UNIVERSITY OF CALIFORNIA,
IRVINE

Developing Spatialomics Platforms to Profile Biomarkers

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Joshua Gu

Dissertation Committee:
Dr. Weian Zhao, Chair
Dr. Peter Donovan
Dr. Anand Ganesan
Dr. Marian Waterman
Dr. Kai Kessenbrock

Oversight Committee:
Dr. Edwin Monuki

2022

DEDICATION

To my family, friends, colleagues, and mentors

who were always supportive of me,

pushed me to reach for my dreams,

and inspire me to do better.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Weian Zhao for being my mentor and providing me with the opportunity to perform research and develop tools in his laboratory during my graduate career. I am grateful for your mentorship and guidance over the years and the collaborative opportunities you provided for me. I am also very thankful I was able to learn about entrepreneurship from you. I would also like to thank Drs. Anand Ganesan, Marian Waterman, and Kai Kessenbrock for serving on my committee and being excellent collaborators over the years. Thank you, Dr. Peter Donovan, for being supportive of me ever since I joined the Cellular and Molecular Biosciences program and for serving as a member of my committee. Thank you, Dr. Edwin Monuki, for your support as my oversight committee and your support. I must also thank Drs. David Haussler, Jason Fernandes, and Sofie Salama for inspiring me to become a research scientist and preparing me for graduate school.

While there are many members of the Zhao laboratory as well as members from our collaborators that I appreciate for their assistance and support in completing these studies, I would specifically like to thank Drs. Enrico Gratton, Per Niklas Hedde, Jan Zimak, Jessica Shiu, and Lorenzo Scipioni who were tremendous mentors for me. I was fortunate to have worked with the LFD team and learn so much about microscopy. Jan, thank you for being a great friend and your continual support. I learned a lot about cloning from you and enjoyed chatting about tech news. Jessica, thank you for teaching me so much about dermatology and letting me work beside you. I would like to thank members of the team, Tam Vu and Alexander Vallmitjana, for working together to get our paper out; Dr. Alan Hauser, for mentoring me in the world of business and startups; Kieu La and Jesus Flores, for helping with such a wide variety of tasks and the good times in lab. Jayapriya Jayaraman, Michael Mellody, and Lexi Crowell, thank you all for being wonderful friends, teammates, and fellow graduate students that supported me during my hardest time of graduate study.

I would like to thank other members of the Zhao lab, Ruchi Desai, Spriha Singh, Hana Schlosser, and Audrey Fung for your friendship, support, and hard work. Thank you, Christine Wei and Hannah Jian, for your diligent work and dedication. To all my friends and those in my cohort, thank you for the friendship and support throughout the years. I would also like to thank Linzi Hosohama for being a great collaborator and helping with the xenograft model and sequencing data. Dr. Arthur D. Lander for the constructive input on transcript detection for our platform. Thanks to the Beckman Laser Institute and Amanda Durkin for allowing me to use the Leica SP8 microscope. I must thank the UCI Institute of Immunology, Drs. Eric Pearlman, Lisa Wagar, and Francesco Marangoni, for the Immunology T32 fellowship that supported me and enlightened me to the complex world of immunology. Thanks to the UCI Stem Cell Center and Drs. Leslie Thompson and Peter Donovan for the Stem Cell Translational Medicine T32 fellowship that supported me and provided insightful seminar events. Thank you, Stanley Behrens and the Behrens family, for the UCI School of Medicine Behrens Research Excellence Award that supported my research and career.

Lastly, I must thank my family and partner, Alyce Oh, for encouraging me to be curious and pursue my passion. I will always appreciate the continuous support I received during my graduate career. Thank you for providing encouragement when I needed it and being there for me through the ups and downs of this journey. Thank you, Archie and Alfie, my wonderful therapy dogs, who provided comfort and companionship through the challenging times. I will forever be grateful for the support and mentorship I have received from everyone.

CURRICULUM VITAE

Joshua Gu

PROFILE

My goal is to expand the scientific frontier and explore solutions to diseases. I am a highly skilled bioengineer with 3 years of experience in Dr. David Haussler's lab and regenerative medicine experience from an internship in Taiwan. During my graduate career, I focused on translational science. In Dr. Weian Zhao's lab, I worked on a project to develop a high-throughput platform for multi-omics spatial analysis in clinical tissue samples and a low-cost imager. I then co-founded Arvetas Biosciences to commercialize the multi-omics spatial analysis platform I developed. Another project involved the development of cell-based mechanosensors, a novel tool to evaluate tumor tissues mechanics and further understand mechanobiology. I have also manufactured a novel organ-on chip plate with human vascularized microtissues that can be used for drug screening applications and disease studies in Dr. Chris Hughes' lab.

EDUCATION

School of Medicine – Biological Chemistry PhD (Feb. 2022)
Biomedical Sciences MS degree awarded: 2021
University of California, Irvine

Bioengineering BS with honors
Minor in Bioinformatics
University of California, Santa Cruz
Degree awarded: March 2018

SKILLS

Molecular Biology: antibody-oligo conjugation, IHC, Immunofluorescence staining, FISH probe design, RNAscope, molecular cloning, Gateway cloning, Gibson assembly, DNA isolation and purification, gel extraction, PCR, qPCR, Luciferase assay, protein gel and western blotting, RNA harvest and isolation, northern blotting, Lentivirus generation

Tissue Culture: Human stem cell culture, Mouse stem cell culture, CRISPR/Cas9, CRISPRi cell line work, mammalian cell culture, retrotransposition assays, nucleofection, PBMC isolation, T cell isolation and expansion

Bioengineering: 3D bioprinting scaffolds, PDMS, microscope development, microfluidics, biosensors, electronic circuits, laser cutting, 3D printing

Software: Python, Java, R, Zeiss Zen Suite, UCSC genome browser, Galaxy, NCBI BLAST, RNAfold, NIS-Elements, ImageJ, CellProfiler, Oligominer, Cura, Graphpad PRISM

Microscopy: Fluorescence lifetime imaging, confocal microscopy, SPIM, Two-photon microscopy, DIVER, Live cell microscopy, electron microscopy

Other experiences: FFPE tissue (Skin biopsy, breast, brain, colon cancer) processing, OCT processing/sectioning, xenograft processing, tissue sectioning, FACS, CyTOF, cryobanking, gas chromatography, lab upkeep and organization

EXPERIENCES

Zhao Lab at UC Irvine – *PhD Candidate* **Aug 2018 – Feb 2022**

- Developing multi-omics spatial analysis platform for detection on cells and tissue. Rapidly located RNA and protein targets in a high-plex fashion while being cost-effective.
    - Filed patent for technology
    - Developed prototype low-cost fast scanning microscope
- Group leader/supervisor of an independent research team @ Zhao lab
    - Led collaborations with 16 groups/centers to answer various biological questions
    - Made experimental plans and assigned tasks
    - Mentored 19 junior graduate students, undergraduate students, interns, and visiting scholars
    - Troubleshoot and optimized assays
    - Mentored technique training and personal development of team members
- Developed cell-based mechanosensors to evaluate tumor tissue mechanics.
- Helped write R01, R21, NIH P30, NIH U54, SBIR grants, several which were funded to support my projects
- Helped manage IBC protocols and chemical inventory, safety training officer, and modified experimental procedures for Zhao lab

Arvetas Biosciences – *Co-Founder* **June 2020 – July 2021**

- Developed platform and benchmarked against other technologies
- Commercializing multi-omics spatial analysis platform
- Helped developed pitch deck
- Recruited early customers and collaborations

Hughes Lab at UC Irvine – *Rotation student* **January 2019 – April 2019**

- Developing vascularized micro-liver to study Hereditary Hemorrhagic Telangiectasia.
- Manufactured microfluidic devices to grow human vascularized microtissues.

Ask-a-Scientist Night @ Irvine Unified School District – *volunteer* **2018 – 2019**

- Helped 6^{th}-12^{th} grade students prepare their science fair projects and answered their scientific questions
- Taught how to structure a scientific project hypothesis and include necessary controls

---

Haussler Lab at UC Santa Cruz – *Undergraduate researcher.* **2016 – 2018**

- My thesis project focused on investigating the role KRAB-Zinc Finger transcription factors (KZNFs) play in regulating retrotransposons and human genome evolution.
- Awarded Dean's, Chancellor's, and Steck Award for the finest senior thesis 2018.

- Used an inducible CRISPRi Gen1C iPSC line to inhibit transcription of KZNFs and retrotransposons to determine if the inhibition of specific KZNFs lead to de-repression of specific TEs.
- Provided lab tour to Harbor high school students, explained my thesis research, and provided college advice.

Wang Lab at National Chung Hsing University – *Intern researcher* **Summer 2017**

- Worked on 3D bioprinting scaffolds to grow fibroblasts, antibacterial materials for wound dressings, biosensors, and automatic system for mass extraction of healthy mitochondria for applications in regenerative medicine.
- Learned to gold plate materials and use the electron microscope to observe the pores.
- Mentored undergraduate and provided suggestions/solutions to problems for graduate students.
- International internship in Taiwan

PUBLICATION/RESEARCH

1. **Gu J**., Fernandes J., Tigyi K., Salama S., Haussler D. *Investigating Interactions between KRAB-Zinc Fingers and Transposable Elements.* Undergraduate Symposium UCSC; June 2017; Santa Cruz, CA.
2. **Gu J**., Fernandes J., Tigyi K., Salama S., Haussler D. *Transposable Element Escapes Repression from KRAB-Zinc Finger Proteins.* Undergraduate Symposium UCSC; June 2018; Santa Cruz, CA.
3. Fernandes JD, Haeussler M, Armstrong J, Tigyi K, **Gu J**, ..., Haussler D, Salama S (2018). *KRAB Zinc Finger Proteins coordinate across evolutionary time scales to battle retroelements.* bioRxiv doi: https://doi.org/10.1101/429563
4. **Gu J**., Zhao W. (Fall 2019) *Developing cell-based mechanosensors to interrogate the tumor microenvironment*. UCI Cancer center RIP talk
5. Argueta D, Aich A, Lei J, Kiven S, Nguyen A, Wang Y, **Gu J**, Zhao W, Gupta K. (2020). *b-endorphin at the intersection of pain and cancer progression: Preclinical evidence*. Neuroscience Letters. https://doi.org/10.1016/j.neulet.2020.135601
6. **Gu J**, Zhao W. (2020) *Spatial profiling the tumor immune microenvironment.* UCI Immunology Department Seminar
7. **Gu J**, Zhao W. (2020) *A robust pipeline for in situ single cell multi-omics spatial analysis*. NIH NCI CSBC poster
8. **Gu J**, Zhao W. (2020) *A robust pipeline for in situ single cell multi-omics spatial analysis*. SOM Grad day poster
9. **Gu J**, Zhao W. (Winter 2021) *Developing an in situ single cell multi-omics spatial analysis platform.* UCI Biological Chemistry School of Medicine Department Seminar
10. **Gu J**, Zhao W. (Spring 2021) *Developing an in situ single cell multi-omics spatial analysis platform.* UCI Biological Chemistry School of Medicine Department Seminar

11. Vu T*, Vallmitjana A*, **Gu J***, …, Gratton E, Zhao W. (2021) *Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis Nat Commun* 13**,** 169 (2022). https://doi.org/10.1038/s41467-021-27798-0 (* = equal contribution)
12. **Gu J**, Zhao W. (Summer 2021) *Developing an in situ single cell multi-omics spatial analysis platform.* UCI Institute for Immunology T32 Training Grant Symposium
13. **Gu J**, Zhao W. (Fall 2021) *Illuminating tissue-wide spatiotemporal transcriptome landscapes in skin development and cancer.* UCI Skin Club
14. **Gu J**, Zhao W. (Fall 2021) *Multi-Omics spatial analysis tool to elucidate the tumor immune microenvironment*. UCI XIX Annual Immunology Fair
15. **Gu J**, Jian H, Wei C, Shiu J, Ganesan A, Zhao W., Hedde P (2022) *A low-cost modular imaging system for rapid, multiplexed immunofluorescence detection in clinical tissues. (Submitted)*

PATENTS

- **Gu J**, Zhao W, Gratton E, Zimak J, Vu T, Vallmitjana A, Hedde P, Scipioni L. 2019. Compositions and methods for spatial profiling of biological materials using time-resolved luminescence measurements. US2019/061282
- **Gu J**, Zhao W, Hedde P. 2022. Low-cost modular imaging system for rapid, multiplexed immunofluorescence detection.

AWARDS/CERTIFICATES

- UCI Beall Applied Innovation/GPS-STEM - Business Concepts for STEM Scientists course
- 2019 NSF GRFP Honorable Mention
- 2020 NIH T32 Immunology Fellowship
- 2020 NIH NCI CSBC – best poster presentation award
- 2020 UCI School of Medicine Grad Day – judge's award for best poster
- 2021 UCI SOM Behrens Research Excellence Award
- 2021 NIH T32 Stem Cell Translational Medicine for Neurological Disorders Fellowship
- 2021 Nominated for ARCS Scholar Award
- 2021 UCI Immunology Fair AAI Young Investigator Award - 1st Place Poster Presentation

MEMBERSHIPS

- American Heart Association Member
- American Association for Cancer Research - Associate member
- Excellence is Science Program AAAS - Member
- UCI Graduate Profession Success-STEM (GPS-STEM) member
- UCI Diverse Educational Community and Doctoral Experience (DECADE) member

ADDITIONAL INFORMATION

Languages: English (native), Mandarin (fluent)

ABSTRACT OF THE DISSERTATION

Developing Spatialomics Platforms
to Profile Biomarkers

By

Joshua Gu

Doctor of Philosophy in Biological Chemistry

University of California, Irvine, 2022

Professor Weian Zhao, Chair

Multiplexed mRNA and protein profiling in the spatial context provides important new information enabling basic research and clinical applications. Unfortunately, most existing spatial transcriptomics and proteomics methods are limited due to either low multiplexing or assay complexity. Here, we introduce a new spatialomics technology, termed Multi Omic Single-scan Assay with Integrated Combinatorial Analysis (MOSAICA), that integrates *in situ* labeling of mRNA and protein markers in cells or tissues with combinatorial fluorescence spectral and lifetime encoded probes, spectral and time-resolved fluorescence imaging, and machine learning-based target decoding. This technology is the first application combining the biophotonic techniques, Spectral and Fluorescence Lifetime Imaging and Microscopy (FLIM), to the field of spatial transcriptomics and proteomics. By integrating the time dimension with conventional spectrum-based measurements, MOSAICA enables direct and highly multiplexed *in situ* spatial biomarker profiling in a single round of staining and imaging while providing error correction removal of background autofluorescence. We demonstrated MOSAICA's capabilities in cell culture and Formalin-Fixed Paraffin-Embedded (FFPE) tissues while obtaining a strong correlation with sequencing data (Pearson's r = 0.96). We then demonstrate simultaneous co-detection of protein and mRNA in colorectal cancer cells. To answer biological questions with a simple 3-4 plex immunofluorescence panel, we developed a low-cost Tissue Imager for under $9,000 and achieved a performance on

par with commercial fluorescence microscopes that cost ~20x more. Additionally, another tool we developed to study the tissue microenvironment was cell-based mechanosensors to quantitively and dynamically assess the tissue mechanics. We have already used MOSAICA to study colon cancer heterogeneity, profile neurological mRNA panels in brain tissue, and profile immuno-oncology panels for skin tissue. MOSAICA represents a simple, versatile, and scalable tool for targeted spatial transcriptomics and proteomics analysis that can find broad utility in constructing human cell atlases, elucidating biological and disease processes in the spatial context, and serving as companion diagnostics for stratified patient care.

CHAPTER 1

INTRODUCTION

Spatialomics is a powerful *in vitro* method that permits the topographical analysis of RNA transcripts and/or protein expression within each cell, tissue, and/or organ. Understanding cells in their morphological context allows us to better understand their functions. Transcripts or proteins are labeled with fluorescent reporter molecules and imaged to reveal their locations, identities, counts, distributions, and interactions in their native environment. This critical technology has garnered increased interest in the biological and clinical field as advances in DNA and RNA sequencing have enabled researchers and clinicians to access an unprecedented amount of genomic, epigenomic, and transcriptomic information (Figure 1.1). An effective analytical pipeline to map out each cell and its function within the body from the molecular to system level resolution is a fundamental requirement for advancing our understanding of developmental biology, computational biology, cancer biology, as well as accelerating the development of precision diagnostics and therapeutics[1-3]. Of critical importance are the enabling technologies which permit the spatial analysis of multi-omics markers, e.g. mRNA and proteins, by revealing their presence, counts, locations, dynamics, and interactions within each tissue and organ. These technologies are key to elucidating the myriad of functional states, cellular processes, and cell-cell interactions which occurs in an integrated 3D spatial context[4-6].

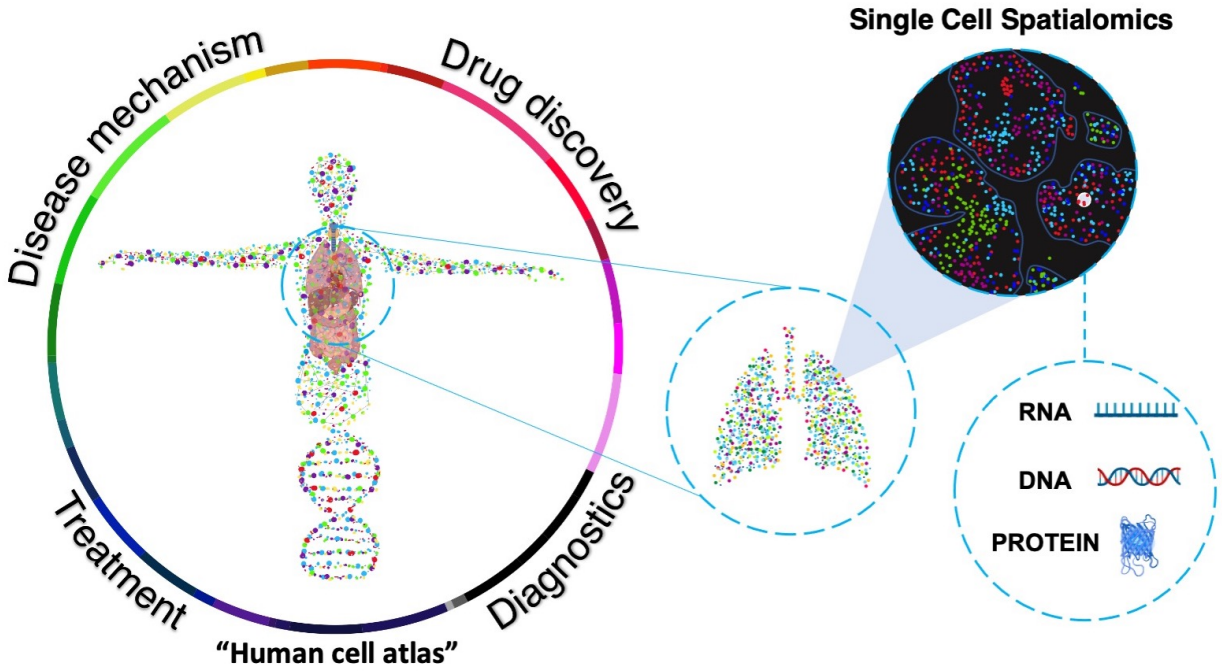


Figure 1.1 Single cell multi-omics
Cell-cell and cell-niche interactions in the tissue microenvironment are tightly regulated in space and time. Spatial analysis provides the precise spatial location and clustering information that NGS cannot. The long-term goal is the generate human atlases using spatialomics data.

Unfortunately, a critical barrier to the enabling of this pipeline is the lack of tools that can rapidly profile multi-omics in space and time that scales from the molecular level to the tissue or organ level in heterogeneous samples[7-11]. Conventional tools for in situ analysis including fluorescence in situ hybridization (FISH) for DNA or RNA analysis and immunohistochemistry (IHC) for protein detection suffer from drawbacks including poor robustness, reproducibility, sensitivity, and the need to screen and empirically test large libraries of potential epitopes and genetic targets against different tissues[12-19]. Moreover, these traditional approaches provide only low throughput and multiplexing capabilities because of the limited number of spectrally available channels[20-22]. Recent single-cell sequencing methods (e.g. Single-cell RNA sequencing (scRNA-seq)) lack the critical spatial context needed to understand complex heterogeneous samples such as tumors[23-

[25]. Spatial transcriptomic methods that are based on sequential labeling, stripping, and imaging (e.g. seqFISH, MERFISH), branched amplification (e.g. RNAscope®, SABER), or barcoded labeling with down-stream sequencing (e.g. Slide-seq) are too complicated, time-consuming, laborious and costly to scale up and often are limited to few specialized laboratories (Table 1)[26-30]. Spatial proteomic methods including Imaging Mass Cytometry (IMC) and CODEX can offer higher multiplexing than conventional immunohistochemistry but suffer from decreased sensitivity, throughput, accessibility and/or high cost[31-33].

We developed a spatial multi-omics technology, that integrates a) in situ labeling of molecular markers (e.g. mRNA, proteins) in cells or tissues with combinatorial fluorescence spectral and lifetime encoded probes, and b) spectra and time-resolved fluorescence imaging and analysis to enable rapid, high-throughput spatial profiling of multi-omics biomarkers. Fluorescence lifetime is a measure of the time a fluorophore spends in the excited state before returning to the ground state and is an inherent

Table 1: *Expected quantitative advances of the proposed MOSAICA technology over state-of-the-art technology.*

Methods	Principle	Multiplexing (# of targets per assay)	Cost (50 genes per sample)	Throughput (assay time)	Multi-omics (e.g. RNA and protein)	Multi-scale analysis (from molecules to large tissue)	Scalability
Conventional FISH	Hybridization with fluorescent oligos	3 to 4	1,000s	1 day	No	No	Poor
RNAScope	Branch amplification with ZZ probes	4 to 12	1,000s	1 day	Inefficient	Inefficient	Poor
Serial labeling (e.g. Seq-FISH, SABER-FISH, MERFISH)	Sequential hybridization, imaging, and stripping	10s to 10,000 (several to 81 rounds)	10s to 100s	Several days to weeks	Inefficient	Inefficient	Poor
In situ sequencing (e.g. FISSEQ, STARmap, etc.)	Barcoded *in situ* Sequencing	Whole transcriptome	100s	Weeks	No	Inefficient	Poor
Proposed Spectral-FLIM FISH	**Spectral and lifetime-based labeling and imaging**	**10s - 100s**	**<10***	**1 day**	**Efficient**	**Efficient**	**Excellent**

*Our approach utilizes primary and secondary probes at 5 nM per assay and requires only a single round of hybridization and imaging. Each 50-plex assay costs $0.38 in primary probes ($0.13 per base for 25 nmoles of each probe with 40 73-mer probes per gene) and $1.60 in fluorescent secondary probes (16 probes at $550 for 5 nmoles of each probe). Hybridization buffers, washing solutions, fixatives, blocking agents, etc. cost <$1.00 per assay. Costs for probes can be further decreased by several magnitudes with pool gene array synthesis and amplification and with manual conjugation of secondary oligos with dyes.

characteristic of the fluorophore and its surrounding environment[34]. By utilizing both time and intensity domains for labeling and imaging, we can discriminate a vast repertoire of lifetime and spectral components simultaneously within the same pixel or image of a sample to enable highly increased multiplexing capabilities with standard optical systems.

As summarized in Table 1, several key capabilities and advantages of our proposed technology for spatial omics compared to existing technologies are expected, including a) direct, in situ spatial profiling of all labeled biomarkers in a single round of hybridization and imaging in contrast to existing approaches where many iterations of sample re-labeling, imaging, indexing, and image registration are often required, b) elimination of sample autofluorescence (by choosing probes with lifetimes different from autofluorescent moieties) and therefore improved detection sensitivity, signal-to-noise ratio (SNR), and detection efficiency as well as elimination of additional complicated, harsh, and time laborious processing steps, c) high multiplexing. We anticipate the ability to achieve simultaneous detection of 10s to 100s of target molecules via one labeling and imaging round with our MOSAICA approach, d) multi-scale analysis from molecules to systems. Our technology can uniquely achieve both high resolution (subcellular features or single molecules) and high imaging throughput, two key parameters which are often trade-offs of each other in existing methods. This feature is enabled by our ability to label and encode biomarkers with unique spectral or lifetime signatures combined with our real-time multicomponent analysis (below), e) multi-omics analysis. As we use standard in situ fluorescence staining reagents and protocols which are compatible across different target molecular species, our platform can simultaneously profile multi-omics (e.g. genome, epigenome, proteome, transcriptome, epigenome, and metabolome) in the same sample.

Integrated multi-omics analysis can be critical to provide greater insights in how molecular information translates to functions in biology and disease, and f) high scalability and generalizability. Most existing spatial technologies are unfortunately employed only by a few laboratories due to their laborious procedures and the need for highly specialized equipment. By contrast, our technology uses standard labeling reagents and protocols and requires only a fluorescence microscope that is either integrated with or equipped with a low-cost lifetime imaging unit which is offered by numerous microscope manufacturers (e.g. Leica, Olympus, Nikon, Zeiss and ISS). Fluorescence imaging remains the most familiar and widely used technique in biological research. Therefore, we believe our technology, once fully developed, will be quickly and broadly adopted in the scientific community to address many unsolved biological questions at an affordable cost.

1.1 In situ Hybridization

In situ hybridization (ISH) is a technique used to detect either DNA or RNA using a probe, typically made with DNA or RNA, within cells and tissue, allowing temporal and spatial information about gene expression to be obtained. During hybridization, the single-stranded probes bind in situ to the expressed mRNA or DNA in the sample that it is designed to target. While some probes are designed with a chemical or radioactive moiety for detection, other probes have readout regions for secondary probes with fluorophores to bind to. This allows scientists to detect target expressions in a 3D space.

The two most common ways to visualize the DNA and RNA targets are fluorescence in situ hybridization (FISH) and chromogenic in situ hybridization (CISH). While both methods use labeled, target-specific probes that hybridize to the sample, the visualization method is different for each. With DNA-FISH and RNA-FISH, fluorescence

microscopy can be used for visualization of multiple targets in the same sample to obtain gene expression/presence and temporal and spatial localization. For CISH, brightfield microscopy is used to visualize the CISH signal and tissue morphology simultaneously for molecular pathology diagnostics. Typically, FISH applications use spectrally distinct fluorophore labels for each different hybridization probe to resolve multiple targets.

There are various single molecule FISH (smFISH) methods, as seen in Figure 1.2, which decorates each mRNA transcript with multiple fluorescent oligonucleotide probes and allows users to visualize and detect labeled transcripts using fluorescent microscopy. For smFISH, multiple probes (10 – 50 per transcript) are required to attain a sufficient signal-to-noise ratios to differentiate specific decorated signals from the nonspecific signals that bind within the sample (Figure 1.2A)[1].

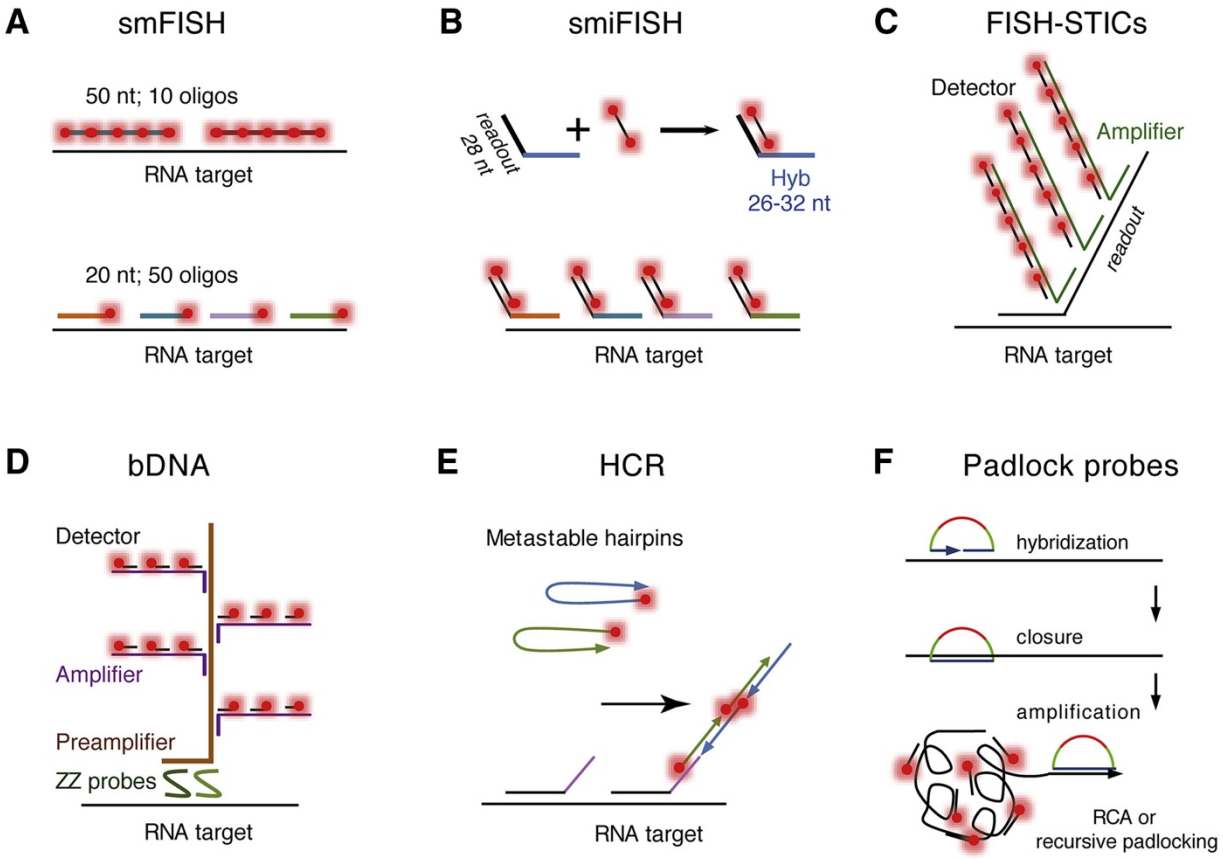


Figure 1.2 Variants of single molecule FISH methods.
(A) Conventional single molecule FISH (smFISH). (B) Single molecule indirect FISH. Nonfluorescent primary probes are first conjugated to mRNA and then subsequently bound to fluorescent secondary probes (C) FISH with Sequential Tethered and Intertwined ODN Complexes (FISH-STICS) utilizes an additional round of labeling to grow out probe tree. (D) Branched DNA (bDNA) utilizes contiguous pairs (ZZ pairs) of oligonucleotides to facilitate specific binding. E) Hybridization chain reaction (HCR) utilizes metastable hairpins to amplify and laterally grow out oligonucleotide labels. F) Padlock probes forms a closed loop upon binding to the transcript and amplifies read out sequences with circle amplification (RCA). Pichon et al. Molecular Cell. 2019[1]

As seen in Figure 1.2B, smiFISH uses indirect labeling schemes where nonfluorescent primary probes are first hybridized to mRNA and then fluorescent secondary probes are hybridized to the primary probe. Since unconjugated oligonucleotide probes can be easily designed to target any complementary regions on the RNA transcript, designing an inexpensive library of probes towards many different genes become highly scalable and cost-effective. The secondary conjugated fluorophores, which are more expensive, can then be designed to hybridize to a “readout”

region on the primary probes and used as a common/shared set to save costs. Furthermore, this indirect labeling method can be further modified to decorate mRNA transcripts with more complex and creative labeling schemes (Figures 1.2B–1E). For this project, we utilize the labeling scheme depicted in Figure 1B to demonstrate how combinatorial labeling with a common panel of fluorophores can be integrated with MOSAICA microscopy to achieve greater multiplexed detection and error-correction. However, our method is not only restricted to this labeling technique and should be compatible with any of the labeling techniques, opening many possibilities for future work on this project.

To rapidly design primary FISH probes for each transcript, the python platform Oligominer was modified. The input sequence is either the mRNA or coding sequence (CDS) FASTA file. The first script is blockParse.py, which screens the input sequence file with the user selected parameters to control for: length, GC content, melting temperature (Tm), spacing, and prohibited sequences like “4GC, 5 AT”, which are important to keep consistent for efficient hybridization. A list of candidate probes is then generated into a FASTQ file, which is then aligned to the genome using an NGS aligner, Bowtie2, to determine the specificity of each candidate probe. Those that are specific are placed into a SAM file and run through the outputClean.py script to keep the unique candidates in a BED file. A seqAligner.py script was developed to align the oligo probes to sequencing data to filter for probes that bind to high expression regions or regions that have higher read counts based on the bigwig histogram. The probeRC.py script performs reverse complement for the probes that require it.

The primary probes comprise of a complementary sequence of 20-35 nucleotides towards the target transcript, preferably within the CDS region which is the most conserved region compared to the untranslated region (UTR). The primary probe has 2 “read-out” regions for the secondary fluorophore probes to hybridize to with “TTT” spacer sequences between the target region – read-out region 1 and read-out region 1 – read-out region 2. Either Oligominer or Sigma Oligoevaluator were then used to screen for secondary structures and primer dimers that would be removed.

1.2 Spectral and Fluorescence Lifetime Microscopy

Fluorescence imaging is an imaging technique used to visualize fluorescent dyes or proteins that label molecular processes or structures. Fluorescence is produced when a process where molecules, such as fluorophores, fluorochromes, or fluorescent dyes, absorb light. After absorbing light, these molecules are raised to a higher energy level (excited state), then emit fluorescent light as they return to their original energy level (ground state) (Figure 1.3). This imaging technique enables a wide range of experimental investigations including the location and dynamics of gene expression, protein expression, and molecular interactions in cells and tissues. However, intensity is only one of the dimensions of fluorescence. Another dimension of each fluorophore is its fluorescence lifetime, which is a measure of how long a fluorophore remains in its excited state before returning to the ground state while emitting a fluorescence photon (Figure 1.3A-B).

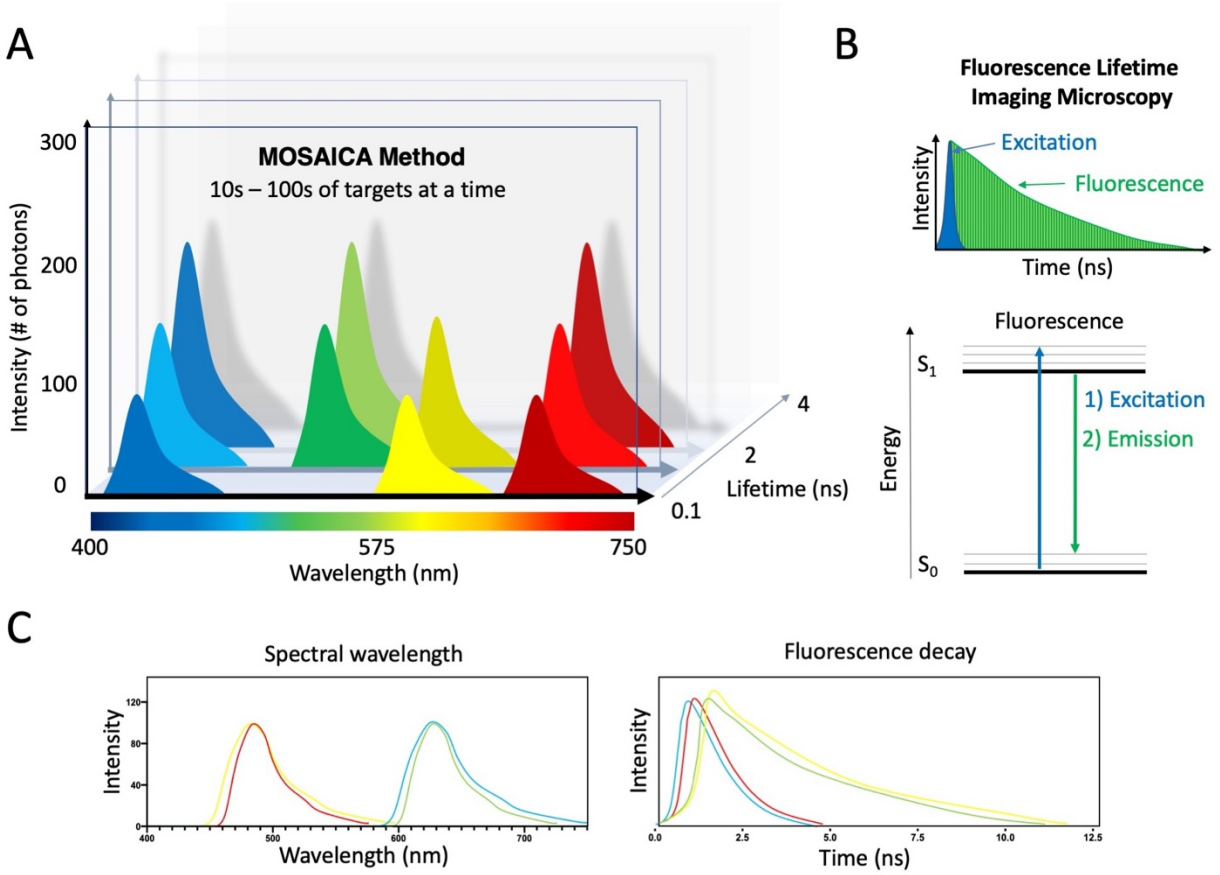


Figure 1.3: Principle of fluorescence lifetime detection.
(A) MOSAICA harnessing the power of the time dimension to multiplex in a single image and detecting multiple targets in the same spectral channel. (B) Fluorescence mechanism demonstrated with Jablonski diagram. (C) Photon distribution in the spectral and temporal dimension.

