An Optically Decodable Bead Array for Linking Imaging and Sequencing with Single-Cell Resolution

Jinzhou Yuan[1], Jenny Sheng[2], Peter A. Sims[1,3,4,*]

[1]Department of Systems Biology, Columbia University Medical Center, New York, NY 10032 USA
[2]Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University Medical Center, New York, NY 10032 USA
[3]Department of Biochemistry & Molecular Biophysics, Columbia University Medical Center, New York, NY 10032 USA
[4]Sulzberger Columbia Genome Center, Columbia University Medical Center, New York, NY 10032 USA
*Correspondence to pas2182@cumc.columbia.edu

Abstract

Optically decodable beads link the identity of an analyte or sample to a measurement through an optical barcode, enabling libraries of biomolecules to be captured on beads in solution and decoded by fluorescence. This approach has been foundational to microarray, sequencing, and flow-based expression profiling technologies. We have combined microfluidics with optically decodable beads to link phenotypic analysis of living cells to sequencing. As a proof-of-concept, we applied this to demonstrate an accurate and scalable tool for connecting live cell imaging to single-cell RNA-Seq called Single Cell Optical Phenotyping and Expression (SCOPE-Seq).

Main Text

Recent advances in microfluidics and cDNA barcoding have led to a dramatic increase in the throughput of single-cell RNA-Seq (scRNA-Seq)[1-5]. However, unlike earlier methodologies[6,7], these new tools do not offer a straightforward way to directly link phenotypic information obtained from individual, live cells to their expression profiles. Nonetheless, microwell-based implementations of scRNA-Seq are compatible with a wide variety of phenotypic measurements including live cell imaging, immunofluorescence, and protein secretion assays[3,8-11]. These methods involve co-encapsulation of individual cells and barcoded mRNA capture beads in arrays of microfabricated chambers. Because the barcoded beads are randomly distributed into microwells, one cannot directly link cellular phenotypes measured in the microwells to their corresponding expression profiles. In SCOPE-Seq, we use optically barcoded beads[12] that allow the sequencing barcode associated with each single-cell cDNA library on the sequencer to be identified by fluorescence microscopy. Thus, we can obtain images, movies, or other phenotypic data from individual cells by microscopy and directly link this information to genome-wide expression profiles.

We have previously demonstrated compatibility of the commercially available "Drop-Seq" beads[1] with our microwell array system for dissecting gene expression heterogeneity in immortalized cell lines[8], patient-derived glioma neurospheres[8], and human glioblastoma resections[13] with single-cell resolution. Here we generate optically barcoded mRNA capture beads for our microwell array system from "Drop-Seq" beads. These beads are conjugated to oligonucleotides with a 3'-poly(dT) terminus and a cell-identifying sequencing barcode upstream. To enable optical identification of the sequencing barcode on each bead, we first attach a unique combination of oligonucleotides selected from a set of 12 in two cycles of split-pool ligation (left panel of **Fig. 1A**). We refer to this combination as an optical barcode, because it can be decoded by sequential fluorescence hybridization, to distinguish it from the sequencing barcode. We discovered that a small fraction of the oligonucleotides detach from the beads after

sonication. This allows us to generate a "bead-free" sequencing library from the supernatant prior to ligation to identify the sequencing barcodes in each optical barcode reaction mixture (right panel of **Fig. 1A**). Split-pool, single-stranded ligation produces a final pool of beads with both optical and sequencing barcodes (dual-barcoded beads, **Fig. 1B**), and sequencing the bead-free library associated with each ligation reaction produces a look-up table linking the optical and sequencing barcodes for each bead in the pool (**Fig. 1C**). Importantly, bead-free library construction and sequencing is done once for each batch of optically barcoded beads, which can be used for many SCOPE-Seq experiments.

In SCOPE-Seq, we load cells into individual wells in the microwell array device described previously[3,8], collect live cell imaging data, and then load dual-barcoded mRNA capture beads into the wells (**Fig. 1D**). We then perform cell lysis, mRNA capture, reverse transcription, and exonuclease digestion with a computer-controlled system in the microwell device. These reactions convert mRNA from individual cells into PCR amplifiable, sequence barcoded cDNA which is covalently attached to the bead that is co-encapsulated with each cell. We perform 12 cycles of sequential fluorescence hybridization to determine the combination of 12 optical barcode oligonucleotides (OBOs) on each bead (**Fig. 1D-E**). We call this process optical demultiplexing. We then cut the microwell array device into multiple pieces, extract the cDNA-bound, optically demultiplexed beads from each piece for pooled scRNA-Seq library construction. Beads from each piece are processed and indexed separately, thereby increasing our multiplexing capacity to $2^{12}*N = 4096*N$ where N is the number of pieces. The resulting scRNA-Seq libraries from different pieces are then pooled and sequenced to obtain the RNA-Seq profile of the individual cells and their sequencing barcodes. To link live cell imaging and large-scale scRNA-Seq, the optical demultiplexing images are first processed to identify the optical barcode on each bead. The associated sequencing barcode is then identified using the look-up table described above to link microscopy data from a cell in a particular microwell to its RNA-Seq profile (**Fig. 1E**).

