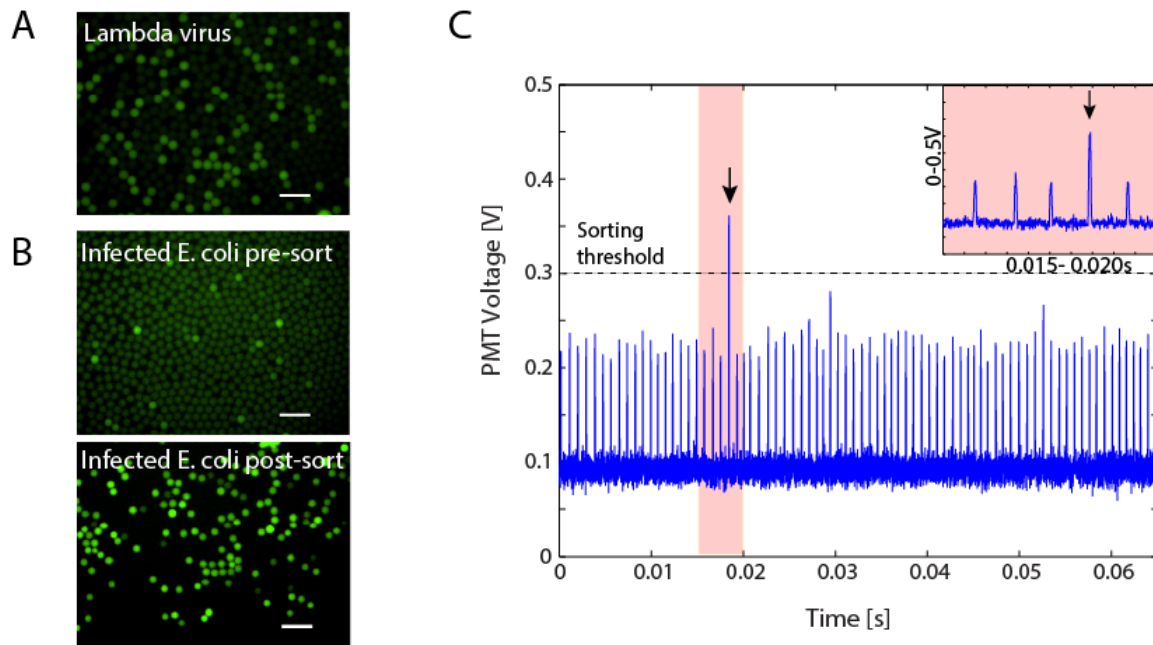


Figure 4.3. A) Digital detection of lambda particles and B) pre- and post- sorted drops containing lambda and its *E. coli* host. Digital detection of lambda using the probes and primers from Figure 4.3A. C) Time trace of fluorescent droplet detection. Droplets are run through a dielectrophoretic microfluidic sorter, and droplets above a set threshold are sorted and collected each peak represents a single droplet. Inset figure shows a magnified view of the lone sorted event in the time period 0.015-0.025 seconds. All scale bars are 100 $\mu$ m.

*Enrichment of phage-infected E. coli genomes.*

Sorting the bacterial cells based on presence of lambda DNA should yield primarily *E. coli* C cells, since only this strain was infected by lambda. To quantify the enrichment for phage-infected cells afforded by PACS, we analyze the fraction of *E. coli* C DNA in the sorted population using qPCR. We perform a second sort in which we lower the concentration of lysogenic *E. coli* C(lambda) to 0.1%, while increasing uninfected *E. coli* BW25113 to 99.9%. Using PACS, we recover all TaqMan positive droplets and the nucleic acids from their cells. To measure the fraction of nucleic acids corresponding to *E. coli* C and *E. coli* BW25113, we perform qPCR on pre- and post-sorted samples using primers specific to the two strains; the results for both the 10% and 0.1% mixtures for both primer sets are provided in Fig. 4.4. The signal from *E. coli* BW25113 is strongly *de-enriched* in the PACS-sorted material, corresponding to curve shifts by ~10  $C_t$  values (Figure 4.4, upper panels). By contrast, the curves for *E. coli* C shift by a much smaller distance in the same direction, demonstrating that PACS for lambda virus nucleic acids *enriches* for these cells. To obtain a quantitative metric of enrichment, we define the enrichment factor as the ratio of host microbe purity in the pre-sorted to the post-sorted samples,

$$e = \frac{C_{\text{post}}(C_{\text{pre}} + K_{\text{pre}})}{C_{\text{pre}}(C_{\text{post}} + K_{\text{post}})}$$

where  $C_{\text{pre}}$  and  $C_{\text{post}}$  are the number of *E. coli* C cells present in the pre- and post-sorted samples, while  $K_{\text{pre}}$  and  $K_{\text{post}}$  are the number of *E. coli* BW25113 present in the pre- and post-sorted samples, respectively. The relationships between  $C_{\text{pre}}$  and  $C_{\text{post}}$ , together with  $K_{\text{pre}}$  and  $K_{\text{post}}$ , can be determined via qPCR, which measures the log-2 fold change of gene copy numbers specific to either microbe,

$$C_{\text{post}} = (2^{\Delta C_{\text{tc}}})C_{\text{pre}} \quad \text{and} \quad K_{\text{post}} = (2^{\Delta C_{\text{tk}}})K_{\text{pre}}$$

