SAND2022-12676

SANDIA REPORT
SAND20XX-XXXX
Printed September 2022

Sandia National Laboratories

Combined Imaging and RNA-Seq on a Microfluidic Platform for Viral Infection Studies

Anna L. Fisher, Kurt C. Sjoberg, Gloria E. Doudoukjian, Elizabeth R. Webster, Raga Krishnakumar

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

The goal of this work was to pioneer a novel, low-overhead protocol for simultaneously assaying cell-surface markers and intracellular gene expression in a single mammalian cell. The purpose of developing such a method is to be able to understand the mechanisms by which pathogens engage with individual mammalian cells, depending on their cell surface proteins, and how both host and pathogen gene expression changes are reflective of these mechanisms. The knowledge gained from such analyses of single cells will ultimately lead to more robust pathogen detection and countermeasures. Our method was aimed at streamlining both the upstream cell sample preparation using microfluidic methods, as well as the actual library making protocol. Specifically, we wanted to implement a random hexamer-based reverse transcription of all RNA within a single cell (as opposed to oligo dT-based which would only capture polyadenylated transcripts), and then use a CRISPR-based method called scDash to deplete ribosomal DNAs (since ribosomal RNAs make up the majority of the RNA in a mammalian cell)[1]. After significant troubleshooting, we demonstrate that we are able to prepare cDNA from RNA using the random hexamer primer, and perform the rDNA depletion. We also show that we can visualize individually stained cells, setting up the pipeline for connecting surface markers to RNA-sequencing profiles. Finally, we test a number of devices for various parts of the pipeline, including bead generation, optical barcoding and cell dispensing, and demonstrate that while some of these have potential, more work is needed to optimize this part of the pipeline.

ACKNOWLEDGEMENTS

We thank our program manager Victoria VanderNoot for her guidance throughout this project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

ACRONYMS AND TERMS

Acronym/Term	Definition
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
rRNA/rDNA	Ribosomal RNA/DNA
cDNA	Copy DNA
scDASH	Single-Cell DASH method
CEL-seq	Single-cell RNA-Seq by multiplexed linear amplification
Smart-seq	Switch Mechanism at the 5' End of RNA Templates
MATQ-seq	Multiple Annealing and dC-tailing-based Quantitative Single-cell RNA-seq
Drop-seq	Droplet microfluidics-based single-cell sequencing method
Seq-well	A simple, portable platform for massively parallel scRNA-seq
SCOPE-seq	Single Cell Optical Phenotyping and Expression Sequencing
polyA	Poly adenine tail (or polyadenylation)
FY	Fiscal Year
CMOS	Complementary Metal Oxide Semiconductor
CAD	Computer-Aided Design
SulfB	Sulforhodamine B
CF	Carboxyfluorescein
PDMS	Polydimethylsiloxane
ABL	Applied Biosciences
PBS	Phosphate-Buffered Saline
LA-4	Mouse lung epithelium cell line
UMI	Unique Molecular Identifier
NTC	Non-Target Control
(q)PCR	(Quantitative) Polymerase Chain Reaction
PEG	Polyethylene Glycol
MSCs	Mesenchymal Stromal Cells
DOE	Department of Energy

1. INTRODUCTION AND MOTIVATION

Pathogens represent a permanent and evolving threat to national security, as vividly demonstrated by the recent COVID19 pandemic. Our ability to rapidly respond to pathogens, from the perspective of detection, diagnostics and countermeasures/therapeutics is highly dependent on our understanding of the mechanisms underlying pathogenesis, both from the perspective of the pathogen as well as the host. Therefore, developing technologies and methods that can be broadly deployed to address these questions across a range of pathogens (including novel ones) are critically in need.

Regardless of whether pathogens are extracellular or intracellular first interface with host cells at the cell surface, where a number of proteins and lipids responsible for either internalizing the pathogen or propagating signals into the cell. Understanding the relationship between the gene expression of host and pathogen cells gives us better insight into the outcomes of their interactions. It may even allow us to predict how novel pathogens are likely to impact both individual and global health, therefore allowing us to prepare countermeasures pro-actively[2-4].

RNA-Seq remains the gold standard for unbiased quantification of gene expression for both mammalian cells and pathogens. However, existing assays of infections of model host systems are limited by the technical difficulty of obtaining multi-omic data at single cell resolution. Conventional studies relying on bulk analysis of large numbers of cells obscure individual cell responses, and do not reveal heterogeneity at the level of cell type and state. The recent advent of single-cell sequencing has allowed us to obtain unprecedented granularity in understanding molecular mechanisms as well as heterogeneity between cells[2,3,5,6]. Plate-based methods such as CEL-seq, Smart-seq, and MATQ-seq have traditionally been used to analyze the transcriptomic differences between individual cells[7-10], but microfluidic methods like Drop-seq and Seq-well are becoming increasingly more prevalent[11,12].

The next step in understanding molecular mechanisms at a single-cell level is the incorporation of a simultaneous optical imaging methodology that allows for the correlation of cell phenotype with gene expression data. The combination of single-cell transcriptomics and imaging will create the potential for combined assays that reveal valuable information such as cell type, protein expression, and enzymatic activity. Some protocols have been developed that combine these two analysis methods, such as SCOPE-seq and the 10x feature barcoding pipeline[13,14]. However, when adapting these protocols for our purposes, we found that they have some limitations. These protocols are not suitable for analyzing non-poly-adenylated RNA, and have not been tested with non-mammalian cells. The 10x pipeline is able to connect a cell's transcriptomic data with some of its surface markers, but is expensive and not easily customizable. It also cannot currently be used to analyze non-poly-adenylated RNA. SCOPE-seq is better optimized for connecting transcriptomic data with cell phenotype, but also lacks the ability to capture non-poly-adenylated RNA[13]. We therefore set out to develop methods that can give us the combination of a single cell with cell surface protein information coupled with non-

polyA-RNA-sequencing information. Further, we sought to determine whether the single-cell imaging and subsequent coupling with a known barcode (to be able to connect the image to the RNA-sequencing data) could be streamlined using in-house developed devices such as microfluidic solutions for coupling primers to beads, and microwells (similar to the DropSeq protocol[11]) for rapid, high-throughput dispensing of cells.

Briefly, we show that using agarose-based beads for optical barcoding (ie. labeling individual cells with unique dye combinations) was not feasible due to the diffusion properties of aqueous dyes in agarose and buffer, resulting in unwanted mixing. We demonstrate that we can establish the necessary protocols and infrastructure for droplet generation (which can then be used to functionalize surfaces) and microwell fabrication. We also show that RNA from single cells can be successfully reverse transcribed into cDNA using random hexamer primers, depleted for ribosomal DNAs and made into a library for sequencing. In addition, we show that individual labeled cells can easily be visualized under a fluorescence microscope, and that in a 96-well format, we can determine which wells have single cells (as opposed to no cells or multiple cells). This imaging can be streamlined using an automated imager such as our CX7 high-throughput imager. In conclusion, this work has established the necessary technology, infrastructure and pipelines to query the molecular states of single cells with ease and granularity, in a customizable manner. Our next steps including sequencing our preliminary results (which will be completed before the end of the FY), and reaching out to both internal and external collaborators to identify use cases for testing our application.

2. METHODOLOGY

2.1. Device Technology

Devices to accommodate the barcoding and imaging of individual cells were proposed for this project. The workflow is shown in Figure 1. Original proposal for microfluidic devices to allow single cell isolation and imaging. Well arrays shown in step 3 are substituted with SEQ-WELL type arrays. Three microfluidic de Initial research indicated technical challenges that needed to be overcome. Each microfluidic device proposed was analyzed and optimized before beginning fabrication or testing. Device prototypes were fabricated and tested independent of the work in imaging protocols and library preparation.

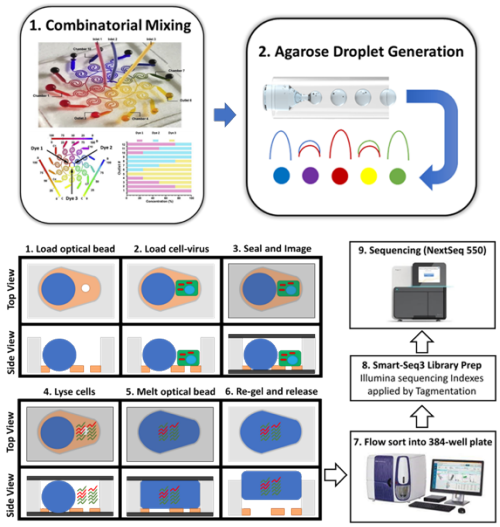


Figure 1. Original proposal for microfluidic devices to allow single cell isolation and imaging. Well arrays shown in step 3 are substituted with SEQ-WELL type arrays.

2.1.1. Dye Mixing

The development of fluorescent dye solutions with varying concentrations is necessary for the protocol that we are developing. In this FY, since we were expecting to only do initial experiments a device was not strictly necessary, as dyes could be mixed by serial dilution pipetting. This would not scale easily to larger numbers of unique dyes. Analysis and initial designs were completed for a future fluidic mixer within this project.

A device for mixing several unique dye combinations was included with this project proposal and was the basis for the devices designed for this project.

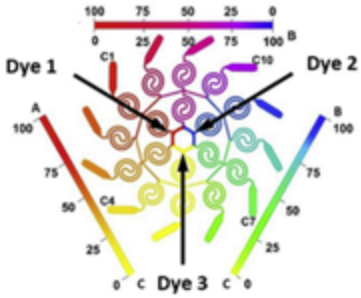


Figure 2. Combinatorial mixer using Dean flow in sequential spiral mixers.

