Open-source Custom Beads for Single-cell Transcriptomics

Thesis by
Taleen Gaied Dilanyan

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended August 25th, 2023

Taleen Gaied Dilanyan
ORCID: 0000-0002-3131-3259

ACKNOWLEDGEMENTS

I dedicate this thesis to my family and friends who supported me through every step of my PhD journey. Your love means everything to me!

Thank you to my advisor Lior Pachter, my committee chair, Doug Rees, and the rest of my committee, Linda Hsieh-Wilson and Mikhail Shapiro, for their guidance and continuous support.

A special thanks to Ken Lau and members of his lab at Vanderbilt University for all their help with troubleshooting and scRNA-seq experiments.

ABSTRACT

Open-source single-cell genomics technologies have helped democratize single-cell genomics and expedite method development. Methods such as inDrops and Drop-seq for single-cell RNA-seq preceded popular technologies such as the 10x Genomics' Chromium platform, however despite initial enthusiasm for open-source methods, their popularity has waned. A major reason has been the lack of availability of low-cost, customizable beads, which are essential for microfluidics based single-cell RNA-seq. We address this challenge by introducing a new method for producing barcoded hydrogel beads for single-cell RNA-seq called HiPER (High-throughput PER-barcoded hydrogel beads) that allows for increasing the diversity of barcode sequences, reducing manufacturing cost, and that can be readily adapted to custom applications. HiPER barcodes are decoupled from the capture sequences and can therefore be configured to capture RNA, DNA, or tailored for specific-gene enrichment.

TABLE OF CONTENTS

NOMENCLATURE

ATAC-seq. The assay for transposase-accessible chromatin with sequencing.

Barcode. a distinct series of DNA bases.

Bead "capping". A reaction to polymerize the capture sequence to barcoded beads..

IVT. In-vitro transcription is a DNA-templated synthesis of long RNA transcripts, including messenger RNA, linearly, with high fidelity. The T7 enzyme requires a promoter site 5' – TAATACGACTCACTATAG – 3' to commence the reaction, and requires DNA to template the synthesis, which keeps the amplification linear. This makes the reaction slower, but less prone to propagating polymerization errors.

Oligo. A short polymer of nucleotides that are typically chemically synthesized for use in a variety of applications.

PER. The primer exchange reaction; a method for synthesizing single stranded DNA without consuming the templating DNA.

Poly dT. a series of poly-deoxythymidine that is commonly used to complement poly adenylated mRNAs to prime reverse transcription.

RT. Reverse transcription of mRNA molecules to complementary DNA (cDNA).

scRNA-seq. Single-cell RNA sequencing.

Toehold. The primer site on the DNA oligo that templates the annealing of the hairpin..

UMI. Unique molecular identifier used to distinguish different molecules of identical barcodes.

Chapter 1

INTRODUCTION

1.1 Background

Single-cell RNA sequencing (scRNA-seq) has recently become a foundational method for parsing heterogeneous expression profiles of cultured cells, tissue biopsies, and whole organisms. The term scRNA-seq now encapsulates dozens of protocols that collectively provide accurate, highly resolved depictions of the clonality, genome evolution, and unique mutations of individual cells in tissues, as well as tracing the impacts of disease progression (Petti et al., 2019; Muyas et al., 2023; Lei et al., 2021). Unlike more traditional bulk sequencing in which the genetic material of thousands of cells are averaged as one output, single cell sequencing allows researchers to more accurately explore the transcriptome of individual cells, revealing unique genomic variations and expression patterns of different cell types, growth stages, and in response to different stressors. The vast proliferation and variation of scRNA-seq, and more generally single-cell genomics technologies, supply a growing repertoire of single cell transcriptomics data facilitating these applications. These data additionally allow for large-scale exploratory research of specific species, tissues, and cell types across many data sets.

Microfluidics-based scRNA-seq technologies have their origins in two open-source methods published contemporaneously: DropSeq and inDrops (Macosko et al., 2015; Klein and Macosko, 2017). This pair of technologies paved the way for in-house productions and utilization of barcoded beads for tagging the transcriptomes of single cells for subsequent sequencing. One of the advantages of open-source technologies is their transparency, making possible the tweaking of assays, and facilitating the combination and customization of different protocols to improve efficiency and usability of protocols. A more recent example of a next-generation open source technology is Hydrop (De Rop et al., 2022), which combines elements of commercial and open-source protocols to offer simplified workflows for both single-cell assay for transposase-accessible chromatin sequencing (ATAC-seq) and scRNA-seq, enhanced detection or modality.

Across microfluidics-based platforms, beads are used to capture and prime the reverse transcription (RT) of messenger RNA (mRNA) transcripts. Single-cell RNA-

seq beads consist of a distinct series of DNA bases (barcode), a unique molecular identifier (UMI) which differentiates individual oligonucleotides (oligos) with the same barcode, and a poly-deoxythymidine (poly-dT) probe (see Figure 1.1).

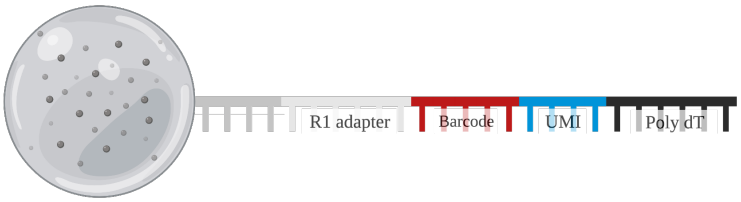

Figure 1.1: Scheme representing a typical bead with a primer site, a barcode or partial barcodes, UMI, and a capture sequence (a poly dT for scRNA-seq beads). The length of barcode and the UMI, as well the order of the these respective parts could vary, but they are present in most beads of microfluidics-based technologies.

The precision required to label thousands or tens of thousands of cells with individual barcodes necessitates hydrogel beads with large numbers of barcode combinations in order to minimize the rate of barcode collisions in subsequent analyses. Consequently, one of the most challenging improvements to scRNA-seq platforms has been the optimization of barcode combinations without increasing the cost of synthesis. The most popular commercial platform for scRNA-seq is the 10x Genomics Chromium system, which enables droplet-based scRNA-seq experiments, requiring use of the proprietary Chromium Single Cell Controller instrument and associated kits. Despite its user-friendliness and reproducibility, oftentimes, the heavy cost of the kits can be prohibitive for academic laboratories, especially for specialized or exploratory studies, impeding the potential research directions and the rate of scRNA-seq progress. This is where open-source alternatives have the potential to accelerate research by lowering costs, increasing accessibility, and facilitating customization. Even though commercial technologies have provided convenience and reproducibility to a large body of research, they offer only predefined options, cannot be customized, and are expensive.

1.2 Single-cell RNA sequencing technologies

InDrops

The inDrops protocol was developed at Harvard Medical School and published by the laboratory of Dr. Allon Klein (Klein and Macosko, 2017). The technology is open source, and was published with detailed protocols for synthesizing, barcoding

and utilizing bis-acrylamide hydrogel beads (HBs) for high throughput scRNA-seq (Zilionis et al., 2017). The method uses primer extension to label beads (see Figure 1.2), which have a photocleavable anchor attached to a primer sequence that allows it to be extended over two rounds of barcodes with 384 × 384 total combinations. This diversity of combinations is sufficient to label 3,000 cells per aliquot, while many thousands could be processed at a time then multiplexed during the library preparation steps using different indexing codes to distinguish the batches. There are, however, significant barriers to using this open source protocol, including some prohibitive expenses. The main cost is the anchor DNA, and the process of DNA barcode synthesis on HBs, which consumes a considerable amount of DNA being stoichiometric (not catalytic). InDrops protocol, for example, describes synthesizing the HBs with a photocleavable DNA anchor, which subsequently serves as a template for barcoding that is accomplished over two rounds of primer extension reactions. Each of the two barcoding rounds is followed by washing and denaturation steps after the extension to successfully create a complete single stranded barcode.

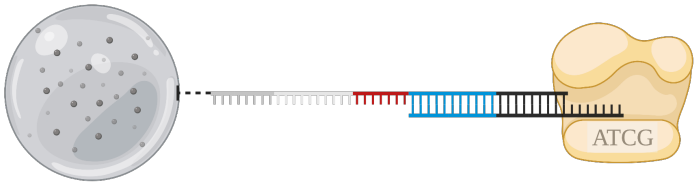

Figure 1.2: Scheme representing primer extension, used by open-source technologies such as inDrops to polymerize the extension of the barcoded sequence templated by the hybridized primer to produce dsDNA. The beads are then denatured to retain the ssDNA.

Drop-Seq

Drop-seq is an open-source single cell sequencing technology developed by the McCarroll Lab at Harvard Medical School. It utilizes droplet-based microfluidics to encapsulate single cells with uniquely barcoded beads and reagents.. Drop-seq is particularly well-suited for analyzing large numbers of cells, making it ideal for capturing the complexity of diverse cell populations and rare cell types. Similar to inDrops, Drop-seq also employs droplet-based microfluidics to enable high-throughput single cell analysis. What sets inDrops apart from Drop-seq is its linear RNA amplification resulting in lower amplification bias and improved sensitivity in gene expression measurements. inDrops' versatility and sensitivity make it suitable for a wide range of applications, including the study of rare cell populations and

low-input samples. These open-source single cell sequencing technologies have significantly contributed to advancing the field of genomics and have been embraced by researchers for their affordability, accessibility, and powerful capabilities in uncovering the intricacies of cellular heterogeneity.

Commercial solutions for scRNA-Seq

Commercially, companies like 10x Genomics have been at the forefront, offering popular platforms like 10x Genomics' Chromium system, which uses microfluidics to capture single cells in droplets for parallel sequencing. On the other hand, open-source technologies have also played a crucial role in advancing the field of single cell genomics, allowing researchers to access cutting-edge tools without proprietary restrictions. Making single cell technologies cheaper and more accessible holds immense importance in advancing scientific research. By reducing the cost and expanding accessibility, researchers from diverse backgrounds, especially in resource-constrained settings, can pursue innovative research questions.

1.3 The Primer Exchange Reaction

First introduced by Kishi et al., the Primer Exchange Reaction (PER) is a method for synthesizing single-stranded DNA (ssDNA) using a DNA hairpin that templates the extension of a targeted DNA segment (Figure 1.3). The single stranded portion of the hairpin anneals to the complementary 3' end of the DNA oligo "toehold" (Figure 1.3-1), which recruits the enzyme to begin polymerization (Figure 1.3-2). The exonuclease-deficient DNA polymerase (Bsu) initiates synthesis of the adjacent DNA region that is the barcode followed by a sequence that is the binding region for the following hairpin. The polymerase terminates the synthesis as it reaches a designated stopping point where the polymerase falls off (Figure 1.3-3), after which the hairpin undergoes branch migration continuously until it detaches, intact, unable to attach to the same molecule, but able to template another extension on another molecule (Figure 1.3-4,5). In the original PER method publication (Kishi et al., 2018), three different stopping mechanisms were introduced to terminate the polymerization of DNA, out of which the most elegant and effective was using a three-nucleotide synthesis regime. This is achieved by using three cytosines CCC at the end of the complement template region in a reaction supplied with dATP, dTTP, dCTP. The deoxyguanosine deficiency causes the polymerase to stall and eventually disengage from the single-stranded product (Figure 1.3-3). This design provides a great advantage as it utilizes unmodified, low cost DNA sequences that

self-assemble into the functional hairpins. DNA-based logic gates, self-assembly,

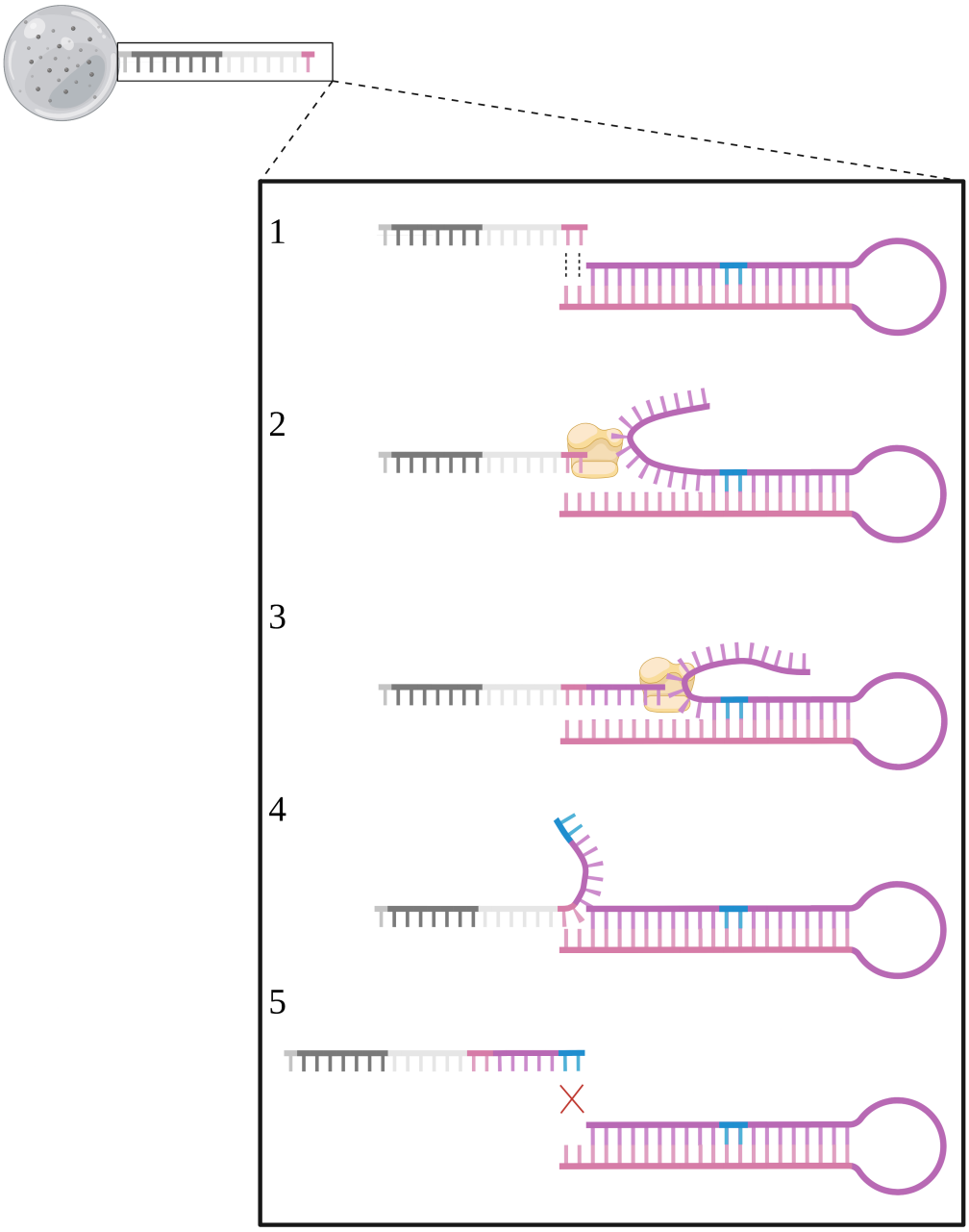

Figure 1.3: Schematic representation of the PER hairpin hybridizing to its complementary template ssDNA to initiate polymerization, strand displacement, and reassembly of the intact hairpin. **1**. The hairpin finds complementary toehold shown in light pink. **2**. The enzyme is recruited to polymerize the complementary sequence templated by the hairpin. **3**. The polymerase stalls when it reaches the stop sequence (in black) and falls off the hairpin-oligo pair. **4**. The hairpin and polymerized DNA strand undergo strand displacement until the hairpin regains its original shape and falls off. **5**. The hairpin should not anneal again because the second toehold (in blue) is a different sequence.

and programmable DNA interactions traces its roots at Caltech (Qian and Winfree,

2011; Rothemund, 2006; Maune et al., 2010). And PER has been used for various synthetic biology applications such as extend long strands of ssDNA designed to construct DNA origami structures (Kishi et al., 2018), which can displace the need for acquiring the DNA scaffolding commercially. PER has also been recently used as a method for micro RNA (miRNA) detection. PER amplifies the miRNA producing ssDNAs, which then mediate DNA-templated silver nanoclusters producing a signal correlating to the concentration of the target miRNA (Ning et al., 2023).
We hypothesized that PER could also be used to synthesizing ssDNA directly on an immobilized surface (hydrogel beads). Our main contribution is the implementation of a PER strategy for building scRNA-seq. Using PER, we have implemented a cost-effective, flexible, and efficient method to synthesize ssDNA anchored in hydrogel beads, making an open source protocol that is accessible and customizable.