To measure a fluorophore's lifetime, specialized software and hardware are required. A typical Fluorescence Lifetime Imaging Microscopy (FLIM) setup is shown in Figure 1.4. A pulsed/modulated light source is used to illuminate the sample for digital frequency domain (DFD) lifetime measurement. Instead of modulating the detector by time gating or gain modulation, the entire signal is collected. A time gate is then applied digitally by splitting the detected signal into several (typically 4, 8 or 16) windows, each covering a specific portion of the excitation pulse period. By applying a slight, incremental phase shift to the position of those windows with respect to the excitation pulse, the fluorescence decay of the dye molecules can be recovered with high (~10 ps) resolution.

Using this data, the position of the lifetime phase and modulation can be calculated and presented as a position on the phasor plot, a 2D histogram of all pixel lifetimes. The phasor method draws from the digital frequency domain hardware and software that permits the use of all photons detected from a sample[35]. Additionally, the representation of the decay data using polar coordinates allows the precise measurement of many lifetime components simultaneously without performing fits of the decay data. This allows automatic detection of a plurality of molecular species in the same field of view.

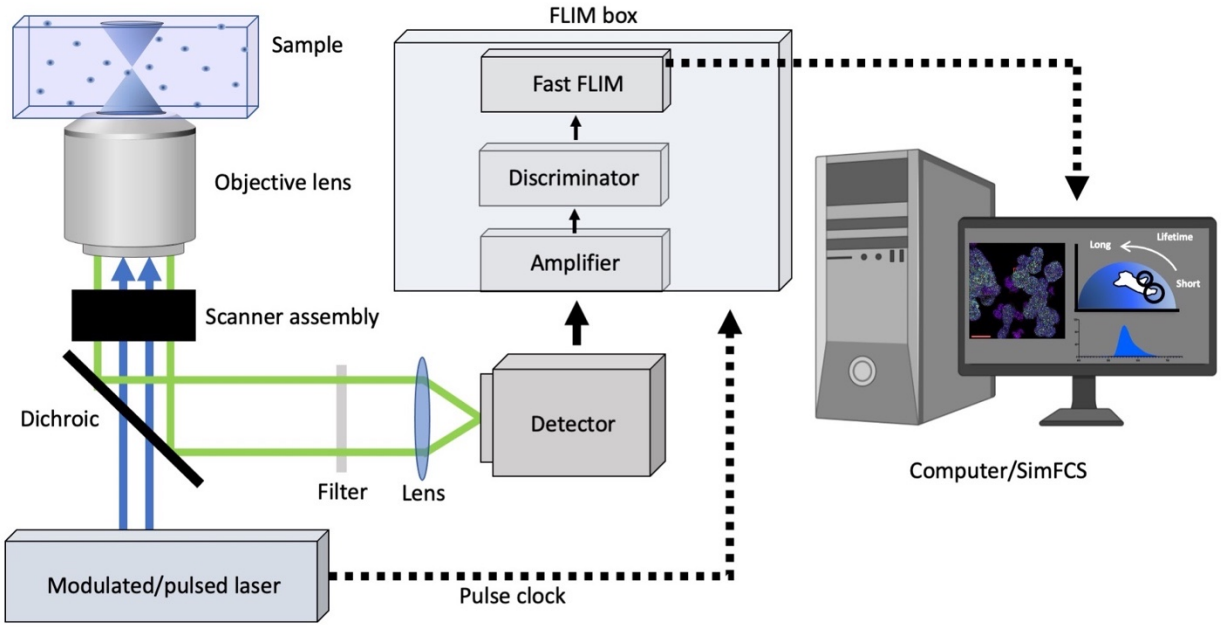


Figure 1.4: Generalized MOSAICA Microscopy setup.
A pulsed/modulated light source is used to illuminate the sample and the fluorescence of the sample is collected by a spectral detector (current resolution around 10 nm). The repetition rate can either be supplied by or delivered to the laser which is used by the electronics in the digital frequency domain to obtain a single photon arrival time using the heterodyne principle (current resolution around 50 ps).

Fluorescence lifetime imaging microscopy (FLIM) is a powerful method that utilizes the time dimension to supplement intensity-based measurements. Fluorophores that have the same excitation and emission spectra but different lifetimes can be separated with FLIM, opening new possibilities for multiplexing capabilities. Table 2 provides a small list of commonly used fluorophores which can be separated by spectral or lifetime

properties. Many multiplexing capabilities are possible with these fluorophores, but this has not been applied to the spatial multiomics field.

Table 2: *List of Commonly Used Fluorophores Differing in Spectra or FLIM Characteristics*

Fluorophore	Exciation Max (nm)	Dichroic (ideal)	Emmision Max (nm)	Extinction Coefficient	Quantum Yield	Brightness	Lifetime (ns)
Alexa 350	346	394	442	N/A	0.55	N/A	N/A
DAPI	358	409.5	461	N/A	N/A	N/A	2.78
Alexa 405	401	411	421	35,000	N/A	N/A	N/A
ATTO 390	390	433	476	24,000	0.9	21600	3.8
ATTO 425	439	462	485	45,000	0.9	40500	3.6
Alexa 430	433	487	541	15000	0.69	N/A	N/A
ATTO 430LS	436	490.5	545	32000	0.65	20800	4
BODIPY 493/503	500	503	506	79000	N/A	N/A	N/A
Dylight 488	493	505.5	518	70000	0.78	54600	3.66
BODIPY FL-X	504	507	510	85000	N/A	N/A	N/A
Alexa 488	495	507	519	73,000	0.92	67160	4
FITC	495	507	519	N/A	N/A	N/A	4.1
BODIPY FL	505	509	513	80000	N/A	N/A	5.7
ATTO 488	500	510	520	90,000	0.8	72000	4.1
ATTO 495	498	512	526	80,000	0.2	16000	1
Oregon Green 488	498	512	526	76000	N/A	N/A	4.1
ATTO 514	511	521.5	532	115,000	0.85	97750	3.9
Alexa 500	503	514	525	N/A	N/A	N/A	N/A
ATTO 520	517	527.5	538	110,000	0.9	99000	3.6
Alexa 514	518	529	540	80,000	N/A	N/A	4.02
Cy3	512	531	550	136,000	0.15	20400	0.3
BODIPY R6G	528	539	550	70000	N/A	N/A	N/A
BODIPY 530/550	524	539	554	77000	N/A	N/A	N/A
ATTO 532	532	542	552	115,000	0.9	103500	3.8
ATTO Rho6G	533	542.5	552	115,000	0.9	103500	4.1
Alexa 532	531	542.5	554	81,000	0.61	49410	2.53
ATTO 542	542	552	562	120,000	0.93	111600	3.7
BODIPY TMR-X	542	558	574	60000	N/A	N/A	6
Alexa 555	555	560	565	155,000	0.1	15500	0.3
Dylight 549	553	561	569	N/A	0.27	N/A	0.65
BODIPY 558/568	558	563.5	569	97000	N/A	N/A	N/A
ATTO 550	556	564.5	573	112,000	0.79	88480	4.06
ATTO 565	564	577	590	120000	0.9	108000	4
ATTO Rho3B	566	577.5	589	120,000	0.5	60000	1.5
ATTO 490LS	496	578.5	661	40,000	0.3	12000	2.6
ATTO Rho11	572	583.5	595	120,000	0.8	96000	4
Alexa 568	578	590.5	603	88,000	0.69	60720	3.6
ATTO Thio12	582	594.5	607	110,000	0.15	16500	2
ATTO Rho101	587	598	609	120,000	0.8	96000	4.2
BODIPY TR-X	589	603	617	68000	N/A	N/A	5.4
Alexa 594	590	603.5	617	92000	0.66	60720	3.9
ATTO 594	603	614.5	626	120,000	0.85	102000	3.9
ATTO 590	593	607.5	622	120,000	0.8	96000	3.7
Alexa 610	612	620	628	144000	N/A	N/A	N/A
ATTO 620	620	631	642	120,000	0.5	60000	2.9
BODIPY 630/650-X	625	632.5	640	101000	N/A	N/A	3.9
ATTO Rho14	625	635.5	646	140,000	0.8	112000	3.7
Alexa 633	632	639.5	647	159,000	N/A	N/A	3.2
Alexa 635	633	640	647	140,000	N/A	N/A	N/A
ATTO 633	630	640.5	651	130,000	0.64	83200	3.3
BODIPY 650/665-X	646	653	660	102,000	N/A	N/A	N/A
ATTO 643	643	654	665	150,000	0.62	93000	3.5
ATTO 647N	646	655	664	150,000	0.65	97500	3.5
ATTO 647	647	657	667	120,000	0.2	24000	2.4
Alexa 647	650	659	668	270,000	0.33	89100	1.04
CY5	650	660	670	250,000	0.3	75000	1
ATTO 665	662	671	680	160000	0.6	96000	2.9
ATTO 655	663	671.5	680	125,000	0.3	37500	1.8
Alexa 660	663	676.5	690	132,000	0.37	48840	1.2
ATTO 680	681	689.5	698	125000	0.3	37500	1.7
Alexa 680	679	690.5	702	205000	0.36	73800	1.2
ATTO 700	700	708	716	120000	0.25	30000	1.6
Alexa 700	702	712.5	723	290000	0.25	72500	1
ATTO 725	728	739.5	751	120000	0.1	12000	0.5
ATTO 740	743	753	763	120,000	0.1	12000	0.6
Alexa 750	749	762	775	290,000	0.12	34800	0.66
Cy7	750	769	788	199,000	0.3	59700	0.8
Alexa 790	784	799	814	N/A	N/A	N/A	N/A

In addition, autofluorescent moieties/artifacts in cells and tissue have characteristic lifetime signatures which occupy different positions that are distinguishable via lifetime (Table 3). Using just intensity measurements, these moieties are inseparable from each other or from a labeled fluorophore. However, by including FLIM measurements, these moieties can be easily corrected or removed out from analysis via the phasor approach.

Table 3: *List of spectral characteristics and lifetimes of endogenous fluorophores[36].*

Endogenous fluorophore	Excitation (nm)	Emission (nm)	Lifetime (ns)
Metabolic coenzymes			
NAD(P)H free	340 (max)	470 (max)	0.4 (free), 1 to 5 (bound)
FAD, flavin	450 (max)	535 (max)	2.3 to 2.9 (free), <0.1ns (bound)
Flavin mononucleotide (FMN)	444 (max)	558 (max)	4.27 to 4.67
Structural proteins			
Collagen	280 to 350	370 to 440	0.2 to 0.4, 0.4 to 2.5
Elastin	300 to 370	420 to 460	0.2 to 0.4, 0.4 to 2.5
Vitamins			
Retinol	327 (max)	510 (max)	1.8, 5.0 (free), 0.7, 3.6, 12 (bound)
Riboflavin	420 to 500	520 to 750	4.12
Vitamin B6	330 (max)	420 (max)	0.6 to 8.4
Vitamin K	335 (max)	480 (max)	—
Vitamin D	390 (max)	480 (max)	—
Vitamin B12	275 (max)	305 (max)	—
Pigments			
Melanin	300 to 800	440, 520, 575	0.1 to 0.2, 0.5 to 1.8, 7.9
Eumelanin	355	520	0.058, 0.51, 2.9, 7
Keratin	277 (max)	382 (max)	1.4
Protoporhorphyrin IX	400 to 450	630, 690, 710	9.7 to 16
Lipofuscin	340 to 395	540, 430 to 460	1.34
Bilirubin	350 to 520	480 to 650	0.02 to 0.09, 1 to 2
Amino acids			
Phenylalanine	258 (max)	280 (max)	7.5
Tryptophan	280 (max)	250 to 310	3.03
Tyrosine	275 (max)	300 (max)	2.5

1.3 Spatialomics in Cancer Research

Our high throughput and high-plex spatial profiling technology will broadly enable scientists and clinicians to better study cancer biology and to develop precision diagnostics and treatments for cancer. Cancer biologists have started to realize, only

recently, how heterogeneous gene (and protein) expression is and how many different cell identities/states there are in tumors. In other words, the dynamic cell fate is defined spatiotemporally by the expression of multiple (rather than a single) genes. Therefore, to fully characterize cells in situ we need to be able to assess multiple transcripts (and proteins) within the same cell, which can be readily addressed by our unique technology through direct, highly multiplexing, in situ biomarker profiling in a single round of staining and imaging. Three applications requiring high-plex in situ analysis that are broadly representative in both basic cancer biology and clinical companion diagnostics (CDx) for stratified care, include: 1) Examining within-cell correlations and location in gene expression sampled among heterogeneous cells will inform gene regulatory mechanisms, which we cannot get from bulk measurements. 2) Single-cell RNA sequencing (scRNA-seq) returns cell identities in the form of rather long "differentially expressed gene lists" that "define" cell types. However, the clustering process loses spatial information and is subjective, variable and error prone. The only way to validate whether a pattern of gene expression really defines a cell type, or conflates multiple cell types, is through multiplex spatial transcriptomics. 3) Patient derived materials are often available in limiting quantity and generating hundreds of sections to test for many markers is tedious and non-feasible. Hence, multiplexing is the only efficient way of doing this.

In cancer diagnosis, prognosis, and patient stratification for combination therapy, especially in immunotherapy, physicians would now want to analyze tumor biopsies for a large number (typically 16-64) of biomarkers of specific tumor cells, infiltrating immune cells, stromal components, and other drug targets[37-39]. Biological and clinical models to develop and validate our technology in the context of these three cancer applications

include colorectal cancer (CRC), melanoma carcinogenesis, and tumor infiltrating lymphocytes.

We collaborated with Dr. Marian Waterman who recently established a xenografted CRC model that exhibits patterns of heterogeneity of Wnt signaling in various cancer cell subtypes in the spatial context[40,41], therefore ideal for developing spatialomics technology. We used accompanying scRNAseq data that confirmed heterogeneity at the transcriptome level and specifically identified a small population expressing markers of cancer stemness ("CSC", e.g. ROBO, SOX2, BMP4, PROX1). This model allows us to develop and validate MOSAICA and evaluate its utility in inferring gene regulatory pathways using quantitative single-molecule in situ RNA analysis, and as a tool for in situ validation of scRNAseq data. Human SW480 CRC line that expresses CSC markers and its lentiviral transduction or shRNA knockdown counterpart control with impaired Wnt signaling, which would result in distinctly different spatial heterogeneity in the xenografted model were used. Xenograft tissue sections (5 μm) were prepared with formalin fixed paraffin embedded (FFPE) preservation due to their wide use in research. For our initial 16-plex panel, we focused on genes in Wnt signaling and implicated in cancer stemness together with housekeeping genes (e.g. POLR2A and mTOR). For 32- and 64-plex, we will expand to other key signaling pathways including FGF, BMP, and TGF-b signaling and to markers that can identify other key constituents of tumors and stroma including subtypes of vessels, fibroblasts, and immune cells.

Simultaneous detection of protein and transcript levels within the same sample will reveal the genotypic and phenotypic heterogeneity and provide enriched information for biology and disease diagnosis. Few technologies can do this due to limitations of detection

and/or sample processing[42,43]. To develop and validate an immuno-oncology marker panel for cancer tissue biopsy analysis for cancer diagnosis, prognosis and treatment stratification applications[37,44], we examined clinical melanoma tissues and characterize tumor infiltrating lymphocytes. Melanoma is chosen as a model because it represents one of the most established tumor types for immune checkpoint inhibitors[45-49]. For this model, we worked with Dr. Anand Ganesan (MD, PhD) who specializes in melanoma carcinogenesis and patient treatment with immunotherapeutics.

Protein targets were chosen based on their immuno-oncology applications and an initial panel comprised of markers for tumor cells (epithelial Pan-cytokeratins, melanoma antigen SOX10), immune cell subsets: T cells (CD3, CD4 and CD8), B cells (CD20), macrophages (CD68), and Tregs (FOXP3), myeloid-derived suppressor cells (CD11b), and immune exhaustion (PD-L1, TIGIT, LAG3). As we increase our multiplex capability, we will expand to markers that cover additional tumor, immune and stromal cell subtypes and checkpoint proteins. For mRNA, corresponding mRNAs for protein markers mentioned above, along with melanoma markers (e.g. PMEL) and housekeeping genes (e.g. POLR2A) were used.

1.4 Spatialomics in Inflammatory Neurodegeneration Disease Research

Alzheimer disease (AD) is the leading cause of age-related dementia. Unfortunately, current therapies are palliative, and several drug candidates have failed in late-stage trials. Hence, there is an urgent need to improve our understanding of the mechanisms that drive the development of AD. Recent studies have identified numerous AD risk genes and highlighted the importance of inflammation in AD and have implicated microglia, the primary immune cell of the brain[50,51]. Yet the precise function of these genes

and proteins, the specific subsets of microglia that express them, and how expression changes with disease progression to influence pathogenesis, remains poorly understood. Multiple sclerosis (MS) is a chronic, autoimmune disease of the central nervous system (CNS) without a cure. Deficits in neurological function result from widespread demyelination and axonal loss caused by infiltrating immune cells[52].

Profiling transcriptome and proteome in space and time at single-cell level at both cellular and tissue level is critical to (a) elucidate stem cell and development biology processes, disease pathophysiology, mechanisms of action of transplanted stem cells, and (b) enable drug discovery, patient diagnosis and prognosis, and stratification for regenerative medicine[5,53]. In inflammatory CNS neurodegenerative diseases including AD and MS, our understanding of the immune system's key role in mediating (a) disease pathophysiology, and (b) immunomodulatory and regenerative mechanisms following stem cell treatments has been hindered by the lack of cost-effective and user-friendly tools that can rapidly perform in situ, high-plex profiling of mRNA and protein markers of the immune components.

Our understanding of underlying mechanisms in neurodegenerative diseases has been hindered by the lack of in situ spatial analysis for target markers to determine their presence, numbers, locations, dynamics, and interactions to reveal key cellular states and processes, cell-cell and cell-niche communications on the community- and tissue-scale in heterogeneous samples. Existing tools for in situ analysis including fluorescence in situ hybridization (FISH) and immunohistochemistry are often confounded by inherent age-related accumulation of autofluorescent lipids. Moreover, these traditional approaches only provide low throughput because of limited separation of spectral

channels. Recent single-cell sequencing methods[54,55] lack the critical spatial context needed to understand complex heterogeneous samples (i.e. brain).

For a collaboration with Dr. Xiangmin Xu, we are working on a neurological mRNA panel to analyze spatial transcriptomics in brain tissues, including neural markers (e.g., APOE, AQP4, BDNF, C1QA, CUX2). In parallel, the panel is run on brain tissue samples through their MERFISH platform, multiplexed error-robust fluorescence in situ hybridization, a commercial platform currently available. We will benchmark the cost-effective MOSAICA platform with MERFISH and run neurological mRNA panels of 15-30 markers on MOSAICA routinely to study neurological diseases and regenerative strategies.

In Dr. Mathew Blurton-Jones's lab, they have established an approach to examine the response of human microglia to AD pathology *in vivo* by generating chimeric AD transgenic mice that exhibit robust forebrain engraftment of human microglia[56,57]. Informed by scRNAseq work identifying human disease-associated microglial (DAM) transcripts in chimeric mice[56] and other recent sequencing studies[58], we examine a panel of 60 genes (including: HLA-DRB1, TREM2, CD9, MAFB, LGALS3, APOE, LPL, SALL1). Human xenotransplanted microglia[56] that expresses numerous microglial markers and exhibit an ex vivo-like human microglia transcriptome will be used to validate probe designs. Tissues were derived from our chimeric human microglial model of AD. This allows us to gain significant insight into the gene expression patterns of human microglia and how these patterns change with disease progression, proximity to AD pathology, and genetic risk.

For future work with the MS model in collaboration with Dr. Craig Walsh, the MOSAICA platform could profile T cell and APC subsets and interacting T cells including Teff: CD45, CD3, CD8; Th1: CD45, CD3, CD4, Tbet, IFNg; Th17: CD45, CD3, CD4, IL-17; Tregs: CD45, CD4, CD25, FoxP3; B cells: CD45, CD19, CD20; Dendric cells (DCs): CD11b, CD11c; Macrophages/ Microglia: CD45, CD11b; NK cells: CD45, CD49, NK1.1; and NK T cells: CD45, CD3. Furthermore, to identify the phenotype of these immune cells (e.g., activated, presenting antigens, naïve, memory, etc.) the MOSAICA platform can co-detect protein markers with antibodies against immune cell profiling markers (CD44, CD127, CD69, MHCI, MHCII, CD80/86, PD-1, and CTLA-4). Other CNS cells indicative of demyelination and regeneration would also be profiled including neurons (bIII Tubulin), astrocytes (GFAP), and Oligodendrocytes (Olig2, PLP). This could delineate the interactions of regulatory T cells and other immune cell types during MS disease and following hNSC transplantation.

1.4 References

1 Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Molecular cell* **71**, 468-480 (2018).

2 Barsoum, I., Tawedrous, E., Faragalla, H. & Yousef, G. M. Histo-genomics: digital pathology at the forefront of precision medicine. *Diagnosis* **6**, 203-212 (2019).

3 Hu, L. *et al.* Fluorescence in situ hybridization (FISH): an increasingly demanded tool for biomarker research and personalized medicine. *Biomarker research* **2**, 3 (2014).

4 Gryglewski, G. *et al.* Spatial analysis and high resolution mapping of the human whole-brain transcriptome for integrative analysis in neuroimaging. *NeuroImage* **176**, 259-267 (2018).

5 Consortium, H. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187 (2019).

6 Ratan, Z. A. *et al.* Application of fluorescence in situ hybridization (FISH) technique for the detection of genetic aberration in medical science. *Cureus* **9** (2017).

7 Gaspar, I. & Ephrussi, A. Strength in numbers: quantitative single-molecule RNA detection assays. *Wiley Interdiscip Rev Dev Biol* **4**, 135-150, doi:10.1002/wdev.170 (2015).

8 Lee, J. H. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip Rev Syst Biol Med* **9**, doi:10.1002/wsbm.1369 (2017).
9 Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* **113**, 11046-11051, doi:10.1073/pnas.1612826113 (2016).
10 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).
11 Mellis, I. A. & Raj, A. Half dozen of one, six billion of the other: What can small- and large-scale molecular systems biology learn from one another? *Genome Res* **25**, 1466-1472, doi:10.1101/gr.190579.115 (2015).
12 Kim, S. W., Roh, J. & Park, C. S. Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. *J Pathol Transl Med* **50**, 411-418, doi:10.4132/jptm.2016.08.08 (2016).
13 Miller, R. T. in *Seminars in diagnostic pathology.*
14 Taylor, C. & Levenson, R. M. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology* **49**, 411-424 (2006).
15 Jensen, E. Technical review: In situ hybridization. *The Anatomical Record* **297**, 1349-1353 (2014).
16 Chu, Y.-H., Hardin, H., Zhang, R., Guo, Z. & Lloyd, R. V. in *Seminars in diagnostic pathology.* (Elsevier).
17 Chae, Y. K. *et al.* Challenges and future of biomarker tests in the era of precision oncology: Can we rely on immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) to select the optimal patients for matched therapy? *Oncotarget* **8**, 100863-100898, doi:10.18632/oncotarget.19809 (2017).
18 Stempel, A. J., Morgans, C. W., Stout, J. T. & Appukuttan, B. Simultaneous visualization and cell-specific confirmation of RNA and protein in the mouse retina. *Mol Vis* **20**, 1366-1373 (2014).
19 Grabinski, T. M., Kneynsberg, A., Manfredsson, F. P. & Kanaan, N. M. A method for combining RNAscope in situ hybridization with immunohistochemistry in thick free-floating brain sections and primary neuronal cultures. *PLoS One* **10** (2015).
20 Blom, S. *et al.* Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Sci Rep* **7**, 15580, doi:10.1038/s41598-017-15798-4 (2017).
21 Kalra, J. & Baker, J. in *Signal Transduction Immunohistochemistry* 237-251 (Springer, 2017).
22 Moffitt, J. R. *et al.* High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc Natl Acad Sci U S A* **113**, 14456-14461, doi:10.1073/pnas.1617699113 (2016).
23 Yuan, G. C. *et al.* Challenges and emerging directions in single-cell analysis. *Genome Biol* **18**, 84, doi:10.1186/s13059-017-1218-y (2017).
24 Medaglia, C. *et al.* Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622-1626 (2017).
25 Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342-357 (2016).

26 Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**, 22-29, doi:10.1016/j.jmoldx.2011.08.002 (2012).
27 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090, doi:10.1126/science.aaa6090 (2015).
28 Eng, C. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239, doi:10.1038/s41586-019-1049-y (2019).
29 Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat Methods* **16**, 533-544, doi:10.1038/s41592-019-0404-0 (2019).
30 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
31 Porta Siegel, T. *et al.* Mass Spectrometry Imaging and Integration with Other Imaging Modalities for Greater Molecular Understanding of Biological Tissues. *Mol Imaging Biol* **20**, 888-901, doi:10.1007/s11307-018-1267-y (2018).
32 Ferguson, C. N., Fowler, J. W., Waxer, J. F., Gatti, R. A. & Loo, J. A. Mass spectrometry-based tissue imaging of small molecules. *Adv Exp Med Biol* **806**, 283-299, doi:10.1007/978-3-319-06068-2_12 (2014).
33 Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e915, doi:10.1016/j.cell.2018.07.010 (2018).
34 Becker, W. Fluorescence lifetime imaging–techniques and applications. *Journal of microscopy* **247**, 119-136 (2012).
35 Colyer, R. A., Lee, C. & Gratton, E. A novel fluorescence lifetime imaging system that optimizes photon efficiency. *Microscopy research and technique* **71**, 201-213 (2008).
36 Datta, R., Heaster, T. M., Sharick, J. T., Gillette, A. A. & Skala, M. C. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. *Journal of biomedical optics* **25**, 071203 (2020).
37 Keren, L. *et al.* A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* **174**, 1373-1387. e1319 (2018).
38 Wagner, J. *et al.* A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell* **177**, 1330-1345. e1318 (2019).
39 Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615-620 (2020).
40 Chen, G. T. *et al.* Disrupting β-catenin dependent Wnt signaling activates an invasive gene program predictive of colon cancer progression. *bioRxiv*, 667030 (2019).
41 Lee, M. *et al.* Mathematical modeling links Wnt signaling to emergent patterns of metabolism in colon cancer. *Molecular systems biology* **13**, 912 (2017).
42 Merritt, C. R. *et al.* High multiplex, digital spatial profiling of proteins and RNA in fixed tissue using genomic detection methods. *bioRxiv*, 559021, doi:10.1101/559021 (2019).
43 Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* **38**, 586-599 (2020).
44 Andrews, A. Treating with checkpoint inhibitors—figure $1 million per patient. *American health & drug benefits* **8**, 9 (2015).

45 Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science* **359**, 1350-1355 (2018).
46 Pedroza-Gonzalez, A. *et al.* Tumor-infiltrating plasmacytoid dendritic cells promote immunosuppression by Tr1 cells in human liver tumors. *Oncoimmunology* **4**, e1008355 (2015).
47 Sade-Feldman, M. *et al.* Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998-1013. e1020 (2018).
48 Amaria, R. N. *et al.* Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nature medicine* **24**, 1649-1654 (2018).
49 Toki, M. I. *et al.* High-Plex Predictive Marker Discovery for Melanoma Immunotherapy–Treated Patients Using Digital Spatial Profiling. *Clinical Cancer Research* **25**, 5503-5512 (2019).
50 Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science*, doi:10.1126/science.aay0793 (2019).
51 Efthymiou, A. G. & Goate, A. M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol Neurodegener* **12**, 43, doi:10.1186/s13024-017-0184-x (2017).
52 Steinman, L. Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. *Cell* **85**, 299-302 (1996).
53 Regev, A. *et al.* The human cell atlas white paper. *arXiv preprint arXiv:1810.05192* (2018).
54 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337, doi:10.1038/s41586-019-1195-2 (2019).
55 Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290 e1217, doi:10.1016/j.cell.2017.05.018 (2017).
56 Hasselmann, J. *et al.* Development of a chimeric model to study and manipulate human microglia in vivo. *Neuron* **103**, 1016-1033. e1010 (2019).
57 McQuade, A. *et al.* Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nature communications* **11**, 1-17 (2020).
58 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).

CHAPTER 2

SPATIAL TRANSCRIPTOMICS USING COMBINATORIAL FLUORESCENCE SPECTRAL AND LIFETIME ENCODING, IMAGING AND ANALYSIS

Authors:
Tam Vu[1,2,†], Alexander Vallmitjana[1,3,†], **Joshua Gu[2,4,†],** Kieu La[1], Qi Xu[1], Jesus Flores[2,5], Jan Zimak[6], Jessica Shiu[7], Linzi Hosohama[8,], Jie Wu[4,9], Christopher Douglas[10], Marian Waterman[8,9], Anand Ganesan[4,7,9], Per Niklas Hedde[3,6,11], Enrico Gratton[1,3,11*], and Weian Zhao[1,2,4,6,9,*]

Author Affiliations:
1. Department of Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States
2. Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, Irvine, California 92697, United States
3. Laboratory for Fluorescence Dynamics, University of California, Irvine, Irvine, California 92697, United States
4. Department of Biological Chemistry, University of California, Irvine, Irvine, California 92697, United States
5. CIRM Stem Cell Research Biotechnology Training Program at California State University, Long Beach, Long Beach, California 90840, United States
6. Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, California 92697, United States
7. Department of Dermatology, University of California, Irvine, Irvine, CA, USA
8. Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California 92697, United States
9. Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, California 92697, United States
10. Department of Pathology and Laboratory Medicine, University of California, Irvine, Irvine, California 92697, United States
11. Beckman Laser Institute & Medical Clinic, University of California, Irvine, Irvine, California 92697, United States

† = These authors contributed equally to this work.
* = corresponding author

Contribution: J.G. developed and optimized the MOSAICA platform's probe design pipeline, sequencing analysis, tissue processing, hybridization process, CellProfiler image analysis and validation/benchmarking with other platforms. Figures 1, 6, S1-3, S7-9 were made by J.G. and J.G. made significant contributions to Figures 2, 4, 5, S6. Imaging on the Nikon and some Leica SP8 imaging were done by J.G.

2.1 Abstract

Multiplexed mRNA profiling in the spatial context provides new information enabling basic research and clinical applications. Unfortunately, existing spatial transcriptomics methods are limited due to either low multiplexing or complexity. Here, we introduce a spatialomics technology, termed Multi Omic Single-scan Assay with Integrated Combinatorial Analysis (MOSAICA), that integrates in situ labeling of mRNA and protein markers in cells or tissues with combinatorial fluorescence spectral and lifetime encoded probes, spectral and time- resolved fluorescence imaging, and machine learning-based decoding. We demonstrate MOSAICA's multiplexing scalability in detecting 10-plex targets in fixed colorectal cancer cells using combinatorial labeling of five fluorophores with facile error-detection and removal of autofluorescence. MOSAICA's analysis is strongly correlated with sequencing data (Pearson's r = 0.96) and was further benchmarked using RNAscope™ and LGC Stellaris™. We further apply MOSAICA for multiplexed analysis of clinical melanoma Formalin-Fixed Paraffin-Embedded (FFPE) tissues. We finally demonstrate simultaneous co-detection of protein and mRNA in cancer cells.

2.2 Introduction

Cell fate and cell-cell, cell-niche interactions are tightly regulated in space at both genetic and tissue and system level to mediate organ development, tissue homeostasis and repair, and disease appearance and progression. Therefore, spatial transcriptomics that profile gene expression landscape at the single-cell level in tissues in a 3D spatial context as shown in this work represents a frontier in biological research and precision medicine[1–8]. For instance, spatial transcriptomics techniques can (a) help realize the

vision of the human cell atlas in generating “high-resolution and comprehensive, three-dimensional reference maps of all human cells in the body”, (b) determine molecular mechanisms that govern cell fate, state, lineage and cell cooperation in tissue formation in developmental biology and regenerative medicine, (c) investigate the biological changes associated with different diseases in a spatial-dynamic fashion and to uncover disease molecular mechanisms and discover disease biomarkers, and (d) characterize the complexities of tissue biopsy (e.g., tumor) in clinical pathology to inform personalized disease diagnosis and therapeutic intervention in the era of precision medicine. Spatial transcriptomics tools need to be able to assess multiple transcripts within the same cell and sample in a highly multiplexed fashion due to the heterogeneous gene expression and many different cell identities/states exist in a particular tissue. Furthermore, patient derived materials are often available in limited quantity and generating many sections to test for different markers separately is tedious and non-feasible.

A major bottleneck in spatial transcriptomics is the lack of tools that can be both easy-of-use and highly multiplexing[7–13]. Conventional tools for in situ analysis including fluorescence in situ hybridization (FISH) (e.g., LGC Stellaris™) can only detect 3–4 targets at a time because of the limited number of spectral channels in fluorescence microscopes[12–14]. Conventional methods for in situ profiling of transcripts are further confounded by the autofluorescent moieties in tissue preparations including clinical biopsies. Recent single-cell RNA sequencing methods provide information on the presence and identity of transcripts in single cells but lack the critical spatial context needed to understand complex heterogeneous tissue[15–17]. Imaging- and FISH- based spatial transcriptomic methods that employ sequential labeling, stripping, and imaging

(e.g., seqFISH, MERFISH) or branched amplification (e.g., RNAscope™, SABER) are often complicated, error-prone, time-consuming, laborious and/or costly to scale up[18–22]. Furthermore, repeated processing of the same sample can in some cases affect tissue structural integrity and target molecules and may not always be amenable for clinical applications such as profiling patient biopsies. Spatial transcriptomics using in situ sequencing (e.g., ISS, FISSEQ, starMAP and ExSeq) or in situ barcoding coupled with ex situ sequencing (e.g., GeoMx, slide-seq, and DBiT-seq) can drastically improve multiplexing but suffer from reduced spatial resolution and detection efficiency especially for low-abundance targets[22–25].

In this work, we report a fluorescence imaging-based spatialomics technology termed MOSAICA (Multi Omic Single-scan Assay with Integrated Combinatorial Analysis) that enables direct, highly multiplexed biomarker profiling in the 3D spatial context in a single round of staining and imaging. MOSAICA employs in situ staining with combinatorial fluorescence spectral and lifetime encoded probes, spectral- and time-resolved fluorescence imaging, and AI-based target decoding pipeline (Fig. 2.1). Fluorescence lifetime is a measure of the time a fluorophore spends in the excited state before returning to the ground state and is an inherent characteristic of the fluorophore and its surrounding environment[26,27]. By utilizing both time and spectral domains for labeling and imaging, we were able to discriminate a repertoire of 10 different fluorescent signatures against autofluorescent moieties and nonspecific binding events within the same sample in this study and expect to scale up to at least 60-plex in the future to enable increased multiplexing capabilities with standard optical systems.

In this study, we describe the MOSAICA pipeline, including automated probe design algorithm, probe hybridization optimization, and validation, combinatorial spectral and lifetime labeling and analysis for target encoding and decoding. Particularly, we developed an automated machine learning-powered spectral and lifetime phasor segmentation software that has been developed to spatially reveal and visualize the presence, identity, expression level, location, distribution, and heterogeneity of each target mRNA in the 3D context. We showcased MOSAICA in analyzing a 10-plex gene expression panel in colorectal SW480 cells based on combinatorial spectral and lifetime barcoding of only five generic commercial fluorophores. Using this model, we illustrated the multiplexing scalability and MOSAICA's ability to correct for stochastic nonbinding artifacts present within the sample. We further demonstrated MOSAICA's utility in improved multiplexing, error-detection, and autofluorescence removal in highly scattering and autofluorescent clinical melanoma FFPE tissues, demonstrating its potential use in tissue for cancer diagnosis and prognosis. To further reveal the potential of MOSAICA, we demonstrated its multiomics capability with simultaneous co-detection of protein and mRNA in colorectal SW480 cells. MOSAICA is rapid, cost-effective, and easy-to-use and can fill a critical gap between conventional FISH and sequential- and sequencing-based techniques for targeted and multiplexed spatial transcriptomics.

2.3 Results

2.3.1 MOSAICA workflow

In a typical MOSAICA workflow (Fig. 2.1), primary oligonucleotide probes designed to specifically bind to mRNA targets with a complementary target region (25–30 base long) are incubated with fixed cell or tissue samples (Fig. 2.1a, b). These primary probes

also contain an adjacent adaptor region consisting of two readout sequences for modular secondary probe binding. In this study, double-ended secondary probes with fluorophores on each end are hybridized to the readout region on the primary probes (Fig. 2.1c). Through combinatorial labeling, each target is encoded with a dye with a distinct spectrum and lifetime signature. The labeled samples are then imaged using a custom built or commercial microscope (e.g., the Leica SP8 Falcon used in this study) equipped with spectral and lifetime imaging capabilities (Fig. 2.1d). Both spectral and fluorescence lifetime data will be captured, and then analyzed using phasor plots (Fig. 2.1e). Our automated machine learning algorithm and a codebook finally reveal the locations, identities, counts, and distributions of the present mRNA targets in a 3D context (Fig. 2.1f).