Our device design was optimized to suite the increased number of mixing steps and a more linear set of dye ratios needed for this application. The initial design relied on Dean flow vortices produced by spiral flow paths. Flow calculations using the Dean flow equation were completed to understand the required flow rates to create fully mixed solutions. Flow vortices form when De is above ~25 and are a function of pipe dimensions, flow rate, and radius of curvature.

$$De = \frac{d\rho v\sqrt{\frac{d}{R}}}{\sqrt{2}\eta}$$

Flow rates necessary for Dean flow would require upstream pressure outside of the range of the available pumping solutions. Each mixer needs to minimize fluidic losses to allow larger arrays. Using a linear mixer made from smaller loops than would be possible in a spiral mixer allows for higher flow rates and a shorter flow path needed for complete mixing. Figure 3 shows the basic design. To generate a device capable of hundreds of unique dye combinations the pattern shown below would be copied concentrically outwards.

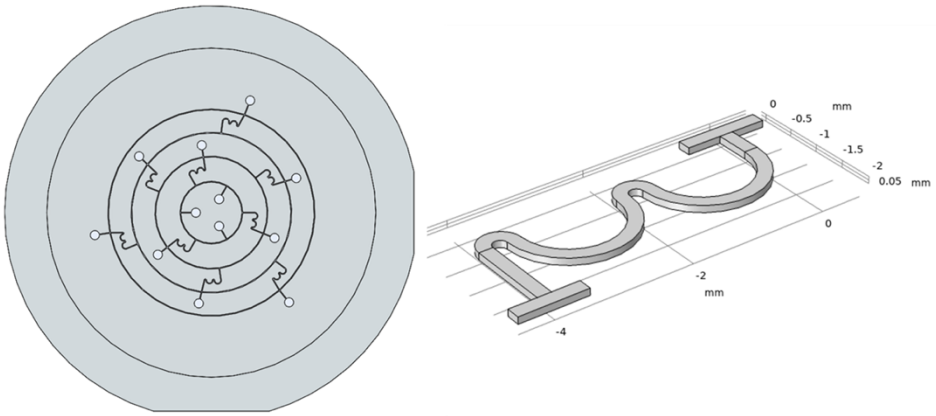


Figure 3. (left) Design approach for a linear combinatorial mixer. Additional mixing steps resulting in more color combinations can be achieved by adding more concentric patterns. (Right) The Dean flow based mixing element used in the combinatorial mixer.

COMSOL Multiphysics® models were generated using transport of dilute species and laminar flow multiphysics. Flow inlets were parameterized by flow rate and outlets were set to 0 psi. Second order discretization of velocity gradients was used in model setup. Fluid and diffusion parameters used were for water and food dyes, as this would be the first tested analog.

2.1.2. Bead Generation

Beads were made from agarose by microfluidic droplet forming devices. Melted 2% agarose solutions (Sigma 39346-81-1) and mineral oil with 1% SPAN 80 (1338-43-8) are flowed through a T-junction and the formation of beads occurs at the pinch point between flow paths. The agarose would quickly gel if the experiment were run at ambient lab temperatures, so the entire flow system is stored in an oven at 40-60C. To facilitate this, a pressure-controlled flow system is used (Elveflow OB1 controller). This allows the drive system to be external to the oven. The flow chip is placed against the glass door of the chamber, allowing for imaging of the droplet formation. Imaging is accomplished with a 40x microscope objective and a simple CMOS sensor from Amscope (MU900). A positional stage allows for quick focusing and alignment to the fluidic chip after opening and closing the oven. Figure 4 shows the experimental setup. Initially, commercial off the shelf droplet chips were procured from chip shop. These proved to be unsuited to the prolonged elevated temperatures, pressures, and mineral oil. Delamination occurred after approximately 3 hours of time in the oven.

Figure 4. Microfluidic chip in oven and imaging system to produce agarose droplets. Pressure based fluidic control allows drive hardware to remain external to the oven.

To create a more durable fluidic chip, design and fabrication was done in house using SU-8 photolithography. Chips were assembled from Dow Corning Sylgard 184 Silicone and Glass, making them more suited to the temperatures and working fluids. Figure 55 shows the fluidic chip design.

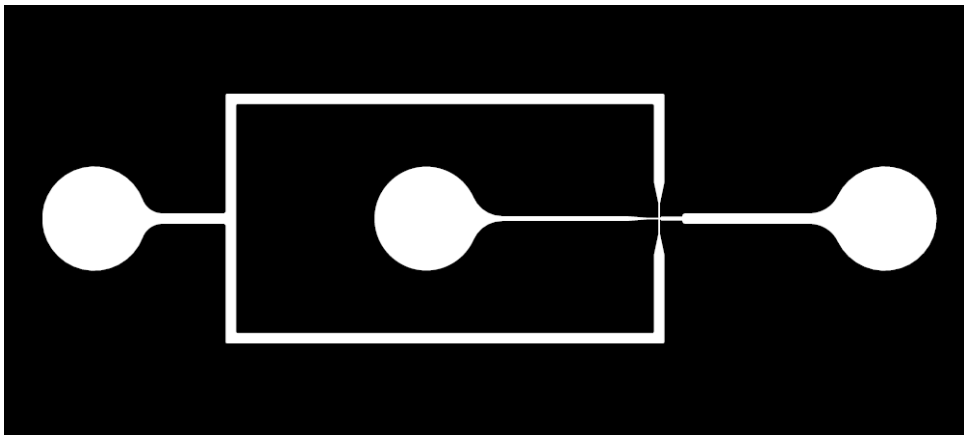


Figure 55. Microfluidic based droplet generator capable of making agarose beads in the 40um range. Mineral oil is flowed into the left port. Melted agarose is flowed into the center port. Droplets form at the T-junction and flow into the collection well through the right port.

2.1.3. Dye, primer, bead fixation

Initial work with dying agarose was carried out to understand the response curves for concentration and fluorescent signal, while referring to prior work in this area[15-17]. Sulforhodamine B (SulfB) and carboxyfluorescein (CF) were selected as test dyes for these experiments. A Tecan plate reader was used to analyze the fluorescence intensity of different concentrations of SulfB and CF. The Tecan plate reader was able to distinguish between different dye concentrations, and appeared to be most sensitive in the 10-100 µM range. There was also minimal overlap between the two dyes.

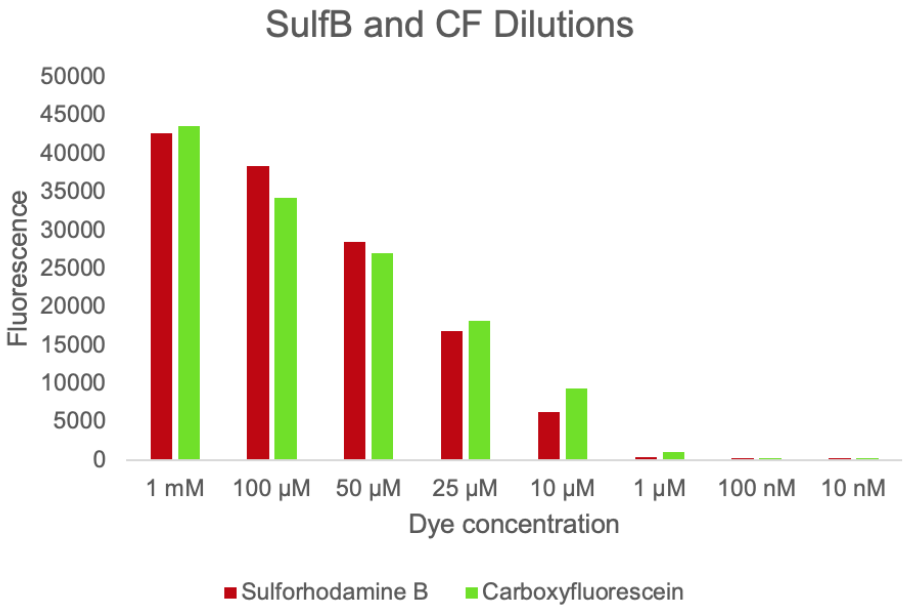


Figure 66. Concentration and fluorescent signal response for two fluorophore dyes. Measured using Tecan plate reader.

5 mm agarose pucks were formed and incubated in different concentrations of SulfB and CF overnight, then analyzed using the Tecan plate reader. The agarose pucks absorbed the dye and had fluorescence readings. There also appeared to be a linear relationship

between dye concentration and fluorescence intensity for both SulfB and CF (Figure 6, 7).

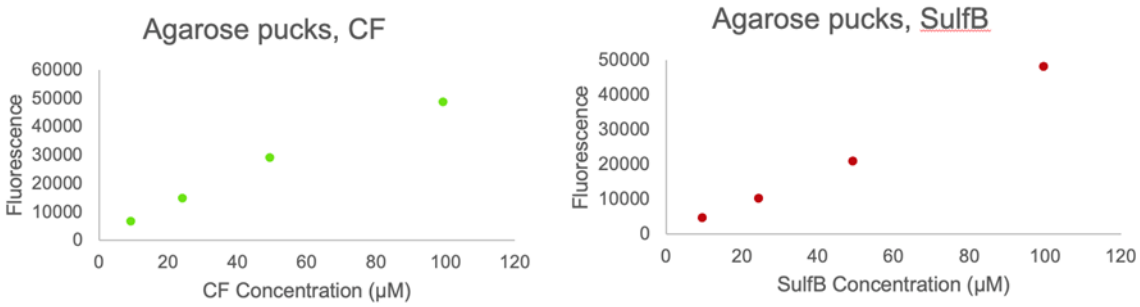


Figure 77. Agarose dye uptake after submersion in a dye solution for two fluorophore dyes.