where  $\Delta C_{\text{tc}}$  and  $\Delta C_{\text{tk}}$  are the differences between the qPCR cross-threshold values for *E. coli* C and BW25113, respectively. This is an estimation, since the actual rate of DNA amplification may be less than 2 per cycle. We know  $C_{\text{pre}}$  and  $K_{\text{pre}}$  because we begin with controlled numbers of each species before sorting, enabling us to define the ratio of the two species with respect to one another,

$$n = \frac{K_{\text{pre}}}{C_{\text{pre}}}$$

and to simplify the enrichment factor to,

$$e = \frac{(n + 1) \left( \frac{1}{2^{\Delta C_{\text{tc}}}} \right)}{\left( \frac{1}{2^{\Delta C_{\text{tc}}}} \right) + n \left( \frac{1}{2^{\Delta C_{\text{tk}}}} \right)}$$

For our initial sort,  $n = 9$  (10% *E. coli* C to BW25113 ratio), we obtain 4.0 for  $\Delta C_{\text{tc}}$  and 13.1 for  $\Delta C_{\text{tk}}$  (Fig. 4.4.), yielding  $e = 9.84$ , indicating that the final sample is enriched to 98.4% for *E. coli* C from an initial concentration of 10%. For the 0.1% spike-in, we obtain 3.0 for  $\Delta C_{\text{tc}}$  and 9.9 for  $\Delta C_{\text{tk}}$  (Fig. 4.4.), providing  $e = 106$ , so that *E. coli* C is enriched by about a hundred times to 10.6% final concentration. Larger enrichment factors can be achieved by further diluting the sample prior

to partitioning into droplets, which reduces the rate of co-encapsulation of the two species and false-positive recovery of off-target cells.

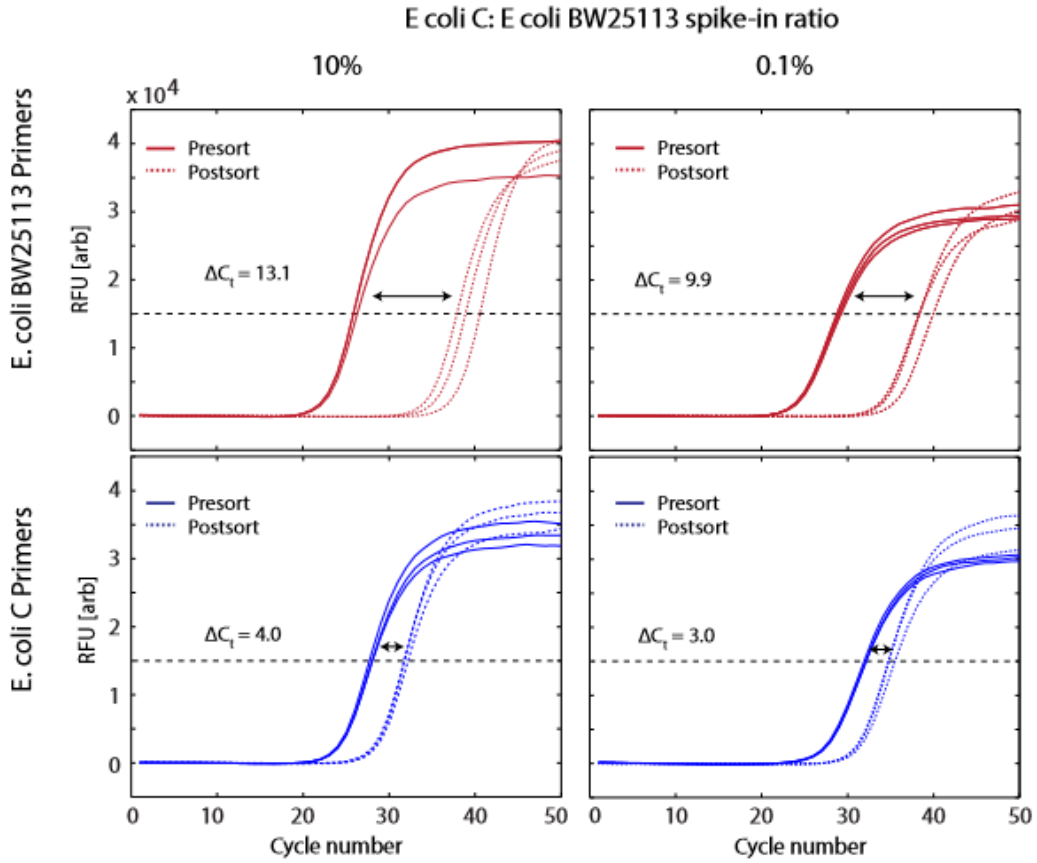


Figure 4.4. qPCR detection of host genomes before and after droplet sorting based on the presence of lambda DNA. Each quadrant shows qPCR amplification curves for DNA extracted from drops before and after sorting. The shifts in the curves reflect the 2-fold change of the DNA quantity according to the specific primers being tested.