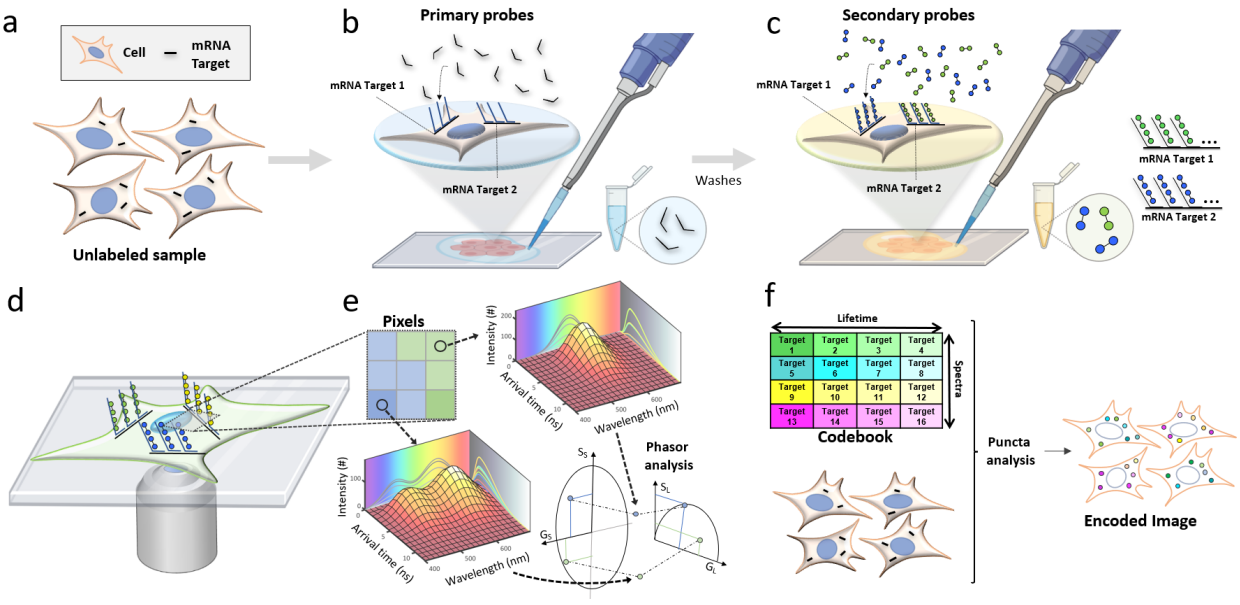


Figure 2.1 Schematic of the MOSAICA approach for labeling and analysis of spectral and time-resolved components.
a) *Sample(s) can be fixed cells or tissues. RNA transcripts from genes of interest are targeted for detection. Protein targets can be stained too in mRNA and protein codetection.* ***b)*** *Primary labeling probes are designed to include two functional regions: a target region which is complementary and can bind to the mRNA target and an adjacent readout region which can subsequently bind to fluorescently labeled oligonucleotides.* ***c)*** *Secondary fluorescent probes are added to bind to the primary probes to form different combinations (combinatorial labeling) through a "readout" domain.* ***d)*** *Labeled targets are measured under*

a fluorescent microscope to interrogate the spectral and lifetime characteristics of the labeled moieties. ***e)*** *Phasor analysis is used to identify which fluorophore labels are present in each pixel and puncta.* ***f)*** *Labeled targets eliciting the encoded intensity-based and time-based signature are decoded to reveal the locations, identities, counts, and distributions of the present mRNA targets in a multiplexed fashion.*

2.3.2 Probe design pipeline

To rapidly design oligonucleotide probes for the transcript of each gene, we modified the python platform, OligoMiner[28], a validated pipeline for rapid design of oligonucleotide FISH probes. Briefly, as shown in Supplementary Fig. 1a, using the mRNA or coding sequence file of the target gene, the blockParse.py script will screen the input sequence and output a file with candidate probes while allowing us to maintain consistent and customized length, GC, melting temperature, spacing, and prohibited sequences. Using Bowtie2, the candidate probes are rapidly aligned to the genome to provide specificity information that is used by the outputClean.py script to generate a file of unique candidates only. The primary probes comprise complementary sequence of typically 27–30 nucleotides and are designed mostly within the coding sequence region, which has fewer variation than the untranslated region[20]. We wrote a script, seqAnalyzer.py, to automate the alignment of primary probes to sequencing data (Supplementary Fig. 1b) so that probes that aligned to regions of lower read counts would be discarded. Furthermore, primary probe “readout” domains and secondary probes (typically 15–20 nucleotides long) are designed to be orthogonal to each other to avoid off-target binding. Libraries and databases of over 200,000 orthogonal sequences are available online and we have simply used those that have been previously validated[29]. Fluorophores exhibiting distinct spectrum (typically with excitation/emission spectra in the 400–700 nm range) and lifetimes (typically in the 0.3–10 ns range) can be conjugated to oligos which were obtained through commercial vendors (see Methods).

2.3.3 Probe labeling validation and optimization

We first investigated the specificity of our labeling condition using a simple cell mixture model comprising wild-type human embryonic kidney (HEK293T-X) cells and HEK293T-X cells engineered with mNeonGreen (Supplementary Fig. 2a) by detecting mNeonGreen mRNA as the gene expression target. The HEK293T-X cells were engineered to express mNeonGreen, an artificial construct inserted that resulted in the cells with the construct to express a green fluorescence. The HEK293T-X cells were transfected with an engineered mNeonGreen plasmid, which contained puromycin and zeocin resistant genes. Three days post transfection, the cells were selected using 0.2ug/ml to 2ug/ml of puromycin and zeocin. The concentration was increased by 0.2ug/ml every 4-6 days. The engineered mNeonGreen HEK293T-X cells were then sorted with FACS, collecting the cells positive for mNeonGreen. The positive cells were then cultured with 2ug/ml of puromycin and zeocin for 2 weeks. The mNeonGreen HEK293T-X cells were then mixed at a 50/50 ratio with wild-type (WT) HEK293T-X cells. The model system was used along with 14 primary probes that targeted the mNeonGreen construct, which is only present in the mNeonGreen cells, but not the WT cells. The secondary probe was the Alexa 647 probe. Puncta in the 647nm channel only present in the cells that expressed a green fluorescence were the probes targeting the mNeonGreen sequence. Since only fluorescent mNeonGreen positive cells can express the corresponding mRNA transcripts, this cell mixture model provides a straightforward tool to assess the specificity and nonspecific binding. Using a Nikon epifluorescence microscope to image the samples following staining with primary and secondary probes (all probe sequences used in this study are provided in Supplementary Data 1), we

detected on average 43.5 puncta per mNeonGreen positive cell (n = 76 cells) and 0.25 puncta per wild-type cell (n = 164) (Supplementary Fig. 2b, c), indicating minimal nonspecific binding with our probe labeling strategy. To further validate the baseline level of nonspecific binding, we included a negative control with the primary probe designed toward dopachrome tautomerase, a gene in the mouse genome that is not expressed in our HEK293T-X model system, along with a condition with secondary probes only. Similarly, an average of 43.5 puncta per cell was detected for the mNeonGreen cells while the wild type and negative controls a mean of 2.5 puncta per cell was detected with a lower signal-to-noise. We next optimized labeling efficiency by testing the number of primary probes and incubation times of primary probes and secondary probes (Supplementary Fig. 3). We determined our optimal condition to comprise a minimal of at least 12 primary probes for each target mRNA (in practice, we always maximize the number of primary probes per mRNA depending on the size of mRNA). Indeed, 40 primary probes per channel per mRNA were subsequently used in this study, with incubation time of 16 h for primary probe hybridization and 1h for secondary probe hybridization, respectively, which were used in subsequent experiments.

2.3.4 Imaging and phasor analysis

Lifetime imaging is a tool that measures the spatial distribution of probes with different fluorescence lifetime. Samples are stimulated with modulated or pulsed lasers at a particular frequency, typically around the 40–80 MHz, which allows the fluorescence to decay within the stimulated period, typically in the ns range. After acquiring for sufficient time, i.e., after enough laser pulses or periods, one can construct a histogram of photon arrival times at each pixel. The shape of this histogram has a rapid rise, followed by a

faster or slower decay which is characteristic of the fluorescent molecule(s) present in the pixel. To model this decay data, an exponential decay model can be fitted or alternatively one can make use of the fit-free phasor approach[30,31]. We used this second approach because it requires no a priori knowledge of an underlying model (i.e. number of fluorescent species at the pixel) and it is computationally inexpensive in virtue of the Fast Fourier Transform algorithm. The phasor transform extracts two values from the decay curve that characterize the shape (and importantly not the size, so that the transform is independent of the amount of photons) and these two values, namely S and G, correspond to the two coordinates of the pixel on the phasor plot (see equations in Supplementary Note 1). The values are obtained by an integral of the product of the decay of the two trigonometric functions, sine and cosine, fit in the stimulation period, and they correspond to the first-order terms of the Fourier Series decomposition of the decay curve.

Similarly, if one uses a spectral detector, i.e., a separate detector for different spectral bands, then for each pixel, one can obtain another histogram, in this case with the number of photons arriving in each channel, i.e., at each wavelength. This curve can also be transformed to an analogous spectral phasor space to map the recorded spectra at each pixel onto the 2D spectral phasor space[32,33]. Combining the lifetime measurement with a spectral detector, one effectively has a 5-dimensional space in which to characterize each pixel. On top of the spatio-temporal coordinates (x,y,z,t), each pixel now carries information in five additional coordinates: its intensity value (however many photons arrived at that pixel), the two phasor coordinates for the lifetime phasor transform, and the two phasor coordinates for the spectral phasor transform[34]. A typical image, on

the order of 10^6 pixels, obtained with this method provides 10^6 points in this 5D space[34]. If the sample presents different populations of fluorescent molecules at different locations, the pixel phasor data at these different locations map to different positions in this phasor space and a clustering technique can be used to resolve each population[35].

There is a direct analogy between the phasor transform in spectral and lifetime fluorescence microscopy (Fig. 2.2). As an example, in this figure, we use a hypothetical experiment where transcripts from 4 different target genes are targeted with 4 fluorescent species. Of the 4 species, we construct the example so that two fluorescent species emit in one color and the other two in another color. At the same time, within each color, one has a short lifetime and the other has a long lifetime. This hypothetical sample is excited, and the individual photons are detected at each pixel (Fig. 2.2a). In each pixel, we accumulate enough photons to build a spectral histogram and a lifetime histogram (Fig. 2.2b). These curves are phasor-transformed to reveal two distinct populations in the phasor space, corresponding to the two colors and the two lifetimes. By means of our previously published automatic clustering using machine learning[35], we identify these populations and return to the image space to label each pixel depending on the group it belongs to in the phasor space (Fig. 2.2c). By combining the spectral and lifetime information, we have automatically segmented the image into regions, i.e., identified the pixels that belong to the different species (Fig. 2.2d). Again, note that in this example in Fig. 2.2, we have chosen the probes to be the most convoluted case possible; one couple shares a similar spectrum and the other couple shares another spectrum. At the same time, one of the members of either couple share a similar lifetime and the other two members of either couple share another lifetime. This is the reason why even if there are

four distinct fluorescent probes, only two spectral populations are detected both in the spectral and lifetime phasor space, and the combinations of these two populations yield to the four distinct groups. The four probes cannot be resolved unless both the lifetime and spectral information are accessed.

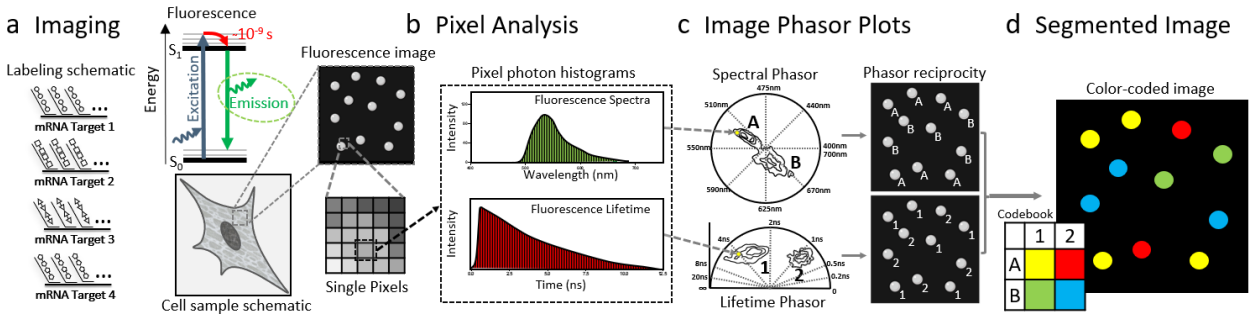


Figure 2.2 Image and phasor analysis with spectrum and lifetime analysis in MOSAICA.
a) *As an example, four different probes are used to target four different genes. The fluorescence is collected using the spectral and FLIM instrument to form images where each pixel carries information of the spectra and lifetime.* ***b)*** *At each pixel we compute the photon distribution in the spectral and temporal dimension. The phasor transform maps these distributions in each pixel to a position on the phasor space.* ***c)*** *The phasor plots reveal the presence of different populations. These populations are identified and then mapped back to the original image.* ***d)*** *We color code the pixels based on the combination of the two properties. This allows us to separate by lifetime probes that were emitting with similar spectra and vice-versa, separate by spectra probes that fluoresce with similar lifetimes.*

2.3.5 Combinatorial target spectral and lifetime encoding and decoding

In the previous section, we showed how by combining the time dimension with the spectral dimension, we can increase the number of possibilities and therefore enhance the multiplexing capabilities squaring the number of targets that can be resolved. To further increase multiplexing and improve detection efficiency, we employ combinatorial labeling, a method in which targets are labeled with two or more unique fluorophores, to greatly increase the base number of targets we can label with a given number of fluorophores/probes. To illustrate this concept, here we demonstrate a minimal exemplary working example of combinatorial labeling where two probes are used to label three targets. In this situation, each probe labels one target and the third target is labeled with

both probes simultaneously. Figure 2.3 shows a real case with such configuration, both for spectra and for lifetime. The cartoon represents the case of using two probes with distinct spectra. When imaging this sample, we can use two spectral channels, Fig. 2.3b, c, where some targets appear in only one channel, other targets appear in only the other channel and the target that is labeled with both probes appears in both channels. All targets are then detected and color-coded depending on their presence in one channel, the other or the two simultaneously (Fig. 2.3d) and the overall counts of each combination in the field of view can be provided (Fig. 2.3e).

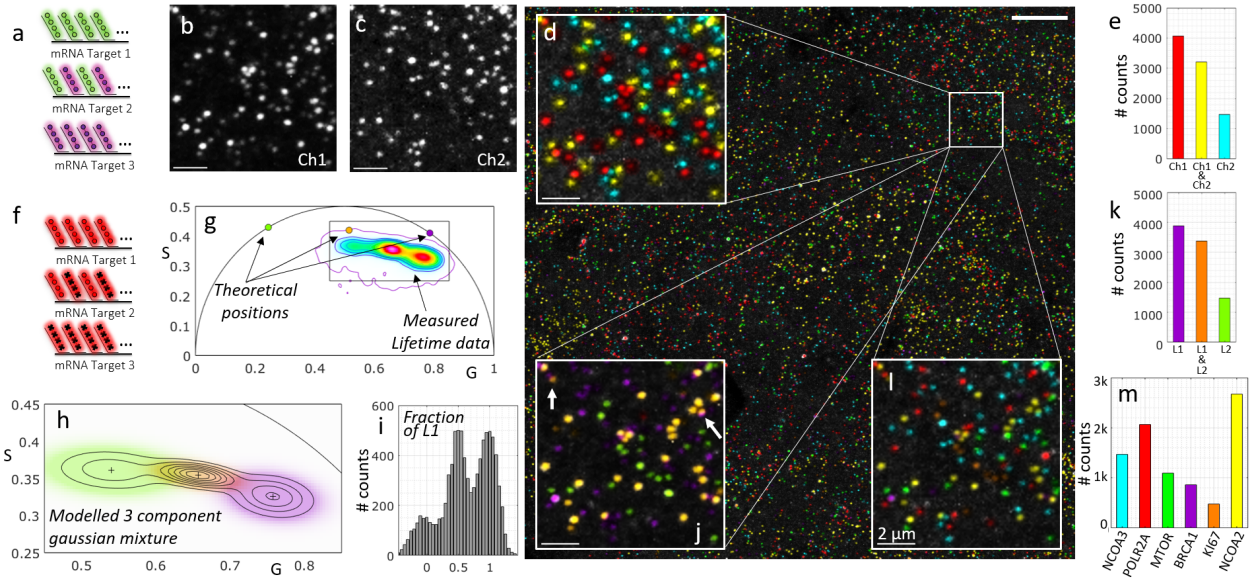


Figure 2.3 Working example of combinatorial labelling of three mRNA targets with two probes. ***a)*** *Three different target genes are tagged using two probes with different spectra. Targets 1 and 3 are tagged each with one probe, Target 2 is tagged with both simultaneously.* ***b,c)*** *The fluorescence is collected in the two expected spectral channels for the known emission of the two probes.* ***s)*** *The maximum projection of the two channels is shown and pseudocolored depending on the presence in the respective channels (as an inset within the whole field of view.* ***e)*** *The actual counts of each target within the whole field of view.* ***f)*** *As a parallel example, three different target genes are tagged using two probes with different lifetime. Targets 1 and 3 are tagged each with one probe, Target 2 is tagged with both simultaneously.* ***g)*** *The phasor plot presents three populations, corresponding to the pixels with the three combinations; the two components by themselves plus the linear combination falling in the middle.* ***h)*** *Machine learning clustering technique is used to identify the groups (Gaussian mixture model).* ***i)*** *The multicomponent method is used to extract the fraction of one of the components in each detected puncta.* ***j)*** *The same inset is shown with the pseudocoloring now depending on the lifetime clustering.* ***k)*** *The counts for each lifetime cluster in the whole field of view.* ***L)*** *The combination of the information in both the spectral and the lifetime dimension yields a final 6-plex.* ***m)*** *The overall counts for the 6-plex detection including POLR2A (Alexa647 & ATTO565), MTOR (ATTO647 & ATTO565), KI67 (Alexa647 & ATTO647), BRCA1 (Alexa647), NCOA2 (ATTO647), NCOA3 (ATTO565) with the appropriate genes that correspond to each combination.*

Experiments were conducted with cultures of mNeon green cells. Scale bar 10μm in large image and 2μm in insets.

Similarly, we show a case in which the targets are now labeled with two probes that have similar spectra but different lifetimes (Fig. 2.3f). In this case, we also introduce the use of the phasor approach to reveal the three expected populations, the pixels that contain both probes appear in the midpoint between the phasor positions of the pixels that contain only one of the probes. Figure 2.3g shows the phasor distribution obtained from the same field of view as in the spectral example, in which we also show the theoretical locations of the probes (corresponding to Alexa647 and ATTO647 with respective lifetimes of 1 ns and 3.5 ns). As is expected in real experimental conditions, there are additional fluorescent components in the sample. We broadly refer to the bulk of these additional components as autofluorescence, which pulls the data away from the expected positions and converges to the mean phasor position of the autofluorescent components. We have previously shown that the Gaussian Mixture Models is the most optimal machine learning clustering algorithm to model phasor data[35], and we use this machine learning technique to infer the phasor locations of the probe combinations (Fig. 2.3h). We can now successfully classify each pixel of the original image into one of the clusters and obtain a probability of belonging to each, i.e., the posterior probability of the model. This allows us to color code the transcripts depending on their assignment to one of the three clusters (Fig. 2.3j) and obtain the counts of the three- lifetime components (Fig. 2.3k). Additionally, we apply our lifetime multicomponent analysis technique[36] in which for each detected puncta, we estimate the presence of one of the lifetime components, in this case lifetime1 (Alexa647, purple in the figure), to obtain the expected

result; that there are clearly three populations with respective fractions centered around [0, 1⁄2, and 1] (Fig. 2.3i).

In the general case, we combine the lifetime and spectral dimensions, and we perform the clustering of the data in a 4D spectral/lifetime phasor space. The clustering technique has the power to not only identify which puncta belong to each cluster but also to assign a probability of belonging to that cluster, which can be used to quantify the certainty of the labeling. For example, in the inset in Fig. 2.3j, we show two cases of puncta that have relatively low confidence in the cluster assignment; they are depicted with blended colors because they fall in the regions of the phasor space where the two clusters are merging.

In this combinatorial example in Fig. 2.3, the three clusters in the lifetime domain multiplexed with the channel-based in the spectral domain yield a 6-plex image using only 3 probes (Fig. 2.3l, m). The specific transcripts for genes targeted for this experiment with the combined probes were POLR2A (Alexa647 & ATTO565), MTOR (ATTO647 & ATTO565), KI67 (Alexa647 & ATTO 647), BRCA1 (Alexa647), NCOA2 (ATTO647), NCOA3 (ATTO565). In the general combinatorial experiment using couples of N probes the total number of possible target genes grows quadratically:

$$\binom{N}{2} = \frac{N!}{2(N-2)!} = \frac{N^2 - N}{2} \quad (1)$$

2.3.6 Simultaneous 10-plex mRNA detection in fixed colorectal cancer SW480 cells using MOSAICA

We next applied MOSAICA to a 10-plex panel of mRNA targets in colorectal cancer SW480 cell culture samples. This cell line was chosen because its xenograft model exhibits spatial patterns of heterogeneity in WNT signaling[37], which will allow us to study

tumorigenesis in the spatial context and potentially identify cancer stem cell populations in colorectal cancer in future studies. Here, we selected this model as a validation platform to demonstrate the multiplexing scalability and error-detection capabilities of our approach. We began by first identifying a set of 10 genes with known expression levels from our bulk sequencing data. Using the aforementioned probe design pipeline, we designed 80 probes (two pairs of 40 probes) for the transcript of each gene: BRCA1, BRCA2, CENPF, CKAP5, POLR2A, KI67, MTOR, NCOA1, NCOA2, and NCOA3. These genes were chosen due to their housekeeping status or involvement in tumorigenesis in colorectal cancer. By encoding the transcript of each gene with a distinct combination of two fluorophores, we generated a codebook of 10 labelling combinations from only five fluorophores following Eq. 1: $\binom{5}{2} = 10$ (Fig. 2.4a) (see Supplementary Table 1 and Supplementary Table 2 for the fluorophores and probes, respectively, used for each target). To assess the baseline nonspecific binding events of our assay, we included a negative probe control sample, which was labelled with primary probes not targeting any specific sequence in the human genome or transcriptome but still containing readout regions for secondary fluorescent probes hybridization (Fig. 2.4a, right). Matching numbers and concentrations of primary and secondary probes that were used in the 10-plex panel were used in this sample.

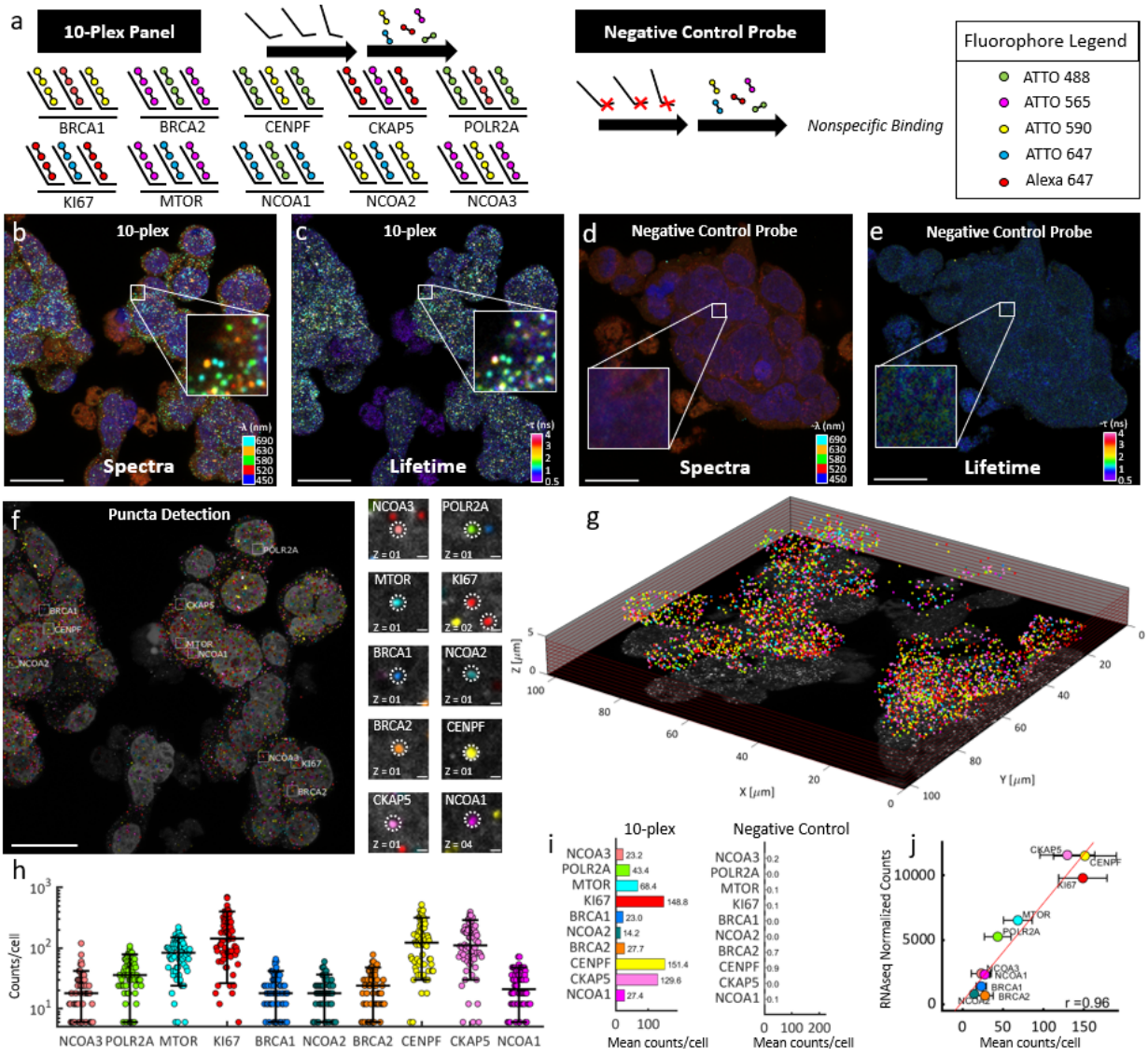


Figure 2.4 Simultaneous 10-plex detection of genes in colorectal cancer SW480 cells in a single round of labeling and imaging.

a) *10 different gene transcripts are labeled with primary probes followed by respective and complementary fluorescent secondary probes. Each gene is labeled with a combination of 2 out of 5 fluorophores for 10 combinations. Negative control probes (mNeonGreen, DCT, TYRP1, and PAX3) targeting transcripts not present in the sample were used with their respective secondary fluorophore probes.* ***b)*** *Spectral image (max-projection in z) of a field of view of the labeled 10-plex sample (5-channel pseudo coloring).* ***c)*** *Lifetime image (max-projection in z) of a field of view of the labeled 10-plex sample (phasor projection on universal circle pseudo coloring).* ***d)*** *Spectral image of the labeled negative control probe sample.* ***e)*** *Lifetime image of the labeled negative control probe sample.* ***f)*** *Final puncta detection after being processed in our analysis software showing highlighted example puncta of each target (insets, right).* ***g)*** *3D representation of the field of view for the 10-plex sample.* ***h)*** *Number of puncta detected for each gene target expression in each cell for the labeled 10-plex samples.* ***i)*** *Mean puncta counts per cell for each gene in the 10-plex samples (left, n=3 experimental replicates, 364 total cells profiled) and negative control probe samples (right, n=3 experimental replicates, 189 total cells profiled).* ***j)*** *Correlation of detected puncta (mRNA puncta count) vs. RNA-bulk sequencing (normalized counts) is shown for each target yielding a correlation (Pearson r) of 0.96. Scale bar 20μm in large images and 1μm in insets.*

An example field of view is shown in Fig. 2.4; first the spectral image overlay (five fluorescent channels including DAPI) of the labeled 10-plex SW480 sample (Fig. 2.4b) and additionally, in the same measurement, the orthogonal lifetime information attained by interrogating each pixel for their lifetime components (Fig. 2.4c). These pixels were phasor-transformed and pseudo-colored based on their projected phasor coordinates on the universal circle. In doing so, both dimensions of data can now be simultaneously accessed to determine which cluster of pixels meet the appropriate and stringent criteria for puncta classification. Similarly, the composite spectral and lifetime images of the corresponding negative control probe sample are shown (Fig. 2.4d, e). Figure 2.4f depicts the now detected pseudo-colored clusters which were successfully classified as one of the RNA markers. A representative inset image for each marker and its targeted detection is provided on the right. Because these are image stacks, the segmentation provides a 3D spatial distribution of the field of view, which can be rendered to visualize the spatial analysis in a 3D context (Fig. 2.4g).

MOSAICA employs an error-detection strategy that gates for specific and pre-encoded fluorophore combinations and rejects any fluorescent signatures which do not meet these criteria. For instance, of the total detected puncta (n = 65,562), we observed a considerable fraction of puncta, n = 25,053 (38%), which was rejected based on their fluorescence emission of only a single channel (Supplementary Fig. 2.4c). We characterize this group as the “undetermined group” because each event can belong to: 1) the nonspecific binding of probes, 2) autofluorescent moieties, or 3) mRNA transcripts, which were not fully labeled with both dyes. For the first case, as previously characterized by several groups, nonspecific binding events is a common inherent issue with single-

molecule FISH techniques which arises from the stochastic binding of DNA probes towards cellular components such as proteins, lipids, or nonspecific regions of RNA and follow a random distribution[14,20]. When combined with events which may be autofluorescence moieties (e.g., porphyrins, flavins), which can exist as isolated diffraction-limited structures and emit strong fluorescence in any particular single channel[38] or mRNA transcripts which were labeled with only one set of fluorophores, these groups represent a confounding issue for standard intensity- based measurements and analysis because they share similar SNR and intensities to real labeled puncta and cannot be differentiated without additional lengthy or complex techniques such as sample clearing or iterative-based labeling and imaging error correction[39]. Therefore, the main benefit of implementing the combinatorial encoded criteria is to ensure target detection fidelity by rejecting stochastic and nonspecific binding labeling events, as well as any event eliciting a lifetime signature that deviated from the utilized fluorophores. Finally, we also observed a relatively small group of puncta emitting fluorescent signal across more than two spectral channels but still eliciting the same spectral and lifetime signatures as the utilized fluorophores; n=2,439. To characterize this population, we performed a simulation running 20,000 iterations of various puncta densities and fitted the corresponding exponential model that characterizes the probability of puncta overlap (described in Methods section and Supplementary Fig. 2.4a,b). We attained an interval for the fraction of lost puncta due to optical crowding ranging from 2.0 to 6.6%, which accounts for the 2,439 puncta (3.7% of the total detected puncta). We name this group the overlapping in Supplementary Fig. 2.4c.

The number of puncta detected of transcripts for each gene in each cell for the labeled 10-plex samples was plotted (Fig. 2.4h) and the mean number of detected puncta per cell split into the different genes classified using MOSAICA phasor analysis with combinatorial labeling. In comparison, we also show the MOSAICA pipeline results with the negative control sample obtaining counts of less than five per thousand mainly due to noise in the images (Fig. 2.4i). To validate these puncta count, we compared them to matching RNA-seq data from the same cell type with n=3 experimental replicates (see replicate comparison in Supplementary Fig. 2.5). Shown in Fig. 2.4j is a scatter plot of the average mRNA puncta count for each cell plotted against the normalized counts from DESeq2 of our bulk RNA-sequencing data for each expressed gene. We obtained a Pearson correlation coefficient of r = 0.96, indicating a significant positive association between the two methods. Furthermore, to assess the rate of false positives and determine if one bright mRNA target could potentially be misidentified as another target, we repeated our experiment by leaving out probes for some expressed genes and then compared the detection rate of remaining targets with the 10-plex data. Specifically, we performed two additional experiments with an 8-plex, as well as two additional experiments with a 2-plex panel to compare the detected transcript abundance values and correlation coefficients against the 10-plex sample (Supplementary Fig. 2.6). We observed that there were no significant differences between these panel sizes in terms of target detection rate, indicating that target misidentification was not an issue for these panel sizes.

To further evaluate the detection efficiency, we performed benchmarking tests with our method against LGC Stellaris™ and RNAscope™ which are commercial gold

standard FISH methods (Supplementary Fig. 2.7). Using the transcript of the housekeeping gene, POLR2A, as an exemplary target, we found a significant association between the number of detected puncta by our method and LGC Stellaris™ (t test p value = 0.4). When compared to RNASCOPE™, we observed that for this cell type and target, both our assays and LGC Stellaris™ did not correlate significantly ($p = 7.8 \times 10^{-4}$ and $p = 3.4 \times 10^{-4}$), indicating a discrepancy in detection efficiency between the two methods. We attribute this difference to MOSAICA and LGC Stellaris™ utilizing a direct labeling and amplification-free method while RNAscope™ utilizes a tyramide signal amplification reaction which generates thousands of fluorophore substrate per transcript and can lead to overlapping puncta or undercounting of detected puncta. Together, these data show MOSAICA can robustly detect target mRNAs of the broad dynamic range of expression levels from single digit to hundreds of copies per cell.

2.3.7 Multiplexed mRNA analysis in clinical melanoma skin FFPE tissues

We next investigated whether MOSAICA can provide multiplexed mRNA detection and phasor-based background correction and error detection to clinically relevant and challenging sample matrices. Assaying biomarkers in situ in tissue biopsies has great clinical values in disease diagnosis, prognosis, and stratification, including in oncology[40–42]. Specifically, we applied a mRNA panel consisting of KI67 (indicative of cell proliferation), POLR2A, BRCA1, MTOR, NCOA2, and NCOA3 to highly scattering and autofluorescent human melanoma skin biopsy FFPE tissues obtained from and characterized by the UCI Dermatopathology Center. Using the same probe design

pipeline, primary probes were encoded with a combination of two fluorophores for the transcript of each gene to exhibit a unique fluorescent signature.

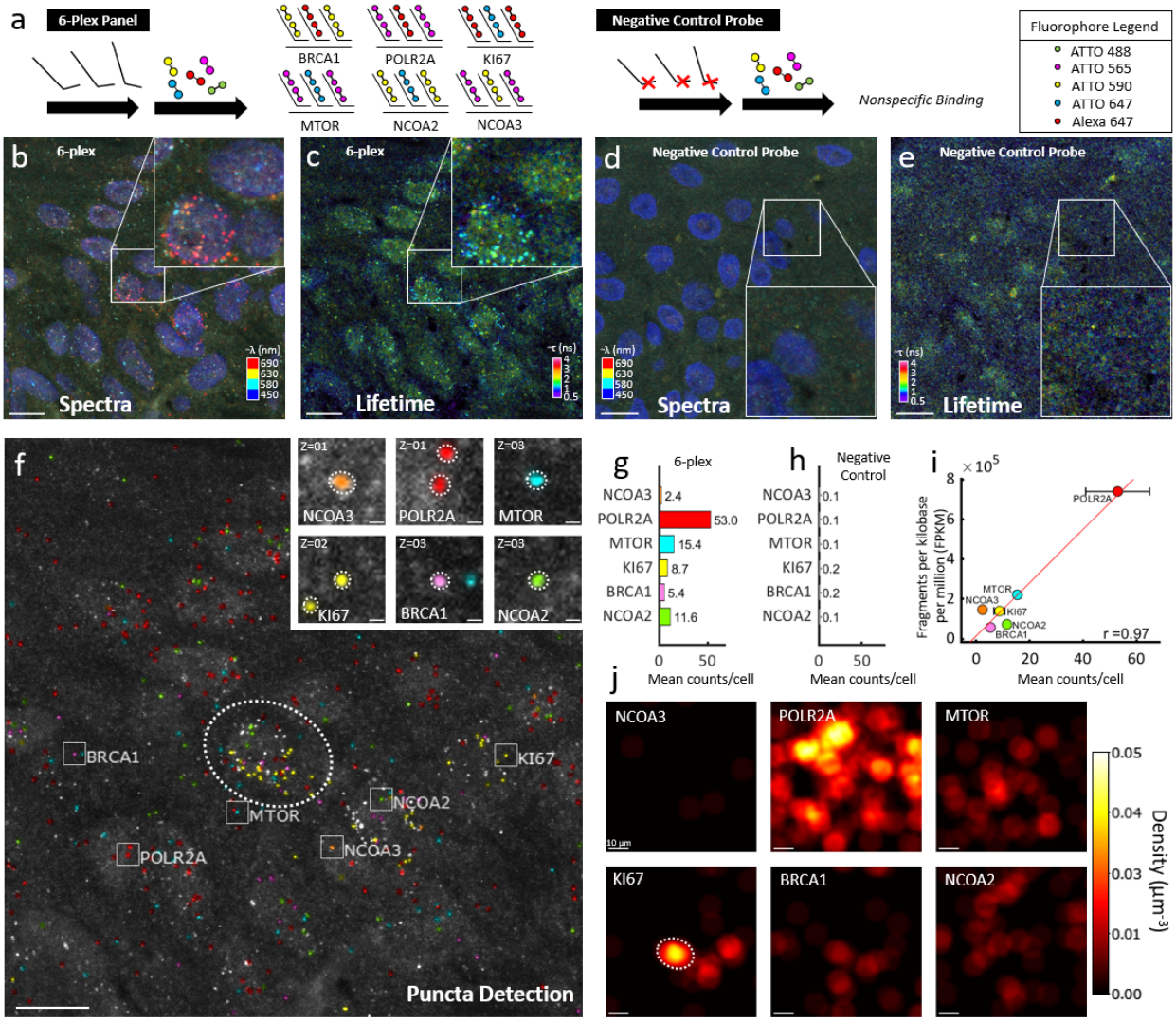


Figure 2.5 Multiplexed mRNA detection in epidermis region of human skin melanoma FFPE tissue.

a) *6 different types of gene transcripts were labeled with primary probes followed by respective and complementary fluorescent secondary probes. Each gene was labeled with a combination of 2 different fluorophores for 6 combinations. Negative control probes targeting transcripts not present in the sample were used with their respective secondary fluorophore probes.* ***b)*** *Spectral image (max-projection in z) of a field of view of the labeled 6-plex sample (3 channel pseudo coloring).* ***c)*** *Lifetime image (max-projection in z) of a field of view of the labeled 6-plex sample (phasor projection on universal circle pseudo coloring).* ***d)*** *Spectral image of the labeled negative control probe sample is depicted.* ***e)*** *Lifetime image of the labeled negative control probe sample.* ***f)*** *Final puncta detection of the 6-plex field of view after being processed in our analysis software showing highlighted example puncta of each target (insets, right).* ***g)*** *Mean puncta counts per cell of each gene in the 6-plex sample (n=2 experimental replicates, 174 cells).* ***h)*** *Puncta count for the negative control probe sample (n=2 experimental replicates, 375 cells).* ***i)*** *Correlation of detected puncta (mRNA puncta count) vs. bulk sequencing (fragments per kilobase per million) is shown for each target.* ***j)*** *Transcript density in the field of view for each of the genes reveals clustering of specific genes, as*

*an example KI67 appears highly expressed in three cells, one of them marked with a dotted ellipse that corresponds to location in **f**). Scale bars 10 μm in large images and 1 μm in insets.*

Figure 2.5b depicts a spectral image overlay (four fluorescent channels including DAPI) of the epidermis region of a labeled 6-plex skin tissue sample. Similarly, as in the previous section, the orthogonal lifetime image was attained after using phasor analysis to create the image depicted in Fig. 2.5c–e depicts the merged composite spectral and lifetime images of the corresponding negative probe sample also in the epidermis region. Figure 2.5f depicts the pseudo-colored puncta which were successfully classified and identified as their assigned mRNA markers. A representative inset image for each marker and its targeted detection is provided on the right. We observed that a population of puncta consisting of nonspecific, autofluorescent, or unknown sample artifacts rejected from analysis, (1,100) or 37.5% of the total detected puncta (2,934). In addition to this group, MOSAICA rejected a small group of puncta that emitted fluorescence in multiple spectral channels (62). This fraction (2.1%) is in concordance with the optical crowding range (2.0–6.6%) that our simulations and models predict (Supplementary Fig. 4). With conventional intensity-based measurements and analysis, both contaminating groups are inherent image artifacts that compromise the integrity of puncta detection unless complicated quenching steps or additional rounds of stripping, hybridization, and imaging are utilized[14,43]. With MOSAICA, these contaminating artifacts can be accounted for with the integration of spectral, lifetime, and shape- fitting algorithms.