High-quality images are crucial for optical demultiplexing. **Fig. 1F** shows two, well-separated "on-off" bead states in the distribution of centered fluorescence intensities of the beads across all demultiplexing cycles. As expected, the distribution of optical barcode usage closely follows Poisson statistics (**Fig. 1G**) suggesting a uniform representation of optical barcodes among the stock beads.

An accurate optical barcode-to-sequencing barcode look-up table is also critical. We generate the look-up table by associating the unique combination of oligonucleotides added to each ligation reaction in each cycle of split-pool synthesis with the sequencing barcodes detected in the corresponding bead-free library. During each round of split-pool ligation, one of $2^6 = 64$ combinations of six oligonucleotides is added to each reaction, generating a 6-bit binary optical barcode for each split. Two rounds of split-pool ligation are performed thus yielding 12-bit binary optical barcodes. The sequencing barcode of each bead should be represented once among the 64 bead-free libraries from the first round of ligation and once among the second set. For each sequencing barcode in each round of split-pool synthesis, we ranked its normalized frequency in the 64 libraries in descending order and averaged the resulting vector across all sequencing barcodes (**Fig. 1H-I**). As expected, the vast majority of sequencing barcodes are represented in only one of the 64 libraries for both rounds of ligation, suggesting only minor imperfections due to sequencing error and low levels of residual DNA between split-pool cycles. The signal-to-background ratios are 17.1 and 9.5 for the first and second rounds of split-pool reactions, respectively.

To further characterize the performance and demonstrate the utility of SCOPE-Seq, we performed a mixed-species experiment where human (U87) and mouse (3T3) cells are labeled with differently colored live stain dyes, mixed together, and then analyzed with SCOPE-Seq. Because human and mouse cells are differentially labeled, we can determine the species of each captured cell by live cell fluorescence imaging in the microwells. The resulting scRNA-Seq libraries, from which we detected ~3,600 unique

transcripts per cell on average, allowed us to independently determine the species associated with each sequencing barcode. We treated these two independent species identifications as a ground truth and considered any inconsistencies between them to result from optical barcode linkage error. We define the linking accuracy as the fraction of single-cell imaging and RNA-Seq data sets that are correctly linked among all linked data sets. **Fig. 2A** shows a scatter plot of the number of reads aligning uniquely to the human and murine transcriptomes for each cell that was linked to imaging data. Importantly, the data points derived from scRNA-Seq are colored using the actual two-channel fluorescence intensity data from live cell microscopy. Species calling was 98.6% concordant between the imaging and scRNA-Seq data, and after correcting for the difference in abundance of human and murine cells, we obtained an imaging-to-sequencing linking accuracy of 96.1% for singlets from which we can unambiguously obtain a species call. From a total of 2,352 RNA-Seq expression profiles, 1,133 of them were linked to imaging data (48.2%). The majority of unlinked cells are due to optical demultplexing errors which usually lead to a failure to link rather than an error. Because the yield of linked optical and sequencing barcodes is high (89.6%), we know that these errors arise from the optical demultiplexing process itself.

The microwell array device is deliberately overloaded to generate a small but non-negligible fraction of multi-species multiplets. Indeed, we observe highly pure human and mouse as well as mixed species RNA-Seq expression profiles (**Fig. 2A**). As expected from the high linking accuracy of SCOPE-Seq, the species purity of RNA-Seq expression profiles are highly consistent with their fluorescent labels from live-cell imaging data (**Fig. 2A**). Interestingly, we were able to remove most of the obvious multi-species multiplets using live-cell imaging data (**Fig. 2B**). We used scRNA-Seq profiles with purity below 70% as a standard for calling human-mouse doublets. We then blinded ourselves to the two-color fluorescence information that distinguishes human-mouse doublets in our imaging data and attempted to identify doublets by manually examining the monochrome images. Imaging- and sequencing-based doublet identifications were conducted by different researchers in a blinded fashion. This resulted in a sensitivity of 66% and a specificity of 99.1% for doublet detection, and a concordance of 87.5% between one- and two-color imaging. Some false negatives likely arise from imperfections in our scRNA-Seq "ground truth", but there could be other issues as well, such as cells that are difficult to distinguish with the low resolution imaging we employed. We anticipate that more sophisticated image processing and better microscopy and cell/nucleus stains will lead to further improvements in sensitivity to make SCOPE-Seq a highly effective approach to detecting multiplets in scRNA-Seq.