The largest problem identified during the research stage of the protocol was the fixation of dyes and primers to the agarose beads. Even for highly concentrated agarose beads, pore size of the agarose will allow the diffusion of dye molecules. This is beneficial for loading dyes into agarose, but detrimental when working with the beads in a secondary solution. COMSOL modeling using the transport of diluted species module was used to predict the diffusion times based on bead density, dye concentration, and temperature. Solutions involving lowering temperature or changing the protocol to minimize the time a bead spends outside of the dye solution were investigated. Two types of models were set up. One investigating dye diffusion from a bead within a bulk solution, and one with the bead contained in a PDMS well. Both models used axisymmetric geometries, with initial concentration of dye in the bead parameterized. The well model contained a boundary layer of PDMS with lower diffusion coefficient while the bulk solution model had a zero-concentration boundary condition. Values for dye diffusion coefficients were parameterized around those found in literature[15-17].

2.1.4. Well arrays

The design of the well arrays was initially intended to follow the designs of the original Seq-Well protocol[12]. Labeled beads and cells are flowed onto a nano-well array that facilitates capture of one bead and one cell, to then be lysed and reverse transcribed with the bead label. This device is optimized to work with beads that can have millions of unique barcodes. The dye mixing strategies in this project are limited to combinations in the hundreds to thousands depending on the number of dyes used. This led to the redesign of the well array to allow for fewer unique barcodes to be used. Figure 8 shows the design of the well array. PDMS is molded onto an SU-8 structure to produce sets of microwells.

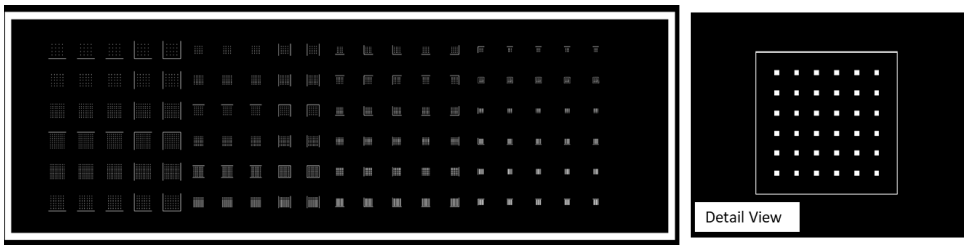


Figure 8 8. Photomask of the well arrays proposed for use in this project. By limiting the number of nano-wells within each array (right) the number of unique dyes needed for this experiment could be reduced from millions to ~36. The number and spacing of nano-wells within an array are parameterized as loading efficiency is unknown.

This layer is bonded to a glass slide for easier handling. A secondary structure is used to mask each set of wells from neighbors allowing filling of multiple cells into each set. The PDMS wells were fabricated using SU-8 photolithography in the ABL microfab laboratory. The secondary structure was molded from PDMS in a 3D printed mold. The top surface of the mold was acrylic to ensure the surface would be sufficiently smooth to temporarily adhere to the well array (Figure 9).

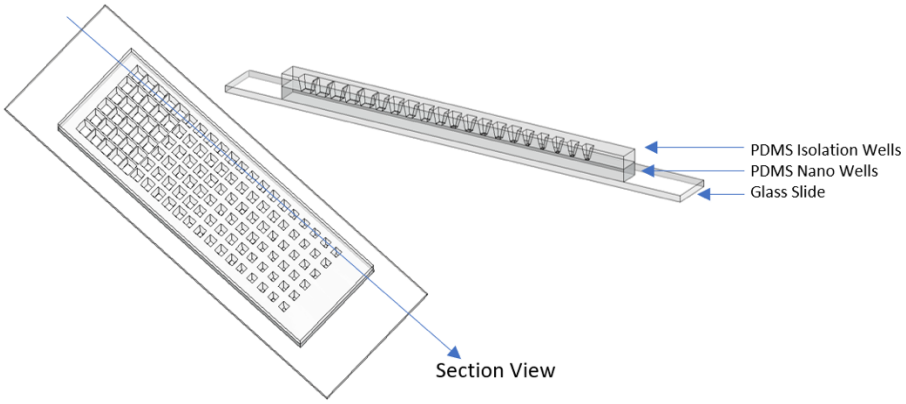


Figure 99. Assembled well array CAD. The glass slide and PDMS nano wells are permanently bonded. The PDMS Isolation wells can be temporarily adhered to allow pipetting of samples onto individual arrays without cross contamination.

Although not completed in this project, the proposed use of the arrays is to load cells and beads into each well. After cells and beads are loaded and allowed to settle to the wells, the secondary structure is removed, and the array is closed with a sheet of polycarbonate film. The array can be imaged from the bottom to determine location of each barcoded bead. At which point the agarose is remelted and cells lysed to allow for reverse transcription reactions to take place.

2.2. Single-cell laboratory protocols

The identification of single cells was achieved via imaging using fluorescent microscopy on diluted cell suspensions. Methodology for obtaining these isolated cells was achieved through the trail of multiple different stains, concentrations, volumes, and well types. Described here is the optimized procedure concluded from these trials (Figure

10). The cell used in this were fully confluent human bone marrow mesenchymal stromal cells. The cells obtained for imaging and subsequent amplification were fluorescently stained using 1X CellMask Deep Red Plasma Membrane Stain. Once fluorescently dyed, the cell suspension was diluted in Phosphate-buffered saline (PBS) to an optimal concentration of 1000 cells/ml. Obtaining a 96-well plate, cells were then transferred in 2 µL droplets per well, with a reference well of 50 µL. At this point each well could be imaged, using a CY5 fluorescent microscope. To identify singlet wells and properly view cells, the reference well of 50 µL was imaged first as a baseline, then the rest of the wells. Once the single cell droplets were identified, two methods were tested for a mode of verification of successful transference of these cells (which will be necessary for amplification). These methods still require further troubleshooting for optimization. The first method was pipetting up the 2 µL cell droplet, then transferring that whole volume to a new well for imaging. The second method was to flush the single cell droplet with 3 µL of transfer the total volume to a new well for imaging.

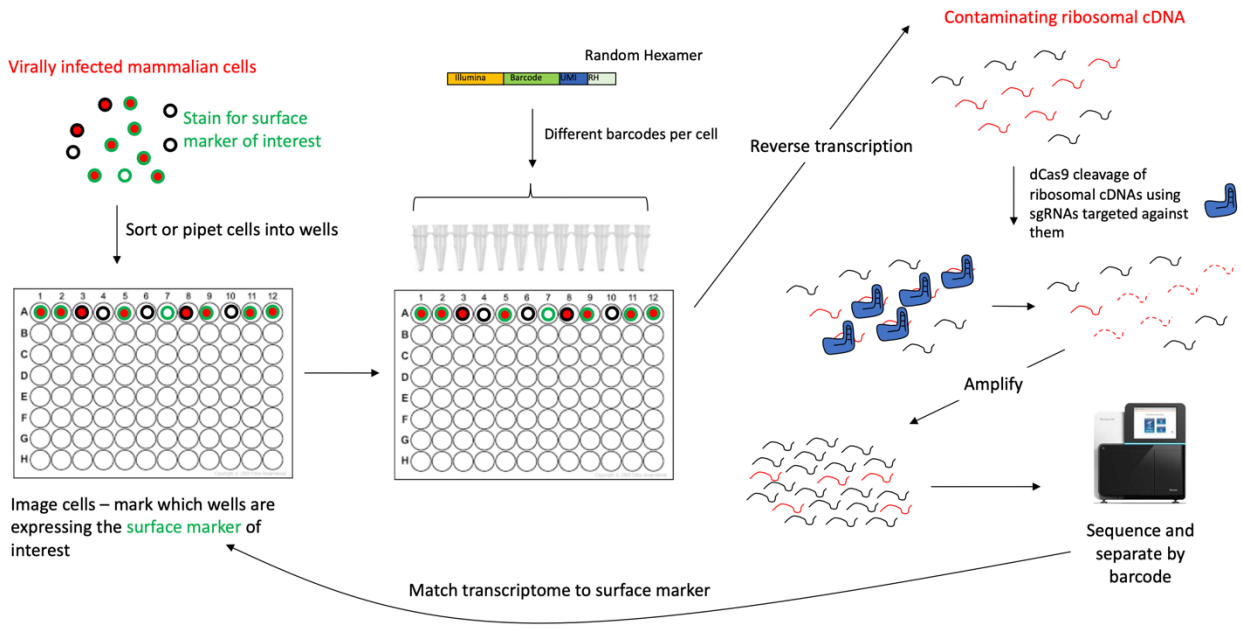


Figure 1010. Project library preparation workflow.

2.2.1. Primers and reverse transcription

Purified mouse lung epithelial cell (LA-4) RNA was used for all optimization experiments to reduce variability related to single-cell differences. Experiments were performed in strip tubes in order to test multiple conditions pers experiment. 20 pg of RNA was added to each strip tube, along with 1 µL of 10 mM dNTPs and 1 µL of 10 µM reverse transcription primer. Custom random hexamer primers containing a cellular barcode and unique molecular identifier (UMI) were designed and used for these experiments. Both barcoded and unbarcoded oligo dT primers were used for positive controls samples. Non-template control (NTC) samples contained water instead of RNA. Samples were mixed thoroughly and briefly centrifuged, then were incubated at 72°C for 3 minutes, 42 °C for 2 minutes, and finally placed on ice.

Reverse transcription mix was prepared and distributed between samples, with each sample receiving 2 µL of 5x Maxima Buffer, 0.25 µL of RNAse Out, 0.75 µL of 20 µM template-switching oligo[8], 2 µL of 25% PEG 8000, 0.5 µL of 200 U/µL Maxima H minus reverse transcriptase, and 1.5 µL of nuclease-free water. Samples were vortexed and centrifuged, then heated using the thermal cycler program described in Table 1.