## 4.5 Conclusion

When confronted with large, heterogeneous populations, cell sorting is invaluable because it allows subpopulations to be systematically isolated and studied. However, currently, specific cell sorting is only possible using affinity reagents, such as antibodies or oligomer probes that specifically bind to the cell type of interest<sup>39,40</sup>; such reagents are rarely available for uncultivable targets. Nonetheless, in many of these cases, sequence data, often partial, is available. In these instances, PACS is uniquely suited for specific cell sorting since it relies on PCR to differentiate between cells based on the presence of a specific nucleic acid sequence. In this application of PACS, this allows bacterial sorting based on the presence of a phage genome within or attached to the host cell (Fig. 4.4). We cannot differentiate between bound phage and infected hosts using techniques presented here as intact virions also provide positive PACS signals (Fig. 4.2), but a stripping step could be included in the PACS workflow, if only hosts with internal viruses are desired. However, previous studies have shown that labeled viruses associated with cells are usually infectious. Clearly, full determination of phage-host interactions will require analysis of the specific phage-host pair and cultivation of both, but PACS allows identification of potential hosts in mixed systems, drastically narrowing the search for elusive hosts.

In its current implementation, PACS differentiates between bacterial cells based on one sequence. However, since PACS relies on TaqMan assays, probes of different color can be used to interrogate multiple sequences simultaneously<sup>41</sup>. This should enable multi-parametric cell sorting similar to current FACS analysis with multiple antibodies. Multi-parametric cell-sorting could recover, for example, the genomes of all cells associated with specific variants of a virus, co-infection of different viral species, or the presence of specific host and virus sequences associated with the same cell.

A limitation of the current system is that the lysis of the bacteria is achieved using the detergents present in the PCR reagents and the high temperature of PCR thermocycling, which may be insufficient for lysing some bacteria or viruses. To broaden the applicability of the method, harsher lysis procedures could be implemented that allow the inclusion of enzymes to digest cellular material. Such procedures can be implemented using recently described agarose emulsion and multi-step droplet merger workflows, which enable the digestion of cell lysates with proteases<sup>19</sup>. Indeed, this is essential for reliably performing PACS on mammalian cells in sub-nanoliter droplets, since undigested mammalian cell lysates can inhibit PCR<sup>19</sup>.

PACS is founded on droplet-based microfluidic technologies whose intrinsic throughput is thousands of droplets per second<sup>33</sup>. This enables facile and rapid sorting of millions of cells<sup>29</sup>. By implementing faster emulsification strategies using air-bubble triggered droplet generation<sup>42</sup> or sequential droplet splitting, and using double emulsion FACS for sorting<sup>43</sup>, it should be possible to enhance the throughput of the system by another order of magnitude; this will further increase the number of microbes that can be sorted for identifying rare phage-host relationships, such as identifying the host of chimeric RNA-DNA virus genomes recently discovered in numerous ecosystems<sup>44,45</sup>. PACS could also be used for screening emerging pathogenic virus sequences to identify possible reservoir hosts.

PACS paired with NGS can be a particularly potent tool for applications in microbiology and virology. PACS-sorted lysates can be recovered and subjected to deep sequencing of cellular DNA and/or RNA, enabling correlation of specific viruses with host cell genomes and their expression patterns. Virus-based PACS can be used, for example, to study how viruses modulate host cell gene expression, or if certain host genetic variants are more susceptible to virus infection. The ability to sort millions of entities based on nucleic acids is valuable for sieving through



complex cell populations to determine specific phage-host interactions. Current available metagenomic datasets can be mined for phage sequences for the design of virus-specific primers.. From a complex population, one could assay the sorted genomic material for host provenance by targeted 16S or whole-genome sequencing . In addition to sorting microbes based on infection by bacteriophage, PACS can be applied to the clinical setting such as the isolation of mammalian cells latently infected by HIV. By combining PACS with sequencing, the presence of a specific pathogen can be correlated with host cell properties, such as somatic mutations or gene expression. Such investigations would be valuable for studying how different pathogens manipulate their hosts.

#### 4.6 Supplemental information

<b>Primers for Taqman droplet PCR</b>	<b>Sequence</b>
T4 probe	5'-/56-FAM/TCGCATTCT /ZEN/TCCTCTGATGGAGCA /3IABkFQ/ -3'
T4 FWD	5'-CCACAACCTAACCGAGGAAGTAA-3'
T4 REV	5'-TGCGATATGCTATGGGTCTTG-3'
PhiX174 probe	5'-/56-FAM/ATG GAA CTG /ZEN/ACC AAA CGT CGT TAG GCC A/3IABkFQ/-3'
PhiX174 FWD	5'- GCGCTCTAATCTCTGGGCAT-3'
PhiX174 REV	5'- CAAAGAAACGCGGCACAGAA-3'
Lambda probe	5'- /56-FAM/AT ACT GAG C/ZEN/A CAT CAG CAG GAC GC/3IABkFQ/-3'
Lambda FWD	5'- GCC CTT CTT CAG GGC TTA AT-3'
Lambda REV	5'- CTC TGG CGG TGT TGA CAT AA-3'
<b>Primers for qPCR</b>	<b>Sequence</b>
E coli C FWD	5'-ACG CAG GGA TTT ACA GCA TAT AG-3'
E coli C REV	5'-GGG TGC TAT ATA ACG GTG TAC TG-3'
E coli BW25113 FWD	5'-GACTACTTGAAGCTGTCCTTCC-3'
E coli BW25113 REV	5'-CGCAGCTTCACCTTG TAGAT-3'