Figure 2.5g, h plots the total number of detected puncta for the labeled 6-plex sample and the negative control probe sample to highlight the final counts obtained using MOSAICA. To validate these puncta counts and their relative expressions, we examined the relationship between the decodified puncta with matching bulk RNA-sequencing

obtained from The Cancer Genome Atlas (TCGA) database (see Methods section). Shown in Fig. 2.5i is a scatter plot of MOSAICA puncta count plotted against fragments per kilobase per million. We obtained a Pearson correlation of r=0.97 for this 6-plex sample, indicating a significant positive association between the two methods. We acknowledge that this strong correlation is particularly dependent on the presence of the highly abundant POLR2A expressed gene. The correlation for the other lower expressed targets excluding POLR2A is r=0.44 which, although still positive, is weaker. We attribute this discrepancy to preanalytical variables typically associated with FFPE sample preservation and pretreatment. For instance, there have been multiple studies, which documented increased variability in quantifying lowly expressed genes in FFPE tissues due to RNA degradation or cross-linking of proteins with nucleic acids[44–46]. Last, the density map of the detected transcripts provides a visual method to identify spatial localization of clusters of genes, such as KI67 (indicative of proliferating tumor cells) being more prevalent in the dermis region while POLR2A is dispersed throughout the region (Fig. 2.5j). Overall, in situ profiling biomarkers, such as KI67 and their spatial clustering can have diagnostic and prognostic values in malignant diseases and MOSAICA provides a robust platform to profile these markers[47].

2.3.8 Simultaneous co-detection of protein and mRNA

Spatial multiomics analysis including especially simultaneous detection of protein and transcript within the same sample can reveal the genotypic and phenotypic heterogeneity and provide enriched information for biology and disease diagnosis. As a pilot experiment to demonstrate MOSAICA's potential for multiomics profiling, we utilized MOSAICA to detect 2 protein targets, Tubulin and Vimentin, and 2 mRNA targets,

POLR2A and MTOR in colorectal cancer SW480 cell culture samples (Fig. 2.6). After staining the sample with the primary antibodies, secondary antibodies were added to fluorescently label the protein targets. After protein labeling, we utilized the same probe design pipeline and labeling strategy for mRNA detection, primary probes were generated and hybridized to the sample after antibody staining. Corresponding secondary probes were hybridized. Figure 2.6a–f depict the individual channels of the sample with Fig. 2.6g showing the merged channels of the 4-plex panel. As both POLR2A and MTOR are assigned to the 647 nm channel and cannot be separated spectrally (Fig. 2.6d), lifetime analysis is used to separate POLR2A (Fig. 2.6e) and MTOR (Fig. 2.6f). Signal-to-noise ratio measured as intensity of the detected puncta over intensity of the surrounding pixels was measured for the two mRNA targets (Fig. 2.6h). In summary, we have demonstrated MOSAICA as a potential spatial multiomics tool, which harmonizes sample treatment between both labeling processes. MOSAICA utilizes staining protocols with efficient target retrieval, blocking, and pretreatment steps where the viability and labeling of both target RNA sequence and protein markers were not compromised after each assay.

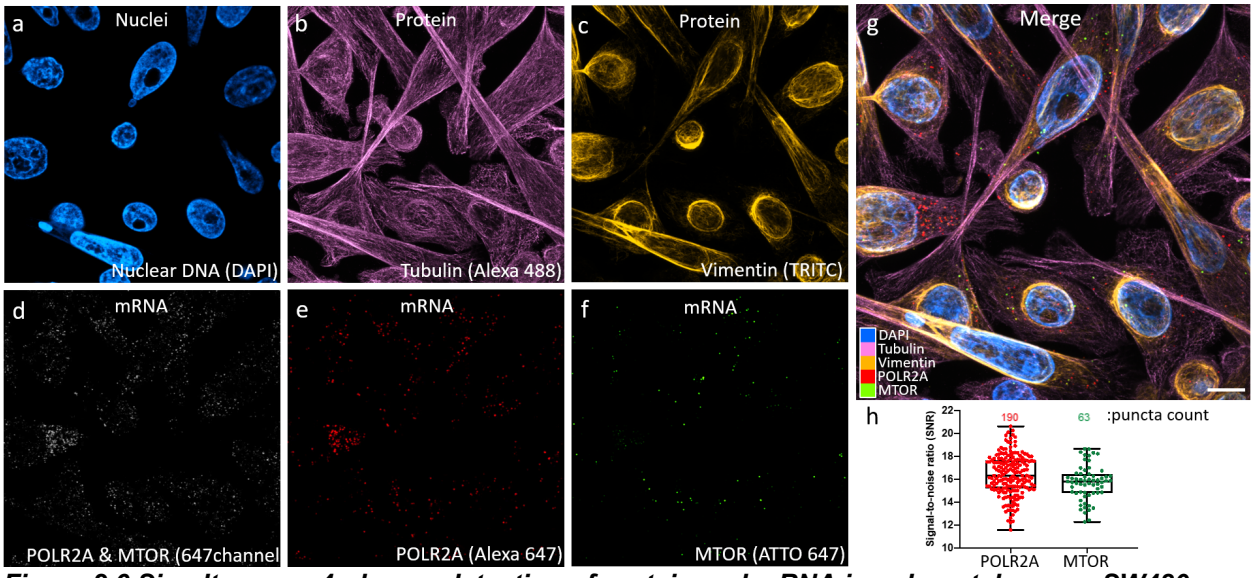


Figure 2.6 Simultaneous 4-plex co-detection of protein and mRNA in colorectal cancer SW480 cells

a) *Intensity imaging showing nuclei labeled with DAPI.* ***b)*** *Intensity image showing Tubulin protein labeled with Alexa488.* ***c)*** *Intensity image showing Vimentin protein labeled with TRITC.* ***d)*** *Intensity image at 647nm showing mRNA targets, POLR2A and MTOR, which were further resolved by lifetime.* ***e)*** *Unmixed lifetime image showing POLR2A puncta labeled with Alexa647.* ***f)*** *Unmixed lifetime image showing mTOR puncta labeled with ATTO647.* ***g)*** *Merged image of all channels. Scale bar is 10 µm.* ***h)*** *Signal-to noise and puncta count analysis for the mRNA targets.*

2.4 Discussion

MOSAICA can fill a gap in the spatialomics field by offering both simplicity and multiplexing through direct in situ spatial analysis of a large number of biomarkers in a single round of staining and imaging (Supplementary Fig. 2.8). By contrast, conventional direct labeling approaches (e.g., RNAscope™, LGC Stellaris™, etc.) are limited to 3 or 4 targets. Other emerging spatial transcriptomics technologies such as seqFISH can offer greater multiplexing capabilities but requires many rounds of sample re-labeling, imaging, indexing, and error-prone image registration. The MOSAICA integrates both the spectral and lifetime dimensions and employs combinatorial target encoding, and phasor- and machine learning-based deconvolution to achieve high-plex analysis without sacrificing assay throughput. MOSAICA's error-detection feature can correct for stochastic nonspecific binding artifacts and autofluorescent moieties, inherent challenges associated with current intensity-based methods. MOSAICA's simple workflow can be particularly important in clinical settings where biopsy samples are limited in quantity and often not amenable for repeated processing. With respect to cost, MOSAICA uses inexpensive DNA primary probes and fluorescently labeled secondary probes which can be shared among many different targets, reducing costs to several dollars per assay. Particularly compared to indirect spatial transcriptomic technologies that interrogate barcoded regions of interest (ROIs) with separate sequencing steps (e.g., 10x Genomics Visium, GeoMx® Digital Spatial Profiler), our direct, imaging-based approach can provide

higher spatial resolution (single molecules or subcellular features), lower cost, simpler workflow, and potentially higher throughput (number of samples analyzed per time unit). Furthermore, our platform uses standard fluorescent probes and fluorescence microscopy and fluorescence imaging remains the most widely used technique in biological research. Several commercial instruments that can acquire both spectral and lifetime information are available including ISS FastFLIM, PicoQuant rapidFLIM, Leica SP8 FALCON, etc., and already exist in numerous shared facilities in industry and academia. Therefore, its minimal requirements of MOSAICA will permit quick and broad adoption in the scientific community.

MOSAICA holds great potential to broadly enable scientists and clinicians to better elucidate biological processes, and to develop precision diagnostics and therapies. Given that gene expression is heterogeneous and many different cell states can exist, one would need to assess multiple expressed genes within the same cell in situ. Therefore, it is anticipated that MOSAICA can enable spatiotemporal mapping in the attempt to construct 3D tissue cell atlas. In addition, MOSAICA can serve as a tool for targeted in situ validation of single-cell RNA sequencing data which reveal cell identities based on “differentially expressed genes” but are subjective, variable and error-prone. Furthermore, we aim to develop MOSAICA as a clinical companion diagnostic tool for stratified care. In particular, insights of the spatial organization and interactions between tumor cells, immune cells, and stromal components in tumor tissues can inform cancer diagnosis, prognosis, and patient stratification[48,49].

One common challenge in imaging-based spatialomics analysis is optical crowding which can limit both the number of molecules that can be detected and the detection

efficiency and accuracy. For instance, as we scale up multiplexing capabilities by labeling more mRNA and proteins with additional fluorophores, more labeled targets and their fluorescent combinations will begin occupying the same voxel, leading to challenges in determining both how many targets there are as well as which type of targets are present within each voxel. We have modeled this phenomenon in Eq. 2 (Methods section) and plotted the results in Supplementary Fig. 2.4. Based on our estimates and in our current transcript density conditions, overlapping accounts for only around 6% of the detected puncta. We currently do not further resolve these cases and, instead, categorize them into the over- lapping group, which do not contribute to total counts. In addition, we intend to further address these cases in the future using our multicomponent approach[36] to unmix spectral/lifetime components within a single voxel by means of higher harmonics of the phasor transform.

With respect to the crowding issue, the phasor analysis method has an additional limitation related to the use of the combinatorial technique. Even if the isolated fluorescent dyes are very far apart on the phasor space, their combinations fall inside the polygon as determined by the positions of the individual dyes. As one increases the number of dyes, the combinations start to overlap creating an ambiguity. For this reason, as we increase our multiplexed panel, our strategy is not only to employ labels which are distinctly separated by both spectral and lifetime properties but importantly to also use more combinations of different labels rather the same labels. The tradeoff between these two counteracting parameters is an exciting endeavor which we look forward to exploring as we progressively build up our repertoire of fluorophores.

Regarding the phasor population overlap, given the imaging settings we have used for the experiments in this paper, the signal-to-noise ratio produces gaussian phasor distributions with 99.7% of the pixels within 0.01 phasor units (6σ). Although the distributions tend to converge due to background autofluorescence, their small covariance matrices guarantee a high level of confidence in assigning each pixel to the correct cluster. As an example, the three gaussian distributions in Fig. 2.3h, have covariance matrices of (coordinate S first, G second) $\Sigma_1 = \begin{pmatrix} 12 & -5 \\ -5 & 89 \end{pmatrix} \times 10^{-5}$, $\Sigma_2 = \begin{pmatrix} 6 & -8 \\ -8 & 51 \end{pmatrix} \times 10^{-5}$ and $\Sigma_3 = \begin{pmatrix} 12 & -8 \\ -8 & 50 \end{pmatrix} \times 10^{-5}$ and mean coordinates of $\mu_1 = \begin{pmatrix} 0.26 \\ 0.40 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} 0.25 \\ 0.49 \end{pmatrix}$ and $\mu_3 = \begin{pmatrix} 0.23 \\ 0.57 \end{pmatrix}$. With these values, the distance between the leftmost and right most cluster is of 0.17 phasor units, with the mean standard deviation from the covariance matrices being 400 times smaller at 3.65×10^{-4}. With numbers like these, we anticipate that our clustering technique can easily resolve even more challenging scenarios such as 6 lifetime phasor clusters and 10 spectral phasor clusters. As a result, our next immediate goal is to scale our multiplexing capability by detecting around 60 mRNA targets simultaneously with 12 different fluorophore species within the same sample. We aim to use seven spectrally distinct fluorophores and an additional five with overlapping spectra but are resolvable by lifetime. A combinatorial scheme of 12 choose two would yield 66 combinations. We could resolve these combinations using a seven-spectral channel instrument where five of the channels would present three populations in the lifetime phasor plot (one for each of the two probes with overlapping spectra in that channel plus the third being the combination of the two). Looking another step ahead, by implementing our recently developed 32-channel spectral-FLIM detector[34] which can

provide 32 independent spectral sources with six lifetime clusters per channel, 192 different fluorophore species can be accessed to provide significantly higher plex detection capabilities.

In the future, we consider expanding our codebook by implementing a Förster resonance energy transfer (FRET)-based barcoding strategy[50] where different FRET fluorophore pairs and various distances between fluorophore donor and acceptor can be used to tune the combinatorial spectrum and lifetime readout. The FRET phenomena can also be used as an additional error-correction mechanism at the nanometer scale to potentially resolve multiple targets in the same voxel. Moreover, the current scanning confocal microscope implemented in MOSAICA can achieve high spatial resolution and z-sectioning but is limited by a relatively longer imaging time. As an example, each z-slice of our MOSAICA images (Fig. 2.4 and Fig. 2.5) took around 1.5 min (1024 × 1024 images at 16 µs pixel dwell time, accumulating an average of 6 frames). However, we anticipate that this approach is compatible with any wide-field imaging technique as long as sufficient image pixel sizes, axial resolution, and photon counts are met. This can be accomplished with our recently developed camera-based light sheet imaging system[51] or a spinning disc confocal system equipped with a FLIM camera to greatly improve imaging throughput[52]. Indeed, MOSAICA is amenable for further integration with other imaging modalities, including expansion microscope, super-resolution, and multiphoton imaging[53–55] to improve subcellular resolution and allow imaging large, scattering tissues. In addition, we will develop user-friendly image analysis software with capabilities enabling classification of single-cell phenotypes, spatial organization and neighborhood relationship among different cell types. Our puncta detection and classifier algorithm can

be improved using convolutional neural networks with clinical training sets to optimize biomarker detection accuracy and efficiency. Finally, we will develop high-plex protein detection component in our multiomics analysis using antibody- DNA conjugates where our combinatorial labeling and barcoding strategy can be used to scale up multiplexing.

2.5 Methods

Primary probe design

A set of primary probes were designed for the transcript of each gene. A python code was used to rapidly design the primary probes while controlling various aspects of the probes such as GC content, length, spacing, melting temp, and prohibited sequences. To begin, probes are designed using exons within the coding sequence region. However, if that region does not provide over 40 probes, the exons from the coding and untranslated regions are used. The candidate probes are then aligned to the genome using Bowtie2, an NGS aligner, to determine if these probes are specific. Probes that are determined specific are then aligned to the RNA sequencing data on the UCSC Genome Browser, further eliminating probes that do not align to regions with an adequate number of reads. While mapping the probes to the genome on the UCSC Genome Browser, the probes are aligned with BLAT (BLAST-like alignment tool). A local BLAST query was run on the probes for the expressed genes in the panel to eliminate off-target hits. For this experiment, each expressed gene had the maximum number of probes that could be designed with our pipeline and requirements. The final primary probe design included two assigned readout sequences of the secondary probe with a “TTT” connector in between, another connector, then one of the probes specific for the transcript of that gene. The primary probes were ordered from Sigma Aldrich and pooled together for the transcript of

each gene. The sequences of all probes used in this study are listed in Supplementary Data 1.

Secondary probe design

Secondary probe structures were based on the design from the Zhuang group[56]. In short, the 20-nt, three-letter readout sequences were designed by generating a random set of sequences with the per-base probability of 25% for A, 25% for T, and 50% for G. Sequences generated in this fashion can vary in their nucleotide content. To eliminate outlier sequences, only sequences with a GC content between 40% and 50% were kept. In addition, sequences with internal stretches of G longer than 3 nucleotides were removed to eliminate the presence of G-quadruplets, which can form secondary structures that inhibit synthesis and binding. To remove the possibility of significant cross-binding between these readout sequences, algorithms from previous reports were used to identify a subset of these sequences with no cross-homology regions longer than 11 contiguous bases[56]. Probes were then checked with BLAST to identify and eliminate sequences with contiguous homology regions longer than 11 nucleotides to the human transcriptome. From the readout sequences satisfying the above requirements, 16 were selected.

Cell culture

Human embryonic kidney (HEK293T) cells (632180; Takara) were cultured in DMEM (10-013-CV; Corning) supplemented with 10% FBS (1500-500; Seradigm) and 1% penicillin (25–512; GenClone). Human colorectal adenocarci- noma (SW480) cells were cultured in DMEM with high glucose (SH30081.02; HyClone) supplemented with 10% FBS (1500-500; Seradigm), 1x L-Glutamine (25–509; GenClone), and 1% penicillin

(25–512; GenClone). SW480 cells were FACS-sorted based on surface marker ROBO-1, and ROBO + and ROBO- cells were used in Fig. 2.4 and Fig. 2.6, respectively. The cells were plated into 8-well chambers and then fixed. The eight-well plates (155409; Thermo Scientific) for HEK293-T and SW480 cells were coated with fibronectin bovine plasma (F1141- 2MG; Sigma Aldrich) before seeding cells onto the 8-well plates. All cultures were grown at 37 °C with 5% CO_2.

mNeonGreen cell engineering

A mNeonGreen construct was transfected into HEK293T-X cells with FuGENE HD Transfection Reagent (E2311; Promega). The cells were then selected with puromycin (NC9138068; Invivogen) and Zeocin (AAJ67140XF; Alfa Aesar) 3 days after transfection. Preparation of fixed cells in cell chambers. When the cells reached 70% con- fluency, cells were fixed for 30 min using 4% paraformaldehyde (15710; Electron Microscopy Science), then washed with PBS 3 times. The cells were then incubated with sodium borohydride (102894; MP Biomedicals) for 5 min and washed with PBS 3 times. 0.5% Triton X-100 (T8787-100ML; Sigma-Aldrich) in PBS was incubated in each well for 5 min and cells were washed with 2x SSCT (2x SSC with 0.1% TWEEN® 20 (P9416-100ML; Sigma-Aldrich). For storage, cells were left in 70% ethanol at 4 °C.

Preparation of FFPE tissues

The University of California, Irvine Institutional Review Board (IRB) approved this study for IRB exemption under protocol number HS# 2019–5054. All human melanoma cases were de-identified samples to the research team at all points and therefore considered exempt for participation consent by the IRB. Fully characterized human patient skin melanoma FFPE tissues with an immune cell score of brisk were obtained

from the UCI dermatopathology center then sectioned to 5 μm slices using a rotary microtome, collected in a water bath at 35 °C, and mounted to positively charged Fisher super frost coated slides. The tissue sections were then baked at 60 °C for 1 h. For antigen unmasking, slides were deparaffinized, rehydrated then followed by target retrieval (with citrate buffer).

Primary probe hybridization

Blocking buffer containing 100 mg/ml Dextran sulfate sodium salt (D8906-100G; Sigma-Aldrich), 1 mg/ml Deoxyribonucleic acid from herring sperm (D3159-100G; Sigma-Aldrich), 0.01% Sodium Azide (S2002- 100G; Sigma-Aldrich), 0.01% tween, and 15% ethylene carbonate (AC118410010; Fisher Scientific) in 2x sodium saline citrate (SSC) was added to the fixed cells or tissues and incubated at 60 °C for 8 min and then at 37 °C for 7 min. Following this preblock step, primary probes with 5 nM of each probe in blocking buffer were added to the samples and incubated at 60 °C for 30 min and then overnight at 37 °C.

Secondary probe hybridization

Once the primary probe solution is removed, the sample is washed with 2x Saline-Sodium Citrate Tween (SSCT) twice. Wash buffer (2xSSCT with 10% ethylene carbonate) is used for 3 washes and incubated in 60 °C for 5 min each time. Blocking buffer is added and incubated at room temperature for 5 min. The sample is then incubated in a solution with 5 nM of the secondary probes in blocking buffer at 37 °C for an hour. The sample is washed with 2x SSCT twice before using wash buffer to wash 3 times and incubated in 42 °C for 5 min each time. For the first wash, 10 mg/mL Hoechst

(H3570; Invitrogen) is diluted 1:1000 in PBS and added to cells. Later, the wash buffer is then removed and replaced with glycerol mounting media and ready for imaging.

Codetection of protein and mRNA

Prior to mRNA labeling, fixed SW480 cells were blocked with 1% Bovine Serum Albumin (RLBSA50; VWR), 0.1% TWEEN® 20, 1:1,000 Sodium Azide, 0.2 U/ml Protector RNase inhibitor (3335399001; Sigma-Aldrich), and 1 mM DTT in RNAse-free PBS (AM9625; Life Technologies) for 30 min at room temperature. These cells were then washed 3 times with 0.1% TWEEN® 20 in RNAse-free PBS for 5 min each wash at room temperature. Antibody solutions containing 1:1,000 Mouse anti-Tubulin (3873BF; Cell Signaling) and 1:200 Rabbit anti-Vimentin (5741BF; Cell Signaling) in the same blocking buffer were subsequently added to the samples and incubated overnight at 4 °C. Following 3 additional washes with 0.1% TWEEN® 20 in RNAse-free PBS for 5 min each at room temperature, antibody solutions containing fluorescently labeled 1:200 Donkey anti-Mouse Alexa-488 (R37114; Fisher Scientific) and 1:200 Donkey anti-Rabbit TRITC (711-025-152; Jackson Laboratories) in the same blocking buffer were added at room temperature for 1 h. After 3 washes with RNAse-free PBS with 0.1% TWEEN® 20 for 10 min each wash at room temperature, 4% PFA in PBS was added for 15 min at room temperature. These cells were then washed 3 times with 0.1% TWEEN® 20 in PBS at room temperature for 5 min. For mRNA labeling, the previously described methods regarding primary and secondary probe hybridization were utilized.

LGC Stellaris™

LGC Stellaris™ RNA FISH probes (Biosearch Technologies, CA, USA) were used, with 48 × 20 mer fluorophore-conjugated oligos tiling the length of the target transcript.

The POLR2A probe set were supplied as predesigned controls conjugated to Quasar 570 fluorophores. Labeling/staining was carried out as described in the LGC Stellaris™ protocol for adherent mammalian cells. The POLR2A probe sets were used at 50 nM.

RNAscope™

The FFPE tissue sections were deparaffinized before endogenous peroxidase activity was quenched with hydrogen peroxide. Target retrieval was then performed, followed by protease plus treatment. The fixed cells pretreatment included treatment with hydrogen peroxide and protease 3. The RNAscope™ assay was then performed using the RNAscope™ Multiplex Fluorescent V2 kit and Akoya Cy5 TSA fluorophore. The positive control (POLR2A) and negative control (dapB) were in C1.

Microscopy Imaging

Our samples can be imaged with any instrument provided that it has spectral and lifetime acquisition capabilities. Our measurements were taken on three separate instruments, a wide-field and two confocal microscopes. A generic MOSAICA scanning confocal instrument setup is depicted in Supplementary Fig. 9.

For validation of fluorophores and their spectral and lifetime signatures, measurements were taken on a 2-channel ISS Alba5 STED platform. This system is equipped with a pulsed white laser (NKT SuperK EXTREME) system where the excitation wavelength(s) can be selected with an acousto-optic tunable filter (NKT SuperK SELECT). Single photons were detected with avalanche photodiode detectors (Excelitas Technologies) and their arrival times with respect to the stimulating frequency (78 MHz) were measured with a FPGA-based electronic board (ISS FastFLIM). Imaging was

achieved by fast beam scanning with galvo mirrors and 3D stacks of images were acquired with a z-piezo mount on the objective.

For measurements of multiplexed/combinatorial labeling and detection experiments (Fig. 2.4 and Fig. 2.5), we utilized a Leica SP8 with the Falcon module. This platform employs a white light laser and an acoustic optic beam splitter dichroic, and the Leica hybrid detectors with excitation band selectable by means of a prism. 3D measurements of cells and tissue samples were taken with a 100x plan apochromat oil objective with a numerical aperture of 1.40, yielding images with an x-y resolution of 100 nm and z-spacing of 500 nm.

For epifluorescence measurements (Supplementary Fig. 2 and Supplementary Fig. 3), images of labeled mRNA transcripts were taken on an inverted Ti-E using a 100× plan apochromat oil objective with a numerical aperture of 1.40. Samples were illuminated with a Spectra-X (Lumencor) LED light source at the 395 nm, 470 nm, 555 nm and/or 640 nm excitation wavelengths. Images were acquired with an Andor Zyla 4.2 sCMOS camera at 4 K resolution with 6.5 μm pixels.

Image Processing

A custom set of scripts were written in MATLAB to process the acquired image stacks, identify individual transcripts and assign each of them to each gene expression target. After reconstructing the images out of the digital list of photons, the analysis runs in parallel a 3D blob detection pipeline on the intensity image stacks to identify each transcript and on the other a clustering pipeline on the phasor-transformed lifetime/spectral phasor data to detect distinct spectral/ lifetime populations. A classifier

then assigns pixels as belonging to a particular expressed gene. The whole pipeline is depicted in Supplementary Fig. 10.

Briefly, the intensity 3D stacks are run through a blob detection algorithm that was developed in order to identify each transcript. The images can be seen as a 3D space where the transcripts appear as spherically symmetric locations with a radial increase in intensity, namely puncta. The algorithm first computes the low- frequency background noise by means of a median filter with a kernel 10 times the size of the expected puncta (the diffraction limit of the instrument, in our case around 250 nm). This low-frequency background is subtracted from the high-pass filtered data obtained by convolving by a gaussian filter of the expected size of the puncta. This on one hand enhances the puncta in the image by giving a prominence value at each pixel with respect to the surrounding regions and on the other suppresses noise in the images. A search for local maxima is performed by finding the locations where the gradient goes to zero and the divergence of the gradient is negative. Once the centers in the 3D coordinate space are obtained the size, absolute brightness and prominence of each puncta is measured.

In parallel, the raw photon counts are used to construct the photon arrival time histogram and photon spectral histogram at each pixel. Phasor transforms are applied to each pixel in each image of the 3D stack in order to construct the stacks' phasor plot. This phasor data is in general a 4-dimensional, each pixel in the intensity image has four additional coordinates; two for the spectral phasor transform plus two for the lifetime phasor transform. The phasor coordinates are clustered using Gaussian Mixture Models[57]. We used an initial experiment tagging the transcripts of housekeeping genes in order to guarantee that all expected populations were present and we trained the

Gaussian Mixture Model using this initial experiment. This pretrained model is then applied to the new sets of data in order to classify each pixel into one of the clusters allowing for the presence of empty clusters. The number of clusters N intuitively should be the number of distinct fluorescent probes or different combinations of probes used to tag the sample, but one must allow for additional populations in the sample, e.g., autofluorescent species. For this reason, in the training of the Gaussian Mixture Models we allowed for one additional cluster to account for autofluorescence and noise.

Finally, by computing the mean phasor coordinates of the pixels within each detected puncta, we can compute the phasor position of each puncta and assign a gene expression label to it by a priori knowing the expected positions of each combination of probes depending on the spectra and lifetime of the probes.

To obtain the number of counts per cell, DAPI image stacks are segmented by means of simple thresholding, estimating the threshold value by hard-splitting of the histogram of photon counts in the channel. The 3D segmented nuclei are then iteratively grown by convolution by a minimal 3x3x3 kernel. This convolution is applied at each pixel of the edge of the segmented volume until no available space is left between the segmented volumes. This yields a division of the imaged volume into polyhedra where each face is exactly the plane bisecting the two closest nuclei edges. This process is analogous to a Voronoi tessellation using the surface of the nuclei instead of points.

In the cell culture experiments, to normalize by cell volume, we used a normalized mean cell volume of 3000 μm^3 since a cell marker was not utilized and the imaged volume thickness (5 μm) was less than the actual cell thickness. To obtain mean counts per cell, total detected puncta counts was divided into the total imaged cellular volume and then

multiplied by an estimated mean cell volume of 3000 μm^3. The total cellular volume was obtained by an intensity threshold segmentation of the background cellular autofluorescence over the gaps between cells.

Simulations

In order to test the detection and classification pipeline, we wrote a set of scripts to simulate spectral/lifetime data which provided a ground truth towards detection and accurate classification debugging. This data generation script allows randomly distributing N diffraction-limited transcripts in an arbitrarily big 3-dimensional space, each with a gaussian intensity profile. We simulated our transcript gaussian profile with a X-Y standard deviation of 200 nm and a Z standard deviation of 500 nm, a peak intensity of 1 ± 0.3 (the intensity becomes relevant when simulating background noise). In the simulation run that we used to test the crowding limitations of the system we simulated tagging the transcripts of genes with couples out of a total of 12 fluorescent probes; 4 distinct spectral probes and 3 distinct lifetimes in each, yielding a total of $\binom{12}{2} = 66$ possible expressed genes.

We generated the simulated images in a cubic space of 10 × 10 × 10 μm, discretized as an image stack of 33 images of 1000 × 1000 pixels (yielding a voxel resolution of 100 × 100 × 300 nm). This volume was generated containing increasing densities of transcripts, ranging from a single transcript of each gene (66 transcripts) up to 2000 transcripts of each gene (132k transcripts) and for each possible value of density a total of 10 iterations each time. These 20k simulated image stack sets were individually processed by our image processing pipeline and the transcript position and labelling obtained by the pipeline was compared to the known ground truth of the generated data.

This simulation provided a benchmark of the density limitations of the method but at the same time giving an idea of the underestimation of the number of transcripts as a function of local density. The simulations allowed us to model the estimated number of overlapping transcripts as a function of density.

A similar set of simulations was run by emulating the conditions in the 10-plex experiment (Fig. 2.4) where the transcripts of genes are tagged with combinations of two out of five probes. The 20k iterations for different densities allowed to plot the density of the classification obtained after detection compared to the real number of transcripts in the simulations. This simulation was fit to the probabilistic model obtained from calculating the number of transcripts that are not overlapped in space (see next section), from which the true number of puncta was extracted (see Supplementary Fig. 2.4).

Overlapping probability

The fraction of puncta that do not overlap with any other puncta depends on the total number of puncta present in the volume of study and the relative proportion between said total volume and the volume of each individual puncta. The following expression is obtained as the product of N-1 times the fraction of available space having removed the volume occupied by one transcript:

$$\frac{n}{N} = \left(1 - \frac{v_i}{V_T}\right)^{N-1} \quad (2)$$

where n is the number of isolated puncta, N is the total number of puncta, v_i is the volume of each puncta and V_T is the total volume (simulated or scanned). The real number of transcripts N cannot be analytically isolated from the previous equation, but one can graphically obtain it. Due to the fact that the transcripts are sub- diffraction limit, the value of v_i is simply the volume of the point spread function of the instrument. Using the detected

number of counts in an experiment n=13.5k and the estimated total imaged cellular volume of 68 kμm^3, both obtained from two image stacks shown in Fig. 2.4, we proceeded to estimate the real number of transcripts present in the sample using the previous expression. Assuming an interval of possible volumes for the transcripts (instrument PSF) between 0.1 and 0.3 μm^3 we obtained an estimated percentage of overlapping puncta in the interval [2.0, 6.6]%. This range of values is in agreement with the number of puncta that we detected in more than two channels in the 10-plex experiments (3.7%) and in the tissue experiments (2.1%). See Supplementary Fig. 2.4 for additional details such as expression (2) plotted as a function of the density of transcripts.

Sequencing Data

Colorectal cancer SW480 cell bulk RNA sequencing (unpublished data) was analyzed with DESeq2. Average expression is then obtained for comparison to the MOSAICA puncta count for each expressed gene. For the human skin melanoma FFPE tissue, the patient sample did not have corresponding sequencing data. RNA sequencing data were obtained from publicly available data from The Cancer Genome Atlas (TCGA), available on the National Cancer Institute (NCI) Genomic Data Commons (GDC) data portal, from 5 human skin melanoma FFPE biopsy thigh punch samples [Entity ID: TCGA-EE-A2GO-06A- 11R-A18S-07, TCGA-EE-A20C-06A-11R-A18S-07, TCGA-YG-AA3N-01A-11R- A38C-07, TCGA-DA-A95Z-06A-11R-A37K-07, TCGA-GN-A26C-01A-11R-A18T-07]. The sequencing data were analyzed with HTseq and normalized for sequencing depth and gene length using Fragments Per Kilobase Million. The average of the 5 patient samples for each transcript were used for correlation graphs with MOSAICA puncta count.

Experimental replicates and reproducibility

Figure 2.3 is a conceptual figure and a single experiment was used as an example without replicates. For the 10-plex cell culture experiments (Fig. 2.4), we ran 3 experimental replicates from which we imaged 6 fields of view (100 × 100 × 5 μm each), with 364 cells in total. In these image stacks, a total of 65,562 puncta were detected where 38,056 were assigned and 27,506 were unassigned to a target. The unassigned counts were further categorized based on assumed overlapping errors (2439) or as undetermined counts (25,053). In the associated negative controls, a total of three experiments were performed, of which we imaged 4 fields of view containing 189 cells total. In these experiments, a total of 2034 puncta were detected, of which 61 were classified as targeted expressed genes due to the expected spectral and lifetime signature and 1959 were classified as undetermined.

The tissue experiments with a 6-plex gene expression panel were replicated a total of 2 times yielding 2 fields of view of 130 × 130 × 3 μm each, together containing 174 cells (Fig. 2.5). A total of 2934 puncta were detected of which 1770 were assigned to a target and 1164 were unassigned, the latter group divided into 62 puncta unassigned due to overlap and 1100 labelled as undetermined. In the associated negative controls, we ran a total of 3 experiments yielding 3 fields of view and 375 cells. In these fields of view, 390 puncta were detected, of which 43 were assigned to the transcripts of targeted genes. Of the other 347, only 4 were assigned to overlap and 339 to undetermined. The protein-mRNA codetection experiment in Fig. 2.6 is a pilot experiment for demonstration purpose and there is no replicate for it. Additional 8-plex and 2-plex experiments were performed on cell cultures, two replicates each, yielding a total of 143 and 130 profiled cells,

respectively. Quantification of the experimental replicates by means of cross correlation is presented in Supplementary Fig. 2.5 and Supplementary Fig. 2.6.