Table 1. Reverse transcription thermal cycler program

Temperature	Time	Cycles
42°C	90 min.	1
50 °C	2 min.	10
42 °C	2 min.	
70 °C	15 min.	1

PCR mix was optimized to improve template amplification. Early experiments used a PCR mix that provided each sample with 4 µl of 5x GC buffer, 0.4 µL of 10 mM dNTPs, 2 µL of 1 µM PCR oligo[8], 0.2 µL of Phusion high-fidelity DNA polymerase, and 3.4 µL of nuclease-free water. Samples were heated following the thermal cycler program described in Table 2. Phusion DNA polymerase thermal cycler program. Later experiments used a PCR mix that provided each sample with 2 µL of 1 µM PCR oligo and 12 µL of 2x KAPA HiFi HotStart ReadyMix. Samples were heated following the thermal cycler program described in Table 3.

Table 2. Phusion DNA polymerase thermal cycler program

Temperature	Time	Cycles
98°C	3 min.	1
98°C	20 sec.	18
67 °C	15 sec.	
72 °C	6 min.	
72 °C	5 min.	1

Table 3. Table 3. KAPA HiFi HotStart ReadyMix thermal cycler program.

Temperature	Time	Cycles
95°C	3 min.	1
98°C	20 sec.	18
67°C	15 sec.	
72 °C	6 min.	
72 °C	5 min.	1

2.2.2. Ribosomal depletion

After amplification, samples were cleaned using a Ampure XP beads. Beads were added to each sample in a 1:1 ratio, and allowed to incubate for 8 minutes at room temperature. Samples were then moved to a magnetic stand for 5 minutes. 30-36 µL of the resulting supernatant was then removed. 150 µL of 80% ethanol was added to each sample and allowed to incubate on-magnet for 30 seconds, and was subsequently removed. This wash was then repeated once more. After removal of the second ethanol wash, the samples were covered and the beads were allowed to dry for 5 minutes. The samples were then removed from the magnet, and each sample's beads were resuspended in 17.5 µL of nuclease-free water. Samples were incubated at room temperature for 2 minutes, then placed back on the magnet for 2 minutes. 15 µL of the resulting supernatant was removed and placed into a clean tube. The DNA concentration of the resulting sample was quantified using a Qubit 4 fluorometer with the dsDNA high-sensitivity kit.

Each sample's ribosomal cDNA was then then digested using the scDASH method described by Loi et al[1]. The Cas9 digestion step was used as originally written, but the subsequent PCR enrichment step was eliminated to help reduce the chance of jackpotting and prevent the introduction of bias.

3. RESULTS AND DISCUSSION

3.1. Dye mixing

The COMSOL models for dye mixing are shown below in Figure 11 and Figure 12. The models were used to understand the minimum pressure drop needed to induce Dean mixing and ensure uniform mixing of the two dyes at the inlet. Verification of these models is still needed in fluidic devices; however initial values show that flow rates of 200mm/s could be achieved with 6 kPa upstream pressure This flow rate results in secondary flow vortices that are able to induce full mixing of the solution. Figure 1111 shows the formation of secondary vortices and the velocity profiles in the cross section of the first bend. Figure 1212 shows the concentration gradient across the flow path of an individual mixer.

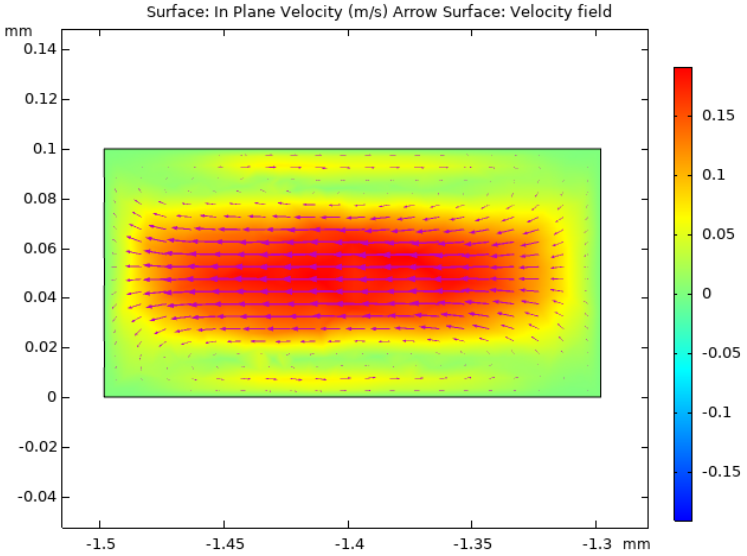


Figure 1111. A cross section view of the flow channel in the first bend. Inertial effects on the flow stream create two counter rotating vortices perpendicular to the channel that aid in mixing.

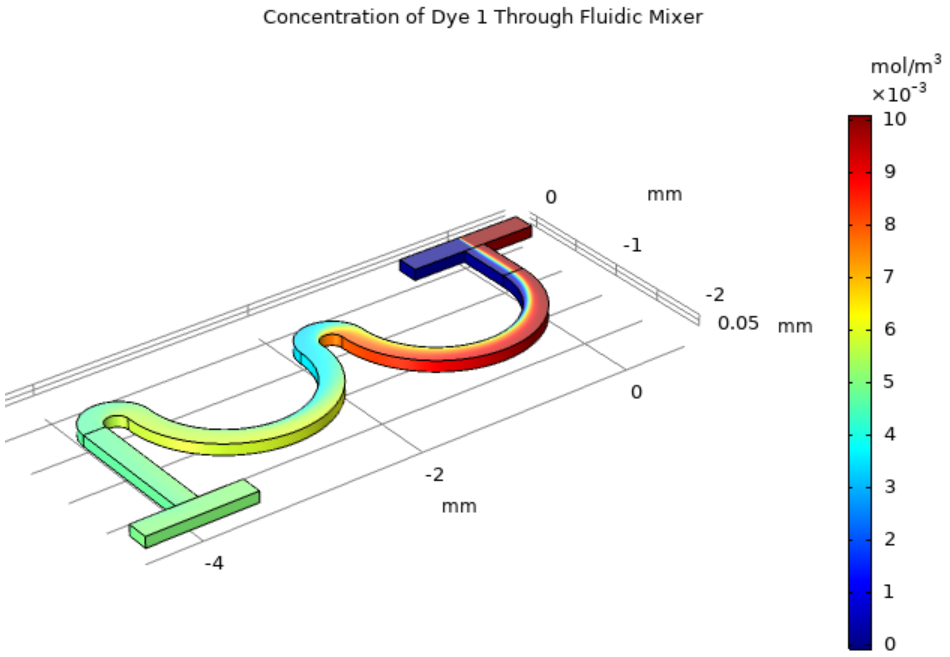


Figure 1212. COMSOL demonstration of a fluidic mixing device for a combinatorial mixer. Concentration of dye 1 loaded into the right inlet is shown. Fluid is mixed within 1.5% at the exits.

3.2. Bead generation

Beads were generated with both COTS mixers and in house fabricated fluidic channels. Figure 1313 shows the generation of agarose beads using the ChipShop 50um fluidic chip. Figure 1414 shows the in house fabricated fluidic chip before and after bonding. Because of delamination in the initial fluidic chips, and the late stage at which in-house developed devices were available, the downstream processing of beads that would include washing and filtering was not completed. Analysis of bead size and morphology was not completed. It is noted that the pressure-based control system allowed for a very wide range of flow rates that would produce droplets. The optimization of the process to ensure droplets were as uniform and as close to 40um would involve varying the flow rates. Having a wide window of flow rates relative to the step size of flow control will make the optimization process easier if this work is continued.

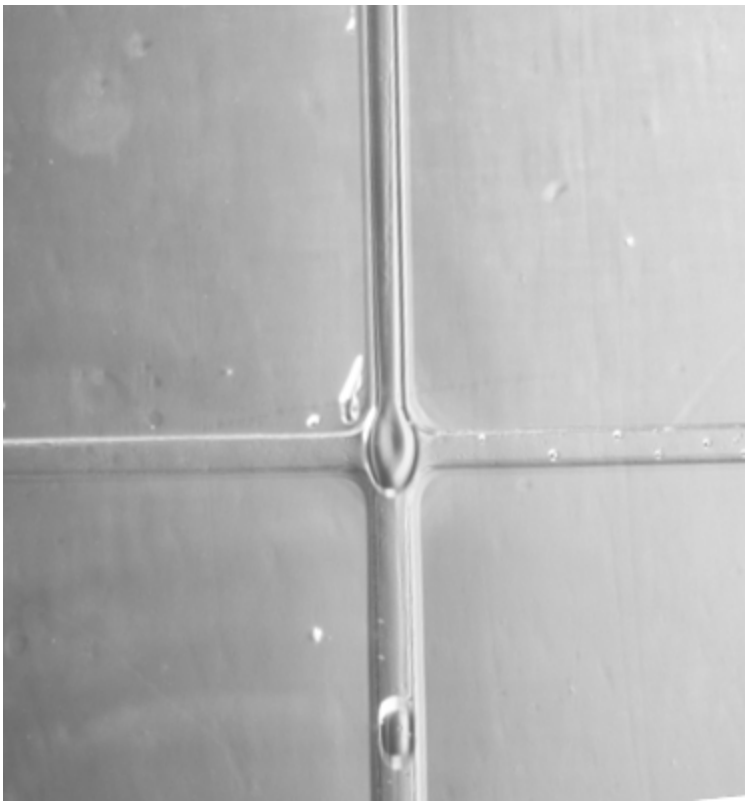


Figure 1313. Droplet formation of melted agarose within mineral oil. Channel dimensions are 30um.