Table 4.1. Primers used for qPCR and TaqMan droplet PCR

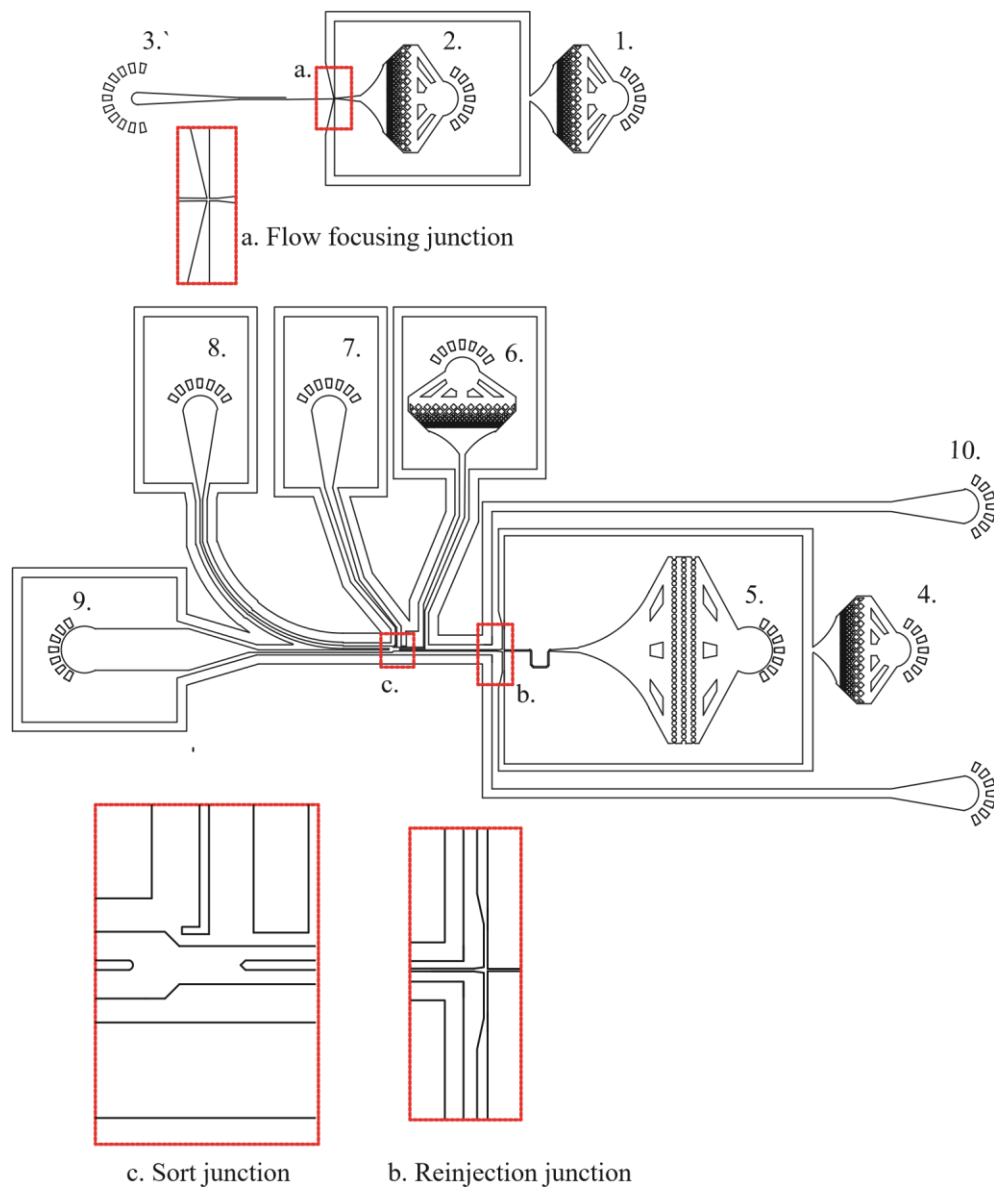


Figure 4.5. Device schematic for flow focus drop maker(top) and sorter (bottom). The numbers correspond to input/outputs for the device— (1) oil input (2) aqueous input (3) output (4) oil input drop spacer (5) drop reinjection (6) oil input drop spacer (7) salt electrode (8) positive sort output (9) negative sort output (10) salt moat input.