More interestingly, additional imaging features obtained from cells may carry information related to gene expression. We measured 19 imaging features from the fluorescence images of cells reflecting various aspects of cell size, shape, and intensity distribution. Principal component (PC) analysis on these features suggests significant heterogeneity among the cells (**Fig. 2C**). The first PC is primarily correlated with cell size-related features (**Fig. 2D**) and the second with shape-related features (**Fig. 2E**). We then ranked protein coding genes by their correlation with these PCs and performed gene set enrichment analysis for each PC (**Fig. 2F**). Perhaps not surprisingly, we found that cell division-related genes are most correlated with the first PC, which is related to cell size, likely because actively dividing cells and mitotic figures tend to be larger. We found enrichment of extracellular matrix vacuolar genes for the second PC, which represents a measure of cell shape and may result from cells in the process of adhering to the microwells and the impact of vacuoles on overall cell morphology.

Many cellular phenotypes are difficult to infer directly from static measurements of the transcriptome, such as protein expression and localization dynamics, organelle dynamics and distribution, morphological features, uptake of foreign objects, and biomolecular secretion. However, there are myriad live cell imaging and microwell-based assays for characterizing these phenotypes and events in individual cells.

We expect that SCOPE-Seq will serve as a highly scalable, accurate, and economical approach to linking live cell microscopy assays to scRNA-Seq, enabling investigation of the transcriptional underpinnings of the resulting cellular phenotypes.

The current implementation of SCOPE-Seq has important limitations, motivating future improvements. For example, the use of separate optical and sequencing barcodes requires an extra step- sequencing of bead-free libraries. Future implementations will use oligonucleotides in which the optical and sequencing barcodes are the same DNA sequence. Ligation of OBOs to the 3'-termini of a subset of mRNA capture sites on the beads also leads to reduced molecular capture efficiency for the corresponding scRNA-Seq libraries relative to beads without optical barcodes. Finally, the multiplexing capacity of the beads in our proof-of-concept experiment is relatively modest, which precludes an error-detecting or correcting code. While our current linking accuracy is high, both the yield of optically demultiplexed cells and accuracy could be improved with such a code.

The scheme presented here for linking cellular phenotypes to sequencing could also be generalized beyond single-cell analysis. Microscopy experiments on small multi-cellular organisms, organoids, or colonies could be linked with sequencing. Additionally, the optically decodable bead array could be used for spatial transcriptomic analysis as demonstrated previously with printed microarrays[14]. One advantage of optically decodable beads over printed microarrays[14,15] is that beads can be prepared in a large batch for use in many experiments starting from commercially available "Drop-Seq" beads with relatively simple tube-based reactions. We hope that this economical approach will serve as a powerful tool for connecting high-throughput microscopy and sequencing on multiple scales.

Methods

Production of dual-barcoded mRNA capture beads

Fig. 2A shows the bead production workflow. Approximately 700,000 commercially available "Drop-Seq" beads (Macosko-2011-10(V+), ChemGenes) are washed and resuspended in 1280 uL of nuclease-free water and then immediately evenly distributed into 64 wells on a 96-well plate. The plate is then sealed with an adhesive cover, vortexed for 1 minute, sonicated in a water batch sonicator (FS-20, Fisher Scientific) for 5 minutes, and incubated at 4°C for 12 hours. The supernatant from each well (contains bead-free DNA) is transferred into a separate plate, stored in a -20°C freezer. Beads in each well are then washed again with nuclease-free water. A unique combination of oligonucleotides (4 uM each) selected from a set of 6 species and ligation reaction mixture (1x T4 RNA ligase reaction buffer, 1 mM ATP, 25% PEG 8000, 0.1% Tween-20, 0.7 U/uL T4 RNA ligase 1 (M0204S, New England BioLabs)) are sequentially added into each of the 64 wells. The plate is then incubated at room temperature for 24 hours on a rotator. 5 uL of 10% SDS and 1 uL of 500 mM EDTA are then added to each well, mixed well and incubated for 5 minutes. The beads from all 64 wells are then pooled into a single tube, washed in TE/TW buffer (10 mM Tris pH 8.0, 1 mM EDTA, 0.01% Tween-20) and subjected to a second round of the above reactions with another set of 6 oligonucleotide species and produces another plate of bead-free DNA containing supernatant.