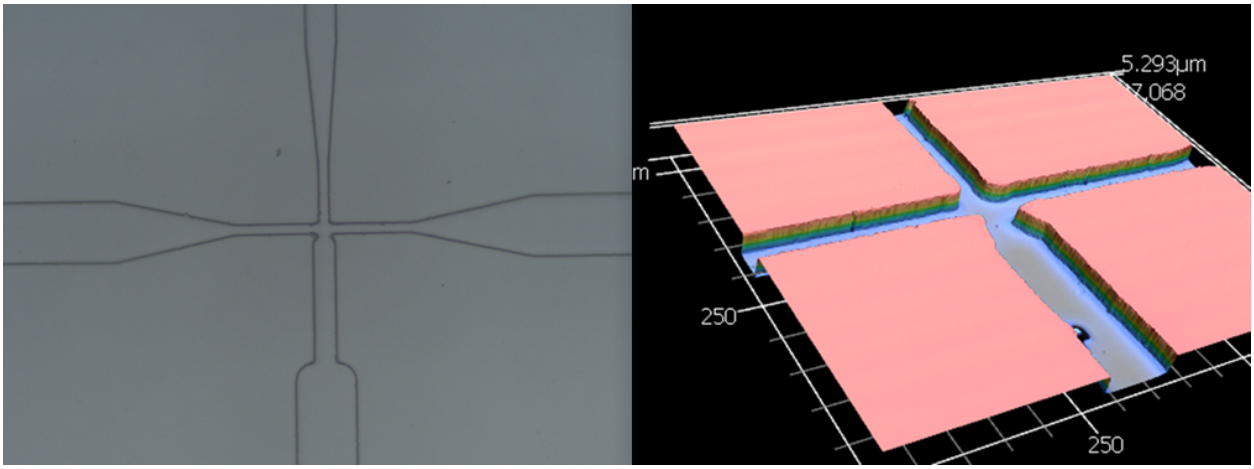


Figure 1414. T-junction of in-house designed and fabricated microfluidic droplet generator. (Left) Optical image. (Right) 3D confocal laser image of the molded PDMS before assembly.

3.3. Dye/primer fixation

COMSOL model results from the study of dye diffusion are shown in Figure 1515. The model shows the dye diffusion time scale on the order of seconds. Since the beads must

be mixed into solution with multiple other beads there is cause for concern if the bead dye is diffusing out before imaging can begin.

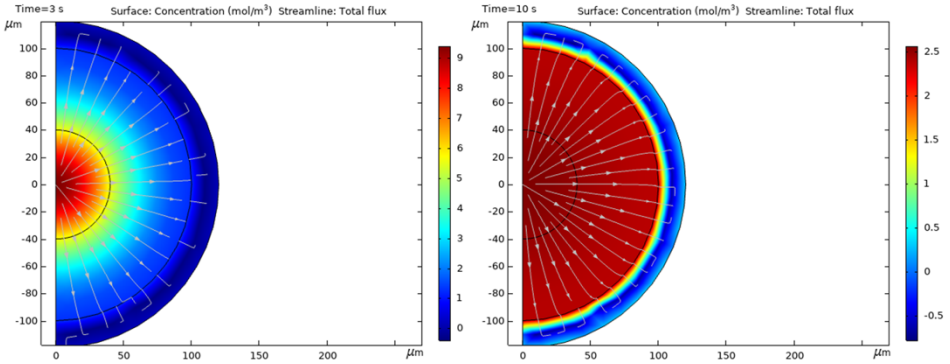


Figure 1515. COMSOL model of a dyed agarose bead in a microwell made from PDMS. (Left) Dye concentration after 3 seconds. (right) Dye concentration after 10 seconds. The concentration equalizes by diffusion out of the agarose rapidly. The PDMS is more impermeable and does not allow further diffusion. A bead in a bulk solution would equalize to the bulk concentration on similar time scales.

An alternative strategy is proposed for future work. Literature review has shown that agarose beads are commonly used in microfluidic devices where the attachment of proteins, small molecules, dyes or transcription primers are needed. Previous work was used to devise a solution for attaching primers that can be used to affix fluorophore dyes as well. Figure 1616 shows the general approach[18]. R-groups referenced are dye or primers.

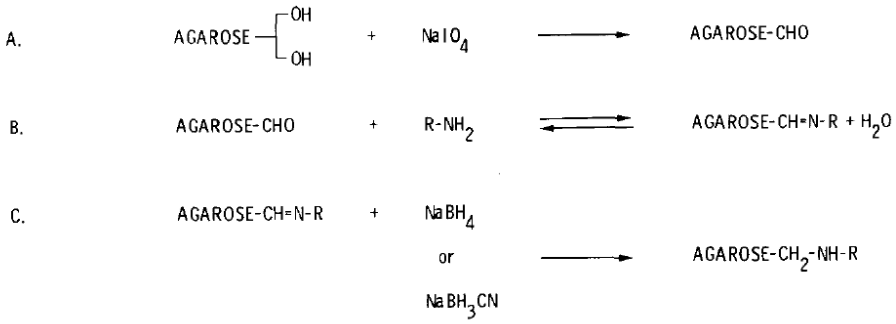


Figure 1616. Chemical process of affixing amine functionalized dyes to agarose beads.

For this application a primer and dye are selected that both are functionalized with a primary amine group (Sigma Aldrich Atto 647N Amine Dye). The agarose bead is oxidized by sodium metaperiodate to generate aldehyde functional groups, these react with the dye and primers and are reduced with sodium borohydride to form stable secondary amine linkages. This approach of functionalization should allow permanently functionalized beads that can be used without diffusing dye or primer from one well in the array to another.

3.4. Well arrays

Well arrays were fabricated from SU-8 molds. Figure 1717 shows a scanning laser confocal image of the array of SU-8 posts that make up the mold. Figure 1818 shows an optical image and confocal height map of the PDMS micro well arrays fabricated.

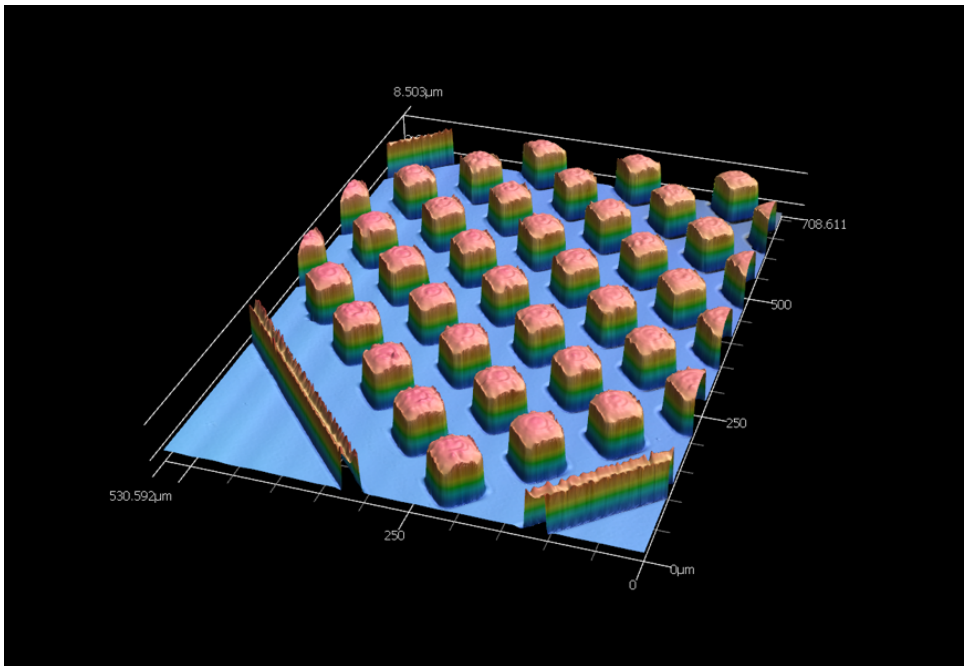


Figure 1717. 3D image of the SU-8 mold used to make the nano well arrays.

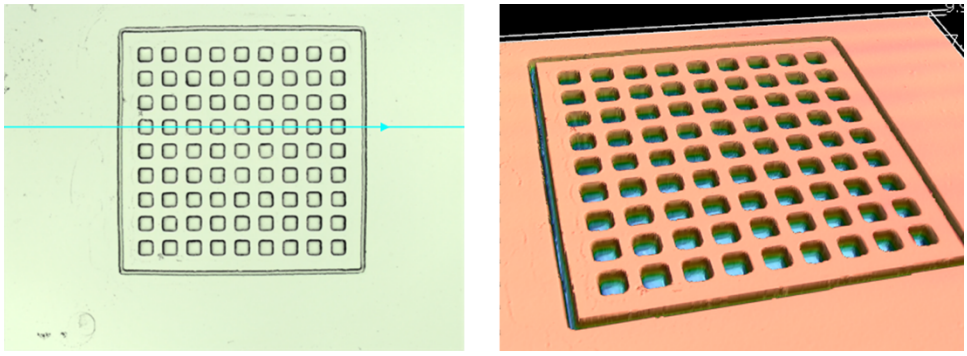


Figure 1818. Optical (left) and Laser map (right) of the PDMS well arrays. Features are replicated from the SU-8 mold into PDMS faithfully.

The process development for SU-8 lithography had multiple pitfalls. Although we had previous experience with the techniques; the facilities in ABL were new to us and required some trouble shooting as there was not an up-to-date protocol. Adhesion of the high aspect ratio posts of these arrays required careful optimization of the pre-bake and post-bake steps as well as surface preparation by oxygen plasma and baking out moisture from the wafer before beginning. The wells were not tested with cells in this project due to time constraints. The efforts of developing the SU-8 protocols and molds will be useful for future projects

Since there were significant delays in the well-array fabrication, it was necessary to investigate other techniques for single cell isolation. These techniques would also be necessary for the isolation of individual dyed beads, so that mixtures of different colors could be made with little or no repeated colors.