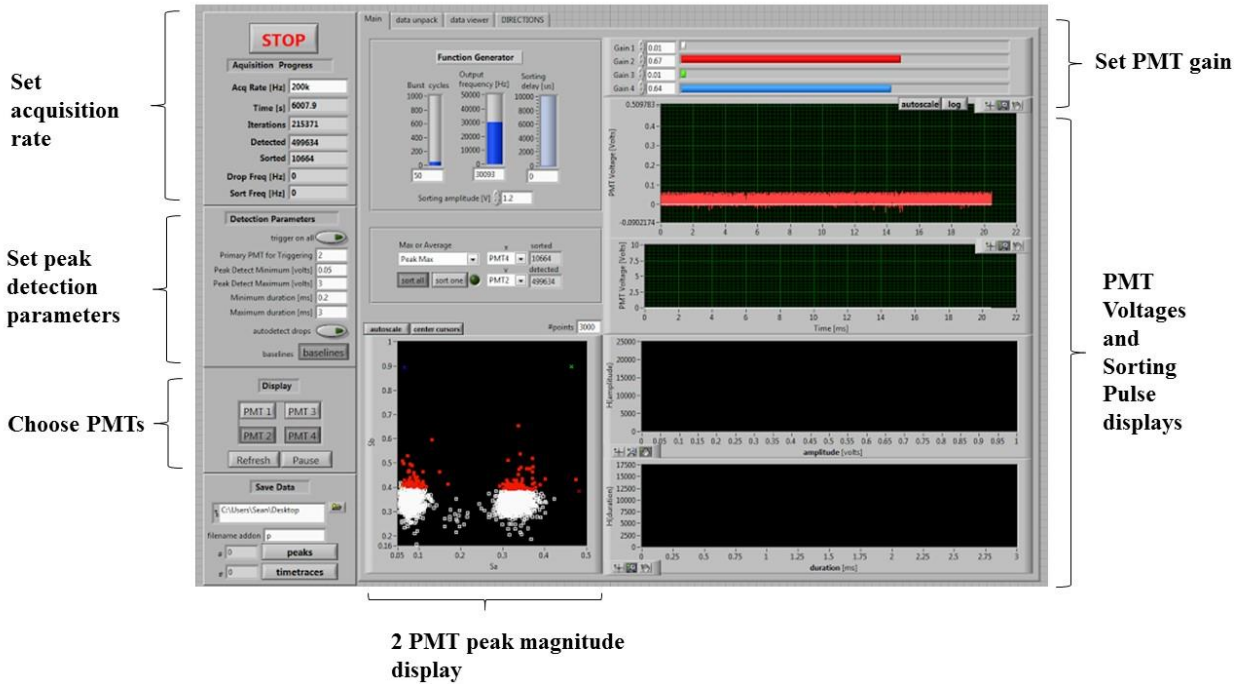


Figure 4.6. LabVIEW interface

Data acquisition was handled by a FPGA Card (National Instruments Corporation) controlled by a custom program written in LabVIEW (National Instruments Corporation). Droplet fluorescence levels are detected by photomultiplier tubes (PMTs) and converted into corresponding levels of signal output voltages. The LabVIEW program receives the fluorescence signals from the optical setup in real time and combines a peak detection algorithm together with user-defined ranges of fluorescence amplitude and width to determine droplet-associated peaks. The data acquisition rate was set at 200 kHz. The software directs the FPGA to output sorting pulses that are amplified by a high-voltage amplifier (Trek). The salt electrode on the sorting device then effects a dielectrophoretic force on any drop that has a desired fluorescence level. The software allows for manipulation of various sorting parameters, such as fluorescence peak level, PMT gain, data acquisition rates, length and magnitude of sorting pulse, etc. Each sorting attempt may require adjustment of aforementioned parameters for optimal sorting.

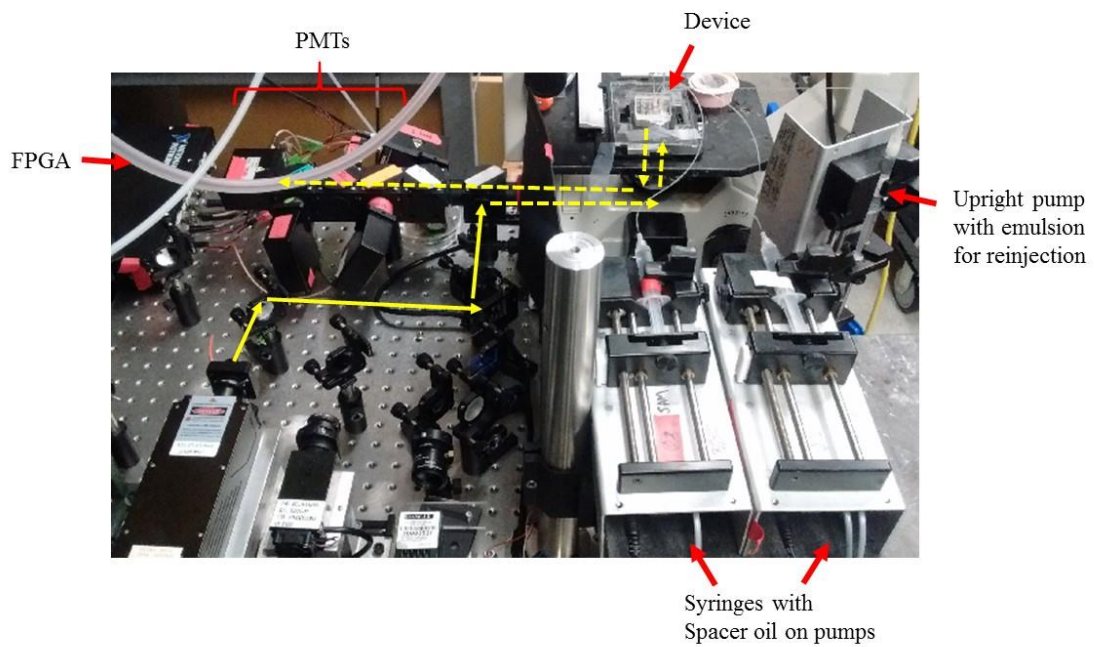


Figure 4.7. Optical setup of sorter. The yellow line depicts the path of photons travelling from the output source(laser) to the device, where it excites the Taqman fluorophore which in turn emits fluorescence that goes to the end-most PMT, where it is detected. Dashed lines represent the path of light that is internal to the setup.