1.4 Contributions of this thesis to scRNA-seq

For more than a decade, single cell sequencing technologies have revolutionized the field of genomics, reshaping our understanding of cellular heterogeneity and complexity. Unlike more traditional bulk sequencing in which the genetic material of thousands of cells are averaged as one output, single cell sequencing allows researchers to more accurately explore the transcriptome of individual cells, revealing unique genomic variations and expression patterns of different cell types, growth stages, and in response to different stressors. Democratization of single cell technologies would also foster collaborations worldwide, ultimately accelerating scientific discoveries and propel the field forward. Moreover, making single cell technologies affordable would facilitate more large-scale studies, enabling researchers to uncover a broader spectrum of cellular diversity and better genome-wide association studies (GWAS) (Jia et al., 2022). Additionally, enhanced accessibility could help incorporate single cell sequencing as a tool in personalized medicine, tailoring treatments to patients based on their unique genetic profiles.

This thesis is organized into six sections starting with this introduction, then presenting the findings of the project and its detailed methodology. I will also be discussing the benefits and challenges of the platform, as well as potential future applications.

Chapter 2

RESULTS

2.1 Bead manufacturing and barcoding with PER hairpins

Bead manufacturing and barcoding with PER hairpins

The hydrogel beads were manufactured similarly to the inDrops protocol with some alterations (Zilionis et al., 2017). Briefly, the bisacrylamide:acrylamide was diluted to 0.15%:6% final concentration with water, Tris-buffered saline–EDTA–Triton buffer (TBSET), ammonium persulfate (APS, 0.3%), and the DNA oligo (25 μM). Post polymerization, the beads were washed into TBSET buffer and filtered using a 70 um filter to eliminate possible bead aggregates. The washed and filtered hydrogel beads were washed into a noEDTA-Tris wash buffer (nWB) and mixed with the remaining barcoding reagents described in the protocol (supplementary info). The barcoding is done by inactivating the hairpins using ExoVIII enzyme which catalyzes the removal of nucleotides from the 5' to 3' direction. As the DNA oligo is anchored to the bead on the 5' end, only the barcoding hairpins are consumed leaving the barcoded oligo intact. Thus the exonuclease can be inactivated to start a new round of barcoding.

Using split-and-pool to increase the barcode space

The barcoding process is done over four separate reactions, with each round of barcoding done with a new set of hairpins that add the next set of partial barcodes thereby exponentially increasing barcode complexity (Figure 2.1). This workflow can be easily modified to encapsulate the desired number of cells by changing the number of barcoding rounds, or the number of partial barcode combinations. For example, the current design involves 38 hairpins, each adding 5 base pairs (bp) partial barcodes over four rounds of split-and-pool barcoding, which yields 38^4 = 2.09 million possible full barcode sequences. For larger scale scRNA-seq experiments, the design was also modified to include a set of 96 hairpins, each adding 6bp partial barcodes (~84M possible barcodes). On the other hand, if encapsulation is on a smaller scale (< 10k cells) the design can also accommodate contracting the barcoding to only three rounds (96^3 = 885 thousand combinations) where the third hairpin is modified to template the capture sequence instead of the fourth hairpin. In summary, the design of the barcoding reaction is flexible, and easy to adapt to

individual research questions.

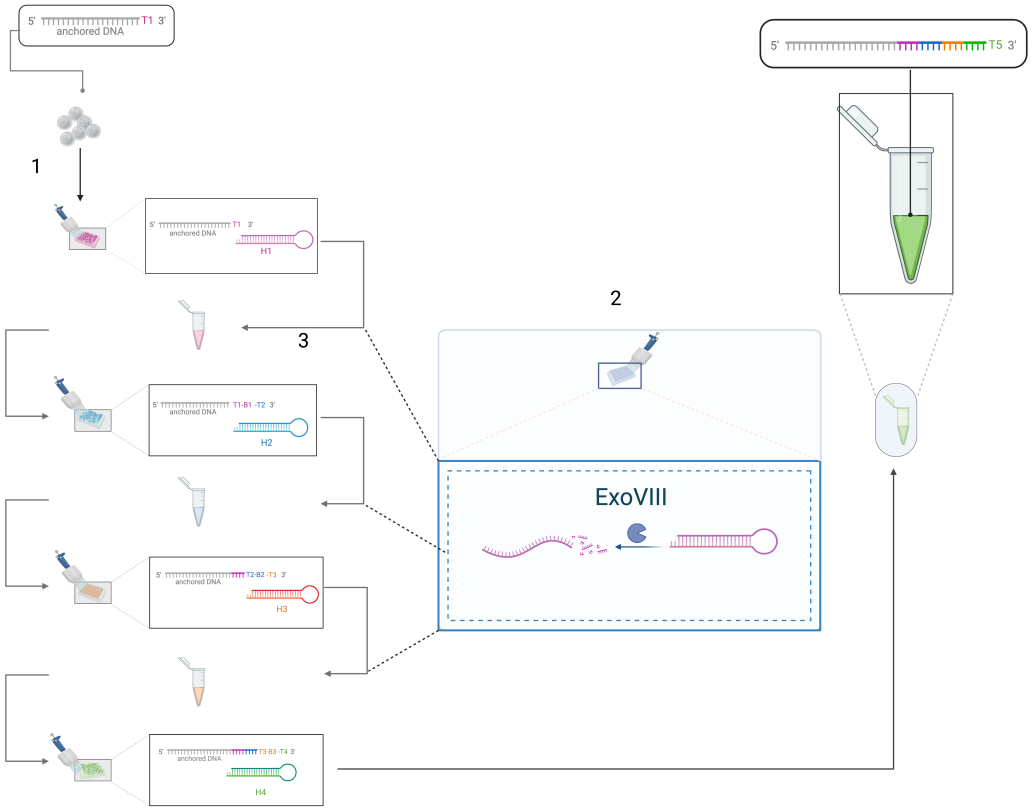

Figure 2.1: Schematic figure explaining the method of split-and-pool barcoding. **1**. The beads are split into separate wells, so that each are incubated with a different barcode sequence. **2**. While the reactions are still separate, ExoVIII is added to them to disable the hairpins before pooling to avoid potential cross contamination of partial barcodes if any DNA molecules are were not barcoded in isolation. **3**. The beads are pooled and mixed thoroughly before they are separated again to add the following partial barcode, and the steps above are repeated.

Intermediate exonuclease treatment minimize barcode cross contamination

To prevent cross contamination of different barcodes during pooling, the hairpins are degraded before pooling. In order to achieve this, an exonuclease was needed to exclusively target dsDNA, leaving the ssDNA intact. The enzyme selected for this step is ExoVIII, which is effective at digesting the hairpins while skipping the anchored ssDNA oligo. A short 15 minutes incubation with ExoVIII at 37C is sufficient to halt the PER reaction, and minimize the risk of inconsistent barcoding. The reaction is then either heat inactivated or the exonuclease consumed by the addition of ProteinaseK. The individual reactions can then be safely pooled, to proceed with the rest of the steps.

Addition of capture sequence separately from the barcode

One of the design elements that makes inDrops beads difficult to adapt is that the second barcode is added at the same time as the UMI and capture sequence. This means that customizing the capture sequence will always require purchasing new 384 plates with the new capture sequence. To make the design more modular and adaptable to changing the capture sequence, we completely separated the barcoding and capping reactions. In addition to the benefit of flexibility, using primer extension instead of primer exchange reaction to add the capture sequence allows for using all four deoxynucleotides, and not be restricted to a 3 nucleotide regime as in the barcoding stage. Thus, the capture sequence is synthesized by targeting the last template sequence added by the fourth barcoding hairpin, and extending the UMI and capping sequence. In summary, the capture sequence template is pre-incubated with the beads at 55℃ with the dNTP mixture and the beads, then T4 polymerase is added and incubated rotating at 37℃ and then heat inactivated and chemically denatured to get a single stranded oligo (see Figure 2.2).

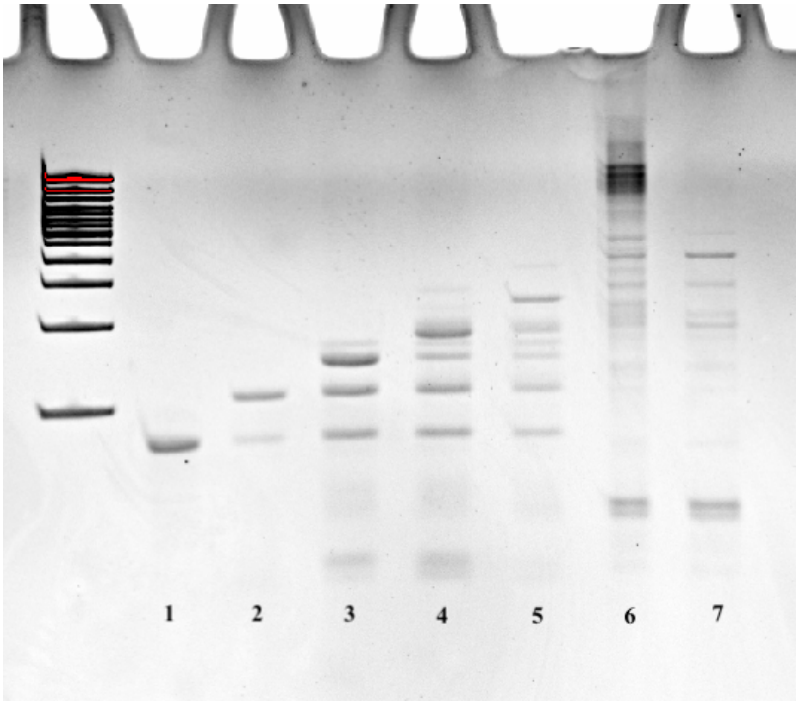

Figure 2.2: An acrylamide gel of a step-by-step barcoding and capping quality control by electrophoresis. After each partial barcoding step, a small aliquot of the beads is kept to track the progress of the reactions, and confirm their quality. The red color indicates signal saturation. **1-5**. The DNA oligo after each consecutive step of partial barcoding. **6**. Oligo after capping before denaturation (double stranded). **7**. The single stranded oligo after capping and denaturation.

2.2 Assessing the quality of the beads

In order to assess the barcoding process, and to confirm the degree of integration of each partial barcode, a batch of around 300 beads were cleaved by Uracil-Specific Excision Reagent (USER) enzyme, and an IDT kit (ssDNA & Low-Input DNA Library Preparation Kit) was used to ligate the illumina adapters for sequencing according the manufacturer's recommendations. The samples were then sequenced on MiSeq, and the data used to assess the quality of the barcoding and the contribution of each barcode plotted (Figure 2.3). The analysis shows that more than 80% of the beads perfectly match the partial barcode at their assigned indices. The distributions of each of the barcodes was relatively equal, with only the fourth partial barcode showing some unequal barcoding. This result confirms a large diversity of barcode sequences on the beads.

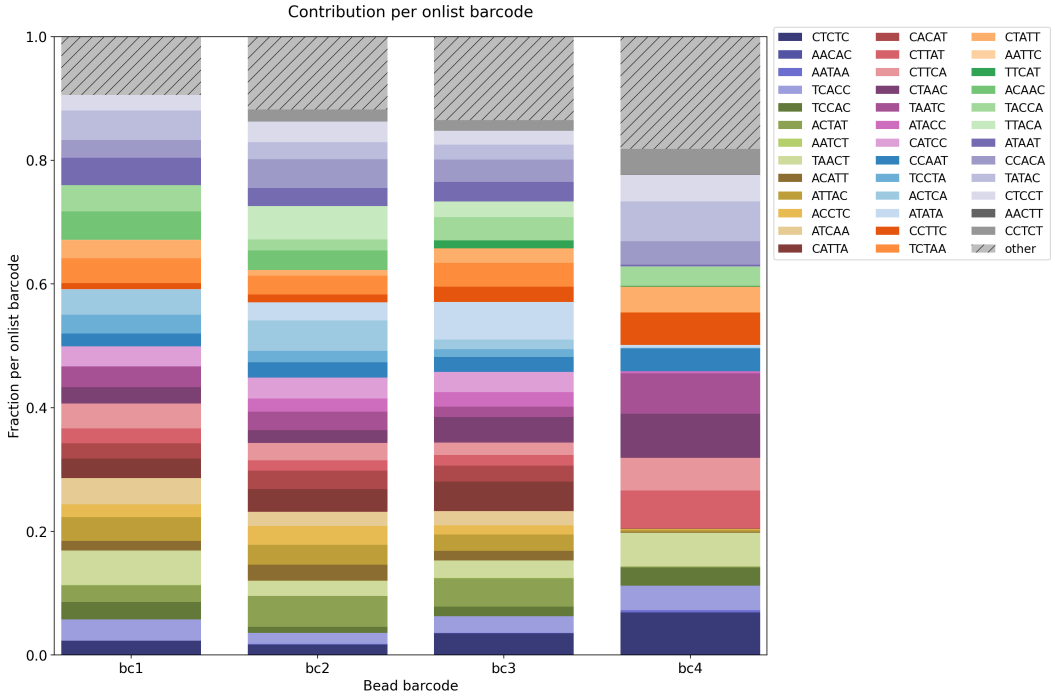

Figure 2.3: Bar plot with colors distinguishing each partial barcode to assess the relative distribution of each sequence in all four of the partial barcodes.

The cells were encapsulated using HiPER beads in the Lau lab in Vanderbilt, following the TruDrop protocol (Southard-Smith et al., 2020). The scRNA-seq libraries were dual-indexed, then sequenced on a NovaSeq6000 using a PE-150 kit. The number of reads ranged between 50-150 million reads depending on the sample. Cell cultures were used to benchmark the HiPER technology, and to be able compare it more easily to data generated by other platforms. The average

quality score of sequencing cycles was evaluated using FASTQC, and the reads were pseudoaligned to a reference index of the respective species of the analyzed cells. Both the reference and the pseudoalignment matrices were generated using the kb-python package where each barcode is validated against a list of valid barcode permutations, and its corresponding genes tallied (Melsted et al., 2021). A knee plot of the barcodes by UMI counts was generated to exclude low-UMI ambient RNA transcripts from valid cell transcripts. Other filtering methods were utilized, including standard filters such as excluding low UMI cells and eliminating the cells with the highest percentage of mitochondrial transcripts.