Statistical Analysis

When comparing distributions of puncta counts, signal-to-noise ratios, and intensity values, Student (two-sided) t-tests were performed against the probability that the measured distributions belong to distributions with equal means. The reported probability values in the figures are symbolized with (* for $p<0.05$, ** for $p<0.01$, *** for $p<10^{-3}$, and **** for $p<10^{-4}$). Pearson correlation coefficient was computed to determine the correlation between the average expression level to the puncta count of each transcript and to compare within replicates of same experiments. For comparison of 2-plex gene expression counts, we implemented a binomial test where we used the obtained proportion of counts from the 10-plex experiments as the reference probability (Supplementary Fig. 2.6).

2.6 Acknowledgement

We thank Dr. Arthur Lander and the UCI Cancer Systems Biology U54 center for their scientific inputs, Dr. Delia Tifrea, Dr. Jeffrey Kim and the UC Irvine Experimental Tissue Resource (ETR) for their help on tissue preparation and pathological characterization, Fairlie Reese for advice about sequencing data processing, Amber Habowski for help processing the bulk-RNA sequencing data, and UCLA's California NanoSystems Institute and UCI's Beckman Laser Institute for their microscope support. Some figures were partly created using BioRender.com. This work was funded by a U54CA217378 grant to the UCI Cancer Systems Biology Center (CaSB@UCI), NIH/NIAMS P30AR075047 Skin Biology Resource Center grant, a P30CA062203 cancer

center support grant to the UCI Chao Family Comprehensive Cancer Center and UCI's Genomics and High Throughput Facility and Experimental Tissue Resource (ETR), NIH/NIGMS R21GM135493 to P.H. and NIH grant P41GM103540 to E.G. Some imaging was performed on the SP8-Falcon at the Advanced Light Microscopy/Spectroscopy Laboratory and the Leica Microsystems Center of Excellence at the California NanoSystems Institute at UCLA with funding support from NIH Shared Instrumentation Grant S10OD025017 and NSF Major Research Instrumentation grant CHE-0722519. T.V. was supported by NSF GRFP (DGE-1839285). A.V. was supported by the Balsells Fellowship, Generalitat de Catalunya. J.G. was supported by a UCI Immunology NIH T32 training grant AI 060573 and National Institute of Neurological Disorders and Stroke (NINDS/NIH) Training Grant NS082174.

2.7 Author Contribution

T.V., A.V. and J.G. designed, conducted and analyzed the experiments. T.V., A.V., J.G. and W.Z. wrote the manuscript. K.L., Q.X., J.F., J.Z. and C.D. conducted experiments. J.S., L.H., J.W., M.W., A.G., P.H. and E.G. provided technical support and consulted on the study. E.G. and W.Z. designed and directed the project.

2.8 Supplemental Material

Supplementary Tables

Table 4. S1 List of Fluorophores Used

Fluorophore	Detection Channel	Excitation Max (nm)	Emission Max (nm)	Extinction Coefficient (M-1cm-1)	Quantum Yield	Brightness	Lifetime (ns)
ALEXA 647	1	650	668	270,000	0.33	89,100	1.04
ATTO 647N	1	646	664	150,000	0.65	97,500	3.5
ATTO 590	2	593	622	120,000	0.8	96,000	3.7
ATTO 565	3	564	590	120000	0.9	108,000	4
ATTO 488	4	500	520	90,000	0.8	72,000	4.1

Table 5. S2 List of Genes Used and their Assigned Fluorophore Combination

Gene	Fluorophore 1	Fluorophore 2	No. of Probes per Gene	Present in Channel(s)	Used in Figure(s)
NCOA3	ATTO 565	ATTO 590	80	2, 3	4
POLR2A	ATTO 488	ALEXA 647	80	1, 4	4
MTOR	ATTO 565	ATTO 647	80	1, 3	4
MKI67	ATTO 647	ALEXA 647	80	1	4
BRCA1	ATTO 590	ALEXA 647	80	1, 2	4
NCOA2	ATTO 590	ATTO 647	80	1, 2	4
BRCA2	ATTO 488	ATTO 565	80	3, 4	4, 5
CENPF	ATTO 488	ATTO 590	80	2, 4	4, 5
CKAP5	ATTO 565	ALEXA 647	80	1, 3	4, 5
NCOA1	ATTO 488	ATTO 647	80	1, 4	4, 5

Table S3. List of Sequences Used
(see link in published paper for file)

Phasor Transform Expressions

The first-order phasor transform of the lifetime intensity photon histogram $I(t)$ is defined as:

$$S = \frac{\int_0^T I(t)\sin(\omega t)dt}{\int_0^T I(t)dt} \quad (1)$$

$$G = \frac{\int_0^T I(t)\cos(\omega t)\,dt}{\int_0^T I(t)\,dt} \quad (2)$$

Where T is the period between excitation pulses (or modulation period) and $\omega = \frac{2\pi}{T}$ is the pulsation frequency such that the period of the trigonometric functions matches the excitation period T.

The first-order phasor transform of the spectral intensity photon histogram $I(\lambda)$ is defined as:

$$S = \frac{\int_{\lambda_0}^{\lambda_f} I(\lambda)\sin(\omega\lambda - \omega\lambda_0)\,d\lambda}{\int_{\lambda_0}^{\lambda_f} I(\lambda)\,d\lambda} \quad (3)$$

$$G = \frac{\int_{\lambda_0}^{\lambda_f} I(\lambda)\cos(\omega\lambda - \omega\lambda_0)\,d\lambda}{\int_{\lambda_0}^{\lambda_f} I(\lambda)\,d\lambda} \quad (4)$$

Where λ_0 and λ_f are the limits of the spectral band of the detector and $\omega = \frac{2\pi}{\lambda_f - \lambda_0}$ is the pulsation frequency such that the period of the trigonometric functions matches the spectral bandwidth $[\lambda_0\ \lambda_f]$.

Supplementary Figures

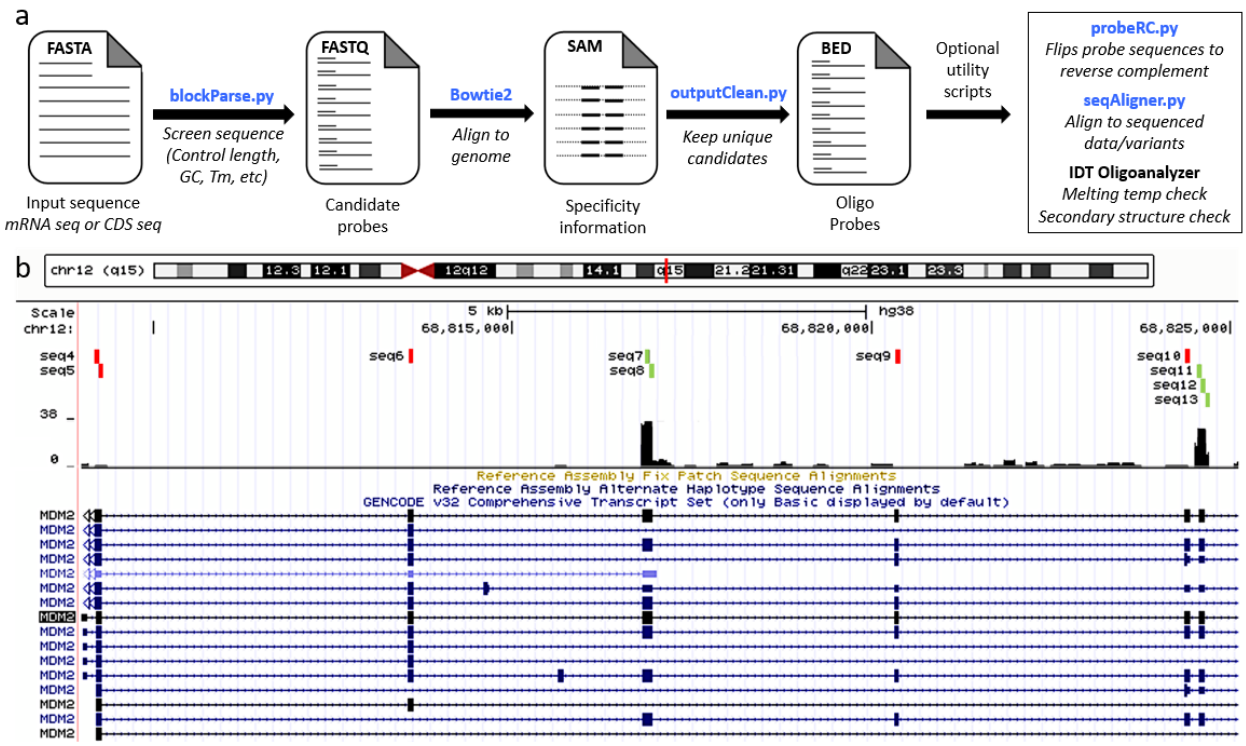


Figure 2.7 S1 Schematic overview of probe design process

a) *A modified version of Oligominer was used to design primary probes.* ***b)*** *In the seqAligner.py script, the probes (labeled "seq#") go through BLAT and are aligned to the RNA sequencing data, as seen on the UCSC Genome Browser. The BigWig browser track shows a histogram of the read counts and probes are aligned and compared to the read counts of the region they overlap with. Probes that overlapped with more than 5% of the highest read count were used (green) and those that aligned with regions considered as "low read count" (5% or less) were removed from the final probe list (red).*

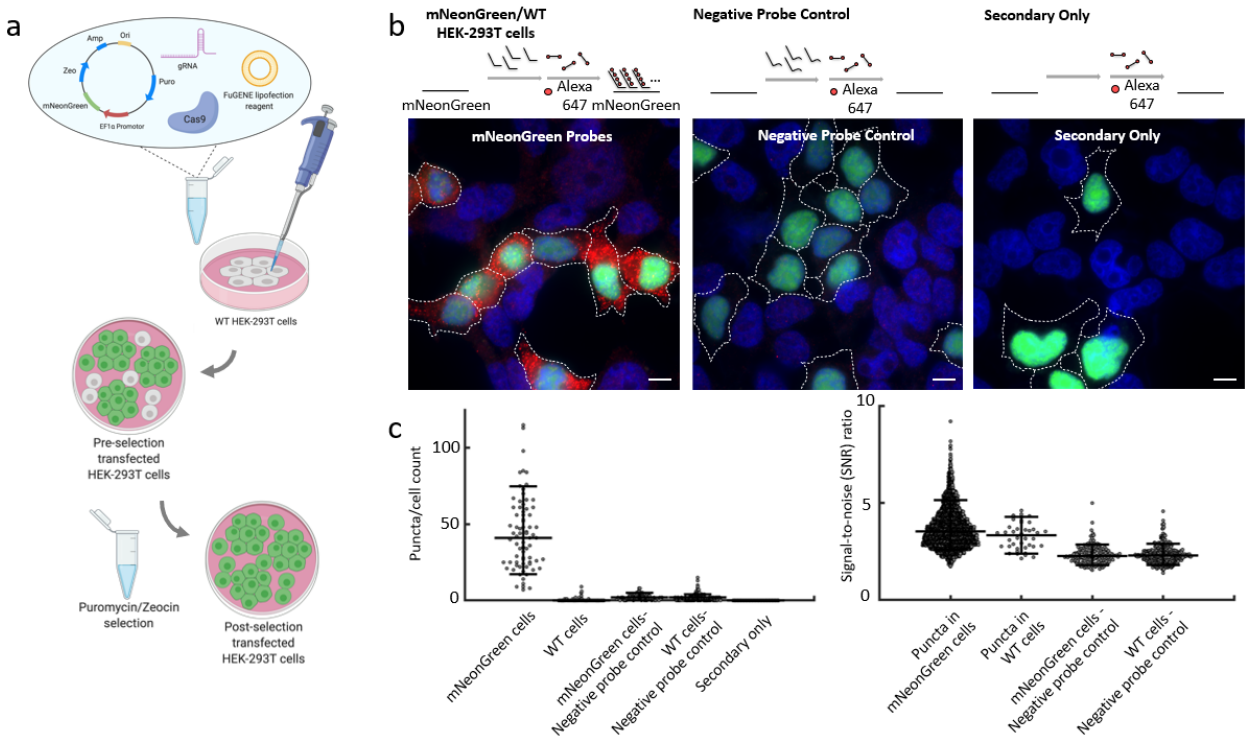


Figure 2.8 S2 Validation of probe hybridization in mNeonGreen cells

a) *Schematic of engineering of mNeonGreen HEK293T-X cells. Engineered mNeonGreen plasmids were transfected into HEK293T-X cells with FuGENE HD Transfection Reagent. Three days after transfection, the cells were then selected with Puromycin and Zeocin.* ***b)*** *Schematic and representative images of each condition. The primary probes were designed to be complementary to mNeonGreen transcripts. A dopachrome tautomerase (DCT) primary probe negative control, which uses primary probes targeting sequences not present in the mNeonGreen HEK-293T-X cells but can still bind to secondary fluorescent oligonucleotides, was used to indicate any nonspecific binding which can occur with primary probe labeling. A negative control where only secondary probes were added but no primary probes were added was used as a reference for nonspecific binding from secondary probes alone. For each condition, the concentration of each primary probe (14 in total) was 1 nM and the secondary probe was 5 nM. Scale bar = 10 μm.* ***c)*** *Plots to quantify the detected puncta per cell and signal-to-noise (SNR) ratio under different conditions. Left, scatter plot showing puncta number per cell (n=755 cells). Right, signal-to-noise ratio (SNR). SNR=each signal intensity/the mean of background noise (n=3,860 puncta).*

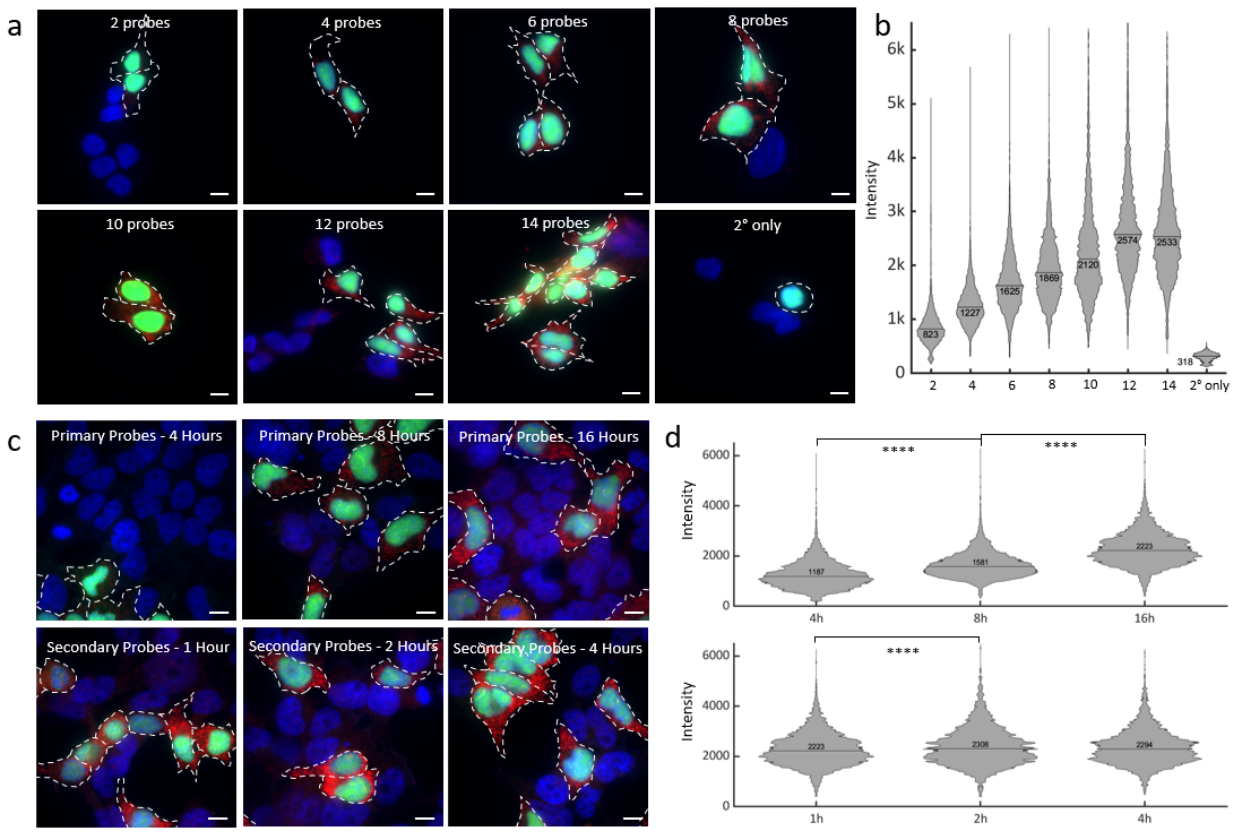


Figure 2.9 S3 Optimization parameters of in situ hybridization conditions.

a) *Representative images of mNeonGreen cells with different numbers of primary probes. Conditions included 2, 4, 6, 8, 10, 12, and 14 mNeonGreen primary probes to the HEK293T-X mNeonGreen and WT model. The concentration of each primary probe is constant (5 nM). Scale bars are 10 µm.* ***b)*** *Intensity distribution of detected puncta shows the effects of the number of primary probes on signal intensity (total n ≈ 64k puncta).* ***c)*** *Representative images of mNeonGreen cells with different incubation time of probes. The primary probes hybridization incubation times consisted of 4, 8, and 16 hours. For secondary fluorophore probes, incubation times tested were 1, 2, and 4 hours. Scale bars are 10 µm.* ***d)*** *Intensity distribution of detected puncta as a function of incubation time. Top, primary probe incubation time (total n ≈ 26k puncta). Bottom, secondary probe incubation time (total n ≈ 20k puncta). Pairwise t-test resulted in p-values <10^4 for conditions marked with ****.*

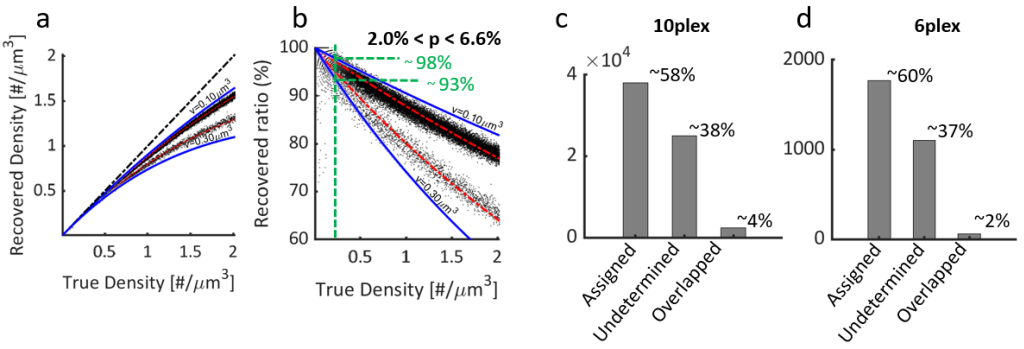


Figure 2.10 S4 Overlapping and inconsistent signal simulations.
Simulations are run at different densities to generate 3D image stacks and puncta are detected using MOSAICA image processing pipeline. ***a)*** *Recovered density from the simulations as a function of the true density (top trend for 1plex bottom trend for 10-plex) in blue the computed limits considering the specified puncta volumes.* ***b)*** *Same data plotted as a percentage of recovered puncta. For the 1plex, there cannot be inconsistent signal due to overlap, the underestimated percentage is solely due to undercounts in the segmentation (top trend). For the 10-plex simulation (bottom trend) the recovered loss is also due to inconsistent signal. Again, in blue, the theoretical limits considering extremal puncta volumes. Green vertical line marks the approximate measured density in the experiments and provides an estimation of the loss due to overlap.* ***c, d)*** *Number of puncta assigned to a particular gene, undetermined puncta and overlapped puncta for the 10-plex experiment (Fig. 4 in the paper) and 6-plex experiment (Fig. 5 in the paper).*

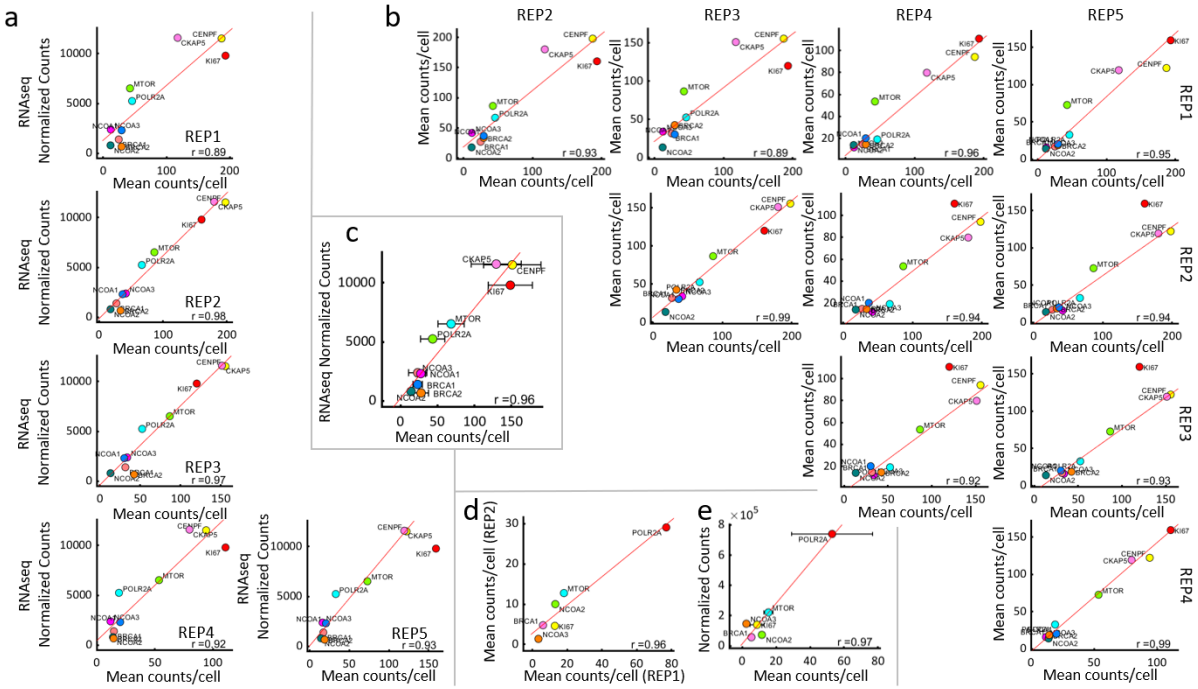


Figure 2.11 S5 Replicates of the experiments.
A total of 5 cell culture sample replicates for the 10-plex detection of genes in colorectal cancer SW480 cells and a total of 2 experimental replicates for the 6-plex melanoma skin FFPE tissues were imaged. ***a)*** *Each of the cell replicates is compared to our RNA sequencing counts.* ***b)*** *Cross comparison against each other for the cell culture replicates.* ***c)*** *The averaged values are compared to RNA sequencing in cell culture experiments.* ***d)*** *Cross comparison against each other for the tissue replicates.* ***e)*** *Averaged values compared to RNA sequencing in tissue experiments. Pearson correlation coefficient is reported in each case.*

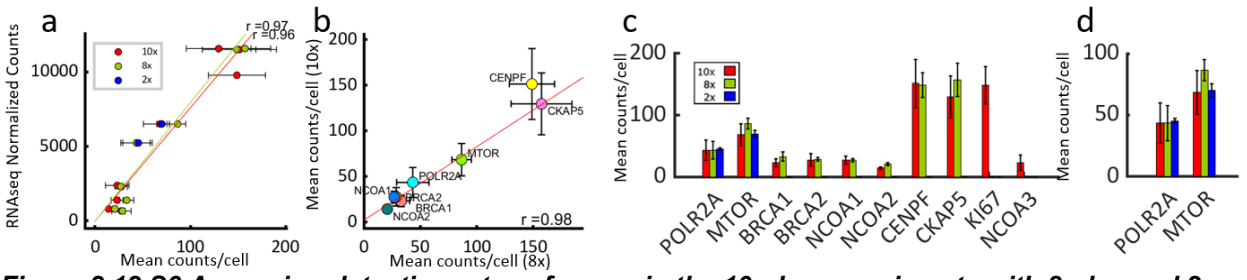


Figure 2.12 S6 Assessing detection rates of genes in the 10-plex experiments with 8-plex and 2-plex experiments.

For these experiments, the 8-plex panel comprises POLR2A, MTOR, BRCA1, BRCA2, NCOA1, NCOA2, CENPF, and CKAP5. The 2-plex panel comprises POLR2A and MTOR. a) Mean puncta counts per cell for each gene in the 10-plex, 8-plex and 2-plex experiments against RNA sequencing data. A Pearson correlation of 0.97 and 0.96 for the 10-plex and 8-plex to RNA sequencing data are shown, respectively. b) Mean puncta counts of the 8-plex experiments were correlated with the 10-plex experiments, Pearson r = 0.98. c) Plotted mean puncta counts per cells for each gene obtained from the 10-plex, 8-plex and 2-plex experiments for comparison. d) Detail of the two genes tagged in all three sets of experiments. Binomial test between the counts of the 2-plex and expected proportion from the 10-plex gives p=0.0796.

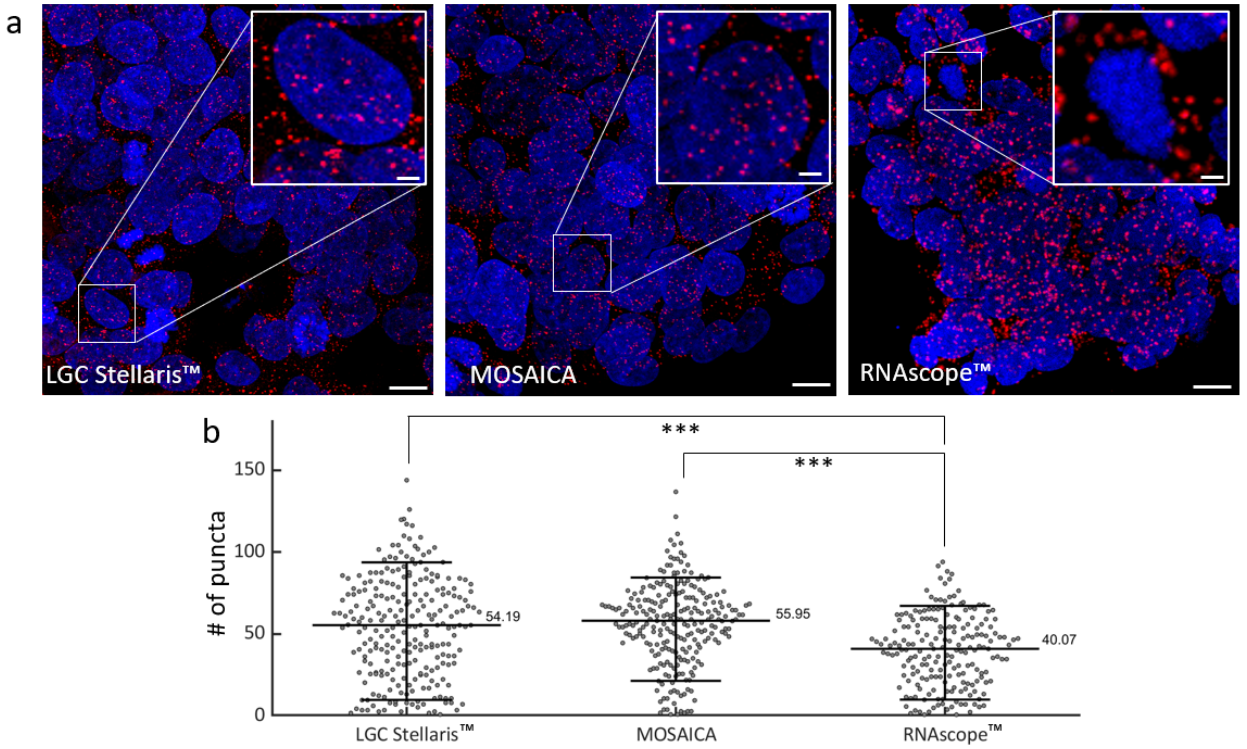


Figure 2.13 S7 Benchmarking MOSAICA against RNAscope™ and LGC Stellaris™

a) *POLR2A gene expression on colorectal cancer SW480 cells following RNAscope™, LGC Stellaris™ and MOSAICA protocols. Scale bars are 10μm.* ***b)*** *Puncta counts per cell volume between three platforms. MOSAICA exhibited comparable puncta per cell count compared to benchmark LGC Stellaris™, whereas RNAscope™ was undercount. Pairwise t-test against null hypothesis that values belong to distributions of equal means were p = 0.4 (LGC Stellaris™ vs MOSAICA), p = 3.4x10^{-4} (LGC Stellaris™ vs RNAscope™ and p = 7.8x10^{-4} (MOSAICA vs RNAscope™). A sliding volume of 3000μm^3 was used throughout the image stacks and the number of puncta counts per volume was obtained. This number was then divided into the average number of cells per volume depending on the 3D segmentation of DAPI nuclei. Scale bars are 10μm in large images and 2μm in insets.*

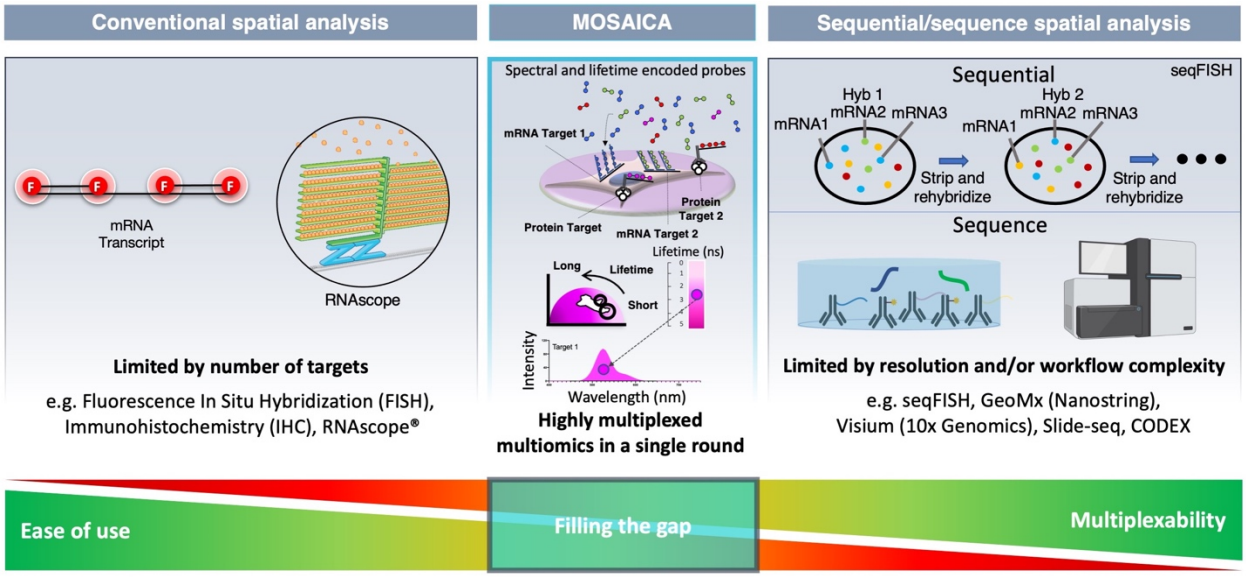


Figure 2.14 S8 Spatialomics field

Multiplexed spatial transcriptomics with MOSAICA, which is rapid, cost-effective, and easy-to-use, can fill a critical gap between conventional FISH and sequential- and sequencing-based techniques.

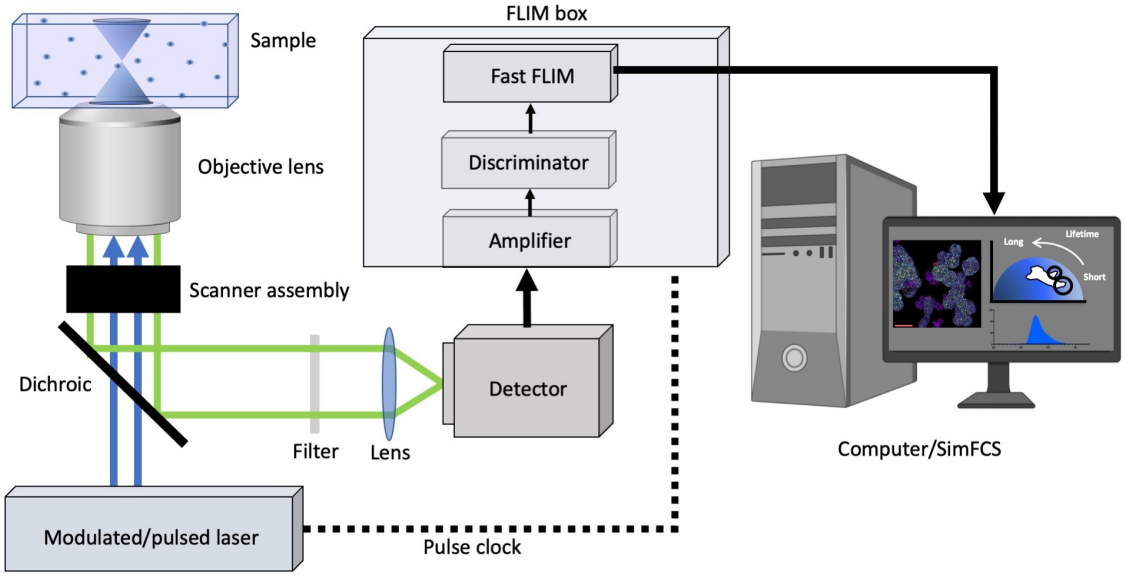


Figure 2.15 S9 Generalized spectral-FLIM Microscopy setup

A pulsed/modulated light source is used to illuminate the sample and the fluorescence of the sample is collected by a spectral detector (current resolution around 10 nm). The repetition rate can either be supplied by or delivered to the laser which is used by the electronics in the digital frequency domain to obtain a single photon arrival time using the heterodyne principle (current resolution around 50 ps)

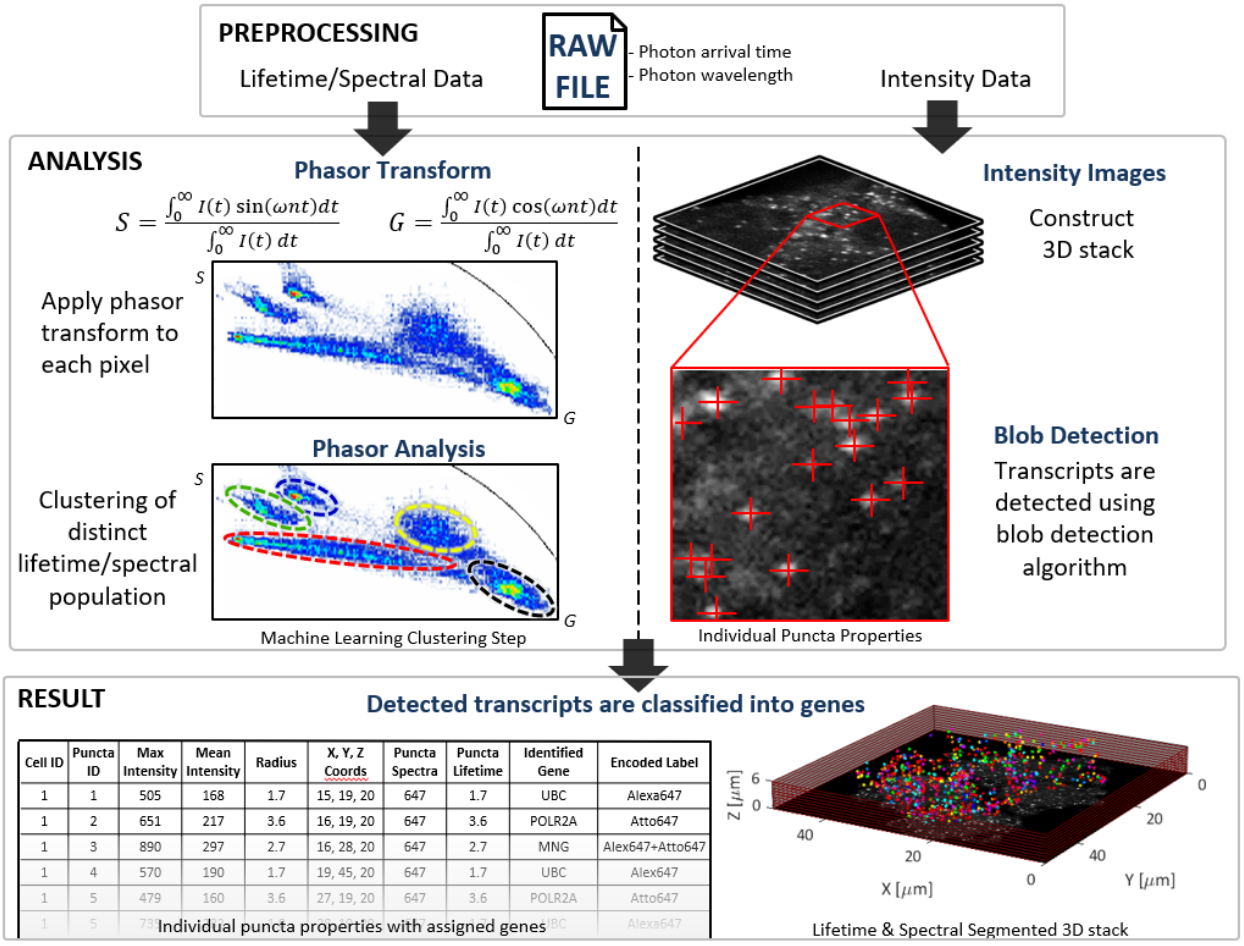


Figure 2.16 S10 Automated pipeline of the processing and analysis
Raw data consists of a list of detected photons with their times of arrivals. Using the acquisition parameters, dwell time, number of pixels, number of repetitions per image etc. the image stacks are reconstructed. Knowing the laser frequency, a photon histogram for each voxel is built and the phasor transform is applied. The two custom made algorithms work in parallel, one identifying clusters in the phasor space, the other identifying puncta in the intensity space. The two then recombine to result in each transcript being identified, assigned to a particular gene and its morphological properties measured.