3.5. Single-cell imaging

In order to properly sequence a single cell, it must first be individually isolated for amplification. To accomplish this, we've refined existing protocols for single cell isolation via imaging with fluorescent membrane stains. The aim here was to optimize concentration, volume, well-type, and stain used. As previously stated, we've optimized the concentration at 1000 cells/ml, fluorescently dyed using CellMask membrane stain, with 2 µl pipetted into 96 well plates, imaged against a refernce well of 50 µl. The 2 µl volume droplets at this concentration most consistently provided singlet wells in comparison to other volumes and concentrations tested, providing approximately 67% with a single cell. As for the size of the wells, while a 384 well plate is applicable, we found that the use of a 96 well plate was more suited for viewing a whole droplet which in turn made identification of a single cell much more efficient (Figure 19). The CellMask membrane stain was sufficient in identifying when a cell is present, therefore a more targeted stain is not necessary for this type of imaging.

Additionally, we tested a few methods in transferring a single cell droplet into another well to mirror what would be done in transferring it to a PCR plate. The transference to another well for secondary imaging was to ensure that the cell can successfully be transferred to a PCR plate without damage to or loss of the cell. Primary methods tested were a) transferring the whole 2 µl droplet, or b) flushing the 2 µl with 3 µl to transfer a total 5 µl suspension containing the isolated cell (Figure 1919). These methods require much more testing to optimize transfer, although preliminary tests suggest the latter to be optimal in preserving the cell while being transferred.

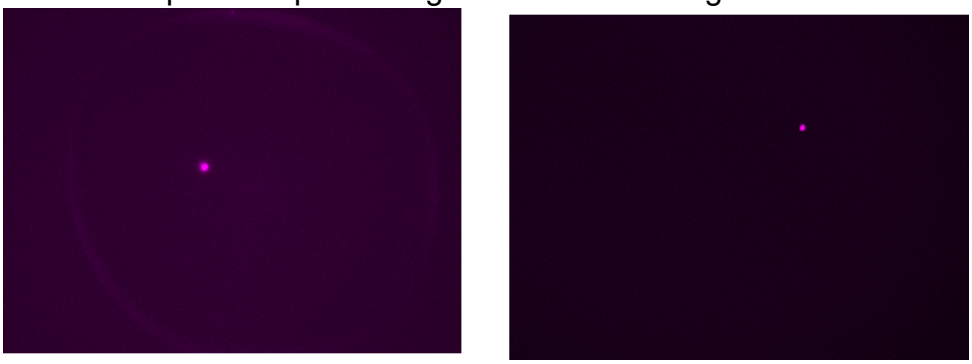


Figure 1919. Fluorescent images of individually-labeled, detached cells. (Left) A fluorescently stained human mesenchymal stromal cell suspended in a 2 µl droplet of a 1000 cell/µl concentration. The clear droplet shape can be observed, which aids significantly in identifying the placement of the cell within a 96 well plate. This is one of multiple images depicting a successful single cell isolation at this volume and concentration. (Right) An isolated cell imaged post-transference via the second method detailed above. Although the droplet formation is lost in the 5 µl volume, that one most frequently produced a successful transfer.

3.6. Library preparation protocol

Our custom-designed random hexamer primers were tested using the protocol described above, and were compared side-by-side to oligo dT primers (Figure 20). RT-qPCR results showed that our random hexamer primers could be effectively used for reverse transcription. Though not as high as oligo dT, likely because many of the sequences being amplified (and therefore using up reagnets/primers) are ribosomal.

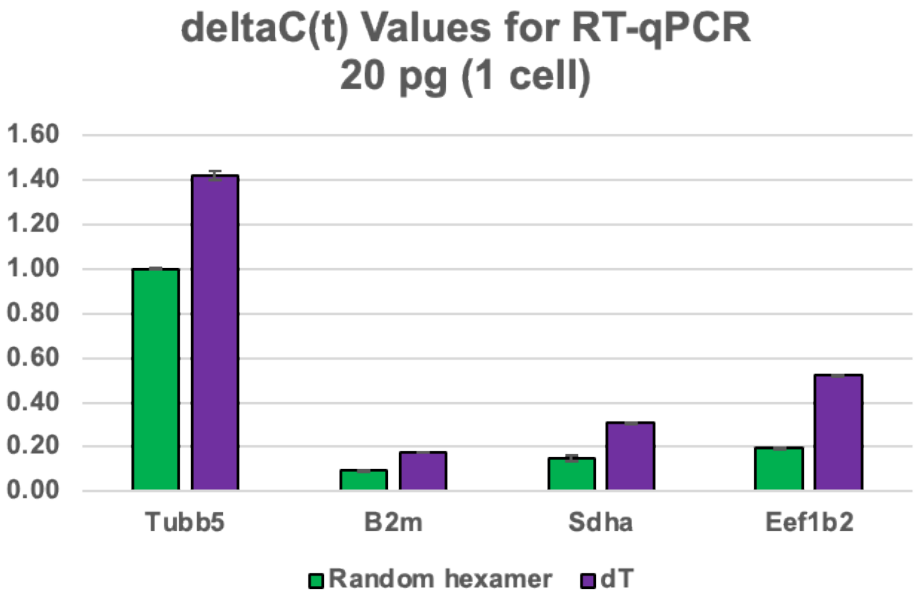


Figure 20. 20 RT-qPCR results for samples using either random hexamers or oligo dTs for reverse transcription priming. Four housekeeping genes were selected.

Though our custom random hexamer primers could be used to successfully generate cDNA from purified RNA, the cDNA yield was lower than expected. This trend was also present in the oligo dT samples, which indicated a problem with the general protocol rather than the primers. In an attempt to troubleshoot this issue, all reagents were replaced, and our starting RNA sample was analyzed for quality. However, the RNA sample was found to be of high quality, and the new reagents had no effect on cDNA yield. After a series of experiments, we discovered that using the KAPA HiFi HotStart Ready Mix instead of the Phusion polymerase kit significantly improved cDNA yield in all samples (Figure 2121). The addition of PEG 8000 before reverse transcription was also found to significantly improve cDNA yield by creating a molecular crowding effect (Figure 22.22).

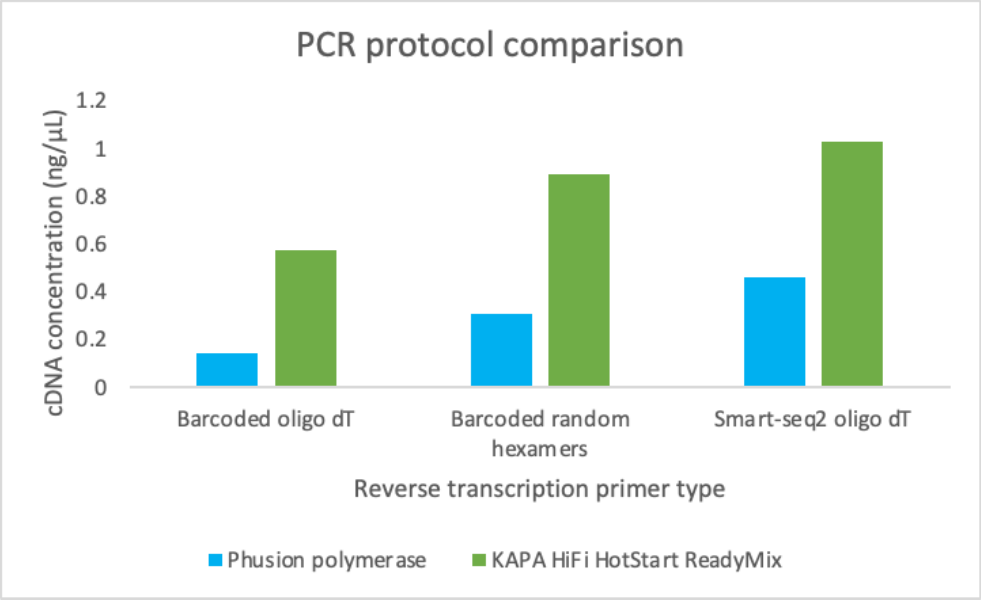


Figure 2121. Comparison between samples amplified using the Phusion polymerase kit and samples prepared using KAPA HiFi HotStart ReadyMix.

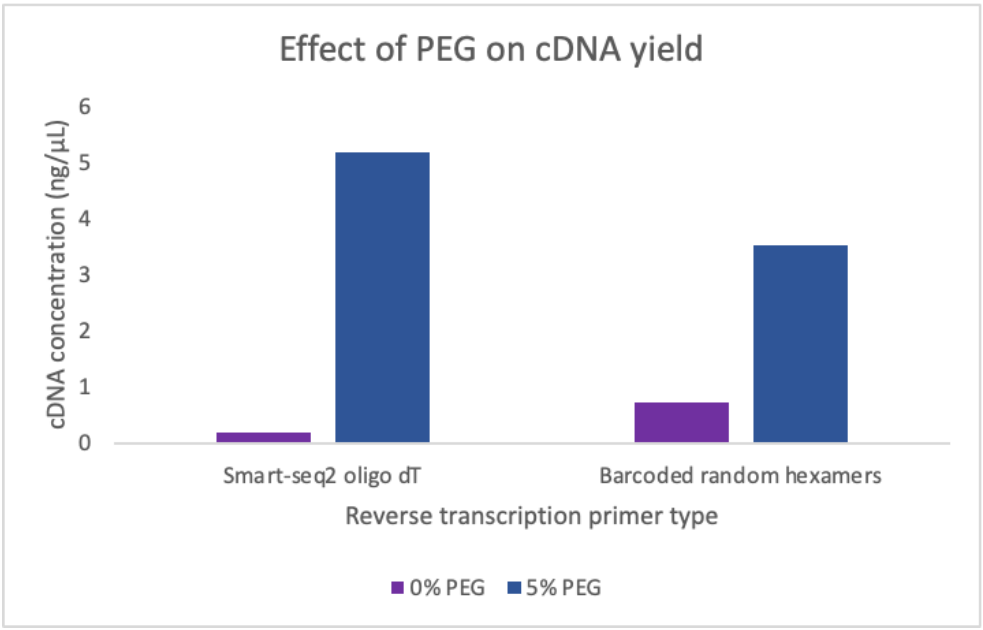


Figure 22.22 Comparison of samples prepared with and without PEG 8000.