Bead-free DNA library construction, sequencing, and analysis

5 uL of each bead-free DNA containing supernatant is mixed with 10 uL 1 uM NexterapolyA oligonucleotide and 2 uL 10X NEBuffer 2 (B7002S, New England BioLabs). The mixtures are incubated at 95°C for 2 minutes and then cooled down to room temperature. 2 uL 10 mM dNTPs and 1 uL Klenow Fragment (M0212S, New England BioLabs) are then added to each mixture and incubated at room temperature for 15 minutes, and at 37°C for 1 hour, followed by heat inactivation at 75°C for 20 minutes,

and then stored at -20°C. 4uL from each reaction is then further mixed with 2 uL 2 uM supernatant-specific P7 primer, 2 uL 2 uM universal P5 primer, and 8 uL of PCR ready mix (KK2601, Kapa Biosystems). The reaction mixtures are then subjected to PCR amplification (95°C 3 minutes, 14 cycles of (95°C 10 seconds, 55°C 30 seconds, 72°C 30 seconds), 72°C 5 minutes). 5 uL of PCR product from each reaction are then pooled and purified first using the solid-phase reversible immobilization (SPRI) paramagnetic bead technology (A63880, Beckman Coulter) with a 2:1 bead-to-sample volume ratio, then using gel electrophoresis with a 10% TBE polyacrylamide gel (4565014, Bio-Rad). 1X SYBR Gold nucleic acid gel stain (S11494, ThermoFisher Scientific) is used to stain the DNA bands and the ~ 180bp bands (**Supplementary Fig. 1A**) were extracted. The gel purified libraries are further characterized on a bioanalyzer (**Supplementary Fig. 1B**). The purified bead-free libraries are spiked with 40% Phix control library (FC-110-3001, Illumina) and sequenced on a NextSeq 500 sequencer (Illumina) with a NextSeq 500/550 High Output v2 kit (75 cycle, FC-404-2005, Illumina) per the vendor's instructions. 26-base long read 1 and 8 base-long index 1 reads are obtained. A custom read 1 sequencing primer is used. 12-nt cell barcode and 8-nt index sequences are extracted from each of the resulting read. For each cell barcode that has an associated RNA-Seq profile, its abundance in each of the 128 bead-free libraries is counted. Error correction is performed on the index sequences when they are 1 hamming distance away from any of the bead-free library index sequence white list.

Single-cell RNA-Seq library construction, sequencing, and analysis

Single-cell RNA-Seq library is constructed as previously described[8] with the following modifications: 1) Cell live stain is performed in a 37°C incubator for 15 minutes prior to cell dissociation; 2) two different colored live stain dyes are used to stain U87 human (stained with Calcein, AM (C3100MP, ThermoFisher Scientific)) and 3T3 mouse (stained with Calcein Red-Orange, AM (C34851, ThermoFisher Scientific)) cells respectively; 3) cells are loaded at 1700 cells/uL; 4) the cell-loaded microwell array is scanned under a microscope in two fluorescence channels and bright field channel before beads are loaded to the microwells; 5) exonuclease I digestion is performed inside the microwell array device; 6) Before the exonuclease I digestion step, a mix of 12 species of unlabeled oligonucleotides whose sequences are complimentary to that of the OBOs (1 uM each in wash buffer (20 mM Tris-HCl pH7.9, 50 mM NaCl, 0.1% Tween-20)) are flowed into the microwell array device, incubated for 1 hour at room temperature, followed by a flush with wash buffer; 7) The device is cut into 10 separate pieces. The beads extracted from each piece are processed in separate reactions for the downstream library construction steps. A different Nextera index primer is used for each piece during Nextera PCR. To facilitate the cutting, a new chip design (**Supplementary File 1**) is used. Microwells are arranged into 10 separated patches with sufficient space between them to accommodate cutting by a razor blade; 8) beads are extracted by soaking the small piece of bead-containing PDMS piece in 100% ethanol, followed by vortexing, water bath sonication and centrifugation in a 1.7 mL microcentrifuge tube. The beads are pelleted to the bottom of the tube while the PDMS piece is stuck at a different level of the tube thus achieving separation between the bead pellet and the PDMS piece (**Supplementary Fig. 2**). The PDMS slab is then removed with a tweezer. Single-cell RNA-Seq library is sequenced as previously described[8]. The resulting sequencing reads were demultiplexed, aligned, and quantified as described previously[13].

Optical demultiplexing workflow, system design, and image processing

After the exonuclease I digestion step, the device is flushed with TE/TW buffer, and 150 mM NaOH solution. 12 rounds of sequential fluorescence hybridization (SFH) are then used to identify the set of OBOs attached to each bead. Each round of SFH consists of four major steps including background scan, probe hybridization, probe scan, and probe stripping. During both background and probe scans, the bead-containing microwell array is imaged in wash buffer through both the fluorescence channel and the bright