2.3 scRNA sequencing of human colorectal cancer human intestine cells (SW620)

This run was used to evaluate the ability of HiPER beads to capture mRNA in droplets. The experiment's sequencing ID is 10203, and could be referred to by this number for convenience. The resulting reads were evaluated as mentioned above. The libraries looked like the typical distribution of cDNA libraries but with an unknown peak at 218 bp that was not eliminated in purification steps (Figure 2.4).

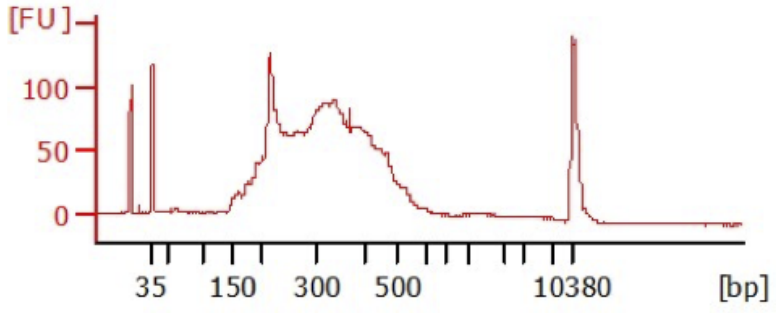

Figure 2.4: A Bioanalyzer trace of cDNA library of a scRNA-seq run using SW620 cells, which is used to assess the quality of the DNA library prior to sequencing.

This peak appears to be either a PCR primer aggregate, or a population of partially barcoded beads that was not observed in the bead sequencing. Even though it was less than a third of the library molarity, it preferentially binds the instrument's flow cell, which then translates to many of the raw reads not to be aligned to cDNA, and the UMI counts/cell to be too low.

Regardless, we are still able to process the UMI counts, and evaluate some of the cell population characteristics. For example, much of the sample seems to have a higher-than-expected fraction of mitochondrial gene counts (Figure 2.5). This is

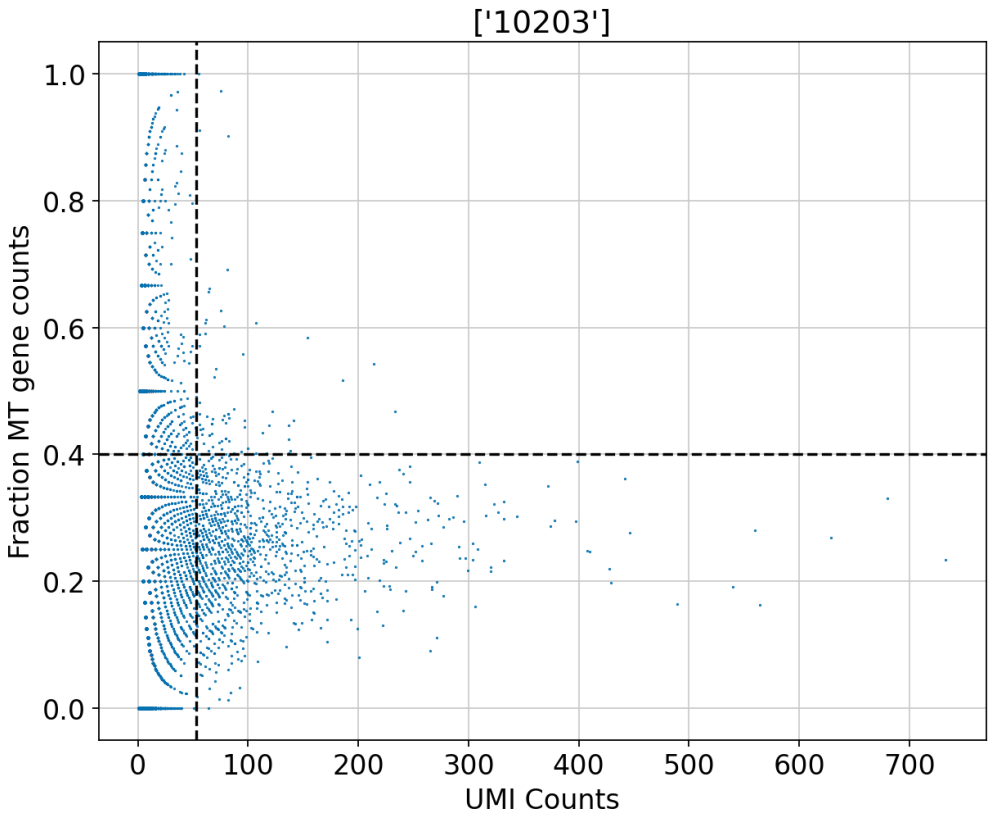

Figure 2.5: A scatter plot showing the fraction of mitochondrial gene counts per UMI count. To produce this plot, the raw reads were aligned to the human genome using kb-python, and the reads with valid barcodes and unique UMIs were compared to a list of mitochondrial genes. The barcodes are plotted against the fraction of their UMIs that correspond to mitochondrial genes.

typically attributed to cells dying, but it could also be due to the biology of the sample being derived of cancer cells correlating with higher metabolic and division rates.

2.4 scRNA sequencing of human HEK293FT and mouse MC-38 cells

Similarly, we ran a 50:50 mixture of human and mouse cells to investigate the beads ability to capture mRNAs on the single cell level. (Experiment sequencing ID: 10271)The cells were thawed and allowed to grow for two days in complete growth medium, but some of the mouse cells were not fully recovered. The samples were processed as described above, and sequenced after confirming the quality of the cDNA library (Figure 2.6).

About 46% of the raw reads mapped to human and mouse genomes. The reads were presented in a knee plot to determine the ambient RNA counts (Figure 2.7).

A barnyard plot of the unfiltered reads shows distinct human and mouse populations

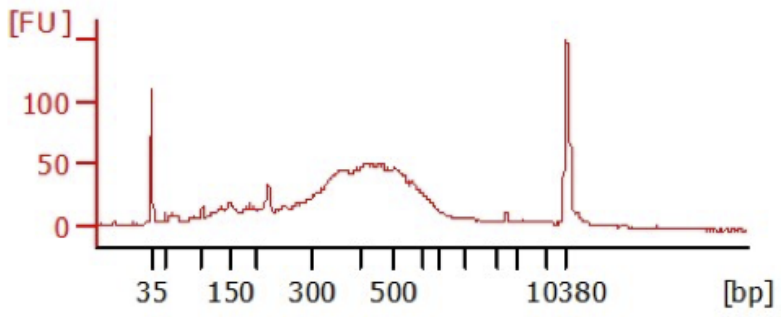

Figure 2.6: A Bioanalyzer trace of a 50:50 mixture of Human Embryonic Kidney(HEK 293FT) cells and Murine colorectal carcinoma (MC-38) scRNA-seq DNA library. The trace is used to assess the quality of the DNA library prior to sequencing.

highlighted in blue and red respectively (Figure 2.8). In Figure 2.9 we observe populations of cell doublets as some irregular cells containing more than the expected levels of ambient RNA especially RNAs of the other organism. This affect might also be related to the double "knee" observed in Figure 2.7 indicating irregularity in the cell mixture. This prompted us to plot the two UMI populations in separate knee plots to determine the quality of each population individually (Figure 2.9).

When plotted separately, it becomes apparent that the human cell population is much more typical, with a single knee and number of cells that is consistent with the expected cell count. The first knee of the mouse cell population, on the other hand, indicates a high number of mouse cells with low UMI counts, as is evident in Figure 2.9. This indicates that the real fraction of mouse to human cells are not equivalent.

2.5 More stringent sample cleanup in downstream library preparation

Upon further examination of the raw sequences of dataset 10271, it became evident that a high percentage of poorly-labeled, short sequences had passed the cleanup steps of the library preparation. To rectify this effect, the sample was SPRI purified with a more conservative 0.65x SPRI volume equivalence. To be able to examine the impact of the more stringent cleanup more directly with the previous dataset (experiment ID 10271), a frozen aliquot of the same encapsulation experiment as the mixed human HEK293T and mouse MC-38 cells was processed. The resulting library was sequenced in experiment ID 10350 (Figure 2.10).

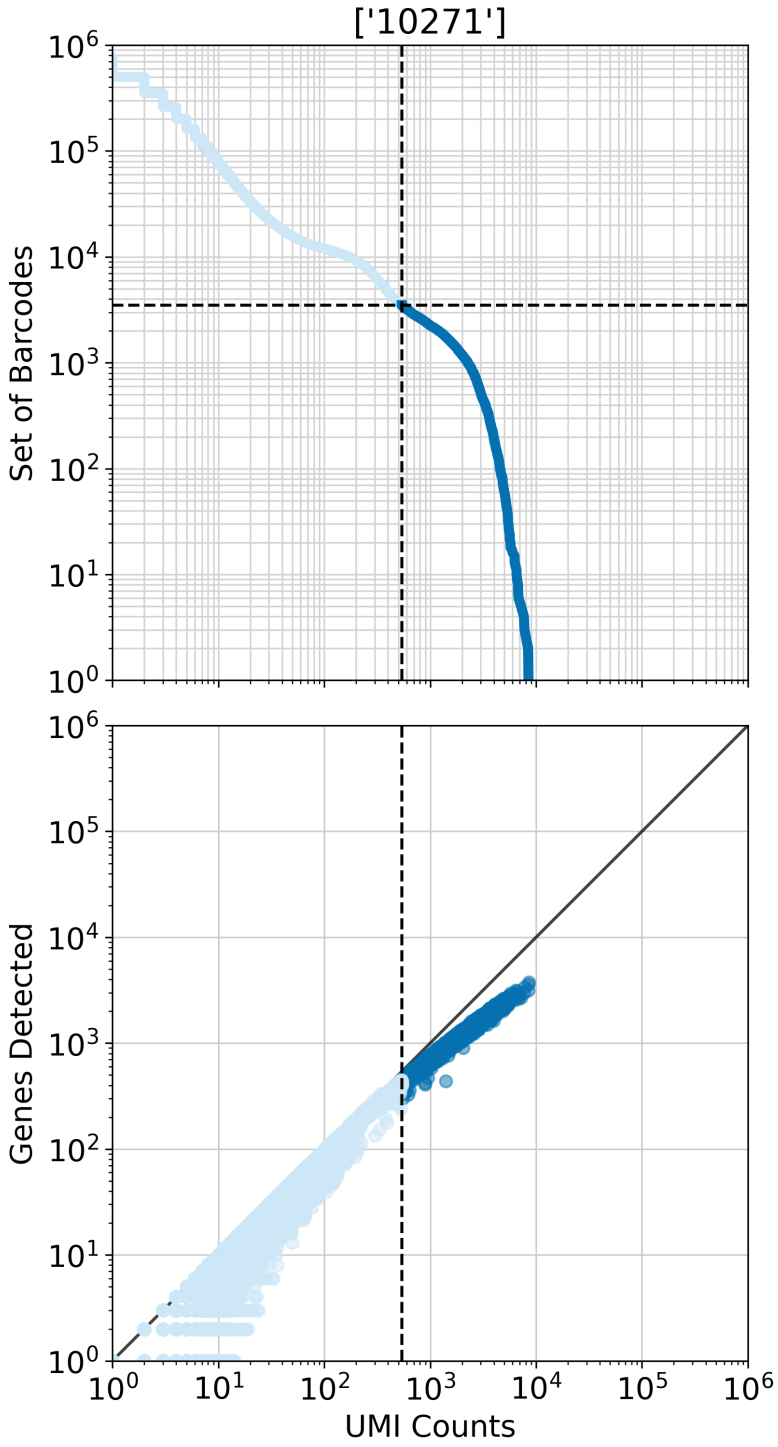

Figure 2.7: top: A knee plot of the sets of barcodes in ex. ID 10271, plotted against their UMI counts to identify significant barcodes. The sample was a 50:50 human-mouse mixed cell culture of HEK293FT and MC-38 cells. The "knee" and inflection point are marked by the dashed line and the significant barcode population highlighted in dark blue. **bottom:** A plot of detected genes in ex. ID 10271 plotted against their UMI counts. The inflection point identified in the knee plot above are also depicted with a dashed line, and the significant genes highlighted in dark blue.

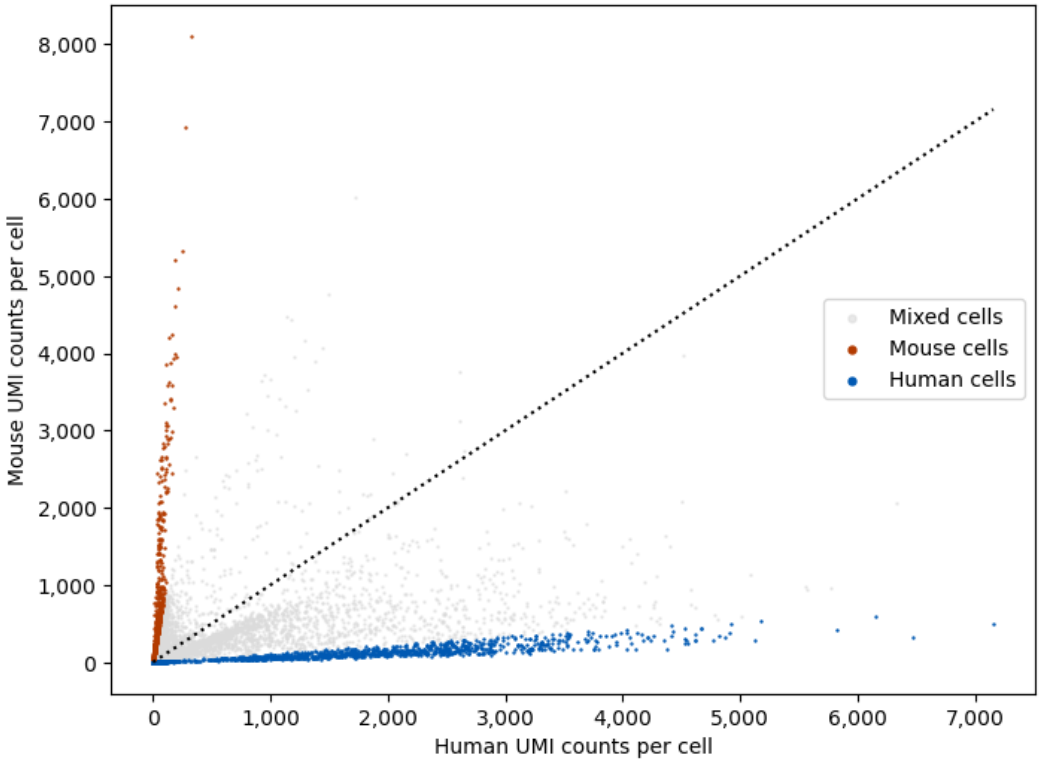

Figure 2.8: Barnyard plot depicting the number of UMIs mapping to the human or mouse transcriptome in a 50:50 human-mouse mixed culture of HEK293FT and MC-38 cells.