One area for troubleshooting in our protocol is our non-template control (NTC) samples. During quality control checks, we found that NTC samples containing no RNA template consistently had nonzero cDNA concentrations after RT and PCR. In some experiments, the cDNA concentrations of these samples looked similar to or even higher than the cDNA concentrations of template samples. While these results are concerning, we found that qPCR analysis revealed minimal amplification of meaningful transcripts in the NTC samples **(**Figure 2323). Further troubleshooting is required to fix this issue, but the qPCR results indicate that it would still be worthwhile to proceed to the digestion step

of the protocol. Sequencing will reveal for certain whether or not there is contaminating DNA in these NTC samples. One theory is that our random hexamer primers are sticking together and forming long strands of DNA that can survive Ampure XP bead cleanup.

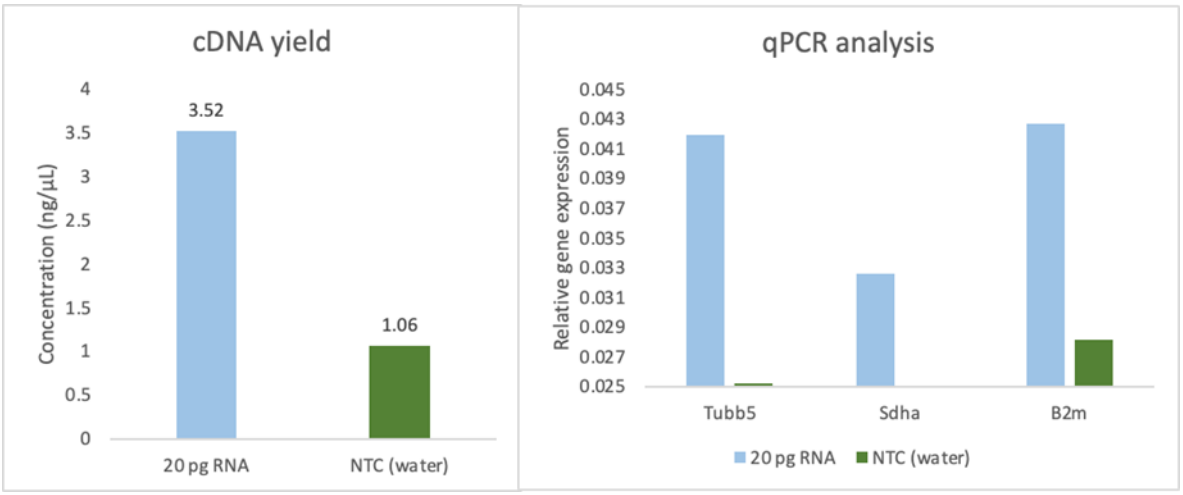


Figure 2323. Comparison of the cDNA yield and qPCR amplification results of samples containing either 20 pg of purified RNA or water.

Once a successful protocol had been established, we tested the CRISPR digestion step of our protocol. We followed the exact protocol described in the scDASH method developed by Loi et al[1], but did not perform the PCR enrichment step immediately after digestion. We decided to rely on the PCR step required for library prep to enrich our undigested transcripts rather than adding an additional amplification step in order to reduce the risk of jackpotting and PCR bias. The CRISPR digestion was found to decrease the amount of 18S cDNA in each sample, but not as much as anticipated. The oligo dT samples also appeared to have ribosomal cDNA in them, which also indicates that further troubleshooting is required **(**Figure 2424). Oligo dT samples theoretically should not have any ribosomal cDNA in them, since oligo dT primers only attach to mRNA transcripts. However, this may be a problem with the qPCR analysis step rather than the actual digestion reaction. Only the 18S primers produced amplification, and not the 5.8S or 28 S primers; this indicates that the problem may be with our qPCR primers. Testing different annealing temperatures would likely be a good first step in troubleshooting this problem. If this is unsuccessful, new primers could be designed that are more specific to these ribosomal cDNA sequences of interest.

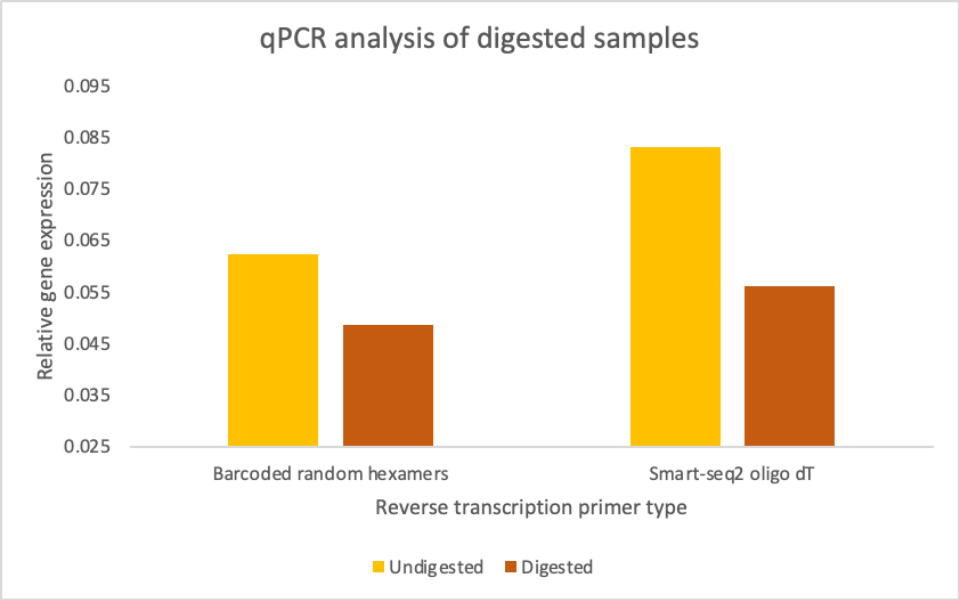


Figure 2424. qPCR analysis of undigested and digested samples.

4. ANTICIPATED OUTCOMES AND IMPACT

Our next steps are to try this protocol with a pathogen-infected cell type. While this was the original goal of this proposal, due to the extensive troubleshooting required to optimize the protocol, we were not able to conduct this experiment within the time frame. We are however, poised to be able to do so as soon as we have some follow-on funding. Part of this next step is also to determine whether the devices we tested can be used to streamline any of the preparation processes (or whether we can identify any commercially-available devices that we may be able to deploy). Our goal is to use either a virally-infected or bacterially-infected lung fibroblast cell line, as these are systems that are well-established at Sandia and are currently being used. We are in conversations with other Bioscience staff to determine the best next steps to conduct this testing.

We anticipate that with some additional troubleshooting, this protocol will be able to use materials and technology that is readily available to solve problems that currently plague diagnostics and countermeasure development for a range of pathogens. Specifically, we will be able to determine how surface markers on cells impact mammalian gene expression across a range of cell types, and in addition, for intracellular pathogens, we will be able to simultaneously understand how pathogen gene expression interplays with host gene expression. This will allow us to focus on specific surface proteins and downstream pathways when determining susceptibility of individuals, detecting and diagnosing infections, and identifying novel therapeutic targets.

In addition, the volume of data that is generated across multiple dimensions (sequencing and high-throughput imaging) provides potential for training algorithms for automated prediction in the future. As discussed later in this section, our goal is not only to understand correlations between pathogens, host surface proteins, and gene expression, but to use this information in a predictive way such that when a new pathogen emerges, we have the data and computational tools available to rapidly determine how to detect and counter the novel threat.

Our goal is to find follow-on funding to finish the troubleshooting so we can publish our findings early next FY. We are perusing opportunities at NIH and DTRA to follow up on this work using a number of avenues. Firstly, we anticipate that this work will be directly relevant to our cell therapy project, which is being reviewed at NIH. Specifically, our cell therapy project (born from a previous bioscience LDRD) aims to engineer cell surfaces in adult mammalian cell types to regulate their responses too specific pathogens. We focus on mesenchymal stromal cells (MSCs) that have highly secretory and immunomodulatory properties, and we have previously demonstrated that engineering the cell surface of MSCs can increase their antimicrobial activity[4]. We believe that this technology will allow us to directly link cell surface composition with intracellular molecular changes, which is part of the larger goal of that project. In addition, we are currently growing our single-cell-based predictive biology capabilities. Specifically, based on single-cell data, can you predict whether individual cells have certain properties, and by inference, can you identify markers to look for in cell populations that are indicative of these properties (allowing cell-based therapies to be used in a more streamlined and reliable fashion). We have previously engaged with DTRA in the space of predictive biology using multi-omic data,

and our hope is to follow up with them using our new single-cell data, showing the increased breadth of our predictive biology capability as a result of this work.

This technology is not limited to use in pathogen-infected mammalian cells. While the current work was focused on optimizing the protocol for mammalian cell lines, this type of information from single-cell information can be useful in several areas that are of interest to DOE and NNSA. Other cell types including plants, algae and fungi, which are extremely important for the emerging bioeconomy are often decimated by pathogen infection, and understanding how to detect and prevent these types of infections is critical for the future of sustainability. In addition, exposure to chemicals or radiation can cause significant changes in human, animal or plant cells among others, resulting in loss of health and/or decimation of agriculture, all of which are extremely consequential to national security. This includes nefarious as well as accidental exposures, and methods such as the one we have developed that can easily be adapted for all these purposes are a critical part of the national infrastructure required to respond to such events. We anticipate some work will be required to adapt these We are engaging with people working at the Joint BioEnergy Institute and Agile Biofoundry to understand how we can use these methods in biomass/biomanufacturing efforts within and outside of Sandia.