field channel. During probe hybridization and probe stripping, a hybridization mixture (probe oligonucleotides at 200 nM in wash buffer) and a melting solution (150 mM NaOH) is flowed into the device respectively, and incubated at room temperature for 10 minutes followed by flushing with wash buffer. Probe oligonucleotides used in each hybridization mixture are Cy5-labeled oligonucleotides whose sequences are complementary to each of the 12 OBOs. An automated reagent delivery and scanning system (**Supplementary Fig. 3**) is designed for automated optical demultiplexing. In the automated reagent delivery system, fixed positive pressure (~1 psi) stabilized by a pressure regular (AW20-F02, SMC Pneumatics) is used to drive fluid flow. Two 10-channel rotary selector valves (MLP778-605, IDEX Health & Science) are connected in parallel to toggle between 14 reagent channels. A three-way solenoid valve (EW-01540-11, Cole-Parmer), located at the downstream of the microwell array device, is used as an on/off switch of the reagent flow. The microwell array device is constantly being pressurized during incubation steps which helps prevent the device from drying and suppresses bubble formation. A microfluidic flow cell which traps bubbles came from upstream is connected in-serial immediate upstream of the microwell array device to further prevent bubbles from entering the microwell array device. A water hammer arrestor consists of a half-filled cryotube is connected immediate downstream of the microwell array device and upstream of the on-off flow switch valve to buffer the hydraulic shock produced by actions of the on-off switch valve. The 10-channel selector valves are controlled by a USB digital I/O device (NI USB-6501, National Instruments). The three-way solenoid valve is controlled by the same USB digital I/O device, but through a homemade transistor-switch circuit. The automated scanning system consists of an inverted fluorescence microscope (Eclipse Ti-U, Nikon) equipped with motorized stage and filter wheel, and a camera (iXon3, Andor). A C program is used to control both the reagent delivery system and the scanning system. A glass slide is attached to the bottom surface of the microwell array device to help keeping microwells on the same imaging plane. The individual image tiles from each scan are stitched together. The stitched images from each scan are then registered to the first demultiplexing cycle scan. The outlines of individual wells are then identified using the bright field images by first thresholding followed by object detection using the particle analyzer plugin in ImageJ. The average bright field and fluorescence intensities of each well in each scan are then measured. Two separate populations of wells are observed based on their average bright field intensities in each demultiplexing cycle (**Supplementary Fig. 4**). The population with higher average bright field intensity corresponds to empty wells whereas the other population corresponds to wells with beads. The mean bright field intensity of each population is identified by fitting the average bright field intensity distribution with a bimodal Gaussian model. The intensity of the shortest bin between the two mean intensities is used as the threshold that separates the two well populations. The device is also scanned immediately after mRNA capture and after reverse transcription. The same analysis is performed on the bright field images to identify wells with beads (**Supplementary Fig. 5**). Further analysis is then performed only on wells with beads. The average fluorescence intensity of each well measured from the background scan is subtracted from that measured from the corresponding probe scan. The resulting background-subtracted fluorescence intensities exhibit bimodal distribution as well (**Supplementary Fig. 6**). The population with higher average fluorescence intensity corresponds to wells with beads in the "on" state whereas the other population corresponds to wells with beads in the "off" state. The thresholds that separate these two populations are identified using the same algorism described above.

Cell live stain image processing/feature extraction

Individual cell live stain image tiles from each color channel are stitched. Rolling ball background subtraction is performed on the stitched images. The background-subtracted images are then registered to the first demultipelxing cycle scan. The well outlines obtained in processing the demultipexing images are used to obtain the average live stain intensities of cells in each well in each of the two live stain color

channels. Two separate populations of wells are observed based on cell live stain intensities (**Supplementary Fig. 7**). The population with higher intensity corresponds to wells with cells stained in the corresponding color whereas the other population corresponds to wells without cells stained in the corresponding color. The thresholds that separate these two populations are identified using the same algorism used in the processing of demultiplexing images. These well-averaged cell live stain intensities are used to decide the species identity of each well. Further analysis is then performed only on wells with either U87 human cells or 3T3 mouse cells. To obtain imaging features of individual cells, adaptive thresholding is performed on the fluorescence live stain images. The adaptive threshold for each well is identified based only on the intensity distribution of the pixels within the smallest bounding square of that well. The default (auto) threshold identification algorithm in ImageJ is used. The outline of each cell is then identified using the particle analyzer in ImageJ. Wells with more than one cell detected are excluded from analysis. 19 imaging features are then obtained for each cell using their live stain image. These features are area, mean intensity, standard deviation of intensity, mode of intensity, minimum intensity, maximum intensity, perimeter, major axis, minor axis, circularity, Feret's diameter, integrated intensity, median intensity, skewness, kurtosis, minimum Feret's diameter, aspect ratio, roundness, and solidity.

Linking imaging and sequencing data

We used our mixed species experiment to assess the accuracy with which we could link sequencing and imaging data with SCOPE-seq. We first defined a set of singlets from the scRNA-Seq data as cells with >90% of reads aligning uniquely to the same species. Next, we extracted the sequencing barcodes for these singlets and identified the "bead-free" sequencing library indices that contained each cell barcode. There are two sets of "bead-free" sequencing library indices corresponding to the two cycles of split-pool synthesis used to generate the dual-barcoded beads. We assigned an optical barcode to each sequencing barcode by taking the index from each of the two sets with the most reads corresponding to the sequencing barcode. Taken together, the two indices identify a specific combination of OBOs comprising the optical barcode. We can link ~89.6% of sequencing barcodes to optical barcodes by this procedure. Finally, we determine whether the optical barcode is found in the optical demultiplexing data, and if it is, we determine the imaging-based species call for the cell in the corresponding microwell. We use the concordance between the scRNA-seq and imaging-based species calls to compute our accuracy.