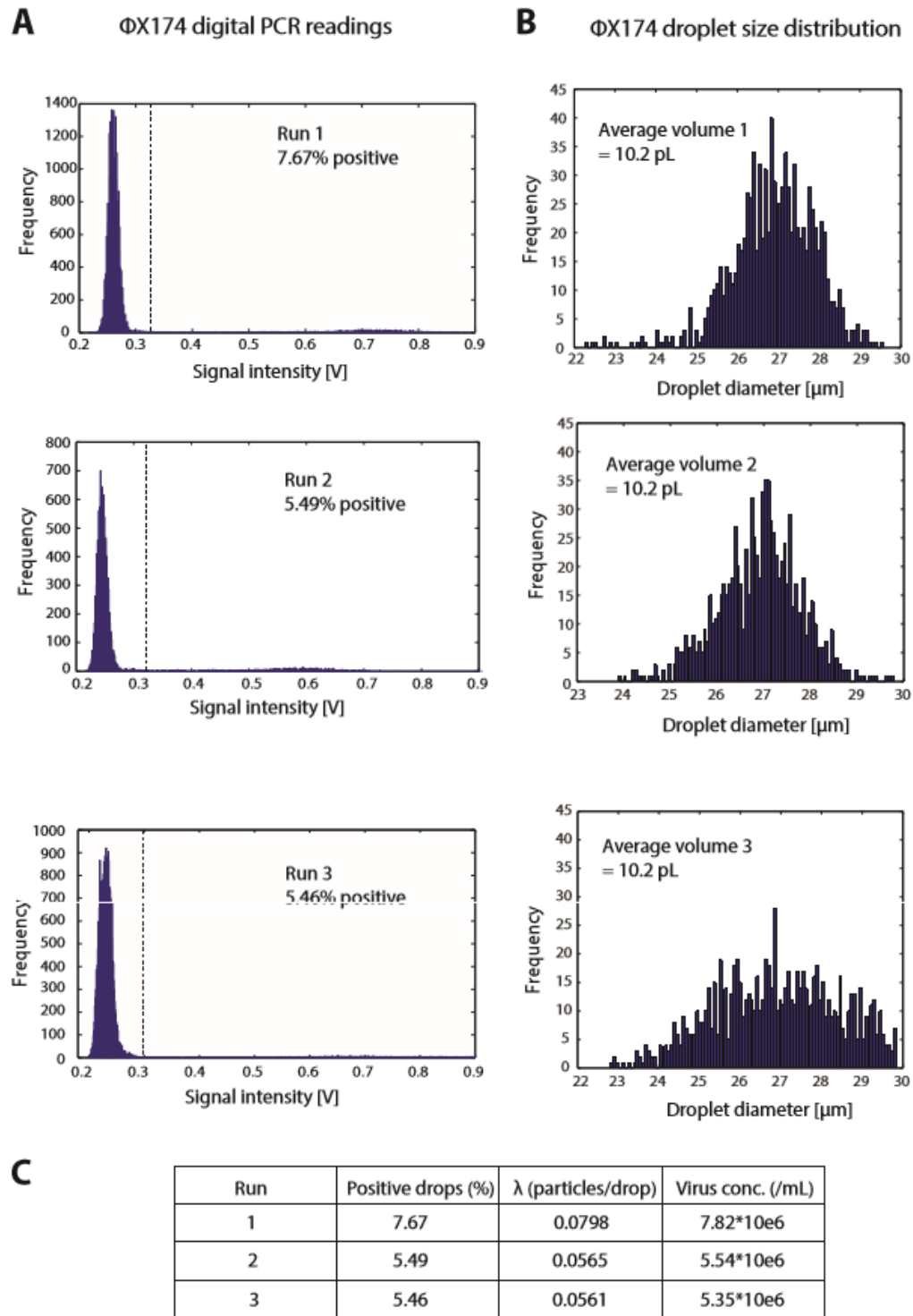


Figure 4.8.  $\Phi$ X174 virus digital droplet assay. A) PMT values recorded for 30,000 drops. B) Droplet diameters recorded for 500 droplets using ImageJ analysis. C) Calculations to derive virus concentrations from the number of positive drops recorded.