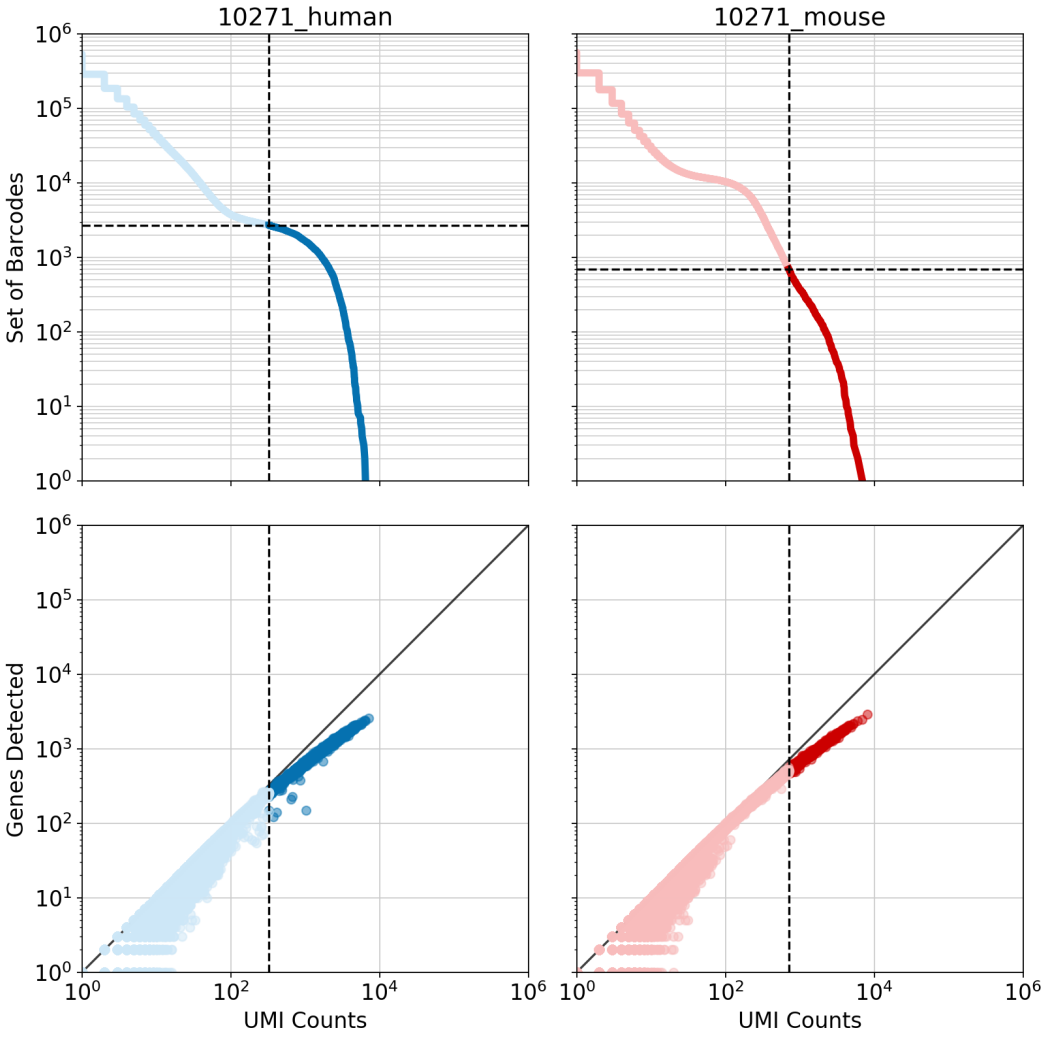

Figure 2.9: Barnyard plot depicting the number of UMIs that are mapped to the human or mouse transcriptome from a 50:50 human-mouse mixture of HEK293FT (blue) and MC-38 (red) cells, with each dot representing a unique barcode (cell).

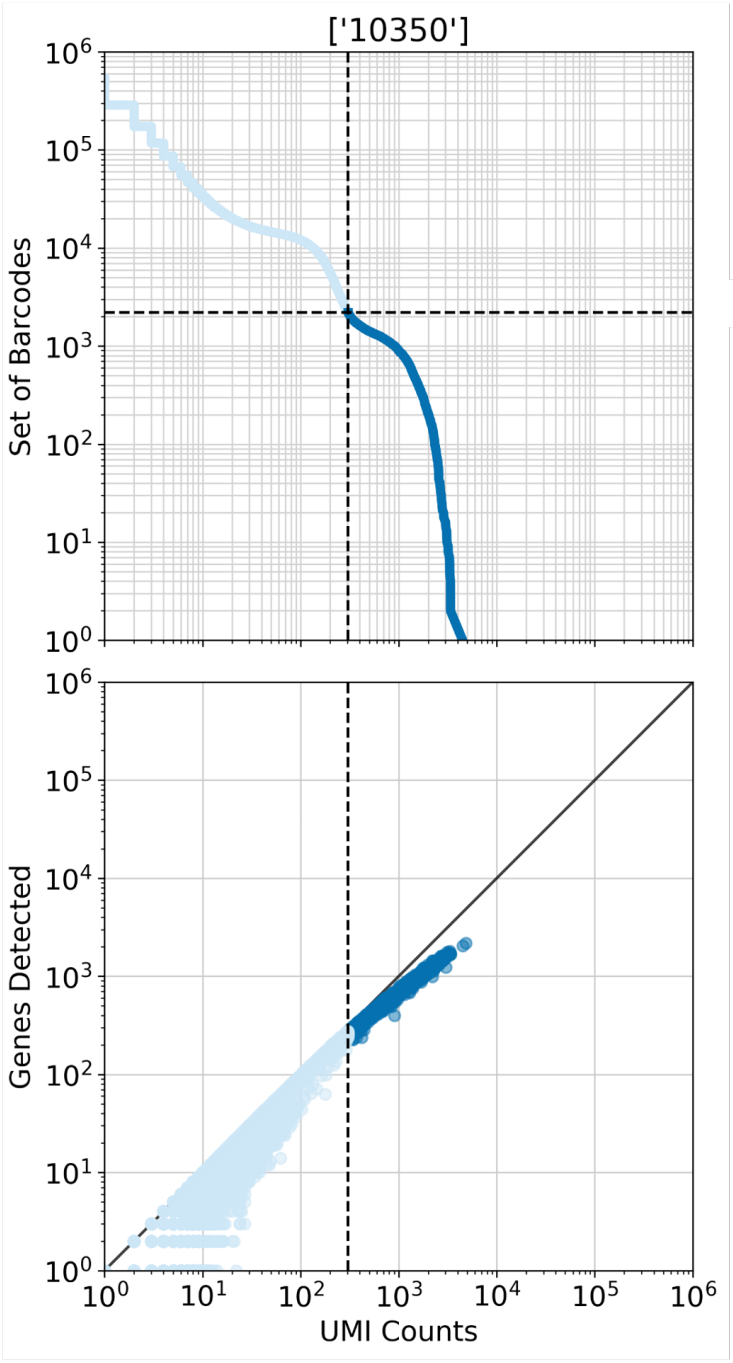

Figure 2.10: top: A knee plot of the sets of barcodes in ex. ID 10350, plotted against their UMI counts to identify significant barcodes. The sample was a 50:50 human-mouse mixed cell culture of HEK293FT and MC-38 cells, . The "knee" and inflection point are marked by the dashed line and the significant barcode population highlighted in dark blue. **bottom:** A plot of detected genes in ex. ID 10350 plotted against their UMI counts. The inflection point identified in the knee plot above are also depicted with a dashed line, and the significant genes highlighted in dark blue.

Chapter 3

DISCUSSION

In this section, I discuss the implications of the results, and some aspects of the development of the HiPER platform. During the development process, I encountered several setbacks that I have overcome or learned from. As it is my belief that it is as important to discuss the successes as the shortcomings, I will detail both aspects here. Moreover, there are some lessons learned that could be valuable in future single-cell technology development. Many of the design elements implemented are shared with other technologies, and were either inspired directly, or have organically converged with more recent technologies, but I believe that HiPER still offers the lowest startup and operational costs, as well as other advantages.

3.1 Anchored DNA oligo design elements

Enzymatic oligo release

As in previous publications, the DNA molecules are anchored to the hydrogel by an acrydite anchor (an acrylic phosphoramidite modification to the 5' end of the DNA) which allows it to be incorporated to the polyacrylamide gel during polymerization. The acrydite anchor is succeeded by a deoxyuridine base (dU), a site with which the enzyme USER can robustly cleave the DNA from the hydrogel at the desired time.

T7 promoter site

An important step of the library preparation includes linear amplification of RNA, for which a T7 promoter site (a specific DNA sequence recognized by the T7 RNA polymerase) is necessary. This element has been used in earlier technologies like CITE-seq, inDrops and MARS-seq (Hashimshony et al., 2012; Zilionis et al., 2017; Keren-Shaul et al., 2019). In an effort to improve the yield of the linear amplification reaction, the sequence surrounding the T7 promoter site was altered, based on the work of Conard et al. (Conrad et al., 2020) the bases proceeding and following the T7 promoter site were modified from the design used in the inDrops design to improve the amplification yield. Upstream of the promoter site was changed to a more AT-rich sequence, while downstream sequence was made more GC rich (see Figure 3.1).

Finally, the oligo's 3' end culminates with the sequence for illumina's Read 1

```
inDrop
--CGATGACGTAATACGACTCACTATAGGGATACC-ACCATGGCTCTTTCCCTACACGACGCTCTTCCGATCT 70
HiPER
GCCGGGAATTTAATACGACTCACTATAGGGAGAGTGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 73
  **   *  ********************* *   * *    ******************************
```

Figure 3.1: Alignment of HiPER DNA oligo with the inDrops counterpart. The alignment highlights the changes made to the sequences that proceed and succeed the T7 promoter site.

sequence which will later on serve as a template to begin R1 sequencing, as well as its last 9 bp serving as the "toehold" region for the first hairpin to anneal and begin barcoding.

3.2 Efficiency of PER in solution vs on bead

One of the difficulties of adapting PER to an immobilized surface is decreased reaction efficiency. The reaction efficiency in-solution is typically >90-95% in solution. However, as PER is used to template barcode polymerization on an immobilized surface (hydrogel bead), there are more elements to take into consideration. First, the beads need to remain in solution and must not be allowed to pellet. This step is crucial for a uniform barcoding of the beads. Also, when the DNA is embedded in the hydrogel matrix, the hairpins do not have equal access to all toeholds as they would in solution. Although the kinetics of the two reactions are relatively similar, a certain fraction of the toeholds on the bead are less accessible to the hairpins, and remain so even with longer reaction incubation. Empirically, it's observed that with each subsequent hairpin, fewer DNA molecules remain accessible to the hairpins affecting the efficiency of the barcoding overall. In order to address these effects and optimize barcoding efficiency, several designs were altered over time.

3.3 Barcoding hairpins

Structure

The hairpins were modeled after the structure of PER hairpins (Kishi et al., 2018). The structure of the barcoding hairpin was designed to consists of three main parts; the "toehold," a 9 bp single stranded region which will anneal to the 3' end of the anchored DNA; the barcode/template, which is double stranded and is designed to template the barcode and the toehold of the following set of hairpins; and the stop sequence, which consists of a string of cytosine bases that will cause the polymerase to stall and initiate the strand displacement activity that will end up

with the hairpin detaching from the anchored DNA, and in its original confirmation, capable of templating the barcode on other DNA molecules. The stop sequence is crucial for the enzyme stalling by excluding deoxyguanosine triphosphate (dGTP) from the deoxynucleotide (dNTPs) mixture provided per reaction, including three consecutive cytosine bases after the barcode and template sequence would stall the enzyme and prevent further polymerization (Figure 3.2). Using primer exchange in lieu of primer extension to construct the barcode allows for using a significantly reduced concentration of the required template. The ratio of template to anchored DNA can be between 1:30 - 1:100 depending on the incubation time. That cuts down on one of the most significant sources of cost in bead manufacturing.

Figure 3.2: A schematic view of the general sequence of a barcoding hairpin. The toehold (n) primes the addition of the barcode as well as the complementary sequence of toehold (n+1) that will prime the next set of hairpins.

Barcode design

The barcodes were designed with particular characteristics. The first design element is for each partial barcode to be designed to have 2 or more hamming distance from each other. The hamming distance allows for each full 20 bp barcode to have at least 8 hamming distances from each other.

Design versatility

The hairpin design file (Appendix B) conveniently divides the structure of the linear hairpin into motifs that can be easily edited to achieve the desired number and order of hairpins. For example, the toehold for the fourth hairpin could be replaced with the template for the capping sequence to decrease the rounds of barcodes from 4 to 3 rounds.

Configuration

The hairpins are ordered as linear oligos from IDT without purification. In order to increase the likelihood of the hairpin assuming the correct secondary structure to template the PER reaction, the hairpins were diluted to the desired concentration with a Tris buffer with NaCl salt content to stabilize the hairpin structure. Then, the diluted hairpins were heat denatured and allowed to cool down to 20C in the PCR machine where the rate of temperature is set to 0.1C/sec. This encourages a larger population of the hairpins to assume the most energetically favorable confirmation and improve barcoding.

3.4 Hairpin sequence

This effect was not accounted for in the initial design of the hairpins, which included 96, 5-nucleotide sequences that begin at the 5' end of the hairpin. When the hairpin assumes its secondary structure, the barcode is then positioned at the beginning of the double-stranded section of the hairpin. It was demonstrated in the original PER publication that hairpins that begin with an adenine base have higher reaction efficiency. This effect was not accounted for in the original design of barcodes, which caused most of the hairpins that did not coincidentally begin with an adenine to have decreased efficiency, affecting the overall efficiency. Consequently, the hairpin design was changed to include an adenine base as the first nucleotide at the 5' end, then followed by the rest of the five nucleotides as this effect, in fact, translates to immobilized PER. Some of these modified hairpins were assessed individually to confirm their improved function, which revealed another unexpected quirk of the hairpins. The barcode sequences that started with more than two consecutive adenines produced side products and lost efficiency. This effect is likely due to the weaker adenine-thymine bonds that decreases the Δ G of the desired confirmation, which in turn causes easier hairpin deformation, and potential faulty barcoding.

3.5 Minimizing barcoding errors and cross contamination

After examining sequences of barcoded beads, some patterns of common faulty barcoding emerged, and several adjustments were implemented to address each type of erroneous barcoding. I will discuss below the most common errors and how they were addressed.

dGTP cleanup hairpin (dGTPH)

One of the hairpin barcoding errors was incomplete barcodes caused by the presence of small amounts of dGTP in the barcoding reaction (the dNTPs used are ≥99% pure, so up to 1% could be dGTP contaminant). To mitigate this issue, a dGTP cleanup hairpin (dGTPH) was added to the protocol. It does not perform any barcoding, instead it is used to consume any traces of dGTP that may be present in the dNTP mixture, which would cause the polymerization to proceed through the stop sequence, and thus rendering the hairpin useless. The design of this cleanup hairpin is described in the original PER publication (Kishi et al., 2018). The function of the cleanup hairpin was confirmed by running a barocding experiment with and without dGTPH (Figure 3.3).

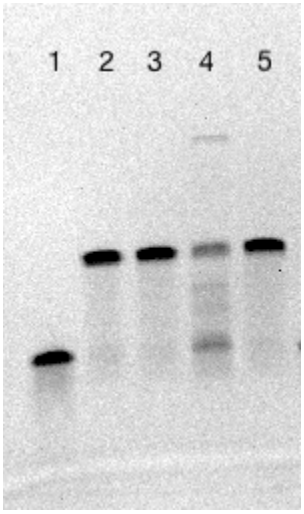

Figure 3.3: A polyacrylamide gel showing an experiment testing the function of the dGTP cleanup hairpin (dGTP).