2.9 References

1. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. Mol. Cell 71, 468–480 (2018).
2. Barsoum, I., Tawedrous, E., Faragalla, H. & Yousef, G. M. Histo-genomics: digital pathology at the forefront of precision medicine. Diagnosis 6, 203–212 (2018).
3. Hu, L. et al. Fluorescence in situ hybridization (FISH): An increasingly demanded tool for biomarker research and personalized medicine. Biomark. Res. 2, 1–13 (2014).
4. Gryglewski, G. et al. Spatial analysis and high resolution mapping of the human whole-brain transcriptome for integrative analysis in neuroimaging. Neuroimage 176, 259–267 (2018).
5. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature 574, 187–192 (2019).
6. Ratan, Z. A. et al. Application of Fluorescence In Situ Hybridization (FISH) Technique for the Detection of Genetic Aberration in Medical Science. Cureus 9, (2017).

7. Gaspar, I. & Ephrussi, A. Strength in numbers: Quantitative single-molecule RNA detection assays. Wiley Interdiscip. Rev. Dev. Biol. 4, 135–150 (2015).
8. Lee, J. H. Quantitative approaches for investigating the spatial context of gene expression. Wiley Interdiscip. Rev. Syst. Biol. Med. 9, 17–19 (2017).
9. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl. Acad. Sci. U. S. A. 113, 11046–11051 (2016).
10. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).
11. Mellis, I. A. & Raj, A. Half dozen of one, six billion of the other: What can small- and large-scale molecular systems biology learn from one another? Genome Res. 25, 1466–1472 (2015).
12. Blom, S. et al. Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. Sci. Rep. 7, 1–13 (2017).
13. Kalyuzhny, A. E. Signal Transduction Immunohistochemistry. (Springer New York, Humana Press, 2017).
14. Moffitt, J. R. et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc. Natl. Acad. Sci. U. S. A. 113, 14456–14461 (2016).
15. Yuan, G. C. et al. Challenges and emerging directions in single-cell analysis. Genome Biol. 18, 1–8 (2017).
16. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science 358, 1622–1626 (2017).
17. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron 92, 342–357 (2016).
18. Wang, F. et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagnostics 14, 22–29 (2012).
19. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, 1360–1363 (2015).
20. Eng, C. H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature 568, 235–239 (2019).
21. Kishi, J. Y. et al. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. Nat. Methods 16, 533–544 (2019).
22. Lewis, S. M. et al. Spatial omics and multiplexed imaging to explore cancer biology. Nat. Methods 18, 997–1012 (2021).
23. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science 363, 1463–1467 (2019).
24. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat. Biotechnol. 38, 586–599 (2020).
25. Liu, Y. et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell 183, 1665-1681.e18 (2020).
26. Becker, W. Fluorescence lifetime imaging - techniques and applications. J. Microsc. 247, 119–136 (2012).
27. Datta, R., Heaster, T. M., Sharick, J. T., Gillette, A. A. & Skala, M. C. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. J. Biomed. Opt. 25, 1 (2020).
28. Beliveau, B. J. et al. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. Proc. Natl. Acad. Sci. U. S. A. 115, E2183–E2192 (2018).
29. Xu, Q., Schlabach, M. R., Hannon, G. J. & Elledge, S. J. Design of 240,000 orthogonal 25mer DNA barcode probes. Proc. Natl. Acad. Sci. U. S. A. 106, 2289–2294 (2009).
30. Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The phasor approach to fluorescence lifetime imaging analysis. Biophys. J. 94, L14–L16 (2008).
31. Ranjit, S., Malacrida, L., Jameson, D. M. & Gratton, E. Fit-free analysis of fluorescence lifetime imaging data using the phasor approach. Nat. Protoc. 13, 1979–2004 (2018).
32. Fereidouni, F., Bader, A. N. & Gerritsen, H. C. Spectral phasor analysis allows rapid and reliable unmixing of fluorescence microscopy spectral images. Opt. Express 20, 12729–12741 (2012).
33. Shi, W. et al. Pre-processing visualization of hyperspectral fluorescent data with Spectrally Encoded Enhanced Representations. Nat. Commun. 11, 1–15 (2020).

34. Scipioni, L., Rossetta, A., Tedeschi, G. & Gratton, E. Phasor S-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging. Nat. Methods 18, 542–550 (2021).
35. Vallmitjana, A., Torrado, B. & Gratton, E. Phasor-based image segmentation: machine learning clustering techniques. Biomed. Opt. Express 12, 3410 (2021).
36. Vallmitjana, A. et al. Resolution of 4 components in the same pixel in FLIM images using the phasor approach. Methods Appl. Fluoresc. 8, 035001 (2020).
37. Chen, G. et al. Disrupting ß-catenin dependent Wnt signaling activates an invasive gene program predictive of colon cancer progression. bioRxiv 667030 (2019).
38. Croce, A. C. & Bottiroli, G. Autofluorescence spectroscopy and imaging: A tool for biomedical research and diagnosis. Eur. J. Histochem. 58, 320–337 (2014).
39. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. Sci. Rep. 8, 1–13 (2018).
40. Gimotty, P. A. et al. Biologic and prognostic significance of dermal Ki67 expression, mitoses, and tumorigenicity in thin invasive cutaneous melanoma. J. Clin. Oncol. 23, 8048–8056 (2005).
41. Chae, Y. K. et al. Challenges and future of biomarker tests in the era of precision oncology: Can we rely on immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) to select the optimal patients for matched therapy? Oncotarget 8, 100863–100898 (2017).
42. Allam, M., Cai, S. & Coskun, A. F. Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. npj Precis. Oncol. 4, 1–14 (2020).
43. Trcek, T., Lionnet, T., Shroff, H. & Lehmann, R. mRNA quantification using single-molecule FISH in Drosophila embryos. Nat. Protoc. 12, 1326–1348 (2017).
44. Klopfleisch, R., Weiss, A. T. A. & Gruber, A. D. Excavation of a buried treasure - DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. Histol. Histopathol. 26, 797–810 (2011).
45. Ripoli, F. L. et al. A comparison of fresh frozen vs. Formalin-fixed, paraffin-embedded specimens of canine mammary tumors via branched-DNA assay. Int. J. Mol. Sci. 17, (2016).
46. Scicchitano, M. S. et al. Preliminary comparison of quantity, quality, and microarray performance of RNA extracted from formalin-fixed, paraffin-embedded, and unfixed frozen tissue samples. J. Histochem. Cytochem. 54, 1229–1237 (2006).
47. Khodaeiani, E. et al. Immunohistochemical evaluation of p53 and Ki67 expression in skin epithelial tumors. Indian J Dermatol. 58, 181–7 (2013).
48. Jackson, H. W. et al. The single-cell pathology landscape of breast cancer. Nature 578, 615–620 (2020).
49. Keren, L. et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell 174, 1373–1387 (2018).
50. Dagher, M., Kleinman, M., Ng, A. & Juncker, D. Ensemble multicolour FRET model enables barcoding at extreme FRET levels. Nat. Nanotechnol. 13, 925–932 (2018).
51. Hedde, P. N., Cinco, R., Malacrida, L., Kamaid, A. & Gratton, E. Phasor-based hyperspectral snapshot microscopy allows fast imaging of live, three-dimensional tissues for biomedical applications. Commun. Biol. 4, 1–11 (2021).
52. Buranachai, C., Kamiyama, D., Chiba, A., Williams, B. D. & Clegg, R. M. Rapid frequency-domain flim spinning disk confocal microscope: Lifetime resolution, image improvement and wavelet analysis. J. Fluoresc. 18, 929–942 (2008).
53. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. Nat. Methods 15, 173–182 (2018).
54. Crosignani, V., Dvornikov, A. S. & Gratton, E. Enhancement of imaging depth in turbid media using a wide area detector. J. Biophotonics 4, 592–599 (2011).
55. Crosignani, V. et al. Deep tissue fluorescence imaging and in vivo biological applications . J. Biomed. Opt. 17, 116023 (2012).
56. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl. Acad. Sci. U. S. A. 116, 19490–19499 (2019).
57. Bishop, C. M. Pattern Recognition and Machine Learning. Information Science and Statistics (2006).
58. Vu, T., Vallmitjana, A. & Gu, J. MOSAICA datasets. https://doi.org/10.6084/ m9.figshare.17072390.v5 (2021).
59. Vallmitjana, A. MOSAICA. https://doi.org/10.6084/m9.figshare.14810820.v4 (2021).

CHAPTER 3

A LOW-COST MODULAR IMAGING SYSTEM FOR RAPID, MULTIPLEXED IMMUNOFLUORESCENCE DETECTION IN CLINICAL TISSUES

Authors:
Joshua Gu[1,2], Hannah Jian[3], Christine Wei[4], Jessica Shiu[5], Anand Ganesan[1,5,6], Weian Zhao[1,2,4,6,7,8,9,*], Per Niklas Hedde[4,10,11,*]

1. Department of Biological Chemistry, University of California, Irvine, California 92697, United States
2. Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, California 92697, United States
3. Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697, United States
4. Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States
5. Department of Dermatology, University of California, Irvine, CA, USA.
6. Chao Family Comprehensive Cancer Center, University of California, Irvine, California 92697, United States
7. Edwards Life Sciences Center for Advanced Cardiovascular Technology, University of California, Irvine, CA, 92697, United States
8. Department of Biomedical Engineering, University of California, Irvine, CA, 92697, United States
9. Institute for Immunology, University of California, Irvine, CA 92697, United States
10. Beckman Laser Institute and Medical Clinic, University of California, Irvine, California 92697, United States
11. Laboratory for Fluorescence Dynamics, University of California, Irvine, California 92697, United States

* = corresponding author

Contribution: J.G. designed/modified and built the Tissue Imager with guidance from P.N.H. Experiments were designed by J.G. and writing and figures were done by J.G. with guidance from W.Z. and P.N.H.

3.1 Abstract

To image 4-plex immunofluorescence-stained tissue samples at a low-cost with cellular level resolution and sensitivity and dynamic range required to detect lowly and highly abundant targets, here we describe a robust, inexpensive (<$9,000), 3-D printable portable imaging device (Tissue Imager) that can immediately be deployed on benchtops for in situ protein detection in tissue samples. Applications for this device are broad, ranging from answering basic biological questions to clinical pathology where immunofluorescence can detect a larger number of markers than the standard H&E or immunohistochemistry (IHC) staining, while the low price point also allows usage in classrooms. After characterizing our platform's specificity and sensitivity we demonstrate imaging of a 4-plex immunology panel in human cutaneous T-cell lymphoma (CTCL) FFPE tissue samples. From those images, positive cells were detected using CellProfiler, a popular open-source software package, for tumor marker profiling. We achieved a performance on par with commercial epifluorescence microscopes that are ~20 times more expensive than our Tissue Imager. This device enables rapid immunofluorescence detection in tissue sections at a low-cost for scientists and clinicians and can provide students a hands-on experience to understand engineering and instrumentation.

3.2 Introduction

Current cancer diagnosis methods are comprised of clinical examination, radiological imaging, and histopathological analysis of tissue biopsies and surgical resections, which provide insight into a patient's type and stage of cancer[1, 2]. Physicians have depended upon histopathology, which is the "gold standard" for visualization and pathological interpretation of tissue biopsies. Pathological analyses of tumor biopsies have broad utility in cancer diagnosis, prognosis, and treatment stratification. Hematoxylin and eosin (H&E) stained histologic sections are considered the standard of care by pathologists and can be used for a variety of applications, such as identifying malignant tumors, segmentation of glands in the prostate, grading of breast cancer pathology, and classification of early pancreatic cancer[3]. The immunohistochemistry (IHC) method, chromogenic immunohistochemistry (CIH), is used to complement H&E staining, which stains the tissue morphology, to detect the presence of specific protein markers for accurate tumor classification and diagnosis. While H&E and CIH stains provide enough information for some applications, there are many cases, such as tumor differentiation and tumor immune microenvironment (TIME) profiling, where more data is needed. In addition, conventional CIH is limited to few markers per tissue section and chromogenic systems used for the staining saturate easily, restricting quantitative analysis[4, 5]. In these cases, labeling the cells with antibodies for immunofluorescence imaging can allow multiplexing, increase the sensitivity and dynamic range, and provide additional information for further characterization[3, 6, 7]. Even though immunofluorescence provides clinical value, it currently requires expensive imaging

hardware (~$100,000-$500,000) and the acquisition of a large field of views to generate sufficient data can be very time intensive.

The ability to multiplex immunofluorescence markers enables studies that investigate cellular co-expression[8, 9], cellular spatial relationships[10], and tissue heterogeneity[11], to name a few. In the field of immunotherapy, understanding the cellular composition and spatial distribution within the sample, which is referred to as spatial biology, has become important[5, 12, 13]. By profiling immune checkpoint inhibitors, which reduce T-cell inhibition and allows them to fight cancer cells, cancer treatments have benefited[14-19].

Cutaneous T cell lymphoma (CTCL) is a type of cancer that starts in white blood cells called T cells (T lymphocytes), which typically help fight pathogens in the immune system[20]. In CTCL, T cells develop abnormalities, causing them to attack the skin and cause rash-like skin erythema, patches of raised or scaly skin, and sometimes skin tumors[21]. Unfortunately, the exact cause of CTCL is still unknown. As CTCL tissue samples contain high levels of T cells, they are a good positive control for T cell markers such as CD3, CD8, and CD14[22]. Hence, we selected CTCL tissue samples to be our model system to demonstrate detection of these T cell markers, which vary from low to high abundance to demonstrate the sensitivity of our imaging platform.

To take full advantage of the clinical value of immunofluorescence, a robust, inexpensive, high-throughput imaging platform that can be deployed immediately to any laboratory or clinic including those in low-resource settings to image clinical tissue samples with immunofluorescence is highly desired. To address this need, we have developed a robust, inexpensive (<$9,000), and portable imaging platform for tissue

samples, the Tissue Imager, that can be placed on the benchtop of any basic laboratory. Our Tissue Imager uses a 3D printable design and widely available components to excite fluorescence of fluorophore-conjugated secondary antibodies that are detected with an inexpensive 20-megapixel CMOS camera module coupled with a long working distance, 10x objective, with sufficient spatial resolution to provide cellular resolution and sensitivity to detect a wide range of protein abundance levels. We demonstrate that, with clinical patient samples, this imaging platform can obtain image resolutions or par with a commercial epifluorescence microscope that is ~20 times more expensive while, at the same time, providing a larger field of view to increase imaging throughput. Our low cost, high throughput, and portable platform can immediately benefit the healthcare and scientific community.

3.3 Results

3.3.1 Tissue Imager design

While existing low-cost microscopy platforms designed for biological fluorescence utilize the camera of cellphones with various illumination schemes such as on-axis epi-illumination[23, 24], off-axis inclined illuminations[25], butt-coupling[26], and total internal reflection[27], there have been limitations regarding spatial resolution, field of view, and the maximum number of spectral channels. To obtain sub-cellular spatial resolution (~1 μm) and multiplexed fluorescence images for clinical tissue biopsy samples on glass microscope slides (25 mm x 75 mm x 1 mm), a device would need to feature an objective lens of reasonably high numerical aperture (~0.3) as well as multiple spectral windows for illumination and detection of several different fluorophores. For the imager to be inexpensive and able to image samples in a high throughput manner, it should be

portable, low cost, and easy to use by technicians with minimal training. Our Tissue Imager meets all these requirements to image tissue samples with multiplexed immunofluorescence staining as illustrated in Figure 3.1A-C. The tissue samples are first stained with antibodies, then imaged with the Tissue Imager, followed by analysis. The overall design of the Tissue Imager can be seen in Figure 3.1D. Clinical tissue sample sections from pathology centers are typically also placed on glass microscope slides that are the same size, hence our Tissue Imager was designed to accommodate this on the sample stage. The images obtained with a 20-megapixel CMOS camera (3648 x 5472 pixels) correspond to a 1.8 x 2.6 mm^2 field-of-view (FOV) with a pixel size at the sample of 0.48 μm (Supp. Fig 1), sufficiently large enough to image a typical human biopsy section such as skin tissue. To validate and benchmark our Tissue Imager, we acquired reference images with a Nikon Ti-1000E widefield microscope using a 10x objective with a sample pixel size of 0.65 μm.

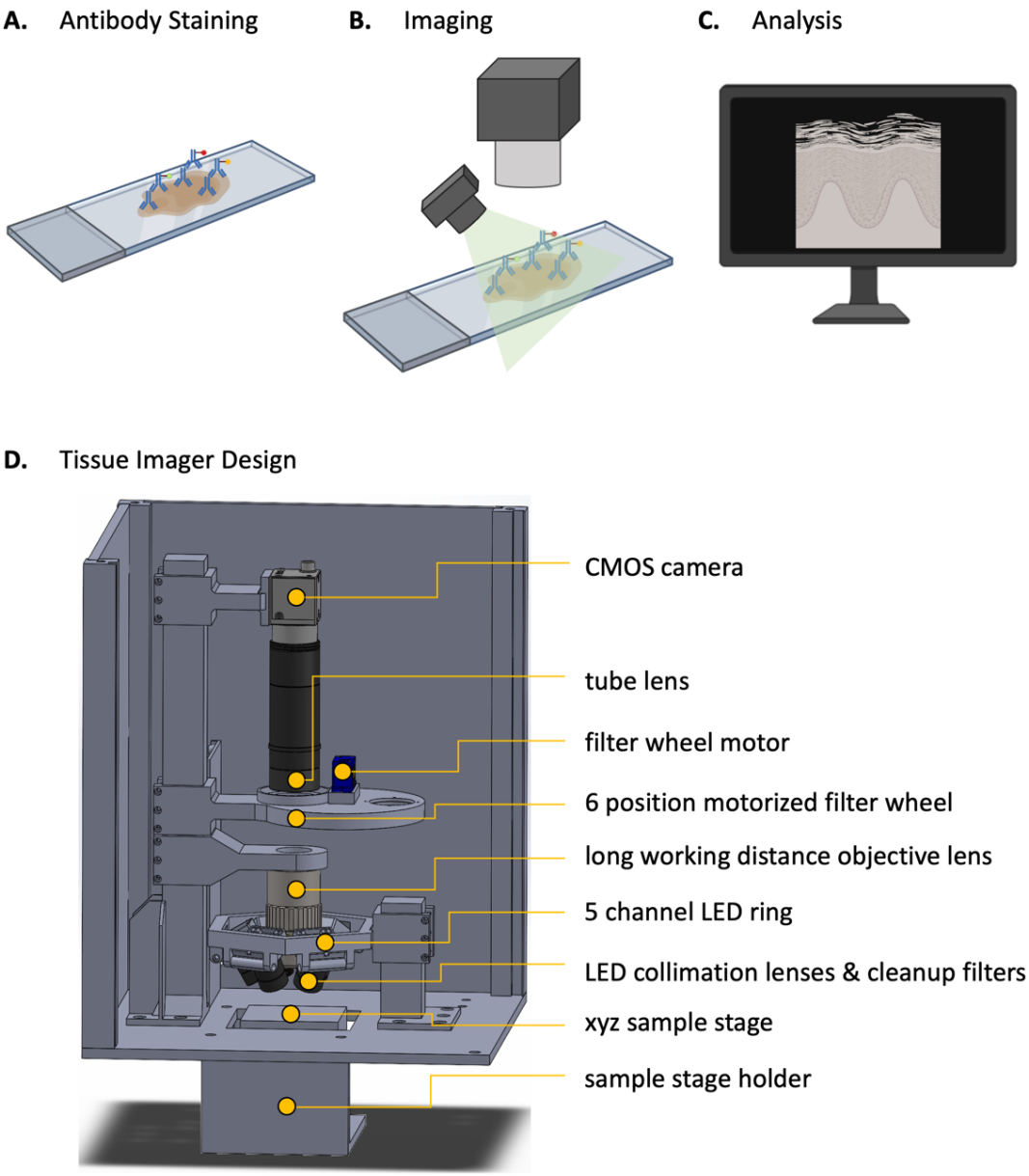


Figure 3.1 Workflow/design schematic
A) Tissue samples are processed, and the protein markers are stained with antibodies. After blocking (e.g., typically with 10% BSA in this study),, the primary antibody cocktail is placed onto the sample, followed by incubation with secondary labeled antibodies. B) The sample is then imaged with the Tissue Imager. C) The images are saved into a folder automatically and the images are then exported into ImageJ and/or CellProfiler for processing and quantification. D) Tissue Imager CAD design with 20-megapixel CMOS camera, motorized filter wheel, 10x objective, 5 LEDs, and sample stage holder.

Tissue sample images were acquired from the top with a 10x Mitutoyo Plan Apochromat Objective (0.28 NA), followed by a 6-position motorized filter wheel (5 bandpass filters currently used with center wavelengths 460 nm, 530 nm, 577 nm, 645

nm, and 690 nm) to spectrally select the fluorescence emission from each fluorophore type on the sample. The fluorescence was then imaged with a tube lens (f = 100 mm) onto the chip of a 20-megapixel CMOS camera and read out via an USB 3 interface compatible with most computers and operating systems. Fluorescence excitation was achieved using a ring above the sample that held 5 LEDs, each coupled to a condenser lens (f = 20 mm) and cleanup filter (center wavelengths 365 nm, 460 nm, 520 nm, 585 nm, and 630 nm). The sample was placed below onto an xyz sample stage allowing for field of view position and focus adjustments. The entire setup was enclosed in a box made from black ¼"-thick laser-cut acrylic boards. This light-tight enclosure prevented external light from contaminating the resulting images. All components including the optics, camera, LEDs, and structural supports were integrated into a CAD model that could be manufactured on a larger scale at a low cost. To adjust and evaluate the illumination homogeneity, images of a reference microscope glass slide were taken for each channel. Remaining variations could be easily corrected through software (Supp. Figure 2). In addition to fluorescence imaging, our device also allowed for the acquisition of brightfield images as required for IHC or H&E-stained samples. For this purpose, separate images with red (630 nm), green (520 nm), yellow (577 nm) and blue (460 nm) illumination were taken and reconstructed into a final RGYB color image.

3.3.2 Evaluation of specificity and sensitivity

To evaluate the performance of the Tissue Imager, we imaged fluorescence beads of various emission/detection ranges to validate all 5 spectral channels. After vortexing and diluting each 1 µm FluoSpheres™ Polystyrene Microspheres (blue/green, yellow/green, orange, red, and crimson) sample 1:2000 in PBS, 10 µl of sample solution

was pipetted into Countess™ Cell Counting Chamber Slides (Invitrogen, C10228). As shown in Figure 3.2, each microsphere population was detected in the expected spectral channel. To evaluate potential spectral crosstalk between channels, each microsphere population was imaged in all channels. We found the fluorescence to be specific to the respective channels demonstrating the specificity of the Tissue Imager and its ability to resolve beads as small as 1 µm in diameter (Supp. Figure 3).

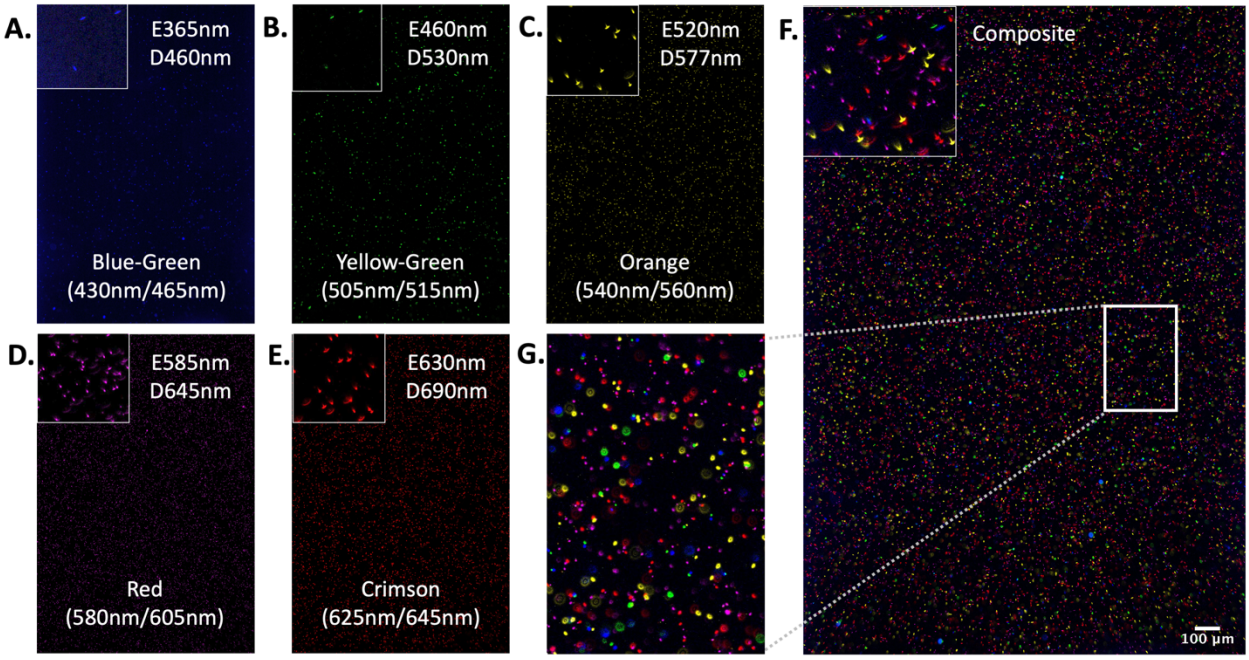


Figure 3.2 Validation of each spectral channel and the specificity of the Tissue Imager
Using 1 µm fluorescent beads at a 1:2000 dilution, each spectral channel was calibrated. A) Blue-Green fluorescence beads with excitation/emission peaks at 430 nm and 465 nm illuminated at 365 nm and detected at 460 nm. B) Yellow-Green fluorescence beads with excitation/emission peaks at 505 nm and 515 nm, illuminated at 460 nm and detected at 530 nm. C) Orange fluorescence beads with excitation/emission peaks at 540 nm and 560 nm, illuminated at 520 nm and detected at 577 nm. D) Red fluorescence beads with excitation/emission peaks at 580 nm and 605 nm, illuminated at 585 nm and detected at 645 nm. E) Crimson fluorescence beads with excitation/emission peaks at 625 nm and 645 nm, illuminated at 630 nm and detected at 690 nm. F) Composite image of all channels. G) Inset of composite image F. Scale bar is 100 µm.

To determine the sensitivity of the Tissue Imager, the Dragon Green Intensity Standard beads (Bangs Laboratories, DG06M) of 5 different intensities were used. This standard bead kit is typically used for fluorescence microscopy and flow cytometry

calibrations. The standard beads were vortexed and diluted 1:10 in PBS-T (0.025% Tween20), then 10 µl of sample solution was pipetted into Countess™ Cell Counting Chamber Slides (Invitrogen, C10228) for imaging. As shown in Figure 3.3A, the Dragon Green (DG) beads DG1-DG5 were excited with the 460 nm LED and detected in the 530 nm channel with fluorescence intensities increasing from DG 1 to DG 5 as expected. In Figure 3.3B, the fluorescence intensity for each bead intensity was quantified using ImageJ and plotted to characterize the sensitivity and wide dynamic range (0.24%-100% intensity).

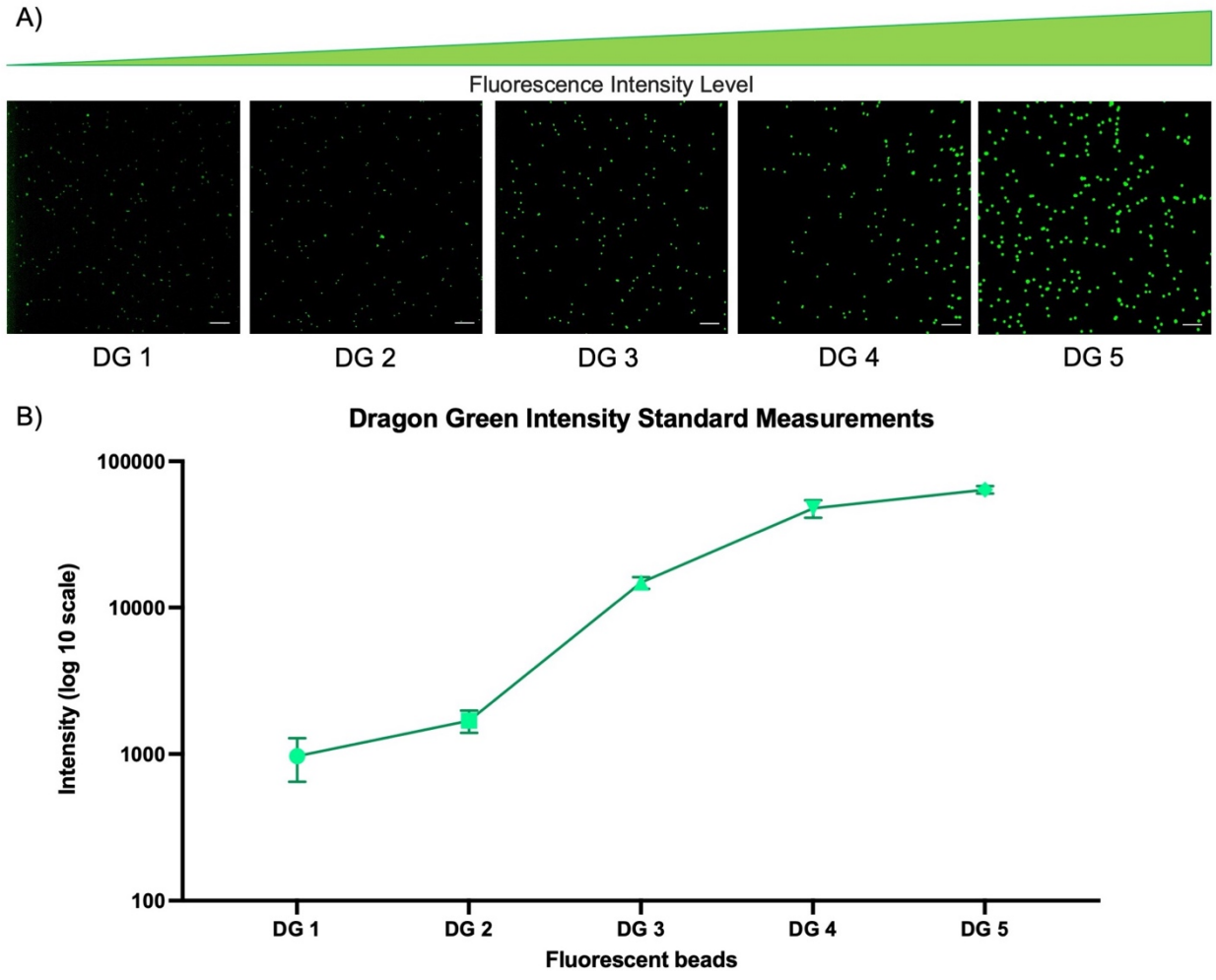


Figure 3.3 Quantifying sensitivity
A) Image of the Dragon Green Intensity Standard beads of ~8 µm polystyrene-based microspheres dyed with increasing amounts of the Dragon Green fluorophore to quantify the sensitivity of the Tissue Imager. 100 µm scale bar. B) Average intensity values of each bead population are plotted. Scatter plot was plotted with the mean and standard deviation.

3.3.3 Evaluation of an immune panel on CTCL tissue samples

The next step was to profile immune markers in clinical tissue samples to demonstrate rapid imaging for a 4-plex protein detection panel. Using our CTCL model, we profiled CD3e, CD8, and CD14 using antibodies. CD3e and CD8 are T-cell markers while CD14 has been used as a marker for monocytes and macrophages[22, 28, 29]. The nucleus was stained with DAPI. The images obtained from the Tissue Imager were compared to H&E and CD3e and CD8 IHC stains from serial sections of the same FFPE block. The CD3e and CD8 from the same section imaged on the Tissue Imager were also imaged on a Nikon Ti-1000E microscope with a 10x objective as a benchmark for immunofluorescence imaging (Figure 3.4). CD14 was not imaged on the Nikon microscope due to the absence of a suitable spectral channel. As confirmed by the IHC staining from the UCI Dermatology Center, CD3e is highly abundant while CD8 is less abundant. This allowed us to demonstrate the Tissue Imager's ability to detect protein markers of various abundance levels. The DAPI (405 nm), CD3e (488 nm), CD14 (594 nm), and CD8 (647 nm) stained CTCL tissue section was imaged within 4 seconds on the Tissue Imager. As negative control, tissues were stained with the secondary antibody only (Supp. Fig 4).

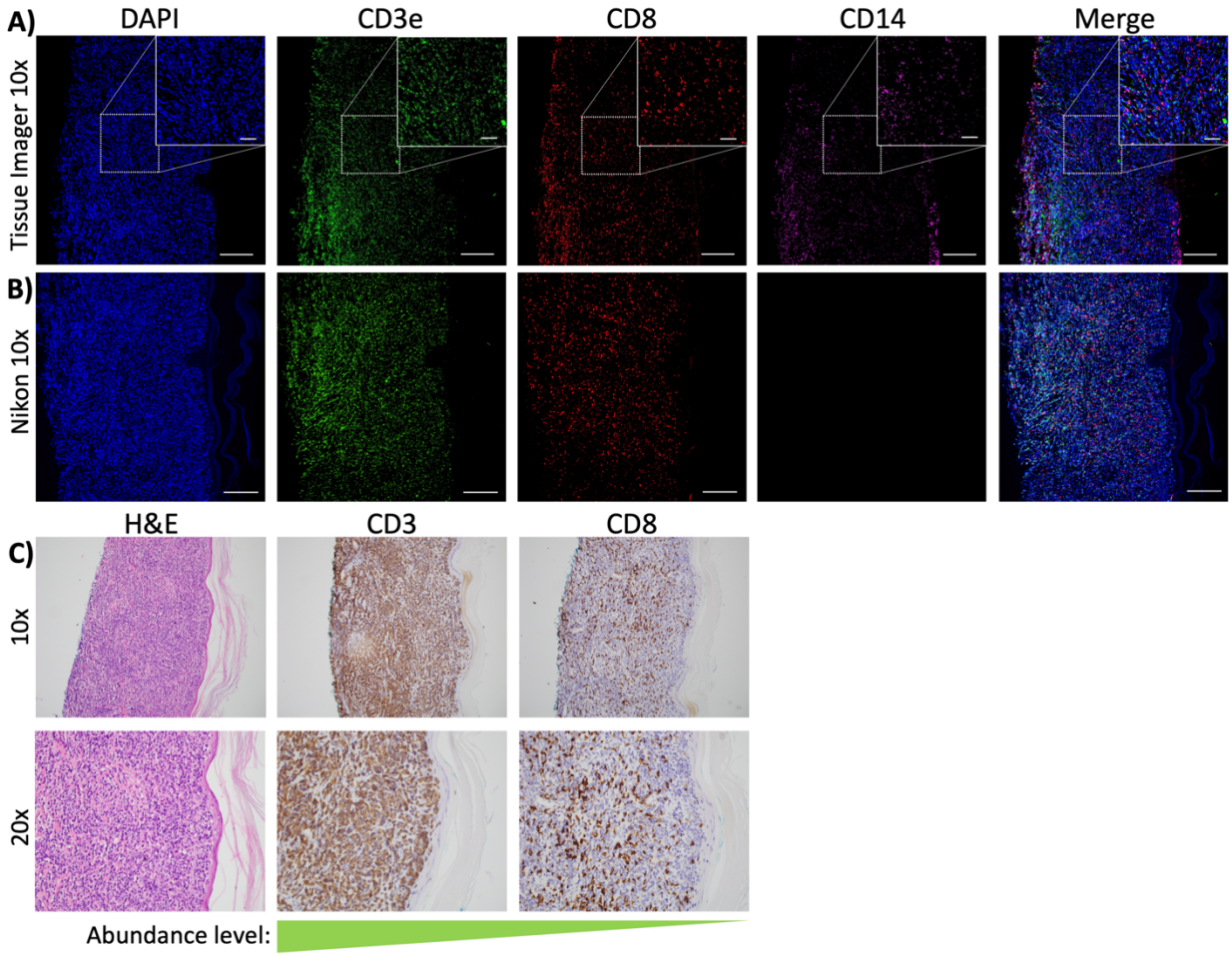


Figure 3.4 A 4-plex immune panel on CTCL with comparison to IHC and Nikon microscope
Immunofluorescence staining of DAPI, CD3e, CD8, and CD14 on Cutaneous T-cell Lymphoma (CTCL) FFPE tissue sections imaged with the Tissue Imager and Nikon. Images were compared to H&E staining, CD3 IHC, and CD8 IHC of serial sections. A) Tissue Imager 10x images cropped for channels (DAPI, 488 nm-CD3e, 647 nm-CD8, 594 nm-CD14) and merged. B) Nikon Ti-1000E 10x images for channels (DAPI, 488-CD3e, 647nm-CD8) and merged. C) 10x and 20x images of H&E, CD3 IHC, and CD8 IHC staining on serial sections of CTCL FFPE tissue sections. Brown staining is positive signal. Scale bars are 200 µm for large images and 50 µm for insets.