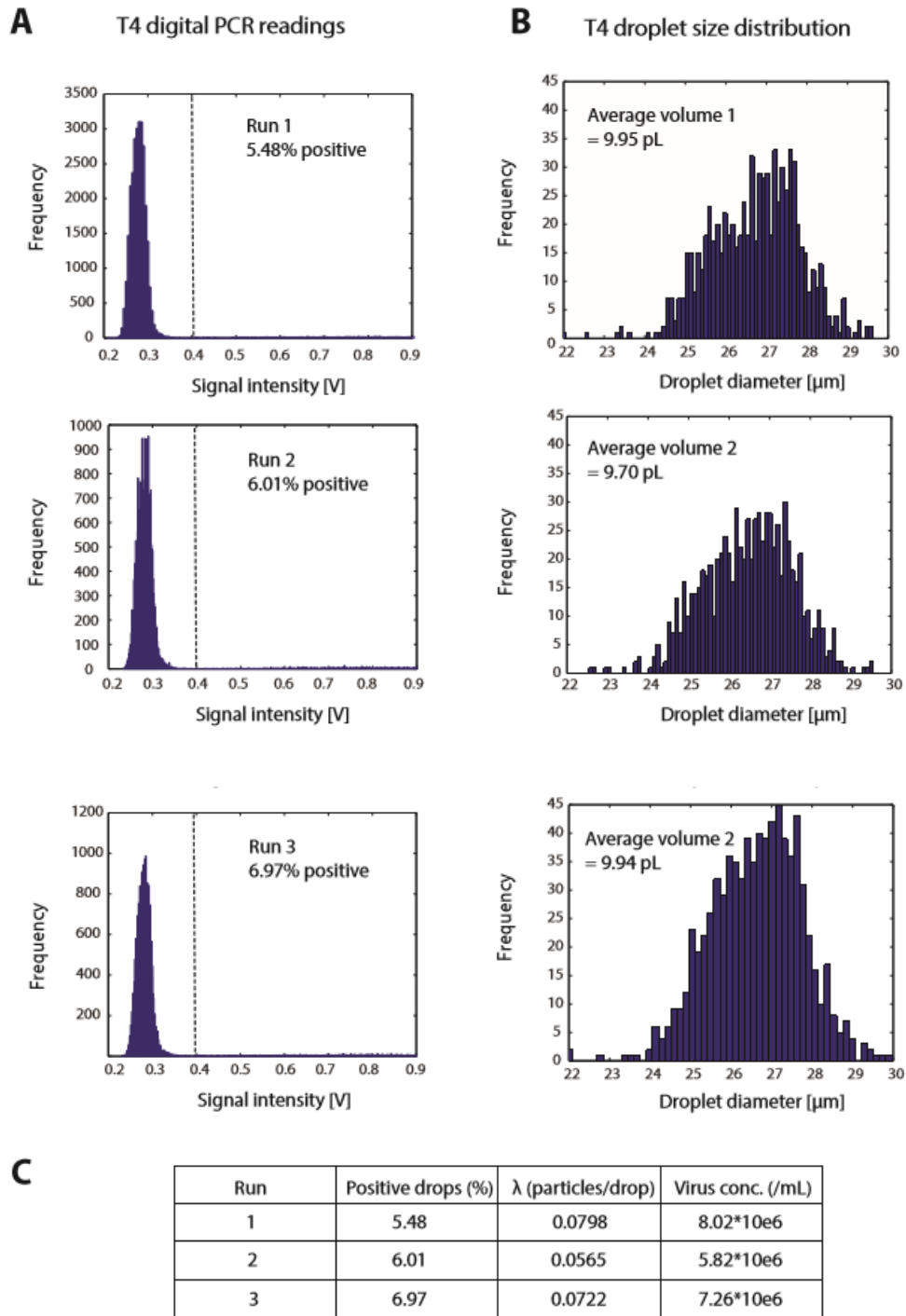


Figure 4.9. T4 virus digital droplet assay . A) PMT values recorded for 30,000 drops. B) Droplet diameters recorded for 500 droplets using ImageJ analysis. C) Calculations to derive virus concentrations from the number of positive drops recorded.

#### 4.7 References

1. B. J. M. Bohannan and R. E. Lenski, *Ecol. Lett.*, 2000, **3**, 362–377.
2. C. a Suttle, *Nature*, 2005, **437**, 356–361.
3. L. R. Thompson, Q. Zeng, L. Kelly, K. H. Huang, A. U. Singer, J. Stubbe and S. W. Chisholm, *Proc. Natl. Acad. Sci.*, 2011, **108**, E757–E764.
4. B. L. Hurwitz, S. J. Hallam and M. B. Sullivan, *Genome Biol.*, 2013, **14**, R123.
5. M. S. Rappé and S. J. Giovannoni, *Annu. Rev. Microbiol.*, 2003, **57**, 369–394.
6. S. R. Vartoukian, R. M. Palmer and W. G. Wade, *FEMS Microbiol. Lett.*, 2010, **309**, 1–7.
7. A. Reyes, N. P. Semenkovich, K. Whiteson, F. Rohwer and J. I. Gordon, *Nat. Rev. Microbiol.*, 2012, **10**, 607–617.
8. V. Dhillon and X. Li, *J. Comput. Sci. Syst. Biol.*, 2015, **8**, 160–165.
9. S. Roux, A. K. Hawley, M. T. Beltran, M. Scofield, P. Schwientek, R. Stepanauskas, T. Woyke, S. J. Hallam and M. B. Sullivan, *eLife*, 2014, **in review**, 1–20.
10. J. M. Labonté, B. K. Swan, B. Poulos, H. Luo, S. Koren, S. J. Hallam, M. B. Sullivan, T. Woyke, K. Eric Wommack and R. Stepanauskas, *ISME J.*, 2015, 1–14.
11. A. F. Andersson and J. F. Banfield, *Science*, 2008, **320**, 1047–1050.
12. A. D. Weinberger, C. L. Sun, M. M. Pluciński, V. J. Denef, B. C. Thomas, P. Horvath, R. Barrangou, M. S. Gilmore, W. M. Getz and J. F. Banfield, *PLoS Comput. Biol.*, 2012, **8**.
13. E. Allers, C. Moraru, M. B. Duhaime, E. Beneze, N. Solonenko, J. Barrero-Canosa, R. Amann and M. B. Sullivan, *Environ. Microbiol.*, 2013, **15**, 2306–2318.
14. L. Deng, A. Gregory, S. Yilmaz, B. T. Poulos, P. Hugenholtz and M. B. Sullivan, *mBio*, 2012, **3**, 1–8.