	1	2	3	4	5
dGTP	+	-	+	+	+
dGTPH	+	-	+	-	+
oligo	+	+	+	+	+
hairpin	-	+	+	+	+

Table 3.1: A table of the reagents of each reaction in the experiment in Figure 3.3

Intermediate exonuclease treatment minimize barcode cross contamination

To prevent cross contamination of different barcodes as the split samples are pooled, the hairpins are consumed before pooling. In order to achieve this, an exonuclease was needed to exclusively target dsDNA, leaving the ssDNA intact. The enzyme

selected for this step is ExoVIII, which is effective at digesting the hairpins while skipping the anchored ssDNA oligo (Figure 3.4). A short 15-minutes incubation with ExoVIII at 37C is sufficient to halt the PER reaction, and minimize the risk of inconsistent barcoding. The individual reactions can then be safely pooled, to proceed with the rest of the steps.

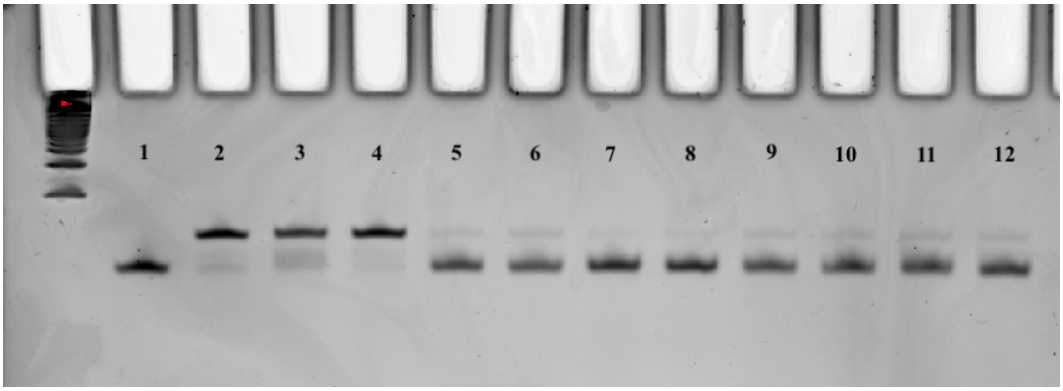

Figure 3.4: A polyacrylamide gel showing an experiment testing the ability of ExoVIII to deactivate barcoding hairpins to selectively stop their function. **1-2** DNA oligo with (2) and without (2) hairpin. **3-12** Before incubating with the DNA oligo, the hairpins were pre-incubated with ExoVIII for 0 min (**3-4**), 5 min (**5-6**), 10 min (**7-8**), 20 min (**9-10**), and 30 min (**11-12**). This demonstrates the speed at which the exonuclease disables the hairpin's function.

3.6 The addition of the capture sequence

Unlike the polymerization of the partial barcode which was achieved by excluding deoxyguanosine from the reaction, the addition of the capture sequence (in most cases) requires all four deoxynucleotides. In the current HiPER design, both the UMI and the capture sequence are added by primer extension, using a T4 polymerase. This particular enzyme was chosen as it demonstrated high polymerization efficiency, as well as a low error rate which could be advantageous in sequencing applications. Its strong exonuclease activity offered the additional advantage of eliminating the need to follow up with an exonuclease treatment to digest the partially barcoded products. The use of this enzyme has been a double-edged sword trying to balance its efficiency and accuracy with its unpredictability. Theoretically, the strong exonucelase activity of the enzyme is advantageous to digest incomplete barcodes which were not extended with the capture sequence. However, some of the sequencing reads reveal that some of the incomplete barcodes were instead capped, which could be due to the low incubation temperature increasing opportunities for non-specific polymerization activity. It is an inconsistent phenomenon that is highly dependant on the concentration of the enzyme, the efficiency of the barcoding, and the duration of the reaction. Despite a significant effort to optimize the current

capping reaction, these capping issues persisted, which prompted the design of alternative methods such as taking advantage of the template switching features of the enzyme Maxima -RH (Figure 3.5). Another polymerase that could be utilized for the addition of the capture sequence could be KAPA HiFi enzyme. It is often used for amplifying NGS libraries due to its efficacy and fidelity, so it is another good candidate to increase the percentage of good quality barcode, and thus the quality of the beads overall.

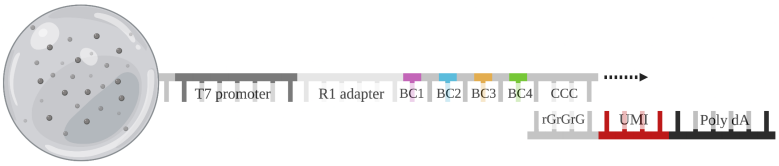

Figure 3.5: Schematic representing using template switching for capping.

3.7 Importance of considering hamming distance in barcode design

The HiPER beads, like other open-source platforms do, depend on combining partial barcodes with split-and-pool to construct a full 20bp barcode. Considering a sufficient hamming distance among the barcode sequences is important to ensure accurate base calling while sequencing. Thus, each partial barcode is designed to be 2 or more hamming distances apart from any other partial barcode. This is important to make sure that the full barcode can be more confidently assigned. The design files of the HiPER barcoding hairpins can be found in Appendix B.

3.8 HiPER beads compared to Hydrop

Hydrop is another microfluidics-based single-cell sequencing platform that diverged out of the inDrops method (De Rop et al., 2022). The method improves upon several elements of the inDrops protocol, starting with eliminating the need for photocleavage to release the oligos from the bead by instead making dissolvable beads (similar to 10x Genomics beads). The barcoding rounds were increased to 3 rounds x 96 barcodes which expanded the theoretical barcode space, as well as incorporating PCR heat cycling steps to increase the efficiency of the barcoding. In comparison to inDrops, Hydrop addresses many cost and usability issues, however, by their own estimation, the initial cost of starting the project is close to $10,000. In contrast, the use of PER hairpins to barcode HiPER beads decreases the cost of barcoding in practice to the cost of the enzymes as the hairpin concentration

needed to barcode the same number of beads is still approximately 30 times less than Hydrop, as well as the fact that separating the capping step from the barcoding further increases the flexibility of HiPER to change its capture target at a moment's notice.

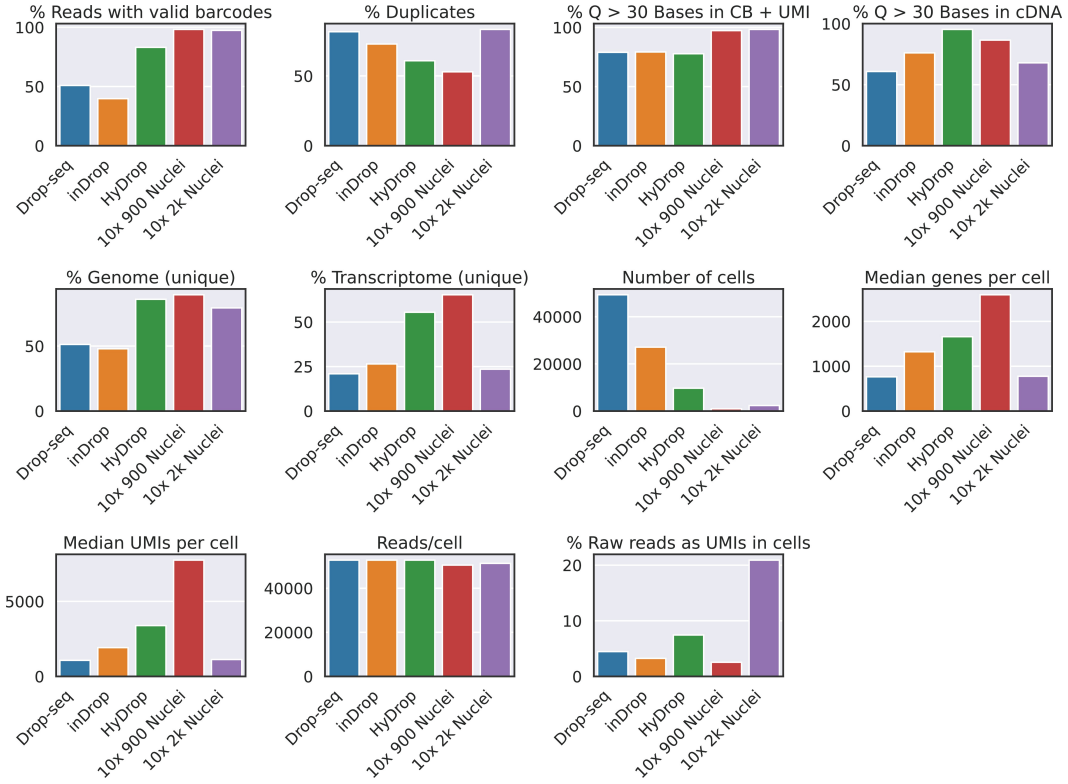

Figure 3.6: Figure from De Rop et al. comparing the Hydrop protocol to inDrops, Drop-seq, and 10x Genomics using several metrics such as %duplicates, reads per cell, and %Q>30.

Despite our desire to further improve HiPER, our platform is already comparable to other open-source technologies in many of the metrics detailed in Figure 3.6. For example, in the dataset of experiment ID 10271, the percentage of raw reads as UMIs in cells is about 4.5%, which is higher than inDrops and equivalent to DropSeq. For the same run, the number of raw reads aquired was about 135 million, out of which 46% uniquely mapped to the transcriptome, close to the percentage reported by Hydrop and higher than both Drop-seq and inDrops. When observing the various quality metrics, even for this early version of HiPER beads, it becomes evident that with deeper sequencing and some modifications of the capping reaction to minimize instances of partial barocdes, the platform has great potential to be a cost-effective and competitive peer of current open-source technologies.

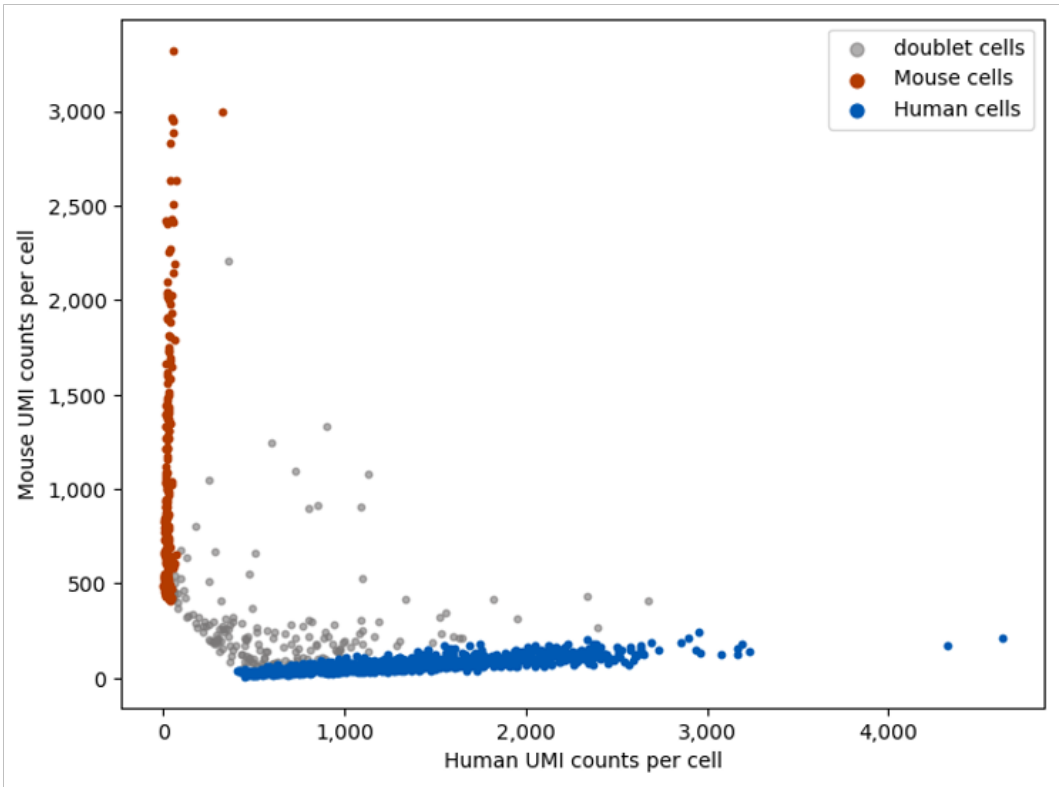

Figure 3.7: Barnyard plot depicting the number of UMIs that are mapped to the human or mouse transcriptome from a 50:50 human-mouse mixture of HEK293FT (blue) and MC-38 (red) cells, with each dot representing a unique barcode (cell).

3.9 The effect of stringent cleanup in the process of library preparation for sequencing

When the raw sequences of dataset 10271 were further examined, a high percentage of sequences have been found that are not labeled with the intended full barcoding sequence. As these sequences consume a significant portion of the total number of reads, it was important to identify and troubleshoot the cause of the issue. As many of these sequences were short, a more conservative cleanup protocol was implemented upstream of the IVT amplification reaction to limit the inherent "noise" in the library sequences. In order to be able to examine the impact of the stringent cleanup more closely with the previous dataset (experiment ID 10271), a frozen aliquot of the same encapsulation of mixed human HEK293T and mouse MC-38 cells was used. The sample was SPRI purified with a more stringent SPRI volume equivalence of 0.65x instead of 0.8x volume equivalence. The resulting library was sequenced in dataset ID 10350.

In Figure 3.7, the barnyard plot of dataset 10350 effectively separates the human and mouse cells, demonstrating sufficient cell identification at the single-cell level.

Compared to the previous dataset 10271, the results of dataset 10350 indicate that more stringent SPRI purification upstream of the IVT amplification is an effective troubleshoot of the decreased specificity of cell identification.

3.10 Designing barcodes better suited for different SBS technologies

While developing the HiPER protocol, several questions slowly emerged on how to best design the barcoding sequences to: first, to minimize the number of sequences with low hamming distances from each other, and second, to optimize the sequences of the barcodes to best complement Illumina's various SBS technologies. For example, 2-Channel SBS Technology features are currently the most prevalent illumina sequencing instruments. It is based on using two fluorescent dyes to represent cytosine and thymine respectively. Two images are taken, one to capture the green filter and the second to capture the red filter wavelength bands. If a fluorescence signal is detected in the first image, but not the second, the base is called to thymine. If it is detected only in the second image, the base is called to cytosine. If signals are present in both images, the base is called for adenine, while the lack of signal is called for guanine. It is important to keep the features of the sequencing instrument in consideration while designing the barcodes as it could dramatically improve sequencing results, especially with smaller scale instruments that are less likely to be multiplexed with other types of samples that could provide the necessary diversity of base pairs.

Chapter 4

PROJECT APPLICATIONS

The HiPER platform builds upon existing elements of open-source microfluidics-based single cell sequencing protocols, while improving design aspects of flexibility as well as upfront and per-cell cost. These features decrease the barrier to adopting the new technology and facilitate exploration of new research questions. In this section, I propose some possible variations to the main HiPER protocol and their associated applications.

4.1 ATAC-Seq

The ability to quickly and affordably customize the number of barcodes, the desired capture sequence, and even the length of UMIs allows the HiPER platform to service a variety of experimental scales and applications. As previously done with technologies like HyDrop (De Rop et al., 2022), HiPER can be adapted to ATAC-seq with the addition of a sequence complementary to the Tn5 transposase adapter to capture chromatin tagmentation product (Jia, Lin, and Chen, 2017).