Imaging-based multiplet detection

Two-color fluorescence images of microwells with cells from both species are merged into a monochrome image by summing the pixel intensities from the two-color channels (**Supplementary Fig. 7**). The monochrome image is then examined by a researcher at multiple brightness levels. Multiplet is identified when there are more than one bright objects separated by a clear dark boundary.

Principal component and gene set enrichment analysis of the cell imaging features

We used the singlet murine cells in our mixed species experiment to analyze the relationship between gene expression and the imaging features described above, because it contained more cells. We started with 805 murine cells and filtered this to 728 cells after removing cells in which the fluorescence containing saturated pixels. We then z-normalized the imaging feature matrix described above for these 728 cells and computed the principal components of the resulting matrix of z-scores. Next, we computed Pearson's correlation coefficient between the principal component vector and the vector of gene expression levels for each gene in log(counts per million+1) for the first two principal components. We then ranked all of the protein-coding genes detected in at least 10% of cells by correlation with each of the two principal components and used this as input for gene set enrichment analysis (GSEA) using the

gene sets from the Molecular Signatures Database C5 collection of cellular component gene ontologies (pre-ranked mode, "classic" enrichment statistic).

Data Availability

The single-cell RNA-Seq data and bead-free optical barcode/sequencing barcode linkage data have been deposited in the Gene Expression Omnibus under accession GSE116011.

Contributions

J.Y. and P.A.S. conceived the study and designed experiments. J.Y. constructed the automated instrumentation for SCOPE-seq and conducted all of the experiments. J.Y., J.S., and P.A.S. analyzed the data. J.Y. and P.A.S. wrote the paper.

Competing Interests

J.Y. and P.A.S. are listed as inventors on patent applications filed by Columbia University related to the work described here.

Acknowledgements

We thank the Sulzberger Columbia Genome Center for assistance and resources for high-throughput sequencing. P.A.S. was supported by NIH/NCI grant R33CA202827, NIH/NIBIB grant K01EB016071, and a Human Cell Atlas Pilot Project grant from the Chan Zuckerberg Initiative.

References

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

2 Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

3 Bose, S. *et al.* Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol* **16**, 120, doi:10.1186/s13059-015-0684-3 (2015).

4 Rotem, A. *et al.* High-Throughput Single-Cell Labeling (Hi-SCL) for RNA-Seq Using Drop-Based Microfluidics. *PLoS One* **10**, e0116328, doi:10.1371/journal.pone.0116328 (2015).

5 Fan, H. C., Fu, G. K. & Fodor, S. P. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367, doi:10.1126/science.1258367 (2015).

6 Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240 (2013).

7 Lane, K. *et al.* Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-kappaB Activation. *Cell Syst* **4**, 458-469 e455, doi:10.1016/j.cels.2017.03.010 (2017).

8 Yuan, J. & Sims, P. A. An Automated Microwell Platform for Large-Scale Single Cell RNA-Seq. *Sci Rep* **6**, 33883, doi:10.1038/srep33883 (2016).

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

10 Love, J. C., Ronan, J. L., Grotenbreg, G. M., van der Veen, A. G. & Ploegh, H. L. A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nature Biotechnology* **24**, 703-707 (2006).

11 Sims, C. E. & Allbritton, N. L. Analysis of single mammalian cells on-chip. *Lab on a Chip* **7**, 423-440 (2007).

12 Gunderson, K. L. *et al.* Decoding randomly ordered DNA arrays. *Genome research* **14**, 870-877, doi:10.1101/gr.2255804 (2004).
13 Yuan, J. *et al.* Single-Cell Transcriptome Analysis of Lineage Diversity and Microenvironment in High-Grade Glioma. *bioRxiv*, doi:10.1101/250704 (2018).
14 Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).
15 Vickovic, S. *et al.* Massive and parallel expression profiling using microarrayed single-cell sequencing. *Nat Commun* **7** (2016).