5. CONCLUSION

In conclusion, we have established the foundations of a protocol that can be adapted for visualizing surface protein expression on single cells, and connecting this with their transcriptome. Specifically, we have achieved the following:

- Visualize single cells that are labeled to specifically detect the expression of one or more surface proteins, move them to then match them with a specific barcode in single-cell sequencing
- Redesign novel random-hexamer-based barcoded primers that can reverse-transcribe single-cell RNA in a quantitatively equivalent manner to oligo dT
- Some depletion of ribosomal cDNA by using a CRISPR-based depletion method (with some additional troubleshooting pending)
- Generate agarose beads using a droplet generator that can be functionalized with primers
- Manufacture well arrays in-house to automate the upstream part of this process

The next steps include putting together the various parts of this protocol, finishing the troubleshooting on a few of the steps, and performing a sequencing run to test the protocol. We are currently conducting a sequencing run to test the reverse transcription and depletion (though the results will not be generated soon enough for this report).
We believe that in the future, this pipeline can be used to interrogate all types of cells that have either encountered pathogens (intracellular or extracellular), or any other environmental changes that are sensed by their surface proteins and reflected in their transcriptome.

REFERENCES

1 Loi, D. S. C., Yu, L. & Wu, A. R. Effective ribosomal RNA depletion for single-cell total RNA-seq by scDASH. *PeerJ* **9**, e10717, doi:10.7717/peerj.10717 (2021).

2 Lin, W. N. *et al.* The Role of Single-Cell Technology in the Study and Control of Infectious Diseases. *Cells* **9**, doi:10.3390/cells9061440 (2020).

3 Huang, W., Wang, D. & Yao, Y. F. Understanding the pathogenesis of infectious diseases by single-cell RNA sequencing. *Microb Cell* **8**, 208-222, doi:10.15698/mic2021.09.759 (2021).

4 Hirakawa, M. P. *et al.* Upregulation of CD14 in mesenchymal stromal cells accelerates lipopolysaccharide-induced response and enhances antibacterial properties. *iScience* **25**, 103759, doi:10.1016/j.isci.2022.103759 (2022).

5 Li, G. *et al.* A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. *Genome Biol* **23**, 20, doi:10.1186/s13059-021-02595-6 (2022).

6 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

7 Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* **17**, 77, doi:10.1186/s13059-016-0938-8 (2016).

8 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

9 Sheng, K. & Zong, C. Single-Cell RNA-Seq by Multiple Annealing and Tailing-Based Quantitative Single-Cell RNA-Seq (MATQ-Seq). *Methods Mol Biol* **1979**, 57-71, doi:10.1007/978-1-4939-9240-9_5 (2019).

10 Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat Methods* **14**, 267-270, doi:10.1038/nmeth.4145 (2017).

11 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

12 Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398, doi:10.1038/nmeth.4179 (2017).

13 Yuan, J., Sheng, J. & Sims, P. A. SCOPE-Seq: a scalable technology for linking live cell imaging and single-cell RNA sequencing. *Genome Biol* **19**, 227, doi:10.1186/s13059-018-1607-x (2018).

14 Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* **19**, 224, doi:10.1186/s13059-018-1603-1 (2018).

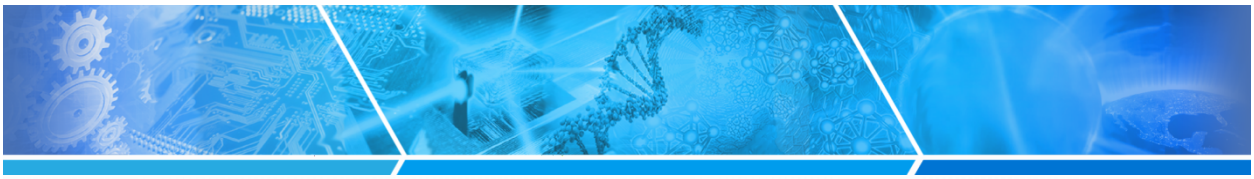
15 Parikh, I., March, S. & Cuatercasas, P. Topics in the methodology of substitution reactions with agarose. *Methods Enzymol* **34**, 77-102, doi:10.1016/s0076-6879(74)34009-8 (1974).

16 Xu, M. T., B; Zhou, W.; Wei, T.; Zhang, H.; Zhou, D. Construction of low melting point agarose emulsion PCR amplification complex gene sequence. *Journal of Genetic Engineering and Biotechnology* **10**, 239-245 (2012).

17 Storen, T. *et al.* Measurement of dye diffusion in agar gel by use of low-coherence interferometry. *Opt Lett* **28**, 1215-1217, doi:10.1364/ol.28.001215 (2003).

18 Cuatrecasas, P. P., I. Polysaccharide matrices for use as adsorbents in affinity chromatography techniques. USA patent US3947352A (1974).

APPENDIX A. POWERPOINT SLIDES



LDRD
Laboratory Directed Research and Development

Bioscience LDRD Ending Project Review
Combined Imaging and RNA-Seq on a Microfluidic Platform for Viral Infection Studies (Project # 226162)

PI - Raga Krishnakumar (8623)

PM - Victoria VanderNoot (8620)

Team - Anna Fisher (8621), Kurt Sjoberg (8621), Gloria Doudoukjian (8623), Elizabeth Webster (8621)

FY22, $100k

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

FY21 Ending Project Summary (PI to complete)

Purpose, Goals & Approach

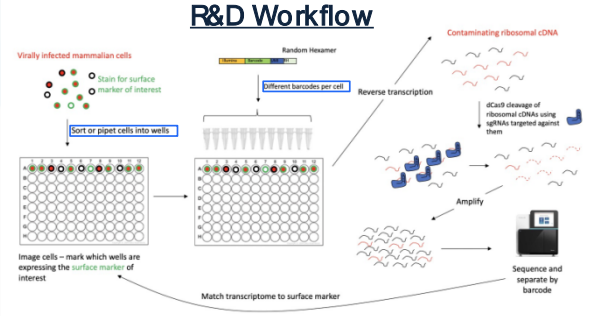
Purpose: Understanding the molecular mechanisms of how individual mammalian cells interface with the environment (eg. pathogens)

Question: In a single mammalian cell, can we connect cell surface protein information with gene expression information from host and pathogen?

Approach: Unlike previous work, our protocol combines:

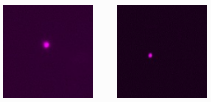
- *Matching surface protein images with transcriptomes without needing feature barcoding*
- *Amplifying both polyA and non-polyA RNA*
- *Device-based generation of functionalized/beads surfaces for matching cells with primer barcodes*

R&D Workflow



Potential for replacement with device-based methods

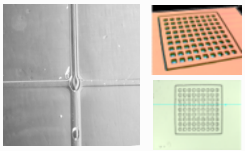
Key R&D Results and Significance

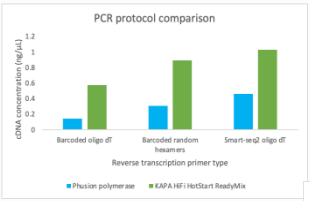


Individually-labeled cells can be visualized for matching with barcodes

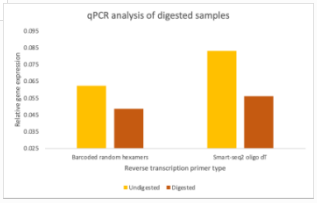
In-house manufactured microwell for cell capture protocols

Droplet-generators for agarose beads





Reverse transcription of single-cell RNA can be performed with barcoded random hexamers as well as oligo-dT



Single-cell ribosomal DNA depletion shows promise, although there is still some non-specific activity

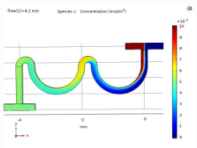
Conclusion: Results show potential for this pipeline, although some troubleshooting is required for end-to-end functionality.

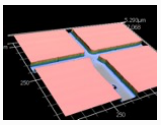
2

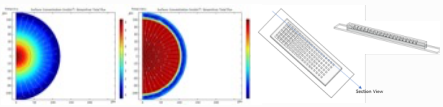
Project Legacy (*Use 1 or 2 slides*)

Notable Technical Outcomes

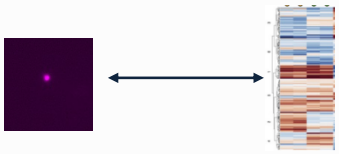
- Capabilities stood up for microfluidic based single-cell imaging. Analysis-Design-Mfg
  - Microfluidic mixers for dye combinations
  - Microwell arrays for cell isolation
  - Agarose droplet generation for dying/labeling captured cells.



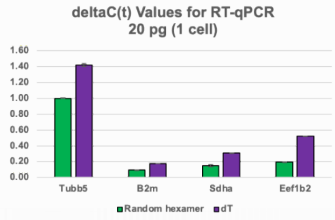




- Labeled single cells can be visualized in this pipeline, and matched with a corresponding barcode



- Random hexamers can be used instead of oligo dT for reverse transcription of single-cell RNA, setting up for ribosomal cDNA depletion



Project Legacy (*Use 1 or 2 slides*)

Important lessons learned

- Challenges of taking on a project conceived by someone else
- Intricate protocol troubleshooting requires higher FTE (including clean room processing)
- Teaming at Sandia is unlike anything I've experienced - so many tangible and intangible positive outcomes!

Mission-relevant impacts

- Developing single-cell sequencing pipelines has relevance across organisms relevant to biodefense and bioenergy
- Developing devices, microwells and protocols that can be applied to other projects
- Generated discussion points on follow-up work in biodefense, biomanufacturing and genome security
- Establishing collaborations across capabilities at Sandia, including discussions with researchers at JBEI

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Email—External

Name	Company Email Address	Company Name

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Hardcopy—External

Number of Copies	Name	Company Name and Company Mailing Address

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.