4.2 Custom capture beads

Custom gene sequence(s) can also be targeted and selectively enriched for either in the form of some or many specific genes, or multi-functional beads that simultaneously target the transcriptome along with the target gene(s). Commercially, customizing gene panels have been offered for limited applications and exorbitant costs, which make it all the more inaccessible. Open-source platforms have also introduced this application, but none have been widely used as each customization requires new sets of barcodes

4.3 Immunoprofiling

An important example of single cell technology integrating into an established field and providing novel research directions is high throughput microfluidic-based immune profiling. Not only do microfluidic-based, high throughput immune profiling methods dramatically increase the number of cells processed, but they facilitate tracing the heavy and light chain sequences to the same cell/Ab. Although these methods may not be suitable to process low count cell samples (>1000 cells), they could still

be valuable in clinical settings for applications such as analyzing biopsies (Landhuis, 2018). Despite rapid developments in the field, and emerging technologies like LIBRA-seq (linking B cell receptor to antigen specificity through sequencing), commercialized immune profiling kits remain limited to human and mouse species (Setliff et al., 2019). This makes profiling other species like non-human primates (NHP) or non-model organisms currently inaccessible, despite their relevance for studying human immunity and disease models (e.g HIV in rhesus macaques) (Messaoudi et al., 2011; Kuhrt et al., 2011). However, this resource gap can be addressed by designing an open source format of customizable, hydrogel barcoded beads with a transferable protocol for immune profiling a variety of organisms.

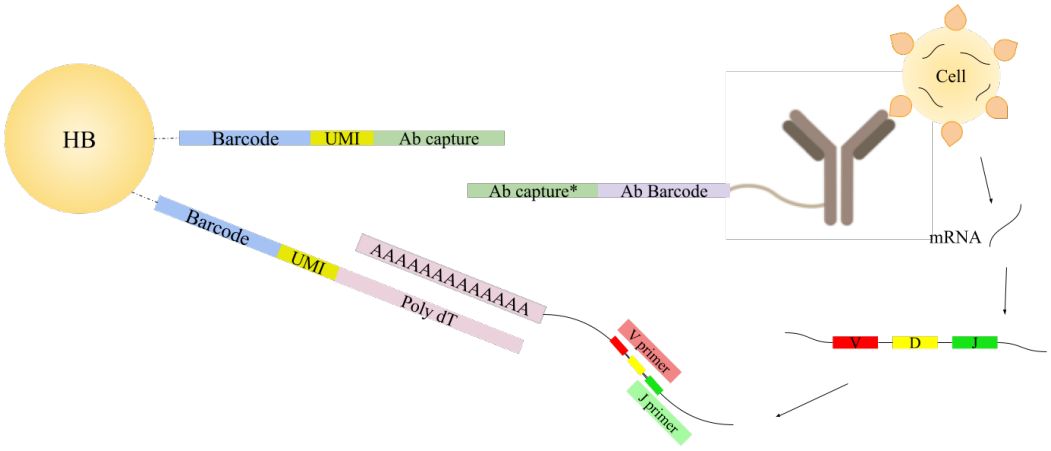

Figure 4.1: Proposed multi-functional beads application.

Immune cells of many studied organisms, especially model organisms, have known cell markers —cell surface proteins that identify the cell type called cluster of differentiation (CD) —the sequences of which also differ between species, allowing for species-specific capture. Fluorescent antibodies are commonly used to bind cell surface proteins and allow for identification or accurate sorting with FACS. For our species of interest, rhesus macaques, naive versus memory B cells for example, express different combinations of CDs, which allow enriching for the desired cell type for the purpose of immune profiling Kuhrt et al., 2011. Similarly, barcoded oligonucleotide-conjugated NHP antibodies can be used to bind cell markers and be sequenced with the cell transcriptome in parallel to help investigate cell surface protein expression. This is achieved by designing a “dual function” bead that can hybridize to polyadenylated mRNA as well as to a custom capture on the Ab oligo. Having both hybridizing "functions", the bead would have the same barcode and could be traced to the same cell. In order to enrich for the V(D)J gene segments

specifically, we can add targeted primers to the resulting cDNA libraries to amplify the desired gene.

We can design beads with dual functionality: first, all oligos on a given bead can be uniformly barcoded; then, using two different primer sequences, a poly dT capture sequence can be specifically added to a percentage of the oligos, and the rest containing the common sequence for capturing the barcoded Abs (Scheme 4). This can yield the specific antibodies' panel for NHP immune profiling, which could then be conjugate to barcoding oligonucleotides by Cu-free Click chemistry detailed by Weiner et al. (Wiener et al., 2020) The antibodies can be functionalized by incubating with the amine reactive Dibenzocyclooctyne N-hydroxysuccinimidyl ester (DBCO-NHS). Meanwhile, we can acquire amine-modified oligos from IDT that could also be activated by adding 3-azidopropionic acid sulfo-NHS ester (3AA-NHS), which could conjugate with the functionalized antibody by Cu free click chemistry. By using these barcoded antibodies, the single-cell analysis would reveal information on both the surface proteins and the differential expression of each cell. We can generalize this protocol and effectively create a template for similar customization such that high throughput technology for immune profiling becomes accessible to many species.

Chapter 5

METHODS

5.1 Hairpin preparation

The hairpins are ordered as linear oligos from IDT without purification. In order to increase the likelihood of the hairpin assuming the correct secondary structure to template the PER reaction, the hairpins were diluted to the desired concentration with a Tris buffer with NaCl salt content to stabilize the hairpin structure. Then, the diluted hairpins were heat denatured and allowed to cool down to 20°C in the PCR where the rate of temperature setting was changed to 0.1C/sec. This encourages a larger population of the hairpins to assume the most energetically favorable confirmation and improve barcoding.

5.2 Hydrogel bead production

A acrylamide/bisacrylamide solution was prepared containing 3.6 mL of acrylamide/bis-acrylamide (AA/BIS), 2.58 mL of AA, and 3.82 mL of nuclease-free water. Filter the solution through a 0.2-μm membrane. The solution can be stored at 4 °C. An aqueous mixture was prepared by mixing 0.25 mL of the solution above, 0.1 mL of TBSET buffer, 0.1 mL of the DNA oligo, and 30 μL of 10% ammonium persulfate (APS) Carrier oil–TEMED mixture was prepared by combining 2 mL of carrier oil and 8 μL of TEMED. Vortex the mixture well. The flow rates used are 600 μL/hr for the aqueous phase and 600 μL/hr for the oil.

5.3 Bead cleanup

The top mineral oil was removed, and 500μL of Tris-buffered saline–EDTA–Triton (TBSET) buffer was added on top of the remaining beads. The bottom fluorous oil phases was carefully removed from the 2-mL collection tubes. To release polymerized HBs from the emulsion, add 1 mL of 20% (vol/vol) PFO in HFE-7500 per tube, vortex well, and centrifuge at 5,000 x g at room temperature for 30 s. HBs should not appear as a milky phase, so if the beads are still not a distinct translucent well-packed phase on top of the 20% (vol/vol) PFO phase, repeat this step. The aqueous phase was carefully removed, taking care not to transfer any fluorous oil and washing with TBSET was continued until all traces of oil have disappeared. The monodispersity of each fraction by visualization was checked on a microscope

slide or hemocytometer. All monodispersed fractions were combined in one 5mL or 15 mL tube depending on the scale of the experiment and cool the HBs to 4 °C. The hydrogel bead diameter was measured using microscope software rulers or by hemocytometer. The beads should have swelled to 70 μm in the aqueous solution. If any fractions contain larger beads, chilled HBs were passed through a 70-μm cell strainer attached to the top of a 50-mL tube and the buffer TBSET buffer gently poured and the beads gravity filtered to achieve optimal filtering. The beads are always stored in TET buffer

5.4 PER barcoding

This protocol demonstrates a typical reaction using close-packed beads mixed with the reagents detailed below. The HBs should be washed 3x in HBW buffer before mixing with the rest of the reagents.

- The hairpins were aliquoted in a 96-well plate or PCR tube strips at 1 μM concentration in hairpin dilution buffer (HDB).
- A master mix (above) is prepared containing the PER reaction without catalytic hairpins. Add no more than 90-100 μL of the above master mix per 200 μL well if scaled up to leave enough room for sufficient mixing of the beads during the PER reaction.
- The hairpin concentration added should be about ~1:30 ratio of the final concentration of beads in the reaction. Incubate at 37°C for 1 h with mixing. We use a thermomixer or a rotating platform for end-over-end mixing.
- Before pooling all reactions, add ExoVIII to each reaction and incubate for 15 minutes to disable the hairpins and added proteinaseK to the reaction and incubated for 15 minutes at 37°C followed by heat inactivation for 10 minutes at 55°C. Pool all reactions in a sterile reservoir (or an Eppendorf tube depending on the final volume) and mix thoroughly.

5.5 Addition of capture sequence

The beads are washed in HBW buffer 3 times. The capping reaction mixture was prepared by pre-incubating the beads with the capping oligo at 55°C for 10 minutes, then the T4 enzyme was added to the capping mixture and incubated at 37°C for 15 minutes. The enzyme activity is stopped by adding 2% of 0.5M of EDTA, and incubated at 75°C for 20 minutes.

5.6 Enzymatic cleanup

- The beads were washed with HBW buffer 3 times.
- The exonuclease cleanup reaction was prepared by mixing ExoI buffer with the beads once with it before starting the ExoI cleanup.
- An equal volume of 1x ExoI buffer and ExoI enzyme were mixed
- The ExoI digest was incubated at 37°C for 30 minutes.
- A 20 μL aliquot was set aside for gel analysis.
- Denaturation solution was prepared fresh while the ExoI reaction is incubating by combining 242 mL of nuclease-free water, 3.75 mL of 10 M NaOH, and 4.2 mL of 30% (wt/wt) Brij-35.
- After the ExoI incubation is complete, the beads were washed 3 times with 3 volumes STOP-10 buffer in an appropriately-sized tube. After the third wash, the supernatant was removed and the fresh-made denaturation solution was added and rotated at RT for 10 minutes.
- The beads were washed with denaturation solution 3 more times without incubation.
- After removing the supernatant of the third wash, 3 volumes neutralization buffer were added and repeated 3 times.
- A 20 μL aliquot was set aside for gel analysis
- The beads were stored in TET buffer for long term storage.

Cell culturing

Two frozen cell cultures (HEK293FT and MC-38) were thawed into DMEM growth medium supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum. The cells were incubated at 37℃ at 5% CO2 for two days, after which they were washed with 1x DPBS and dissociated into a single-cell suspension with 0.25% Trypsin at room temperature. The cells were then centrifuged at 300x rcf for 5 minutes and washed another time with fresh 1x DPBS, before the resulting pellet was suspended into 15% OptiPrep in DPBS and quantified with a hemocytometer. The concentration of cells was approximately 90 cells/ μ L.

5.7 Cell encapsulation

The cells were encapsulated with the HiPER beads using the inDrops platform microfluidics chip design (Atrandi Bioscience, MCN-C5) and the encapsulation performed with a 1Cell-Bio pump. Along with the beads, the cells were co-encapsulated with a cell lysis/reverse transcription premix consisting of buffer premix (5x RT buffer, Tris-HCl, dNTP mix, MgCl2, DTT, and Igepal CA-63), RNaseOUT, USER II Thermolabile enzyme, and Maxima H - RT enzyme. The bead occupancy rate per run ranged between 60-80% and fractions of approximately 3000 cells were collected on ice. Before the reverse transcription reaction, the emulsion is incubated at 37℃ for 15 minutes in order for the USER enzyme to release the barcoded DNA from the bead by catalyzing the removal of the dU base. The reaction is then incubated at 50℃ for 1 hour followed by a 5 minutes heat inactivation at 85℃.

5.8 Library preparation

Reverse transcription and second strand synthesis

The emulsion is broken by addition of 20% PFO in HFE7500 oil, and the aqueous phase is carefully transferred to a 0.45 μm tube filter where the solution is centrifuged at 16000x rfc for 5 minutes at 4℃. The filtered mixture is then incubated with ExoI enzyme at 37℃ for 30 minutes, then cleaned by 0.8x magnetic size selection beads (SPRI) using the manufacturer's standard protocol and eluting with 17 μL of nuclease free water. To the cleaned cDNA: mRNA hybrid, Second Strand Synthesis buffer and enzyme was added which was incubated for 2.5 hours at 16℃ and heat inactivated at 65℃ for 20 minutes.

T7 IVT linear amplification

After the SSS synthesis is completed, the HighScribe T7 high yield RNA amplification reagents are added to it (10x reaction buffer, NTPs, and T7 enzyme) to commence the IVT linear amplification of the RNA for 15 hours at 37℃. The mixture is purified with 1.0x volume of SPRI beads post incubation and eluted into 20 μL. The concentration of the IVT product is then quantified using an RNA Qubit kit, and optionally, the quality of the library (transcript distribution) can be visualized using a Bioanalyzer High Sensitivity RNA kit, or the equivalent ScreenTape kit.

RNA fragmentation

The fragmentation of RNA is performed using NEBNext Magnesium RNA Fragmentation Module adhering to the manufacturer's recommendations. As the RNA

molecules captured were typically shorter, the incubation time varied between 1-2 minutes before stopping and SPRI purifying the fragmented product with 1.25x volume equivalence of SPRI.

Random hexamer reverse transcription

In order to add the Read 2 sequencing template to the fragmented product, aDNA oligo with the reverse complement R2 sequence is designed with a random hexamer in its 3' end (R2-N6 primer), which will prime the RT reaction. R2-N6 is preincubated with dNTP mix and the purified fragmented RNA for 3 minutes at 70℃ and cooled on ice for 2 minutes. The Prime Script buffer and enzyme is added to the cooled mixture, incubated at 30℃ for 10 minutes, 42℃ for 1 hour, then heat inactivated for 10 minutes at 70℃. The RT product is SPRI purified with a 0.8x volume equivalence to prepare for qPCR.

qPCR and PCR amplification of DNA library

In order to accurately estimate the number of PCR cycles to sufficiently amplify the DNA library without over amplification, a diagnostic qPCR is run with minimal volume of the library, with the resulting Ct value informing the number of required cycles. The final product is dual size-selected to eliminate short sequences, primer dimers, and possible large concatemers or unfragmented DNA. Before sequencing the libraries, their bioanalyzer trace is evaluated to confirm the quality of the library.

BIBLIOGRAPHY

Adil, Asif et al. (2021). "Single-Cell Transcriptomics: Current Methods and Challenges in Data Acquisition and Analysis". In: *Front. Neurosci.* 15.

Buggenum, Jessie A G L van et al. (Mar. 2016). "A covalent and cleavable antibody-DNA conjugation strategy for sensitive protein detection via immuno-PCR". en. In: *Sci. Rep.* 6.1, pp. 1–12.

Conrad, Thomas et al. (Aug. 2020). "Maximizing transcription of nucleic acids with efficient T7 promoters". en. In: *Communications Biology* 3.1, pp. 1–8.

De Rop, Florian V et al. (Feb. 2022). "Hydrop enables droplet-based single-cell ATAC-seq and single-cell RNA-seq using dissolvable hydrogel beads". In: *Elife* 11. Ed. by Naama Barkai and Klaas Mulder, e73971.

Efremova, Mirjana and Sarah A Teichmann (Jan. 2020). "Computational methods for single-cell omics across modalities". en. In: *Nat. Methods* 17.1, pp. 14–17.