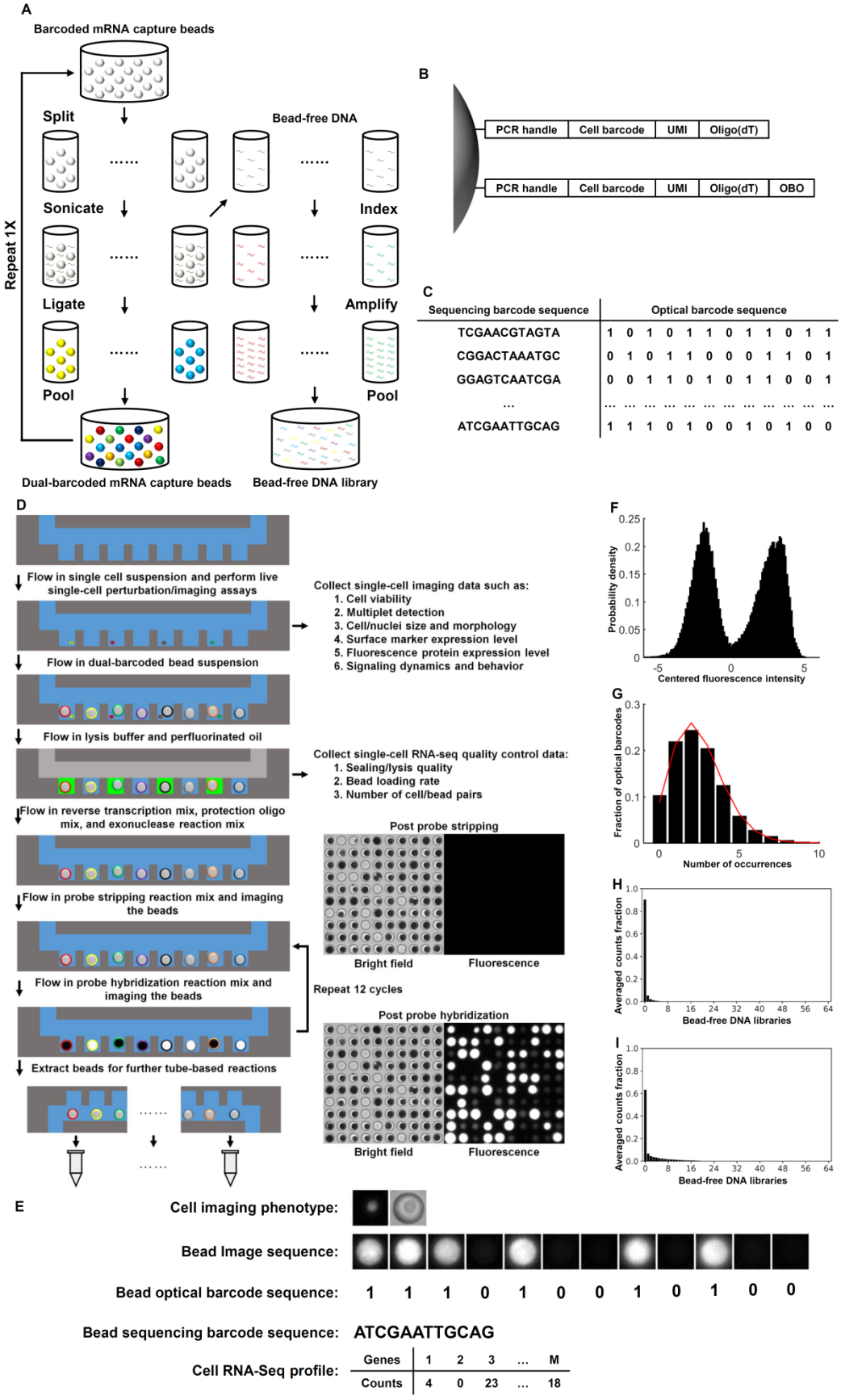
A
Barcoded mRNA capture beads
Split
Bead-free DNA
Sonicate
Index
Ligate
Amplify
Pool
Pool
Repeat 1X
Dual-barcoded mRNA capture beads
Bead-free DNA library
B
PCR handle
Cell barcode
UMI
Oligo(dT)
PCR handle
Cell barcode
UMI
Oligo(dT)
OBO
C
Sequencing barcode sequence
Optical barcode sequence
TCGAACGTAGTA 1 0 1 0 1 1 0 1 1 0 1 1
CGGACTAAATGC 0 1 0 1 1 0 0 0 1 1 0 1
GGAGTCAATCGA 0 0 1 1 0 1 0 1 1 0 0 1
ATCGAATTGCAG 1 1 1 0 1 0 0 1 0 1 0 0
D
Flow in single cell suspension and perform live single-cell perturbation/imaging assays
Collect single-cell imaging data such as:
1. Cell viability
2. Multiplet detection
3. Cell/nuclei size and morphology
4. Surface marker expression level
5. Fluorescence protein expression level
6. Signaling dynamics and behavior
Flow in dual-barcoded bead suspension
Flow in lysis buffer and perfluorinated oil
Collect single-cell RNA-seq quality control data:
1. Sealing/lysis quality
2. Bead loading rate
3. Number of cell/bead pairs
Flow in reverse transcription mix, protection oligo mix, and exonuclease reaction mix
Post probe stripping
Flow in probe stripping reaction mix and imaging the beads
Bright field
Fluorescence
Flow in probe hybridization reaction mix and imaging the beads
Repeat 12 cycles
Post probe hybridization
Extract beads for further tube-based reactions
Bright field
Fluorescence
F
Probability density
Centered fluorescence intensity
G
Fraction of optical barcodes
Number of occurrences
H
Averaged counts fraction
Bead-free DNA libraries
I
Averaged counts fraction
Bead-free DNA libraries
E
Cell imaging phenotype:
Bead Image sequence:
Bead optical barcode sequence: 1 1 1 0 1 0 0 1 0 1 0 0
Bead sequencing barcode sequence: ATCGAATTGCAG
Cell RNA-Seq profile:
Genes 1 2 3 … M
Counts 4 0 23 … 18