15. L. Deng, J. C. Ignacio-Espinoza, A. C. Gregory, B. T. Poulos, J. S. Weitz, P. Hugenholtz and M. B. Sullivan, *Nature*, 2014, **513**, 242–245.
16. A. D. Tadmor, E. A. Ottesen, J. R. Leadbetter and R. Phillips, *Science*, 2011, **333**, 58–62.
17. V. T. Dang and M. B. Sullivan, *Front. Microbiol.*, 2014, **5**, 1–8.
18. S. W. Lim, T. M. Tran and A. R. Abate, *Plos One*, 2015, **10**, e0113549.
19. D. J. Eastburn, A. Sciambi and A. R. Abate, *Nucleic Acids Res.*, 2014, **42**, 1–10.
20. D. Paez-Espino, E. A. Eloë-Fadrosch, G. A. Pavlopoulos, A. D. Thomas, M. Huntemann, N. Mikhailova, E. Rubin, N. N. Ivanova and N. C. Kyrpides, *Nature*, 2016, **536**, 425–430.
21. J. D. Karam, *Molecular Biology of Bacteriophage T4*, ASM Press, 1994.
22. W. Arber, L. Enquist, B. Hohn, N. Murray and K. Murray, *Experimental methods for use with lambda*, 1983, vol. 1.
23. S. W. Lim, T. M. Tran and A. R. Abate, *PLOS ONE*, 2015, **10**, e0113549.
24. S. Hafenstein and B. a Fane, *J. Virol.*, 2002, **76**, 5350–5356.
25. Y. Xia and G. M. Whitesides, *Annu. Rev. Mater. Sci.*, 1998, **28**, 153–184.
26. L. Mazutis, J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths and J. A. Heyman, *Nat. Protoc.*, 2013, **8**, 870–891.
27. L. Mazutis, J. Gilbert, W. L. Ung, D. a Weitz, A. D. Griffiths and J. a Heyman, *Nat. Protoc.*, 2013, **8**, 870–91.
28. J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klivanov, A. D. Griffiths and D. a Weitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4004–4009.
29. A. Sciambi and A. R. Abate, *Lab Chip*, 2015, **15**, 47–51.

30. A. Sciambi and A. R. Abate, *Lab. Chip*, 2014, **14**, 2605–2609.
31. Z. Nie, M. Seo, S. Xu, P. C. Lewis, M. Mok, E. Kumacheva, G. M. Whitesides, P. Garstecki and H. a. Stone, *Microfluid. Nanofluidics*, 2008, **5**, 585–594.
32. G. F. Christopher and S. L. Anna, *J. Phys. Appl. Phys.*, 2007, **40**, R319–R336.
33. T. M. Tran, F. Lan, C. S. Thompson and a R. Abate, *J. Phys. Appl. Phys.*, 2013, **46**, 114004.
34. a Huebner, M. Srisa-Art, D. Holt, C. Abell, F. Hollfelder, a J. deMello and J. B. Edel, *Chem. Commun. Camb. Engl.*, 2007, **2**, 1218–1220.
35. B. Anderson, M. H. Rashid, C. Carter, G. Pasternack, C. Rajanna, T. Revazishvili, T. Dean, A. Senecal and A. Sulakvelidze, *Bacteriophage*, 2011, **1**, 86–93.
36. D. Refardt, *Bacteriophage*, 2012, **2**, 98–104.
37. R. Pethig, *Biomicrofluidics*, 2010, **4**, 22811.
38. A. Sciambi and A. R. Abate, *Lab Chip*, 2015, **15**, 47–51.
39. E. H. Turner, S. B. Ng, D. a Nickerson and J. Shendure, *Annu. Rev. Genomics Hum. Genet.*, 2009, **10**, 263–284.
40. L. Mamanova, A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure and D. J. Turner, *Nat. Methods*, 2010, **7**, 111–118.
41. V. Taly, D. Pekin, L. Benhaim, S. K. Kotsopoulos, D. Le Corre, X. Li, I. Atochin, D. R. Link, A. D. Griffiths, K. Pallier, H. Blons, O. Bouché, B. Landi, J. B. Hutchison and P. Laurent-Puig, *Clin. Chem.*, 2013, **59**, 1722–1731.
42. A. R. Abate and D. A Weitz, *Lab. Chip*, 2011, **11**, 1713–1716.
43. S. W. Lim and A. R. Abate, *Lab. Chip*, 2013, **13**, 4563–72.
44. G. S. Diemer and K. M. Stedman, *Biol. Direct*, 2012, **7**, 13.

45. K. Stedman, *J. Mol. Evol.*, 2013, **76**, 359–364.

**Publishing Agreement**

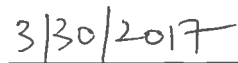
*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

***Please sign the following statement:***

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*



\_\_\_\_\_  
Author Signature



\_\_\_\_\_  
Date