After acquisition, the images were processed using ImageJ and analyzed using a CellProfiler[30] image analysis pipeline (Figure 3.5A). The CellProfiler pipeline was validated by manually counting six 700 x 700-pixel regions that were randomly selected throughout the tissue section. The manual counting was used to obtain the percentage of cells positive for each marker and compared to the counts detected in the CellProfiler pipeline. As shown in Supplementary Figure 5, there were no significant differences

between the manual counts and CellProfiler counts for all 3 markers (CD3e, CD8, and CD14). By detecting the DAPI-stained nuclei 8,238 cells were found in this image (Figure 3.5B). The cells positive for each marker were then detected and quantified, with 51.0% cells expressing CD3e, 16.1% cells expressing CD8, and 17.7% cells expressing CD14 (Figure 3.5C). The percentages of cells positive for each marker was then plotted for all images (n = 7), resulting in an average of 48.6%, 14.9%, and 12.4% cells positive for CD3e, CD8, and CD14 respectively (Figure 3.5D). The CellProfiler pipeline also detected cells that co-expressed both CD3e and CD8. On average, 9.3% of cells were CD3e/CD8-positive, 39.6% of cells were CD3e-positive/CD8-negative, and 5.7% of cells were CD3e-negative/CD8-positive (Figure 3.5E). The ability to detect co-expression of multiple protein markers on the same cell at single-cell level is of interest in clinical pathology such as profiling both the presence of T cells and the abundance of immune checkpoint proteins for patient stratification.

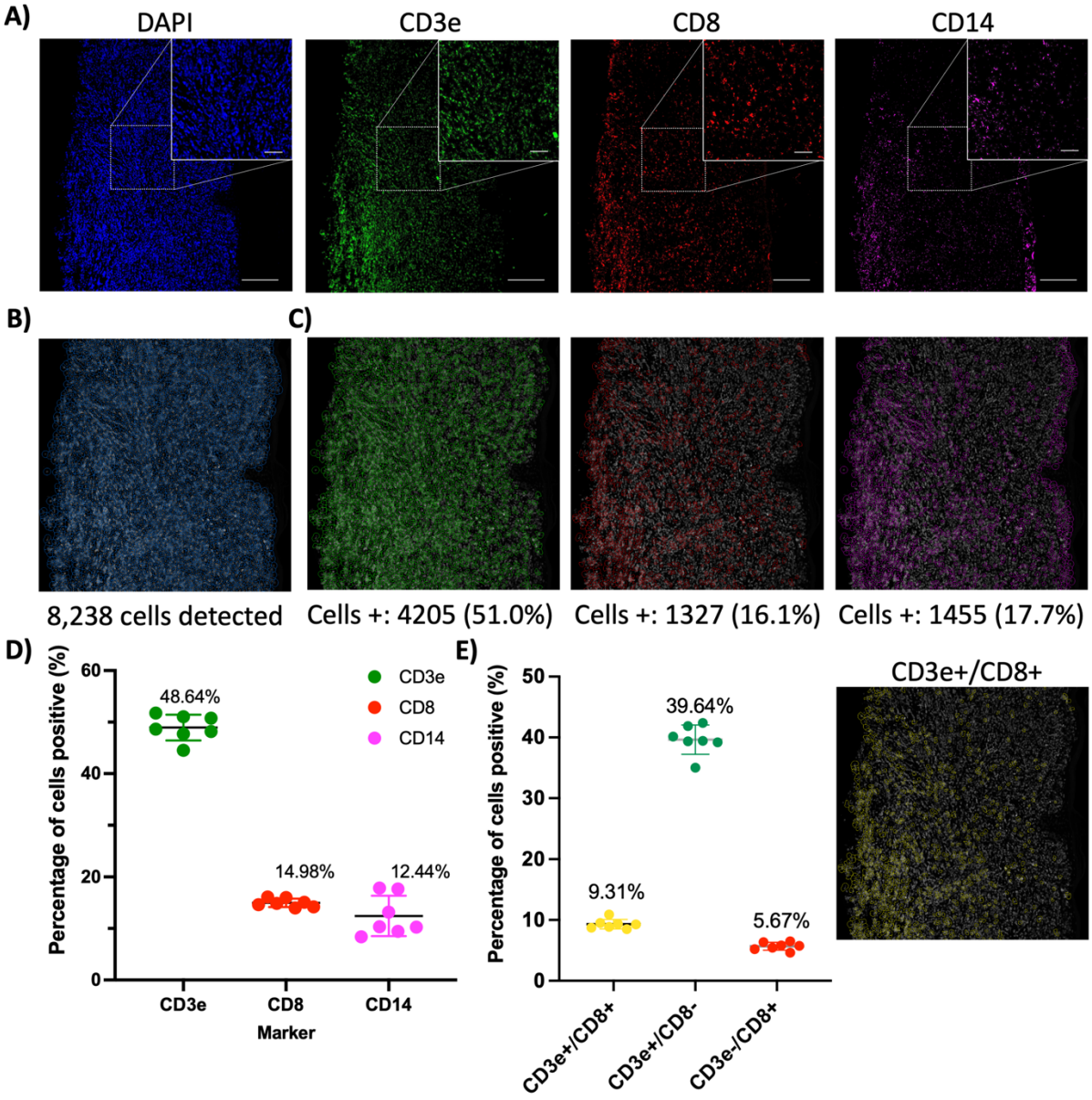


Figure 3.5 CellProfiler quantification of images and cell segmentation
A) The fluorescence images obtained from the Tissue Imager (DAPI, CD3e, CD8, CD14). B) The cell segmentation outline from the CellProfiler pipeline. C) Cells identified as positive for the staining for each marker were outlined and the cells positive percentage were calculated (n=7). D) Percentage of cells positive for each marker with the average of 48.64%, 14.98%, and 12.44% for CD3e, CD8, CD14 respectively. E) Percentage of cells CD3e+/CD8+, CD3e+/CD8-, CD3e-/CD8+ were on average 9.31%, 39.64%, and 5.67% respectively. Outline of cells CD3e+/CD8+. Scatter plots were plotted with the mean and standard deviation.

3.4 Conclusion/discussion

In summary, the Tissue Imager described here represents a low-cost instrument (<$9,000) which is a simple yet sensitive and highly versatile (5 fluorescence channels + RYGB brightfield) design that could be produced easily, thus being a useful tool in settings such as academic laboratories. This device provides a low-cost platform for scientists to rapidly image clinical samples on lab benchtops or any location with little space available as well as an opportunity for students to gain the knowledge and experience in engineering, instrumentation, and software development. This device could also be used for other applications such as tissue microarray imaging with minimal modifications to enable high-throughput batch sample analysis. In the future, additional features could be incorporated such as different and/or additional spectral channels and potentially hyperspectral detection, a motorized/automated sample stage, and possibly even fluorescence lifetime detection with time-of-flight-resolving consumer cameras. Additionally, the light path could be modified to include a broadband light source and (phase) masks to enable dark-field and phase-contrast imaging. Fluorescence quenchers such as TrueBlack could be used to quench tissue autofluorescence thus increasing sensitivity, and novel computational tools could be leveraged to maximize the information obtained from the images[31]. Basic analysis modules are also available on ImageJ, providing users the opportunity to learn about these algorithms and create their own Tissue Imager workflows.

3.5 Materials and Methods

Tissue Imager design

The sample slides were illuminated with five different 5 W LEDs (365 nm, 460 nm, 520 nm, 585 nm, 630 nm, 120° angle of emission) using a custom designed 5 channel LED ring mounted above the sample stage. After collimation with aspherical lenses of 20 mm focal length (Thorlabs, Newton, NJ), the LED emissions were spectrally cleaned with bandpass filters (365/10 nm, 460/30 nm, 520/20 nm, 585/20 nm, 630/20 nm) (Chroma, Bellows Falls, VT). LEDs were driven by individual DC-DC driver circuits to adjust the current (max 1,000 mA each). Fluorescence was collected with a long working distance (WD 34 mm) 10x Mitutoyo Plan Apochromat Objective (Thorlabs, Newton, NJ) coupled to a 1” diameter achromatic tube lens of 100 mm focal length (Thorlabs, Newton, NJ). The custom 6-position filter wheel was actuated with a servo motor controlled with an Arduino Nano microcontroller, that was also used to control the LED drivers. Before imaging with a 20 MP monochrome CMOS camera (FLIR Blackfly, FLIR Systems, Goleta, CA), bandpass filters were used to block scattered excitation light (450/50 nm, 530/30 nm, 577/25 nm, 645/30 nm, 690/50 nm) (Chroma, Bellows Falls, VT). All electronics were powered by a 5 V, 3.5 A power supply. After 3D printing of the model (Ender-3 pro, Creality3D, Hong Kong and Ultimaker S5, Ultimaker B.V., Netherlands), all relevant optical components were inserted and attached.

Resolution measurements

A 10-mm ruler (R1L3S1P, Thorlabs) was imaged with RGB settings for a brightfield image. The image was then quantified using ImageJ by measuring the distance in pixels

between 1 division (50 μm) or 2 divisions (100 μm). The μm/pixel value was then calculated, and the average value was obtained (Supplementary Fig. 1).

Fluorescent beads

1 μm FluoSpheres™ Polystyrene Microspheres of various colors (Invitrogen, F13080, F13081, F13082, F13083, F8816) were vortexed and diluted at 1:2000 with PBS before being pipetted into a Countess™ Cell Counting Chamber Slide (Invitrogen, C10228). The Dragon Green Intensity Standard Kit with 5 Intensities (Bangs Laboratories, DG06M) (DM1-5) around 8 μm were vortexed, diluted 1:10 in PBS-T (0.025% Tween20), and pipetted into a Countess™ Cell Counting Chamber Slide (Invitrogen, C10228).

Preparation of FFPE tissues

The University of California Irvine IRB approved this study for IRB exemption under protocol number HS# 2019-5054. All human cutaneous T-cell lymphoma (CTCL) cases were de-identified samples to the research team at all points and therefore considered exempt for participation consent by the IRB. Fully characterized human patient skin CTCL FFPE tissues were obtained from the UCI Dermatopathology Center, then sectioned to 5 μm thick slices using a rotary microtome, collected in a water bath at 35°C, and mounted to positively charged Fisher super frost coated slides (Fisher Scientific, 12-550-15). The tissue sections were then baked at 60°C for 1 hour. For antigen unmasking, slides were deparaffinized, rehydrated then followed by target retrieval (with citrate buffer).

Antibody staining

The samples were blocked with 10% BSA in PBS for 2 hours at room temperature. Antibody solutions containing Rabbit anti-Human CD3e (Abcam, ab52959), Mouse anti-Human CD8 (Abcam, ab75129), and Goat anti-Human CD14 (LifeSpan, LS-B3012-50)

antibodies and 1% BSA in PBS were subsequently added to the samples and incubated overnight at 4℃. Following a PBS wash, antibody solutions containing fluorescently labeled Donkey anti-Rabbit Alexa 488 (ThermoFisher, A-21206), Donkey anti-Mouse Alexa 647 (ThermoFisher, A-31571), and Donkey anti-Goat Alexa 594 (ThermoFisher, A32758) antibodies in 5% secondary raised serum and 1% BSA in PBS were added at room temperature for 1 hour. Three 5-minute washes at room temperature with RNAse-free PBS were then performed, with the second wash containing 1:1000 Hoechst stain.

Image acquisition and data transfer

1 μm fluorescence beads were imaged with a camera exposure time of 1,000 ms at 365/460 nm excitation/emission and a camera exposure time of 100 ms was used for all remaining channels. The Dragon Green Intensity Standard Kit (Bangs Laboratories, DG06M) was imaged with an exposure time of 1,000 ms in the 460/530 nm channel (excitation/emission). Tissue sections were imaged with an exposure time of 300 ms for DAPI staining (excitation 365 nm, detection 460 nm), 1,500 ms for the 488 nm (excitation 460 nm, detection 530 nm) and 647 nm (excitation 630 nm, detection 690 nm) channels, and 2,000 ms for the 594 nm (excitation 585 nm, detection 645 nm) channel. For all images taken, the camera gain was set to 26 dB. Images were saved in 16-bit TIFF format for further processing. For analysis, the tissue images were cropped 2,300 x 2,300 pixels.

Validation images were acquired with an inverted Nikon Ti-1000E epifluorescence microscope using a 10x plan apochromat oil objective with a numerical aperture of 0.45. Samples were excited with a Spectra-X (Lumencor) LED light source at 395 nm, 470 nm, and/or 640 nm. Images were acquired with an Andor Zyla 4.2 sCMOS camera. The H&E

image in Supplementary Figure 2 was taken on a Nikon Eclipse E400 with the Nikon Plan Fluor 10x/0.30 DIC objective and a QImaging MicroPublisher 6 camera.

Immunofluorescence quantification

CellProfiler

Fluorescence signal intensity was quantified using the open-source software CellProfiler. Raw .nd2 images from Nikon and composite Tissue Imager images created with another open-source software, ImageJ, were fed into a CellProfiler pipeline. In the pipeline, the nuclei were identified using the IdentifyPrimaryObjects module then expanded to represent the cell bodies. Protein fluorescence was also identified with the IdentifyPrimaryObjects module. Raw channel images were rescaled with the RescaleIntensity module for accurate protein and background intensity measurements that were obtained using the MeasureObjectIntensity and MeasureImageIntensity modules, respectively. Positive cell determination was done using the RelateObjects module.

The SNR (Signal-to-Noise Ratio) was calculated by dividing the intensity value of the protein by that of the background. Protein and background intensity values were averaged for each FOV (Field of View). The percentage of positive cells was calculated by dividing the number of cells positive for the protein of interest by the total number of cells detected in the FOV. All values used in the data were taken from the CellProfiler pipeline.

Manual counting

To validate the CellProfiler pipeline, positive cells were manually counted with the Cell Counter plugin on the open-source software ImageJ. Tissue Imager imagers were

cropped to 700 x 700 pixels with a total of 6 field-of-views in various regions of the sample. The DAPI channel and the fluorescent channel labeling the protein of interest were merged using ImageJ. Any cell with the fluorescent signal indicative of the presence of the protein was manually marked as a positive cell with a single dot in the image and counted by the Cell Counter. The percentage of cells positive obtained via manual counting was then compared to the percentage of cells positive from the CellProfiler pipeline.

Statistical Analysis

Student (two-sided) t-tests were performed for the comparison between manual counts and CellProfiler counts. For Figure 3.3, each fluorescent bead population had 13 beads selected randomly throughout the image for quantification on ImageJ. For Supplementary Figure 5, the P values for the student t-test between counting methods were 0.76, 0.83, and 0.22 for CD3e, CD8, and CD14 respectively.

3.6 Acknowledgement

We would like to thank Michael Neel for help with imaging H&E. This work was funded by a U54CA217378 grant to the UCI Cancer Systems Biology Center (CaSB@UCI), NIH/NIAMS P30AR075047 Skin Biology Resource Center grant, a seed grant from UCI Precision Health through Artificial Intelligence Initiative, a P30CA062203 cancer center support grant to the UCI Chao Family Comprehensive Cancer Center, and NIH/NIGMS R21GM135493 to P.N.H. J.G. was supported by a UCI Immunology NIH T32 training grant AI 060573 and National Institute of Neurological Disorders and Stroke (NINDS/NIH) Training Grant NS082174. We thank Arvetas Biosciences for contributing

instrument parts for the prototype development. Figure 1 was partially created using BioRender.com.

3.7 Author contribution

J.G. designed, conducted, and analyzed the experiments. P.N.H. designed and J.G. and P.N.H constructed the Tissue Imager. J.G., W.Z., and P.N.H. wrote the manuscript. H.J. and C.W. helped conduct experiments. J.S. and A.G. provided technical support and consulted on the study. W.Z. and P.N.H. directed the project.

Competing interests

A.G. is a founder of Alyra Therapeutics. W.Z. is a cofounder of Velox Biosystems Inc., Amberstone Biosciences Inc., and Arvetas Biosciences Inc. P.N.H. holds a part-time position at Amberstone Biosciences Inc.

3.8 Supplementary Material

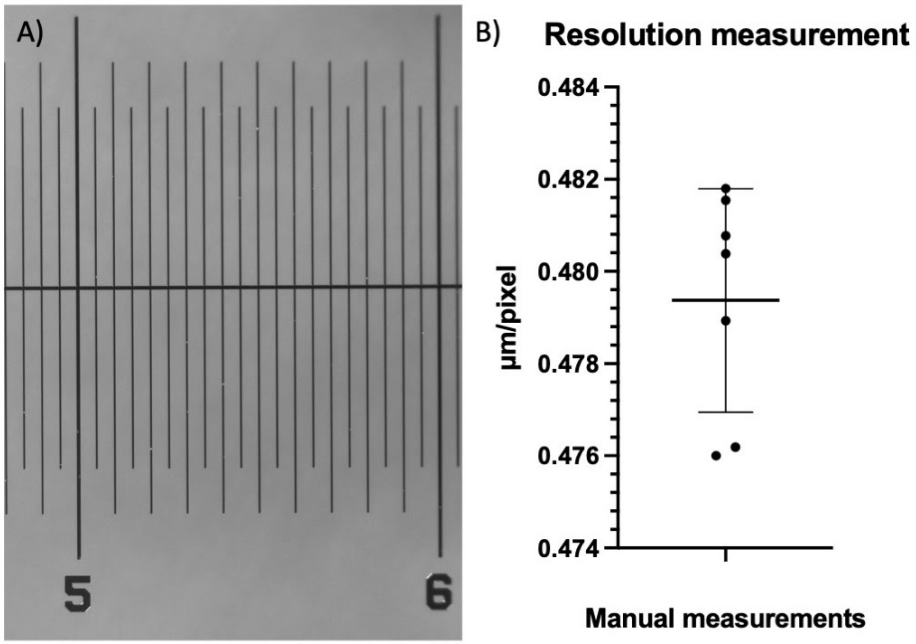


Figure 3.6 Supplementary Figure 1: Resolution calibration for Tissue Imager
A) Image of the Thorlabs 10mm ruler with 50µm spacing between the divisions B) Measurements plotted with an average manual measurement (n=7) of 0.48µm/pixel on the Thorlabs 10mm ruler. Scatter plot was plotted with the mean and standard deviation.

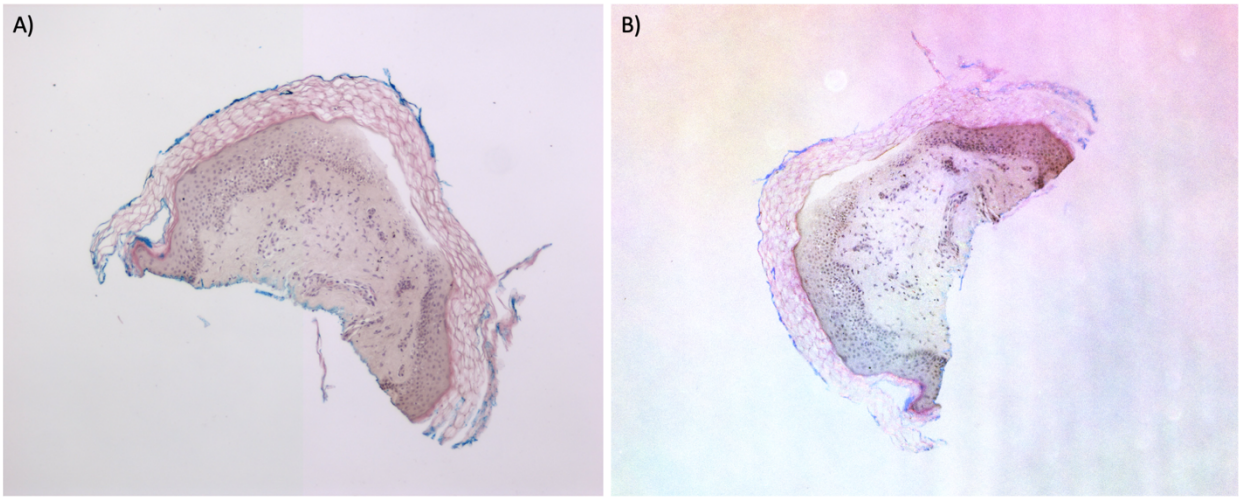


Figure 3.7 Supplementary Figure 2: Brightfield image on Tissue Imager with RGYB imaging
H&E staining on human CTCL skin FFPE tissue. A) Imaged on Nikon Eclipse E400 microscope. B) RGYB imaging on Tissue Imager.

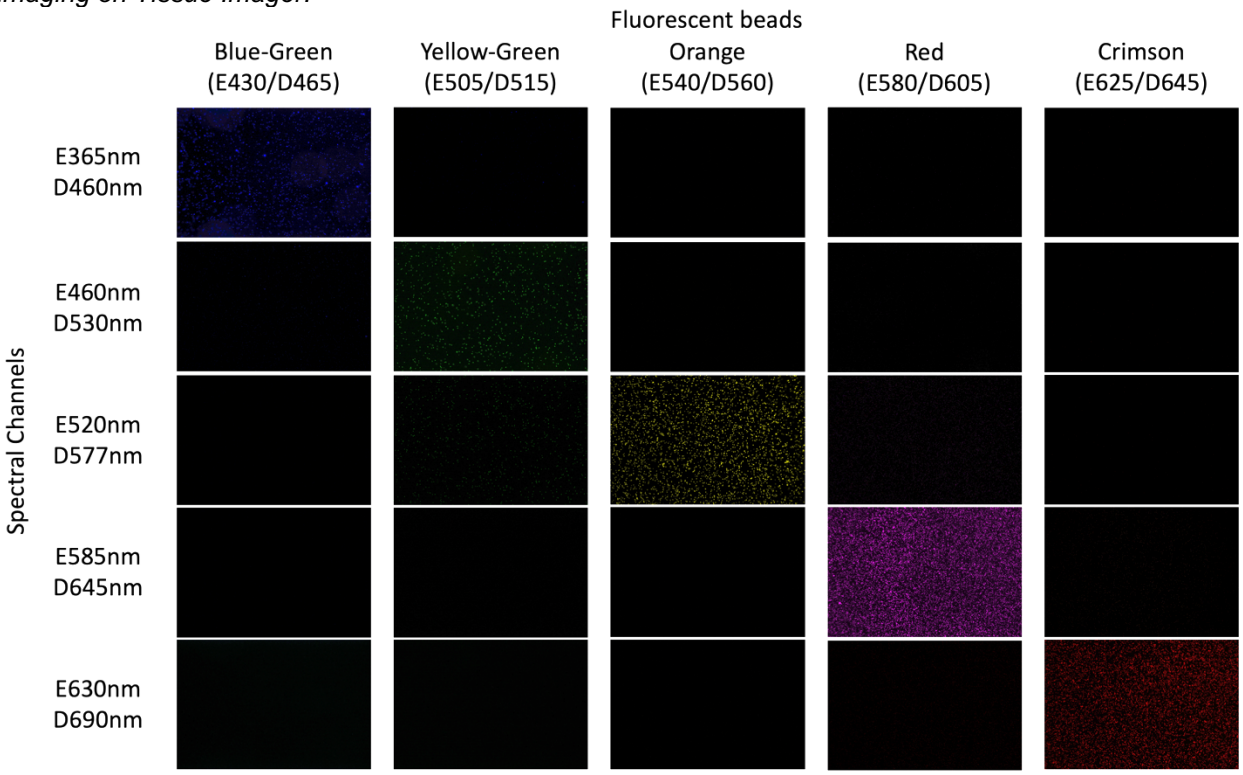


Figure 3.8 Supplementary Figure 3: Fluorescence bead and spectral channel calibration for Tissue Imager
Each 1 µm fluorescent bead population (column) was imaged in each spectral channel on the Tissue Imager (row) to validate the absence of crosstalk. The beads were Blue-Green (E430nm/D465nm), Yellow-Green (E505nm/D515nm), Orange (E540nm/D560nm), Red (E580nm/D605nm), and Crimson (E625nm/D645nm)

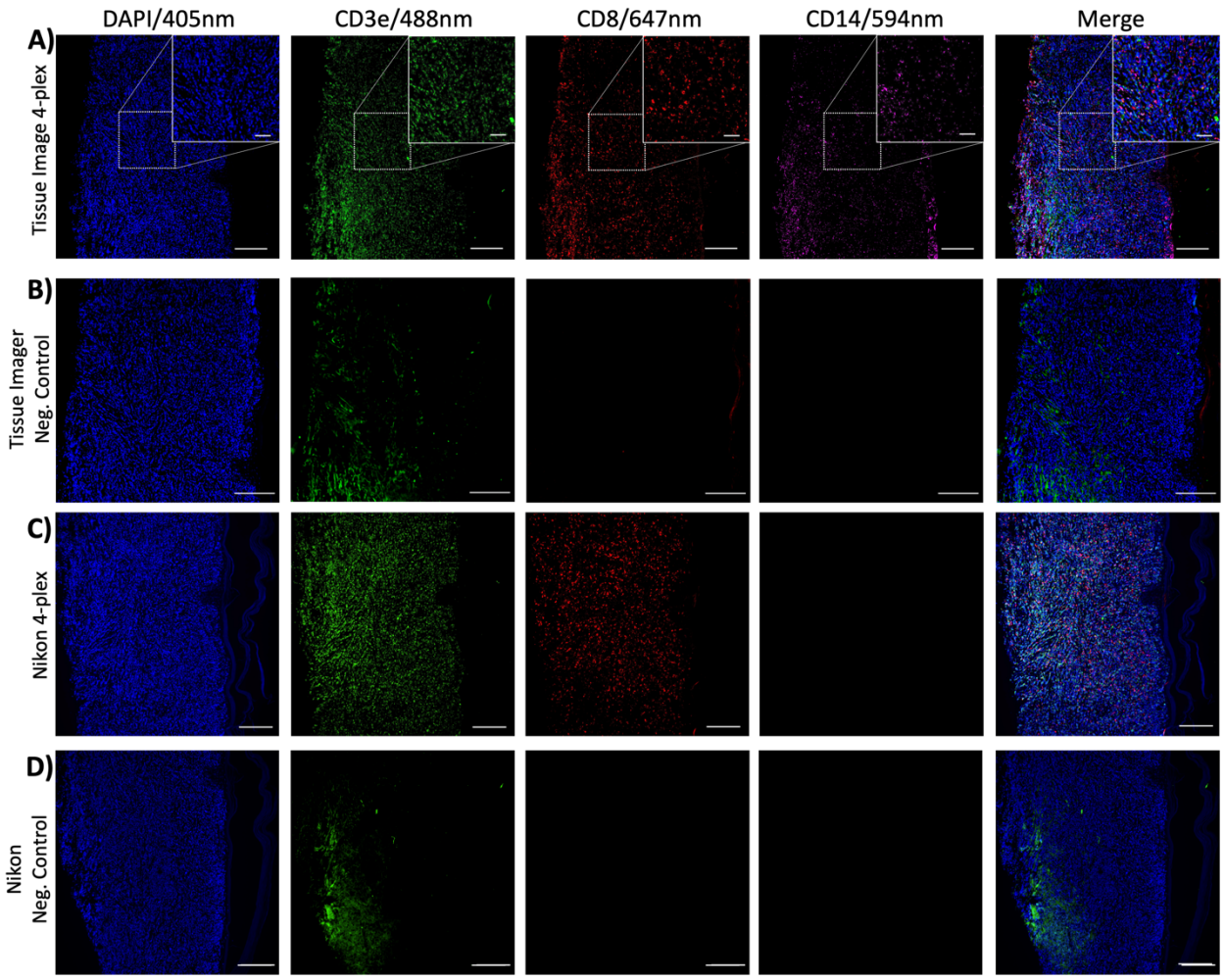


Figure 3.9 Supplementary Figure 4: 4-plex panel with corresponding controls

A) The 4-plex detection of DAPI, CD3e, CD8, CD14 in human CTCL skin FFPE tissue imaged on the Tissue Imager. B) The negative control of secondary fluorophore antibody staining only on imaged on the Tissue Imager. C) The 4-plex detection of DAPI, CD3e, CD8, CD14 in human CTCL skin FFPE tissue imaged on the Nikon microscope. B) The negative control of secondary fluorophore antibody staining only on imaged on the Nikon microscope. Scale bar is 200 μm.

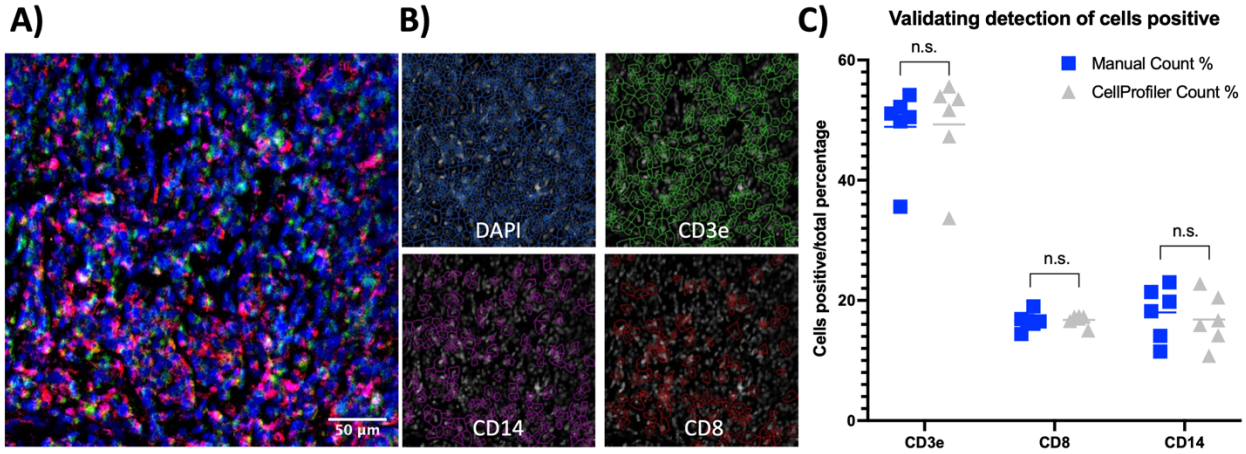


Figure 3.10 Supplementary Figure 5: Manual counting to validate CellProfiler detection
A) Cropped 700 x 700 pixel merged image of 4-plex (DAPI, CD3e, CD8, CD14) on human skin cutaneous T-cell lymphoma FFPE tissue. B) CellProfiler outline of detected cells with DAPI, CD3e, CD8, and CD14. C) Plot comparing manual counts to CellProfiler counts of CD3e, CD8, and CD14 in cropped images (n=6). The mean was plotted as a line. Pairwise T-test between manual counting and CellProfiler counting for each marker was not significant (P=0.7549, 0.8246, 0.2237 for CD3e, CD8, CD14 respectively).

3.9 References

1 Pollack, L. A. *et al.* Melanoma survival in the United States, 1992 to 2005. *J Am Acad Dermatol* **65**, S78-86, doi:10.1016/j.jaad.2011.05.030 (2011).

2 Bohndiek, S. E. & Brindle, K. M. Imaging and 'omic' methods for the molecular diagnosis of cancer. *Expert Rev Mol Diagn* **10**, 417-434, doi:10.1586/erm.10.20 (2010).

3 Burlingame, E. A. *et al.* SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning. *Scientific Reports* **10**, 17507, doi:10.1038/s41598-020-74500-3 (2020).

4 Blom, S. *et al.* Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Scientific reports* **7**, 1-13 (2017).

5 Kalra, J. & Baker, J. in *Signal Transduction Immunohistochemistry* 237-251 (Springer, 2017).

6 Duraiyan, J., Govindarajan, R., Kaliyappan, K. & Palanisamy, M. Applications of immunohistochemistry. *J Pharm Bioallied Sci* **4**, S307-309, doi:10.4103/0975-7406.100281 (2012).

7 Hester, C. A. *et al.* Comparative outcomes of adenosquamous carcinoma of the pancreas: an analysis of the National Cancer Database. *Journal of surgical oncology* **118**, 21-30 (2018).

8 Parra, E. R. *et al.* Immuno-profiling and cellular spatial analysis using five immune oncology multiplex immunofluorescence panels for paraffin tumor tissue. *Scientific reports* **11**, 1-15 (2021).

9 Stack, E. C., Wang, C., Roman, K. A. & Hoyt, C. C. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods* **70**, 46-58 (2014).

10 Feng, Z. *et al.* Multispectral imaging of T and B cells in murine spleen and tumor. *The Journal of Immunology* **196**, 3943-3950 (2016).
11 Sanchez, K. *et al.* Multiplex immunofluorescence to measure dynamic changes in tumor-infiltrating lymphocytes and PD-L1 in early-stage breast cancer. *Breast Cancer Research* **23**, 1-15 (2021).
12 Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature biotechnology* **38**, 586-599 (2020).
13 Manesse, M., Patel, K. K., Bobrow, M. & Downing, S. R. in *Biomarkers for Immunotherapy of Cancer* 585-592 (Springer, 2020).
14 Pedroza-Gonzalez, A. *et al.* Tumor-infiltrating plasmacytoid dendritic cells promote immunosuppression by Tr1 cells in human liver tumors. *Oncoimmunology* **4**, e1008355 (2015).
15 Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science* **359**, 1350-1355 (2018).
16 Sade-Feldman, M. *et al.* Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998-1013. e1020 (2018).
17 Amaria, R. N. *et al.* Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nature medicine* **24**, 1649-1654 (2018).
18 Toki, M. I. *et al.* High-plex predictive marker discovery for melanoma immunotherapy–treated patients using digital spatial profiling. *Clinical Cancer Research* **25**, 5503-5512 (2019).
19 Brede, C. *et al.* Mapping immune processes in intact tissues at cellular resolution. *The Journal of clinical investigation* **122**, 4439-4446 (2012).
20 Bagherani, N. & Smoller, B. R. An overview of cutaneous T cell lymphomas. *F1000Research* **5** (2016).
21 Tarabadkar, E. S. & Shinohara, M. M. Skin directed therapy in cutaneous T-cell lymphoma. *Frontiers in Oncology* **9**, 260 (2019).
22 Berti, E. *et al.* Primary cutaneous CD8-positive epidermotropic cytotoxic T cell lymphomas: a distinct clinicopathological entity with an aggressive clinical behavior. *The American journal of pathology* **155**, 483-492 (1999).
23 Breslauer, D. N., Maamari, R. N., Switz, N. A., Lam, W. A. & Fletcher, D. A. Mobile phone based clinical microscopy for global health applications. *PloS one* **4**, e6320 (2009).
24 Collins, J. T. *et al.* Robotic microscopy for everyone: the OpenFlexure microscope. *Biomedical Optics Express* **11**, 2447-2460 (2020).
25 Wei, Q. *et al.* Fluorescent imaging of single nanoparticles and viruses on a smart phone. *ACS nano* **7**, 9147-9155 (2013).
26 Yelleswarapu, V. R., Jeong, H.-H., Yadavali, S. & Issadore, D. Ultra-high throughput detection (1 million droplets per second) of fluorescent droplets using a cell phone camera and time domain encoded optofluidics. *Lab on a Chip* **17**, 1083-1094 (2017).
27 Sung, Y., Campa, F. & Shih, W.-C. Open-source do-it-yourself multi-color fluorescence smartphone microscopy. *Biomedical optics express* **8**, 5075-5086 (2017).
28 Tokura, Y., Yagi, H., Seo, N., Takagi, T. & Takigawa, M. Nonerythrodermic, leukemic variant of cutaneous T-cell lymphoma with indolent clinical course: Th2-

type tumor cells lacking T-cell receptor/CD3 expression and coinfiltrating tumoricidal CD8+ T cells. *Journal of the American Academy of Dermatology* **43**, 946-954 (2000).

29 Pujol, B. F. *et al.* Endothelial-like cells derived from human CD14 positive monocytes. *Differentiation* **65**, 287-300 (2000).

30 Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, 1-11 (2006).

31 Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Next-generation computational tools for interrogating cancer immunity. *Nature Reviews Genetics* **20**, 724-746 (2019).

CHAPTER 4

TISSUE OPTIMIZATION

Contribution: J.G. designed and performed the experiments for this chapter.

4.1 Introduction

With transcriptomic studies becoming more common in the fields of biological sciences and clinical research, the success of all transcriptomic studies on tissue samples depends on proper tissue preservation and processing. If the tissue samples are not properly processed or stored, the integrity of the RNA within the sample will degrade and impact the results of the transcriptomic detection [1]. For many years, formalin-fixed, paraffin-embedded (FFPE) tissue samples have been standard in the field of histology and immunostaining. Another common method is optimal cutting temperature (OCT). The goal of both methods is to preserve the tissue sample. However, many factors, such as aim of study, storage conditions, available equipment, preparation time, and more, should be considered prior to deciding which method to use. Table 6 compares these 2 methods.