Fessenden, Marissa (Sept. 2015). "The cell menagerie: human immune profiling". en. In: *Nature* 525.7569, pp. 409–411.

Gong, Haibiao et al. (Jan. 2016). "Simple Method To Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Protein Detection in Single Cells". In: *Bioconjug. Chem.* 27.1, pp. 217–225.

Haile, Simon et al. (2021). "A Scalable Strand-Specific Protocol Enabling Full-Length Total RNA Sequencing From Single Cells". In: *Front. Genet.* 12.

Hashimshony, Tamar et al. (2012). "CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification". In: *Cell reports* 2.3, pp. 666–673.

Hollenstein, Marcel (2018). "DNA Synthesis by Primer Exchange Reaction Cascades". In: *Chembiochem* 19.5, pp. 422–424.

Hong, Tae Hee and Woong-Yang Park (2020). "Single-cell genomics technology: perspectives". en. In: *Exp. Mol. Med.* 52.9, pp. 1407–1408.

Ilicic, Tomislav et al. (Feb. 2016). "Classification of low quality cells from single-cell RNA-seq data". en. In: *Genome Biol.* 17, p. 29.

Islam, Saiful et al. (2011). "Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq". In: *Genome research* 21.7, pp. 1160–1167.

Jariani, Abbas et al. (May 2020). "A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast". In: *Elife* 9. Ed. by Antonis Rokas and Naama Barkai, e55320.

Jia, Peilin, Ruifeng Hu, et al. (Oct. 2022). "scGWAS: landscape of trait-cell type associations by integrating single-cell transcriptomics-wide and genome-wide association studies". en. In: *Genome Biol.* 23.1, p. 220.

Jia, Xianbo, Xinjian Lin, and Jichen Chen (Nov. 2017). “Linear and exponential TAIL-PCR: a method for efficient and quick amplification of flanking sequences adjacent to Tn5 transposon insertion sites”. In: *AMB Express* 7.1, p. 195.

Kashima, Yukie et al. (Sept. 2020). “Single-cell sequencing techniques from individual to multiomics analyses”. en. In: *Exp. Mol. Med.* 52.9, pp. 1419–1427.

Keren-Shaul, Hadas et al. (June 2019). “MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing”. en. In: *Nat. Protoc.* 14.6, pp. 1841–1862.

Kharchenko, Peter V, Lev Silberstein, and David T Scadden (2014). “Bayesian approach to single-cell differential expression analysis”. en. In: *Nat. Methods* 11.7, pp. 740–742.

Kishi, Jocelyn Y et al. (Feb. 2018). “Programmable autonomous synthesis of single-stranded DNA”. en. In: *Nat. Chem.* 10.2, pp. 155–164.

Klein, Allon M and Evan Macosko (2017). “InDrops and Drop-seq technologies for single-cell sequencing”. In: *Lab on a Chip* 17.15, pp. 2540–2541.

Kuhrt, David et al. (Jan. 2011). “Naıve and memory B cells in the rhesus macaque can be differentiated by surface expression of CD27 and have differential responses to CD40 ligation”. In: *J. Immunol. Methods* 363.2, pp. 166–176.

Landhuis, Esther (May 2018). “S ingle-cell approaches to immune profiling”. en. In: *Nature* 557.7706, pp. 595–597.

Lei, Yalan et al. (June 2021). “Applications of single-cell sequencing in cancer research: progress and perspectives”. In: *J. Hematol. Oncol.* 14.1, p. 91.

Li, Haikuo and Benjamin D Humphreys (July 2021). “Single Cell Technologies: Beyond Microfluidics”. en. In: *Kidney360* 2.7, pp. 1196–1204.

Li, Jifen and James Eberwine (May 2018). “The successes and future prospects of the linear antisense RNA amplification methodology”. en. In: *Nat. Protoc.* 13.5, pp. 811–818.

Liao, Jinling et al. (Jan. 2020). “Single-cell RNA sequencing of human kidney”. In: *Sci Data* 7.

Liu, Siyuan John et al. (Apr. 2016). “Single-cell analysis of long non-coding RNAs in the developing human neocortex”. en. In: *Genome Biol.* 17.1, p. 67.

Macosko, Evan Z et al. (May 2015). “Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets”. en. In: *Cell* 161.5, pp. 1202–1214.

Marx, Vivien (Jan. 2021). “Method of the Year: spatially resolved transcriptomics”. en. In: *Nat. Methods* 18.1, pp. 9–14.

Matuła, Kinga, Francesca Rivello, and Wilhelm T S Huck (2020). “Single-Cell Analysis Using Droplet Microfluidics”. en. In: *Advanced Biosystems* 4.1, p. 1900188.

Maune, Hareem T et al. (Jan. 2010). “Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates”. en. In: *Nat. Nanotechnol.* 5.1, pp. 61–66.

Melsted, Páll et al. (Apr. 2021). “https://doi.org/10.1038/s41587-021-00870-2Modular, efficient and constant-memory single-cell RNA-seq preprocessing”. In: *Nature biotechnology*. DOI: https://doi.org/10.1038/s41587-021-00870-2.

Mereu, Elisabetta et al. (June 2020). “Benchmarking single-cell RNA-sequencing protocols for cell atlas projects”. en. In: *Nat. Biotechnol.* 38.6, pp. 747–755.

Messaoudi, Ilhem et al. (Jan. 2011). “Nonhuman Primate Models of Human Immunology”. In: *Antioxid. Redox Signal.* 14.2, pp. 261–273.

“Method of the Year 2013” (Jan. 2014). en. In: *Nat. Methods* 11.1, pp. 1–1.

Minev, Dionis et al. (Dec. 2019). “Rapid in vitro production of single-stranded DNA”. In: *Nucleic Acids Res.* 47.22, pp. 11956–11962.

Muyas, Francesc et al. (July 2023). “De novo detection of somatic mutations in high-throughput single-cell profiling data sets”. en. In: *Nat. Biotechnol.*

Ning, Linhong et al. (Mar. 2023). “Primer Exchange Reaction Coupled with DNA-Templated Silver Nanoclusters for Label-Free and Sensitive Detection of MicroRNA”. en. In: *Appl. Biochem. Biotechnol.*

Ntranos, Vasilis et al. (Feb. 2018). “Identification of transcriptional signatures for cell types from single-cell RNA-Seq”. en. In: *bioRxiv*, p. 258566.

Petti, Allegra A et al. (Aug. 2019). “A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing”. en. In: *Nat. Commun.* 10.1, p. 3660.

Petukhov, Viktor et al. (June 2018). “dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments”. en. In: *Genome Biol.* 19.1, p. 78.

Pimentel, Harold et al. (July 2017). “Differential analysis of RNA-seq incorporating quantification uncertainty”. en. In: *Nat. Methods* 14.7, pp. 687–690.

Pisco, Angela Oliveira, Bruno Tojo, and Aaron McGeever (2021). “Single-Cell Analysis for Whole-Organism Datasets”. In: *Annual Review of Biomedical Data Science* 4.1, pp. 207–226.

Qian, Lulu and Erik Winfree (2011). “Scaling up digital circuit computation with DNA strand displacement cascades”. In: *science* 332.6034, pp. 1196–1201.

Rakszewska, Agata et al. (Oct. 2014). “One drop at a time: toward droplet microfluidics as a versatile tool for single-cell analysis”. en. In: *NPG Asia Mater* 6.10, e133–e133.

Ravindran, Resmi et al. (Apr. 2014). “Plasma antibody profiles in non-human primate tuberculosis”. In: *J. Med. Primatol.* 43.2, pp. 59–71.

Rothemund, Paul W K (Mar. 2006). “Folding DNA to create nanoscale shapes and patterns”. en. In: *Nature* 440.7082, pp. 297–302.

Setliff, Ian et al. (2019). “High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity”. en. In: *Cell* 179.7, 1636–1646.e15.

Sha, Ying, John H Phan, and May D Wang (2015). “Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data”. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, pp. 6461–6464.

Shapiro, Ehud, Tamir Biezuner, and Sten Linnarsson (Sept. 2013). “Single-cell sequencing-based technologies will revolutionize whole-organism science”. en. In: *Nat. Rev. Genet.* 14.9, pp. 618–630.

Shivram, Haridha and Vishwanath R Iyer (Sept. 2018). “Identification and removal of sequencing artifacts produced by mispriming during reverse transcription in multiple RNA-seq technologies”. en. In: *RNA* 24.9, pp. 1266–1274.

Southard-Smith, Austin N et al. (July 2020). “Dual indexed library design enables compatibility of in-Drop single-cell RNA-sequencing with exAMP chemistry sequencing platforms”. en. In: *BMC Genomics* 21.1, p. 456.

Tanay, Amos and Arnau Sebé-Pedrós (May 2021). “Evolutionary Cell Type Mapping with Single-Cell Genomics”. en. In: *Trends Genet.*

Tang, Fuchou, Catalin Barbacioru, Ellen Nordman, et al. (2010). “RNA-Seq analysis to capture the transcriptome landscape of a single cell”. In: *Nature protocols* 5.3, pp. 516–535.

Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, et al. (2009). “mRNA-Seq whole-transcriptome analysis of a single cell”. In: *Nature methods* 6.5, pp. 377–382.

Tavakoli-Koopaei, Reyhaneh, Fatemeh Javadi-Zarnaghi, and Hossein Mirhendi (Apr. 2022). “Unified-amplifier based primer exchange reaction (UniAmPER) enabled detection of SARS-CoV-2 from clinical samples”. en. In: *Sens. Actuators B Chem.* 357, p. 131409.

Wang, Yongcheng et al. (n.d.). “Dissolvable Polyacrylamide Beads for High-Throughput Droplet DNA Barcoding”. en. In: *Adv. Sci. Lett.* n/a.n/a (), p. 1903463.

Wen, Wen et al. (Mar. 2020). “Immune Cell Profiling of COVID-19 Patients in the Recovery Stage by Single-Cell Sequencing”. en. In: *medRxiv*, p. 2020.03.23.20039362.

Wiener, Julius et al. (Jan. 2020). “Preparation of single- and double-oligonucleotide antibody conjugates and their application for protein analytics”. In: *Sci. Rep.* 10.

Yeh, Hsin-Sung and Jeongsik Yong (Apr. 2016). “Alternative Polyadenylation of mRNAs: 3-Untranslated Region Matters in Gene Expression”. In: *Mol. Cells* 39.4, pp. 281–285.

Zhang, Xiannian et al. (Jan. 2019). “Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems”. en. In: *Mol. Cell* 73.1, 130–142.e5.

Zilionis, Rapolas et al. (Jan. 2017). “Single-cell barcoding and sequencing using droplet microfluidics”. en. In: *Nat. Protoc.* 12.1, pp. 44–73.

Appendix A

SUPPLEMENTARY INFORMATION

A.1 Materials

- Mineral oil (Sigma-Aldrich, cat. no. M841-100ML)
- Nuclease-free water (Life Technologies, cat. no. 10977-023)
- EDTA (0.5 M; Life Technologies, cat. no. 15575-038, 100 mL)
- UltraPure Tris–HCl, pH 8.0 (1 M; Life Technologies, cat. no. 15568-025)
- Tween-20 (Sigma-Aldrich, cat. no. P7949-100 ML)
- Triton X-100 (Fisher, cat. no. BP151-100)
- NaCl (Sigma-Aldrich, cat. no. S3014-1KG)
- KCl (Fisher, cat. no. P217-500)
- Carrier oil (RAN Biotechnologies, cat. no. 008-FluoroSurfactant-2wtH-50G;)
- 1H,1H,2H,2H-Perfluorooctanol (Sigma Aldrich, cat. no. 370533-25G)
- Exonuclease I (ExoI, 20,000 U/ml; NEB, cat. no. M0293L)
- Deoxynucleotide solution mix (dNTP; 10 mM each; NEB, cat. no. N0447L)
- 3M Novec 7500 Engineered Fluid (HFE-7500 oil, 3 M; Novec, cat. no. Novec 7500)
- Acrylamide solution (AA; 40% (wt/wt); Sigma-Aldrich, cat. no. A4058-100ML)
- AA/bis-acrylamide solution, 40% (wt/wt), molar ratio 19:1 (AA/BIS; Sigma-Aldrich, cat. no. A9926-100ML)
- Span-80 (Sigma-Aldrich, cat. no. S6760-250ML)
- Hexane (Sigma-Aldrich, cat. no. 227064-1L)

- N,N,N',N'-Tetramethylethylenediamine (TEMED; Sigma-Aldrich, cat. no. T9281-25ML)
- APS (Sigma-Aldrich, cat. no. A3678-25G)
- Bst 3.0 DNA polymerase (Bst; 8,000 U/ml; NEB, cat. no. M0374L)
- Bsu polymerase (8,000 U/mL; NEB, cat. no. M0374L)
- Brij-35 (30% (wt/wt); MCLab, cat. no. DB35-100)
- NaOH solution (50%(w/w); fisher, cat. no. SS254-500)
- USER (1U/μL) enzyme; (NEB, cat. no. M5505L)
- TBE-Urea Precast Gels; (BioRad, cat. no. 4566036)
- ssDNA & Low-Input DNA Library Preparation Kit (IDT, cat. no. 10009859)
- Thermolabile ProteinaseK(NEB, cat. no. P8111S)
- Exonuclease VIII, truncated (NEB, cat. no. M0545S)
- Magnesium RNA Fragmentation Module Protocol (NEB, cat. no. E6150S)
- Tris EDTA–Tween buffer (TET): Combine 480 mL of nuclease-free water, 5 mL of 1 M Tris–HCl (pH 8.0), 10 mL of 0.5 M EDTA, and 5 mL of 10% (vol/vol) Tween-20 in nuclease-free water. Filter the buffer through a 0.2-μm membrane.
- Hydrogel bead wash buffer (HBW): Combine 980 mL of nuclease-free water, 10 mL of 1 M Tris–HCl (pH 8.0), 200 μL of 0.5 M EDTA, and 10 mL of 10% (vol/vol) Tween-20 in nuclease-free water. Filter the buffer through a 0.2-μm membrane.
- Tris-buffered saline–EDTA–Triton buffer (TBSET): Combine 822 mL of nuclease-free water, 10 mL of 1 M Tris–HCl (pH 8.0), 137 mL of 1 M NaCl, 1.35 mL of 2 M KCl, 20 mL of 0.5 M EDTA, and 10 mL of 10% (vol/vol) Triton X-100 in nuclease-free water. Filter the buffer through a 0.2-μm membrane.
- 1% (vol/vol) Span-80 in hexane: Combine 99 mL of hexane and 1 mL of Span-80 in a fume hood. Store the solution in a glass bottle in a fume hood.

- STOP-25 buffer: Combine 880 mL of nuclease-free water, 10 mL of 1 M Tris–HCl (pH 8.0), 50 mL of 0.5 M EDTA, 10 mL of 10% (vol/vol) Tween-20, and 50 mL of 2 M KCl. Filter the buffer through a 0.2-μm membrane.
- STOP-10 buffer: Combine 1,820 mL of nuclease-free water, 20 mL of 1 M Tris–HCl (pH 8.0), 40 mL of 0.5 M EDTA, 20 mL of 10% (vol/vol) Tween-20, and 100 mL of 2 M KCl. Filter the buffer through a 0.2-μm membrane.
- Denaturation solution: Prepare fresh while the ExoI reaction is incubating. Combine 242 mL of nuclease-free water, 3.75 mL of 10 M NaOH, and 4.2 mL of 30% (wt/wt) Brij-35. Filter the buffer through a 0.2-μm membrane.
- Neutralization buffer: Combine 385 mL of nuclease-free water, 50 mL of 1 M Tris–HCl (pH 8.0), 10 mL of 0.5 M EDTA, 5 mLof 10% (vol/vol) Tween-20, and 50 mL of 1 M NaCl. Filter the buffer through a 0.2-μm membrane.