Figure 1: A) Workflow for generating the dual-barcoded beads and the associated bead-free DNA sequencing libraries. B) Sequence composition of the bead-bound mRNA capture oligonucleotides and optical barcode oligonucleotides (OBOs). C) A look-up table linking sequence barcodes and optical barcodes. D) Workflow for SCOPE-Seq. E) Workflow for linking live, single-cell imaging data with single-cell RNA-Seq profile. F) Probability density of the centered bead fluorescence intensity during demultiplexing. G) Optical barcode usage histogram. H) Averaged counts fraction distribution in the first round I) and second round of split-pool reactions.

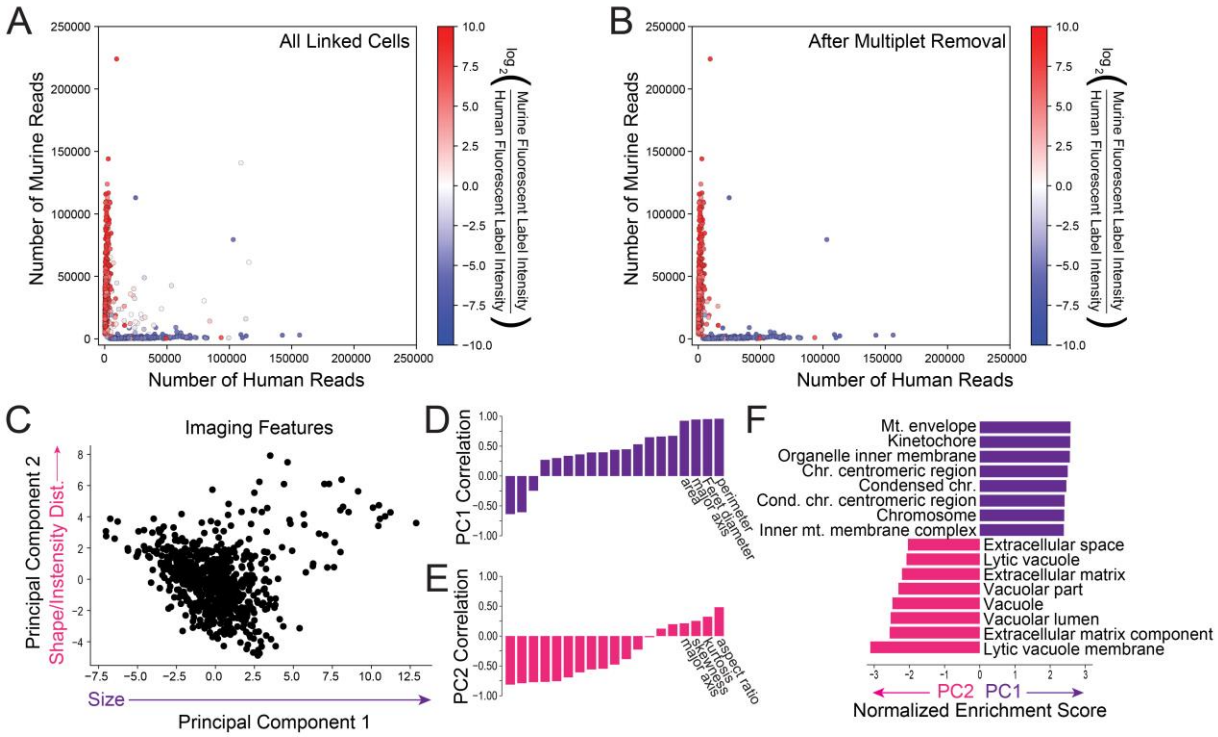

Figure 2: A) Scatter plot of the number of uniquely aligned human and murine reads for each linked sequencing barcode colored by the relative fluorescence intensity of the human and murine labels before and B) after imaging-based multiplet removal. C) Principal component analysis of the 19 fluorescence imaging features for the murine cells. D) Correlation between the first PC and each imaging feature. E) Correlation between the second PC and each imaging feature. F) Normalized enrichment scores for MSigDB gene sets enriched in genes that are correlated with the first (purple) and second (magenta) PCs.