A.2 Hydrogel bead Production

- Prepare 10 mL of the acrylamide/bisacrylamide solution which contains: 3.6 mL of acrylamide/bis-acrylamide (AA/BIS), 2.58 mL of AA, and 3.82 mL of nuclease-free water. Filter the buffer through a 0.2-μm membrane. The solution can be stored at 4 °C.

Reagent	[stock]	volume (μL)	[final]
Acrylamide/bis-acrylamide(1:19)	40%	3 mL	12%
Acrylamide solution	40%	3.1 mL	12.6% AA
N.F. water		3.9 mL	
total	-	10 mL	-

Table A.1: Table of reagents for the aqueous acrylamide mixture.

- Prepare the aqueous mixture by mixing 0.25 mL of the solution above, 0.1 mL of TBSET buffer, 0.1 mL of the DNA oligo, and 30 μL of 10% ammonium persulfate (APS).
- Prepare 2 ml of carrier oil–TEMED mix by combining 2 ml of carrier oil and 8 μl of TEMED. Vortex the mixture well.
- Transfer the carrier oil–TEMED mix to a 3-ml syringe. To transfer, load the oil–TEMED mix into a p1000 pipette tip, insert the tip firmly into the Luer tip of the syringe, and suck the liquid into the syringe by withdrawing the plunger.

Reagent	[stock]	volume (μL)	[final]
Nuclease-free water		520	
1× TBSET buffer		100	
4× AB solution	4x	250	1x
DNA oligo in IDTE	250μM	100	25 μM
APS	10% (wt/vol)	30	0.3% (wt/vol)
Total		1000	

Table A.2: Table of reagents of the aqueous phase to manufacture the hydrogel beads.

- Attach a 25-gauge needle and connect a piece of tubing (40 cm (16 inches)) using tweezers. Hand-prime the syringe: hold the syringe upright, and then slowly push the plunger until the liquid fills half of the tubing. Then mount the syringe on a syringe pump.
- flow rate: 600 μL/hr for aqueous phase 600 μL/hr for oil.
- Prepare 1 ml of acrylamide–primer mix by combining the following reagents: Note that APS should not be added until immediately before loading the mixture into the syringe.
- Prepare 1 mL of carrier oil–TEMED mix by combining 1 mL of carrier oil and 4 μL of TEMED. Vortex the mixture well and filter the solution through a 0.2-μm membrane. (The TEMED might start precipitating in the oil with time, and definitely overnight, so try to not prepare in advance especially if making a larger batch of beads.)
- Transfer the carrier oil–TEMED mix to a 3-mL syringe. To transfer, load the oil–TEMED mix into a p1000 pipette tip, insert the tip firmly into the Luer tip of the syringe, and suck the liquid into the syringe by withdrawing the plunger without breaking contact as the oil can spill out of the top. Attach a 26-gauge needle and connect a piece of tubing (at an appropriate length to connect the chip to the pump) using tweezers. Hand-prime the syringe: hold the syringe upright, and then slowly push the plunger until the liquid fills most of the tubing. Then mount the syringe on a syringe pump.
- Load and prime a second 3-mL syringe with acrylamide-DNA aqueous solution (gel mix) and mount on a syringe pump. Pump-prime the syringe: run the syringe pump until the liquid reaches the end of the tubing.
- Connect the syringes (via tubing, using a tweezers to aid insertion) to corresponding inlets of the hydrogel bead generation device and place it on the

stage of a microscope. Use flow rates of 600 μL/hr for the carrier oil and 600 μL/hr for the acrylamide–primer mix.

- Begin droplet generation by operating the syringe pumps at the stated flow rates. Collect eluent from the droplet generation device via tubing connected to the outlet. Collect eluent in Eppendorf tubes pre-filled with 250 μL mineral oil, submerging the outlet tubing in the mineral oil.

- Ensure stable flow before collecting emulsion for polymerization, and keep a waste collection tube on hand in case the on-chip flow profile becomes unstable. Stable flow is characterized by a static "cone" emanating from the T-junction. Monodisperse beads can be observed at the oulet port during device operation.

- Droplet diameter should be approximately 60-μm. Droplet size can be checked during device operation by "spotting" the eluent on a microscope slide. While flowing, expel 20 μL of eluent on a clean microscope slide, and quickly determine droplet diameter with calibrated microscope software. Adjust flow rates accordingly —higher aqueous flow rates relative to the oil will give larger droplets, while a slower aqueous flow rate will give smaller droplets. It is strongly recommended to collect fractions throughout the course of droplet generation as the flow profile can destabilize due to ambient vibrations or wearing out the hydrophobic coating of the microfluidics channel.

- Remove the top mineral oil, and add 500 μL of Tris-buffered saline–EDTA–Triton (TBSET) buffer on top of the remaining beads. Carefully remove the bottom fluorous oil phases from the 2-mL collection tubes. To avoid loss of beads, use a thin pipet tip (Rainin P20 tips, or gel loading tips) for the bottom phase, as beads stick to the outside surface of the tip.

- To release polymerized HBs from the emulsion, add 1 mL of 20% (vol/vol) PFO in HFE-7500 per tube, vortex well, and centrifuge at 5,000 x g at room temperature for 30 s. HBs should not appear as a milky phase, so if the beads are still not a distinct translucent well-packed phase on top of the 20% (vol/vol) PFO phase, repeat this step.

- Carefully remove the aqueous phase taking care not to transfer any fluorous oil and keep washing with TBSET until all traces of oil have disappeared.

- Check monodispersity of each fraction by visualization on a microscope slide or hemocytometer. Combine all monodispersed fractions in one 5mL or 15 mL tube depending on the scale of the experiment and cool the HBs to 4 °C. Measure the hydrogel bead diameter using microscope software rulers or by hemocytometer. The beads should have swelled to 70 μm in the aqueous solution.

- If any fractions contain larger beads, pass the chilled HBs through a 70-μm cell strainer attached to the top of a 50-mL tube. Also use an ice-cold TBSET buffer to facilitate the passage of the beads through the mesh. Don't try to force the beads through the mesh by pushing through the buffer. Instead, pour the buffer gently and let the beads gravity filter to achieve optimal filtering.

- Always store beads in TET buffer.

A.3 Barcoding

- Aliquot the hairpins in a 96-well plate or PCR tube strips at 1 μM concentration in hairpin dilution buffer (HDB). set up PCR to the following program:

Temp. (°C)	Time (Min)	Ramp-up rate (°C/sec)
98	2	4
20	HOLD	0.1

Table A.3: The PCR program steps for heat-treating the HiPER barcoding hairpins.

- Note that this protocol demonstrates a typical barcoding reaction using close-packed beads mixed with the reagents detailed below. The HBs should be washed 3 times in HBW buffer before mixing with the rest of the reagents.

- A master mix (above) is prepared containing the PER reaction without catalytic hairpins.

- Note that no more than 90-100 μL should be added of the above master mix per 200 μL well if scaled up to leave enough room for sufficient mixing of the beads during the PER reaction.

- The hairpin concentration added should be about ~1:30 ratio of the final concentration of beads in the reaction.

- Incubate at 37°C for 1 h with mixing using a thermomixer or a rotating platform for end-over-end mixing.

Reagent	[stock]	volume μL	[final]	unit
T7AAopt-default (25 μM)	25	300	7.5	μM
100 mM magnesium sulfate	100	100	10	mM
NEB 2.0	10	100	1	1x
dHTP	10	75	0.75	mM
spermidine	100	100	10	mM
Bsu LF	5	50	0.25	U/μL
dGTP_cleanup_hairpin	100	10	1	μM
hairpin 1	1000		265	nM
N.F. water		0		
total		735		

Table A.4: Table of reagents of the barcoding reaction.

- Before pooling all reactions, add ExoVIII to each reaction and incubate for 15 min to disable the hairpins. Add proteinaseK to reaction incubating for 15 minutes at 37C and heat inactivating for 10 minutes at 55C. The exonuclease can also be heat inactivated at 75C for 20 minutes.
- Pool all reactions in a sterile reservoir (or an Eppendorf tube depending on the final volume) and mix thoroughly.
- Note that 96-well plates covered with a foil seal have caused volume loss during the experiment, thus, we recommend using strip tubes with individual caps.
- Wash beads with Wash Buffer.

Capping Barcoded Hydrogel beads

- Wash the beads in HBW buffer 3 times in a 15 mL tube. Note that the beads pellet better when using lower brake speed on swinging bucket centrifuge. This step decreases loss of beads when discarding the supernatant between washing steps.
- Prepare the capping reaction mixture as follows:
- Incubate the capping reaction mixture at 37C for 15 minutes. Add 3 volumes of STOP-25 buffer to stop the reaction.
- Remove a 20 μl aliquot for gel analysis.
- Store the capped beads in TET buffer overnight or proceed with the enzymatic cleanup immediately.

Reagent	[stock]	volume (μL)	[final]	unit
barcoded beads	25	100	8.3	μM
polydT_cap primer (100μM)	100	30	10	μM
T4 polymerase	3	30	0.3	U/μL
NEB r2.1 buffer	10	30	1	1 x
10mM dNTP	10	30	1	mM
water		80	-	-
total		300	-	-

Table A.5: Table of reagents of the capping reaction.

- Wash the beads with HBW buffer 3 times as before. Prepare 1x ExoI buffer and wash the beads once with it before starting the ExoI cleanup. Add an equal volume of 1x ExoI buffer and ExoI enzyme, as detailed below.
- Perform ExoI digest 37C for 4 hours. Add 3 volumes of STOP-25 buffer and heat inactivate at 80C for 20 minutes.
- Remove a 20 μl aliquot for gel analysis.
- Prepare denaturation solution fresh while the ExoI reaction is incubating by combining 242 mL of nuclease-free water, 3.75 mL of 10 M NaOH, and 4.2 mL of 30% (wt/wt) Brij-35. Keep on ice.
- After the ExoI incubation is complete, wash the beads 3 times with 3 volumes STOP-10 buffer in an appropriately-sized tube. After 3rd wash, remove the supernatant and add the fresh-made denaturation solution and rotate at RT for 10 minutes. Wash the beads with denaturation solution 3 more times without incubation.
- After removing the supernatant of the third wash, add 3 volumes neutralization buffer and repeat three times. Remove a 20 μL aliquot for gel analysis Finally, wash beads with TET buffer for long term storage.
- To quality control by electrophoresis, take the 20 μL aliquots taken throughout the barcoding and capping steps, take 7 μL (3 μL hard-packed beads) into a separate tube and add 1 μL of each of: 0.1 M DTT, 10x Cutsmart, and USER or USERII enzyme. (USER enzyme cannot be heat inactivated while USERII enzyme can be heat-inactivated at 85°C for 20 minutes)
- Incubate at 37°C for 15-30 minutes.

- Add a 2x denaturing dye and denature for 10 minutes at 95°C in preparation for electrophoresis.
- Load samples into gel with appropriate size markers (for barcoded and capped beads). Stain with SYBR gold to visualize the efficiency of the reaction.
- wash beads into BHM buffer 3 times.
- loaded 23μL MM to each well (should be 32 reactions), then add 8.11 μL of H1 hairpins incubated at 37°C for 60 minutes added 10 μL of ExoVIII MM and incubated for 15 minutes at 37°C added 10 μL of ProteinaseK MM and incubated for 15 minutes at 37°C then inactivated for 10 minutes at 55°C.
- combined all 30 reactions and washed with HBW for 3 times.

A.4 Library preparation

- Break emulsion by adding 20% PFO in HFE7500 oil, and carefully transfer the aqueous phase is to a 0.45 μm tube filter where the solution is centrifuged at 16000 xrfc for 5 minutes at 4°C.
- Add the mixture from Table A.6.
- Incubate the filtered mixture with ExoI enzyme at 37°C for 30 minutes, then cleanup by 0.8x magnetic size selection beads (SPRI) using the manufacturer's standard protocol and eluting with 17 μL of nuclease free water.

Reagent	volume (μ)
water	79
HinfI HF 10x buffer	9
HinfI enzyme	7
ExoI enzyme	5
total	100

Table A.6: Table of reagents of the ExoI digest reaction.

- To the cleaned cDNA: mRNA hybrid, add Second Strand Synthesis buffer and enzyme (see Table A.7) and incubate for 2.5 hours at 16°C and heat inactivate at 65°C for 20 minutes.
- After the SSS synthesis is completed, add the HighScribe T7 high yield RNA amplification reagents to it (see Table A.8) to commence the IVT linear amplification of the RNA for 15 hours at 37°C.

Reagent	volume (μ)
purified hybrid RNA:cDNA	17
10x SSS buffer	2
SSS enzyme mix	1
total	20

Table A.7: Table of reagents of the Second Strand Synthesis reaction.

- Purify the reaction with 1.0x volume of SPRI beads post incubation and eluted into 20 μL.
- Quantify the concentration of the IVT product by using an RNA Qubit kit, and optionally, visualize the quality of the library (transcript distribution) using a Bioanalyzer High Sensitivity RNA kit, or the equivalent ScreenTape kit.

Reagent	volume (μ)
purified cDNA	20
T7 HiScribe IVT buffer	8
NTP	32
T7 HiScribe IVT enzyme	8
water	12
total	80

Table A.8: Table of reagents of the IVT linear amplification reaction.

- Perform the fragmentation of RNAusing NEBNext Magnesium RNA Fragmentation Module adhering to the manufacturer's recommendations (Table A.9).
- Purify the fragmented product with 1.25x volume equivalence of SPRI.

Reagent	volume (μ)
Fragmented IVT	10
10x STOP frag buffer	2
CleanNGS	24
water	8
total	34

Table A.9: Table of reagents of the RNA Fragmentation reaction.

- Pre-incubate R2-N6 with dNTP mix and the purified fragmented RNA for 3 minutes at 70℃ and cooled on ice for 2 minutes (Table A.10).

Reagent	volume (μ)
eluted RNA	8
dNTP (10mM)	1
TD_Read1-N6	2
total	11

Table A.10: Table of reagents of the Random Hexamer RT reaction.

- Add the Prime Script buffer and enzyme to the cooled mixture, incubate at 30℃ for 10 minutes, 42℃ for 1 hour, then heat inactivate for 10 minutes at 70℃ (Table A.11). Purify the RT product with a 0.8x volume equivalence of SPRI to prepare for qPCR.

Reagent	volume (μ)
RNA mixture	11
5x PrimeScript Buffer	4
RNaseOUT	1
PrimeScript enzyme	0.5
water	3.5
total	20

Table A.11: Table of reagents of the Prime Script RT reaction.

- PCR amplify the RT product with the minimum number of cycles to avoid over amplification (Table A.12).

Reagent	volume (μ)
eluted cDNA	9.5
water	0.5
HiFi mix	12.5
p5/p7 primer mix (5μM)	2.5
total	25

Table A.12: Table of reagents of the PCR amplification reaction with indexed PCR primers.

Appendix B

EXTERNAL LINKS

- Analysis pipeline: https://github.com/tdilanyan/thesis
- Data: 10.22002/4z7fd-ryp16
- Design files: 10.22002/kpgvp-g8j05