Table 6 Tissue preservation methods comparison

	FFPE	OCT
Tissue preservation	Fixation through perfusion or immersion in 4% paraformaldehyde (PFA) or 10% neutral buffer formalin (NBF)	Snap freezing
Equipment	Microtome and tissue cassettes	Cryostat microtome
Preparation time	~2 days; dehydration in ethanol, clearing in xylene or histoclear, paraffin wax infiltration	<1 day; freeze in OCT media on dry ice, acclimate in cryostat before sectioning
Standard section thickness	4-7 μm	1-100 μm
Storage	Dry on glass microscope slide at room temperature with desiccate or 4°C	Fresh in -80°C freezer
Antigen masking	Medium	Low
Pros	Easy storage and versatile	Doesn't require fixing and low antigen masking
Cons	Harder to produce thinner sections	Delicate sections and requires cold temperature

The RNA integrity of tissue samples can be impacted by many factors during the tissue processing steps, including the fixation process, embedding, and storage over long periods of time [2-6]. During the fixation process, fragmentation of nucleic acid, modification by chemical reactions between formaldehyde and nucleic acids, and crosslinking with proteins and biomolecules are some of the factors that can impact the RNA integrity [2]. Cellular processes and tissue autolysis can cause the degradation of RNA. High temperatures during the embedding process and extended storage of embedded samples can lead to increased fragmentation. Prolonged storage of sectioned tissue can lead to oxidation of RNA on the exposed surface as well.

4.2 Colon cancer SW480 xenograft optimization

In collaboration with Dr. Marian Waterman, we investigated the heterogeneity of Wnt signaling within colon cancer SW480 cells. The RNA-seq data from Dr. Waterman's lab indicated that there were 2 populations, one positive for the gene ROBO1 and one negative for ROBO1. A panel of markers were then identified to be highly expressed in

either the ROBO1+ or ROBO1- population. We worked on developing an mRNA panel using MOSAICA to spatially profile gene expression and detect the patterns of heterogeneity.

The Waterman lab developed a SW480 xenograft model that would be used for this heterogeneity study. For the MOSAICA platform to accurately detect mRNA transcripts in the xenograft FFPE model, the xenograft tissue processing required optimization. In addition to using POLR2A and UBC housekeeping genes and a negative control probe with our MOSAICA platform, RNAscope was also used to benchmark and assess RNA integrity of the tissue sample. Preliminary data showed certain regions of the negative control sample having clusters of bright signals brighter than the positive control, indicating that optimization in the fixation process, tissue processing, target retrieval, and protease treatment. The goal was to obtain good signal of target probes and decrease non-specific binding.

To optimize the fixation and tissue processing steps, various time points for each step were tested with the batch of subcutaneous SW480 xenograft tumors harvested. With 4 mice harvested for the 26- and 30-day post-injection timepoints and 2 (1 with only 1 tumor) mice harvested at the 33-day post-injection timepoint and 2 tumors per mice, a total of 19 tumors were harvested (Figure 4.1a-c). Tumors were fixed in 10% NBF for 16.5hrs, 21hrs, 24hrs, 27.5hrs, or 28.5hrs. Each condition has 2 tumors as replicates. Tissue samples were then embedded into the tissue cassette and sectioned with a microtome. The tumors were measured and weighed when they were harvested, and the volume and mass were calculated. As seen in Figure 4.1d-f, the tumors were similar in mass and volume.

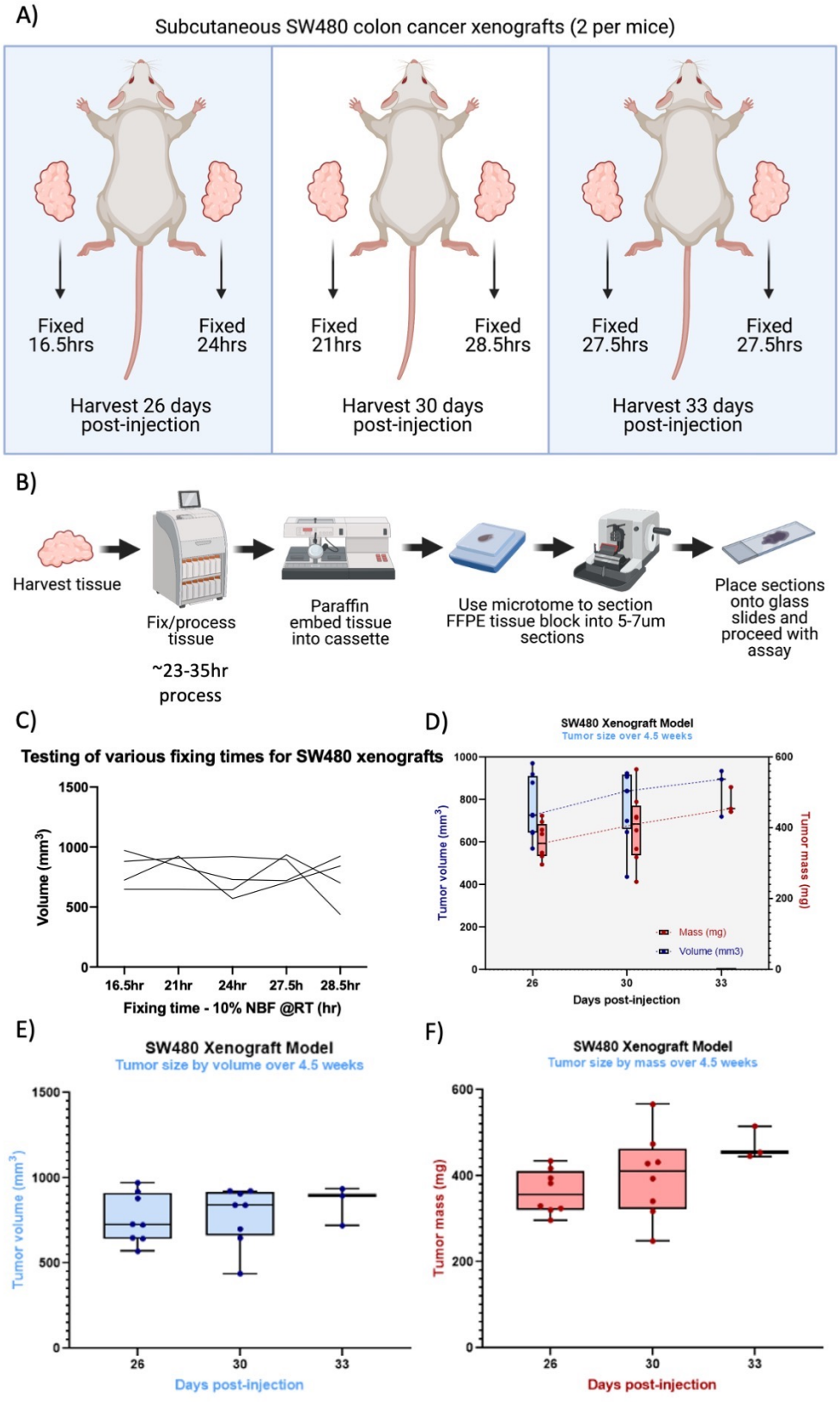


Figure 4.1 Harvesting subcutaneous colon cancer SW480 xenografts for FFPE processing

Using POLR2A as the positive control and dapB as the negative control on the RNAscope platform, the 21hour fixation time was determined to be the most optimal in terms of signal-to-noise (SNR) ratio (Figure 4.2). However, the SNR was lower than the SNR obtained on previous FFPE tissue samples from the pathology center. This indicated that further optimization would be needed downstream, including target retrieval and protease treatment (Figure 4.3). Using Table 7, the images were analyzed to optimize the target retrieval and protease treatment based on the morphology and background signals. After performing various combinations of target retrieval and protease treatment times (Figure 4.4), the optimal condition of 10 minutes target retrieval and 15 minutes protease treatment was selected for the colon cancer xenografts.

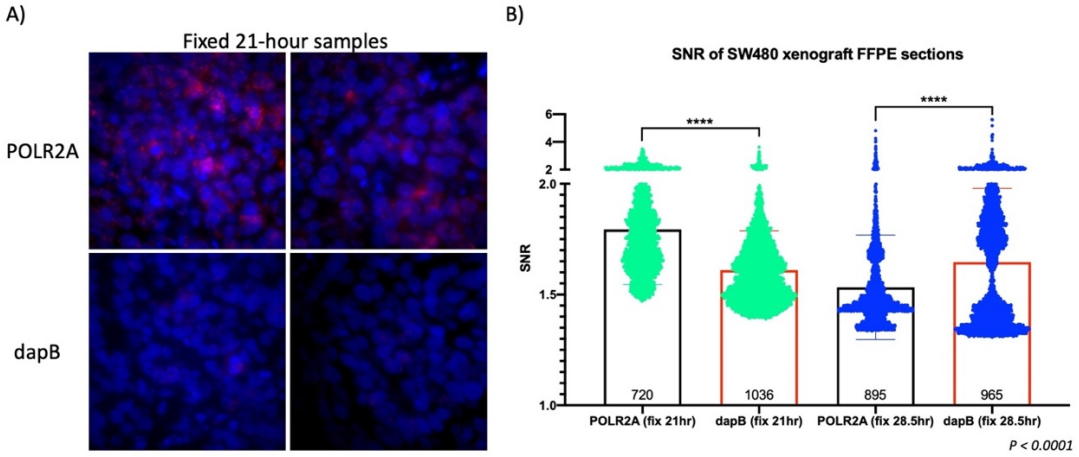


Figure 4.2 Fixation optimization results signal-to-noise analysis
A) Images of POLR2A (positive control) and dapB (negative control) on the RNAscope platform for the 21-hour samples in 133x 133 µm FOV. B) Signal-to-noise ratio (SNR) of 21-hour sample compared to 28.5-hour samples with student-t test analysis (P<0.0001).

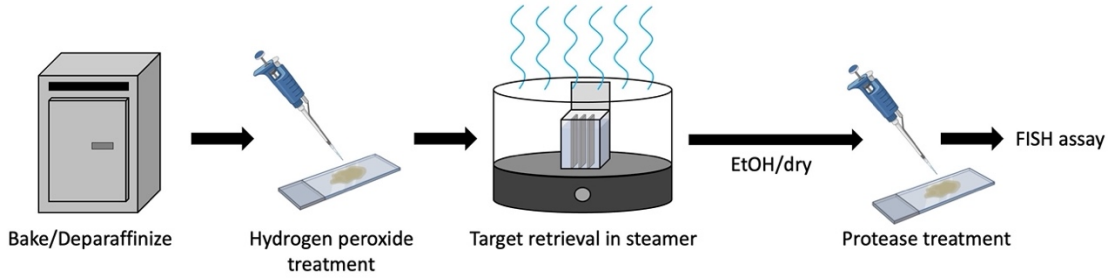
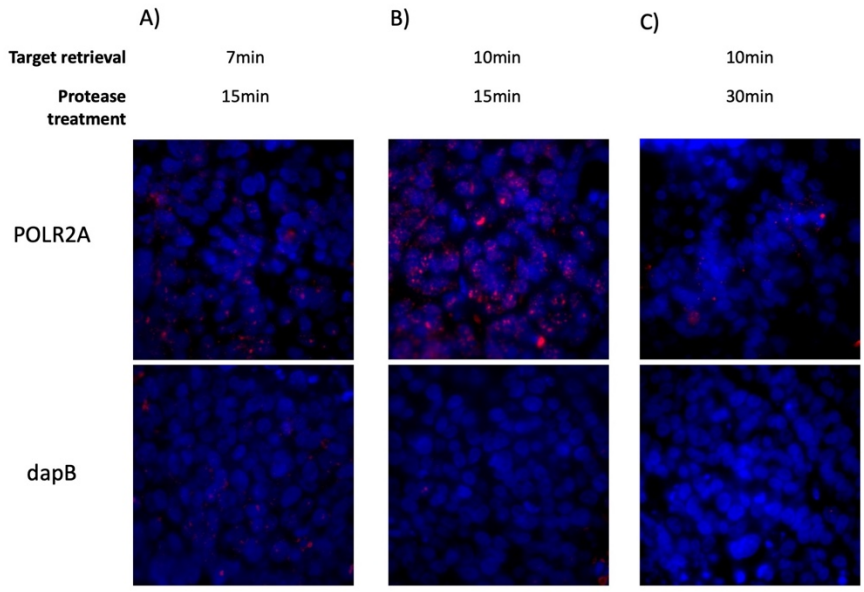


Figure 4.3 Tissue processing workflow

Table 7 Tissue processing optimization guidelines

Image description	Possible cause	Possible effect/solution
Only strong nuclear signal	Under-digested	Weak/no signal due to poor probe accessibility. Increase boiling and/or protease.
Destroyed tissue morphology, donut/ghost nuclei	Over-digested	High background, nonuniform strong/weak signal. Decrease boiling and/or protease.
Intact tissue morphology and nuclei, homogenous H&E staining	Optimal digestion	High SNR, strong positive control with no/negligible background
Good morphology, weak/no signal and low SNR	Over-fixed/under-digested tissue	
Poor tissue morphology (faded with loss of cell borders), loss of RNA	Under-fixed/over-digested tissue	Protease over-digestion leads to loss of RNA
Cytoplasmic background	Tissue dried up during assay	Make sure tissue is completely covered with solution and humidity box is wet.
Extracellular background	Incomplete paraffin removal	Use fresh/new EtOH and histoclear. Agitate slides during incubations.
Tissue detached from slide	Suboptimal tissue prep or wrong type of slide	Bake longer (up to overnight), reduce boiling time, and use SuperFrost Plus slide.

Modified from RNAscope manual



D)

Target retrieval (min)	Protease plus (min)
7	15
7	30
10	15
10	30
15	15
15	30

Figure 4.4 Target retrieval and protease treatment optimization
A-C) Various target retrieval and protease treatment times were tested with POLR2A (positive control) and dapB (negative control) on the RNAscope platform. D) The conditions that were tested.

4.3 Colon cancer SW480 xenograft EGFR staining demonstrating heterogeneity

The colon cancer SW480 xenografts were known to express both ROBO+ and ROBO- cell types. To spatially visualize the heterogeneity of these 2 populations, an EGFR staining was performed using antibodies. Based on the sequencing data, the ROBO- population was known to express EGFR whereas the ROBO+ population did not. This was confirmed with the colon cancer SW480 xenograft staining of EGFR in the FFPE tissue fixed for 21 hours (Figure 4.5). For future work, the ROBO+ population will also be stained at the same time with FGFR4, with is specific to the ROBO+ population based on sequencing data.

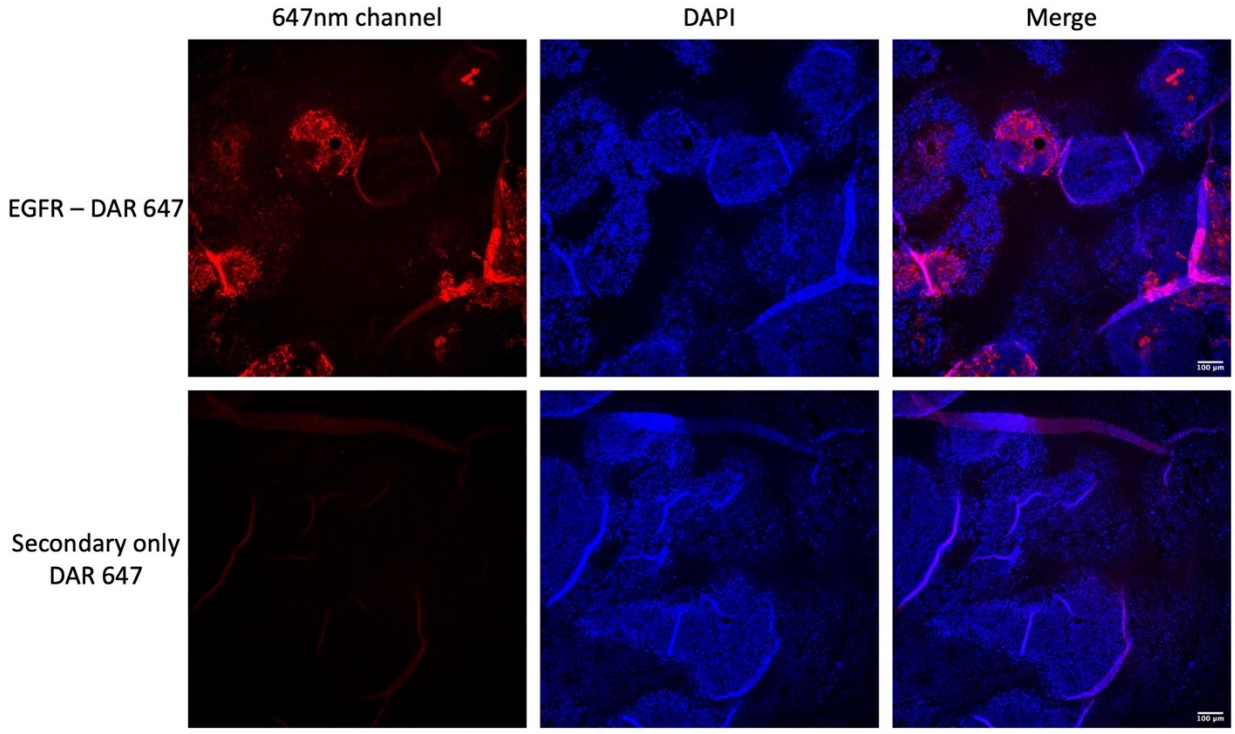


Figure 4.5 EGFR staining in colon cancer SW480 xenograft FFPE tissue
The top row with the EGFR antibody staining paired with donkey anti-rabbit 647 and bottom row with the negative control of secondary antibody only.

4.4 Reducing autofluorescence in tissue samples

Autofluorescence, a common issue when performing immunofluorescence in tissue samples, is caused by sources such as aldehyde fixatives, tissues components with endogenous fluorescence (ECM proteins, red blood cells, macrophages), and lipofuscins (autofluorescent granules of oxidized proteins and lipids that build up in lysosomes of cells with age). The autofluorescence is typically most abundant in the blue and green wavelengths, masking the fluorescence signals from the antibodies stained. Tissue fixation can be a major factor that causes autofluorescence if the process is not optimized. Hence optimizing the fixation process can reduce a significant amount of autofluorescence. To further reduce autofluorescence, autofluorescence quenchers are typically used. While there are several methods to reduce autofluorescence in tissue samples, most methods like autofluorescence quenchers introduces non-specific background fluorescence in the red and far-red channels. For example, Sudan Black B, a lipophilic dye, masks autofluorescence from lipofuscins, but also introduces significant background fluorescence in the red and far-red channels. An alternative is TrueBlack, a commercial autofluorescence quencher that does not introduce as much background fluorescence in the red and far-red channels.

As autofluorescence has its own unique lifetime characteristics, it can be detected on the phasor plot. To determine if fluorescence lifetime imaging microscopy (FLIM) could help reduce the autofluorescence in the image after processing, human cutaneous T-cell lymphoma (CTCL) skin FFPE tissue, known to have significant autofluorescence in the stratum corneum was stained with antibodies, with and without TrueBlack, and processed with and without lifetime analysis (Figure 4.6A-D). As seen in Figure 4.6A and C,

TrueBlack reduces autofluorescence. However, removing autofluorescence with lifetime analysis removed more autofluorescence (Figure 4.6B and C). The populations of each fluorophore and autofluorescence could be detected on the phasor plot with the fluorophore populations near the expected lifetime (Figure 4.6 E-G). Interestingly, TrueBlack separated the fluorescence and autofluorescence populations on the phasor plot more, making lifetime analysis easier.

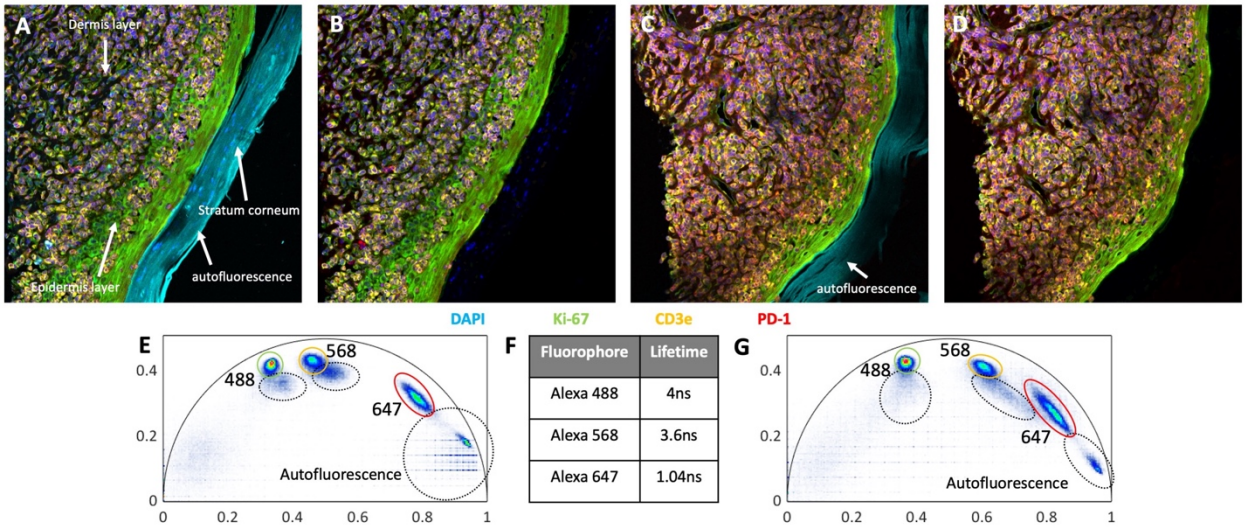


Figure 4.6 Removing autofluoresence in tissue samples with TrueBlack and FLIM
Human Cutaneous T-cell Lymphoma (CTCL) FFPE tissues were stained with Ki-67, CD3e, and PD-1 antibodies and DAPI staining. Images were taken on the Leica SP8 Falcon microscope with FLIM capabilities. The images were taken contain a view of the stratum corneum, epidermis, and dermis layer. A) A 4-plex detection without TrueBlack or FLIM analysis. B) A 4-plex detection without TrueBlack, with FLIM analysis to remove autofluorescence. C) A 4-plex detection with TrueBlack, without FLIM analysis. D) A 4-plex detection with TrueBlack and FLIM analysis to remove autofluorescence. E) Phasor plot of A/B with each population labeled (black circles = autofluorescence). F) Expected lifetime of each fluorophore in solution. G) Phasor plot of C/D with each population labeled (black circles = autofluorescence).

4.5 Methods

FFPE tissue fixation

The harvested tumor xenograft tissues were measured and weighed, placed into a pre-labeled tissue cassette, then submerged in 100mL of 10% neutral buffered formalin (NBF). Parafilm was used to wrap and seal the opening of the container with NBF to

prevent excess oxygen reactions. Once the fixation time was complete, the sample was immediately transferred to the tissue processor with the program below.

Tissue processing with Leica TP1020 automated processor (has 12 bins only):

Step	Reagent	Time (min)	Temp
1	70% EtOH	30	25°C
2	80% EtOH	30	25°C
3	90% EtOH	30	25°C
4	95% EtOH	30	25°C
5	100% EtOH	30	25°C
6	100% EtOH	30	25°C
7	100% EtOH	30	25°C
8	Histoclear	30	25°C
9	Histoclear	30	25°C
10	Histoclear	30	25°C
11	Paraffin	45	60°C
12	Paraffin	45	60°C
13	Paraffin	45	60°C

FFPE tissue processing

The FFPE tissues were sectioned to 5 μm slices using a rotary microtome, collected in a water bath at 35 °C, and mounted to positively charged Fisher super frost coated slides. The tissue sections were then baked at 60 °C for 1 hr 10 mins. For antigen unmasking, slides were deparaffinized, rehydrated then followed by target retrieval (with citrate buffer).

RNAscope™

The fixed tissue pretreatment included treatment with hydrogen peroxide and protease plus. The RNAscope™ assay was then performed using the RNAscope™ Multiplex Fluorescent V2 kit and Akoya Cy5 TSA fluorophore. The positive control (POLR2A) and negative control (dapB) were in C1.

Protein staining

The sample was blocked in 10% BSA for 2 hours at room temperature, incubated with the EGFR antibody (Clone D38B1) for 16 hours at 4 °C at a concentration of 0.2µg/ml, then incubated with the donkey anti-rabbit 647 secondary antibody for 1 hour at room temperature at a concentration of 10µg/ml.

Imaging

Images were acquired with an inverted Nikon Ti-1000E using a 10x plan apochromat oil objective with a numerical aperture of 0.45. Samples were illuminated with a Spectra-X (Lumencor) LED light source at the 395 nm, 470 nm, and/or 640 nm excitation wavelengths. Images were acquired with an Andor Zyla 4.2 sCMOS camera.

Image analysis

The images obtained from the Nikon microscope were then analyzed in CellProfiler to detect the puncta and quantifying the intensity of each while also obtaining the background intensity. The signal-to-noise ratio (SNR) was calculated with the values of fluorescence intensity from each puncta and the background.

4.6 References

1. Von Ahlfen, S., et al. (2007). "Determinants of RNA quality from FFPE samples." PLoS One **2**(12): e1261.
2. Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M., & Okubo, K. (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic acids research*, *27*(22), 4436-4443.
3. Bresters, D., Schipper, M. E. I., Reesink, H. W., Boeser-Nunnink, B. D. M., & Cuypers, H. T. M. (1994). The duration of fixation influences the yield of HCV cDNA-PCR products from formalin-fixed, paraffin-embedded liver tissue. *Journal of virological methods*, *48*(2-3), 267-272.
4. Macabeo-Ong, M., Ginzinger, D. G., Dekker, N., McMillan, A., Regezi, J. A., Wong, D. T., & Jordan, R. C. (2002). Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. *Modern pathology*, *15*(9), 979-987.
5. Bhudevi, B., & Weinstock, D. (2003). Detection of bovine viral diarrhea virus in formalin fixed paraffin embedded tissue sections by real time RT-PCR (Taqman). *Journal of virological methods*, *109*(1), 25-30.
6. Cronin, M., Pho, M., Dutta, D., Stephans, J. C., Shak, S., Kiefer, M. C., ... & Baker, J. B. (2004). Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *The American journal of pathology*, *164*(1), 35-42.

CHAPTER 5

DEVELOPING A CELL-BASED MECHANOSENSOR TO DYNAMICALLY AND QUANTITATIVELY ASSESS THE TISSUE MECHANICS

Contribution: J.G. designed and performed the experiments for this chapter.

5.1 Introduction

Physicians have long used physical palpation to detect cancer nodules [1], but we now know that this stiffness difference is primarily due to collagen density, crosslinking, and fiber linearization within the tumor microenvironment (TME) [2-4]. Increased tissue stiffness facilitates tumor growth, invasion, and metastasis and can affect treatment efficacy and resistance [1, 5]. However, the mechanisms that underlie how tumor cells interact reciprocally with their mechanical environment during cancer progression remain incompletely understood [6, 7]. One critical barrier to further understanding stiffness-mediated tumor progression lies in the limitations of current technologies to characterize physical properties of tissues at a cellular resolution in their native environment. Imaging modalities such as ultrasound and elastography are capable of longitudinal measurements in vivo but have limitations in spatial resolution and feature size [8]. Conversely, AFM, microrheology, and traction force microscopy have high spatial resolution but cannot be used in vivo and only provide surface measurements [9, 10]. Techniques such as FRET-based molecular sensors, magnetic tweezers and optical traps tend to interrogate molecular pathways but fail to characterize mechanobiology at a tissue level [11-14]. Importantly, most existing technologies measure substrate mechanics but cannot directly interrogate cell responses.

Cells constantly sense their biochemical and biophysical surroundings with unparalleled spatiotemporal resolution and precision and can rapidly transduce that information to regulate downstream gene expression through transcriptional programs. For instance, the transcriptional factor YAP(Yes-associated protein)/TAZ(transcriptional co-activator with PDZ-binding motif) preferentially translocates to the nucleus to trigger downstream gene expression in a stiffness-dependent manner [15]. Leveraging cell's innate mechanosensing mechanisms, we developed cell-based mechanosensors that can dynamically and quantitatively assess tumor tissue mechanics in their native environment in vivo. Using CRISPR/Cas9-mediated site-specific genomic integration of modular genetic circuits, we will repurpose known mechano-sensitive transcriptional regulators such as YAP/TAZ to drive downstream expression of reporter fluorescent proteins (Figure 5.1). These fluorescent sensors will be dynamic and reversible and can quantitatively interrogate a wide range of distinct biologically relevant stiffness in a multiplex and continuous fashion.

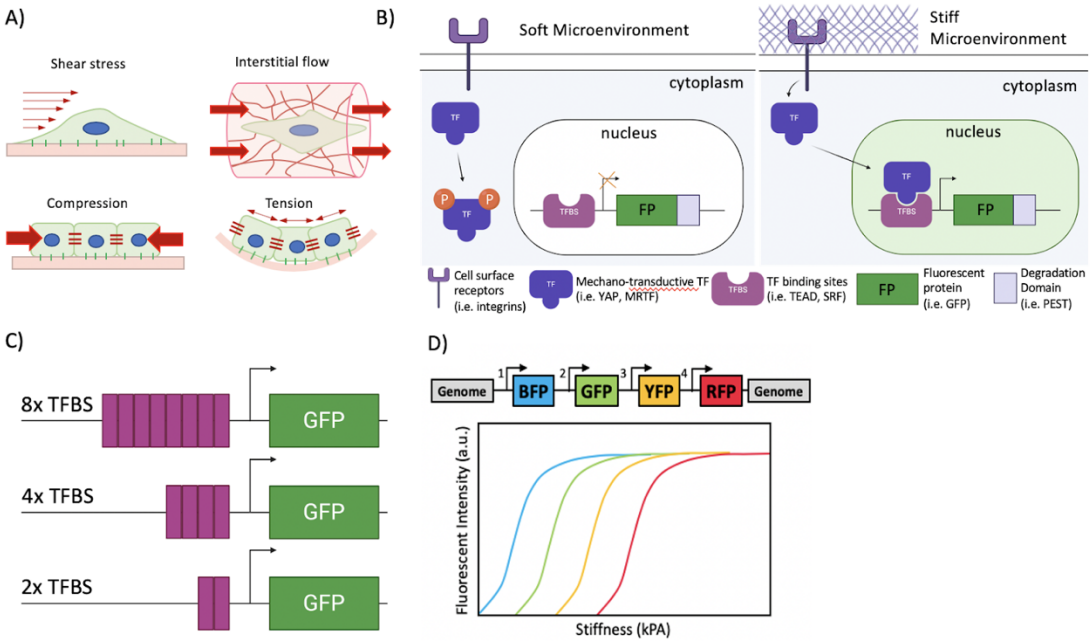


Figure 5.1 Overview schematic of cell-based mechanosensors
A) Examples of mechano forces that cells sense. B) Transcription factor activation in soft and stiff microenvironment. C) Varying number of TFBS repeats to construct singleplex sensors with different stiffness response, D) Long-pass filter multiplex sensor design with inhibition by protease cleavage. P1, P2, P3, proteases for fluorescent proteins; BFP, GFP, YFP, and RFP: blue, green, yellow and red fluorescent proteins, respectively.

A cell-based mechanosensor will reveal what cells actually "feel" in their TME and represent a paradigm-shifting method of dynamically interrogating tumor mechanics during disease progression or treatment over time at a cellular resolution in vivo. With these new capabilities offered by our technology, we can potentially answer outstanding questions in cancer biology and treatment that were previously not possible or difficult to address due to lack of tools. For instance, the cell-based mechanosensor will allow us to construct an ongoing "stiffness map" of the heterogeneous and evolving TME as cancer progresses and to reveal how the dynamic cell-ECM interactions affect the timing and evolution of physiological events such as growth, invasion, and metastasis. New insights of cancer mechanobiology provided by this cell-based mechanosensor will therefore have far reaching clinical significance in 1) using the aberrant tumor mechanical properties ("mechanophenotype") as a diagnostic and prognosis marker to detect small metastases and pre-metastatic legions and to predict cancer invasiveness, respectively, and 2) in improving efficacy and reducing resistance of current drugs by normalizing the tumor niche as well as in developing new drugs that directly target pathological stiffening. In fact, our cell-based mechanosensor may allow functional screening for drugs that perturb ECM mechanics or cancer cell mechanosensing[16] directly in vivo.

Our approach to study cancer mechanobiology shifts the paradigm from characterizing ECM mechanics as most existing techniques due to equipping cancer cells with mechanosensors to reveal how they sense and respond to TME cues. Incorporating

mechanosensors with transcriptional programs allows us to examine how cells integrate mechanical and other cues together and translate that information into functional readouts. To our knowledge, our cell-based mechanosensors, for the first time, will allow one to interrogate dynamic cell-ECM interactions in the native, 3D TME in vivo in longitudinal studies. It also enables dissection of cellular level variability and survey of the biomechanical landscape across the entire tumor tissue through multiscale, temporal and spatial analysis.

This project integrates several innovative tools from synthetic biology, cell engineering, biomaterials, and noninvasive optical imaging to address important and complex questions in cancer mechanobiology. First, we will adapt emerging tools from mammalian synthetic biology to construct multi-transcriptional unit vectors [17, 18], which will be integrated into specific sites in the genome via CRISPR/Cas9. Genetic circuits composed of multi-transcriptional units allow us to incorporate a series of fluorescent reporters into the same cell where each sensor activates under distinct levels of stiffness. By changing the number of repeats for transcription factor binding sites (TFBS) in each transcriptional unit, we can readily modulate their cooperative interactions with transcription factors and tune their signal amplitude and stiffness sensitivity. Furthermore, proteolytic elements that specifically degrade reporter proteins can be incorporated into the circuit, which allows us to fine-tune the stiffness sensor's dynamics and resolution. Second, our stiffness-tunable hydrogel system enables precise manipulation of mechanical stimuli within the natural TME, which allows us to deconvolute the roles of biophysical cues from other signals in cancer progression. Third, state-of-the-art techniques including noninvasive intravital imaging (longitudinal measurements in vivo),

second-harmonic generation imaging (SHG) (characterizing ECM structure) and combined AFM–fluorescence microscopy (analyzing stiffness and sensor signal of tumor tissue ex vivo) uniquely enable us to evaluate, calibrate and optimize our mechanosensors.

The Zhao lab previously demonstrated mesenchymal stem cells (MSC) engineered with YAP/TAZ synthetic promoter can drive expression of downstream reporters or therapeutics, in a stiffness-dependent manner on tunable hydrogels in vitro and in breast tumor niche in vivo [19]. Dr. Adam Engler and his colleagues demonstrated, in a landmark paper, that tissue and matrix stiffness alone is sufficient to direct cell fate [20]. Recently, Dr. Engler's laboratory has developed tunable and reversible hydrogels both in vitro and in vivo to study how matrix stiffness drives epithelial–mesenchymal transition and metastasis in breast cancer [21]. Our understanding of transcription factors (e.g. YAP/TAZ, MRTFα and TWIST) in mechanotransduction pathways especially in breast cancer, which is the focus of this application, has greatly improved through multiple recently studies [6, 15, 21], allowing us to harness them for transcription-level sensor design. The modular multi-transcriptional unit vector design along with CRISPR-mediated genome insertion, both coming to fruition in synthetic biology, allow us to incorporate a series of fluorescent reporters into the same cell with great control and precision for multiplex sensing.

Our stiffness-inducible synthetic promoters are typically comprised of a specific number of repeat TFBS upstream of a minimal promoter (e.g. YB-TATA). By varying TFBS consensus sequences, number of binding site repeats, and repeat spacing we can control the dynamics and long-term or steady-state expression level of our sensors. TFBS

sequences were identified using a custom computational algorithm that uses widely available ChIP-Seq data to pad the consensus sequence. Each variant will induce nuclear-localized GFP expression and will be assembled using landing pads and integrated using CRISPR/Cas9 to the safe harbor locus, AAVS1 [17]. This precise chromosomal integration of genetic payloads ensures copy number consistency of our sensors [17, 18].

5.2 Results

Our sensors are constructed by repurposing the cell's innate mechanosensing machinery. Triggered by mechanical stimuli (e.g., increased ECM stiffness), mechano-sensitive transcription factors (e.g. YAP/TAZ) translocate to the nucleus and bind to synthetic promoters to drive downstream expression of reporter fluorescent proteins (Figure 5.1). Therefore, individual sensors function by changing the expression level of reporters (fluorescence intensity) in response to different nuclear concentrations of the mechano-sensitive transcription factors (i.e. dosage of mechanical stimuli). Using modular genetic circuits, we will engineer cells with multiple transcriptional units, each with distinct number of repeats for transcription factor binding sites (TFBS). This design augments transcription factor-TFBS interactions in a cooperative fashion and allows our mechanosensors to quantitatively sense distinct stiffness in a multiplex assay. Stiffness reporter specificity and resolution can be improved using a band-pass filter where reporters of high-stiffness will also co-express protease that specifically degrades reporters of lower stiffness.

Using human breast cancer MDA-MB-231 cells and MSCs as model systems, they were each transfected with cloned YAP/TAZ sensors that contained 2x, 4x, or 8x repeats

of the TFBS (Figure 5.2A). The differences in green fluorescence intensities of cells between the different constructs were quantified (Figure 5.2B). As seen in Figure 5.2 A, B, increasing YAP/TAZ TFBSs in the MBA-MB-231 model system led to higher sensitivity. The 8x YAP/TAZ sensors in MDA-MB-231 cells were tested on 1kPa and 42kPa MeHA hydrogels (Figure 5.2 C,D), resulting in the 8x YAP/TAZ sensor being significantly brighter on the stiffer substrate (42kPa). This indicated that the 8x YAP/TAZ sensor was a good candidate for detecting stiff substrate. Figure 5.2 E further demonstrates the 8x YAP/TAZ in MBA-MB-231 being sensitive for the 40kPa region, but not in the 1-20kPa region.

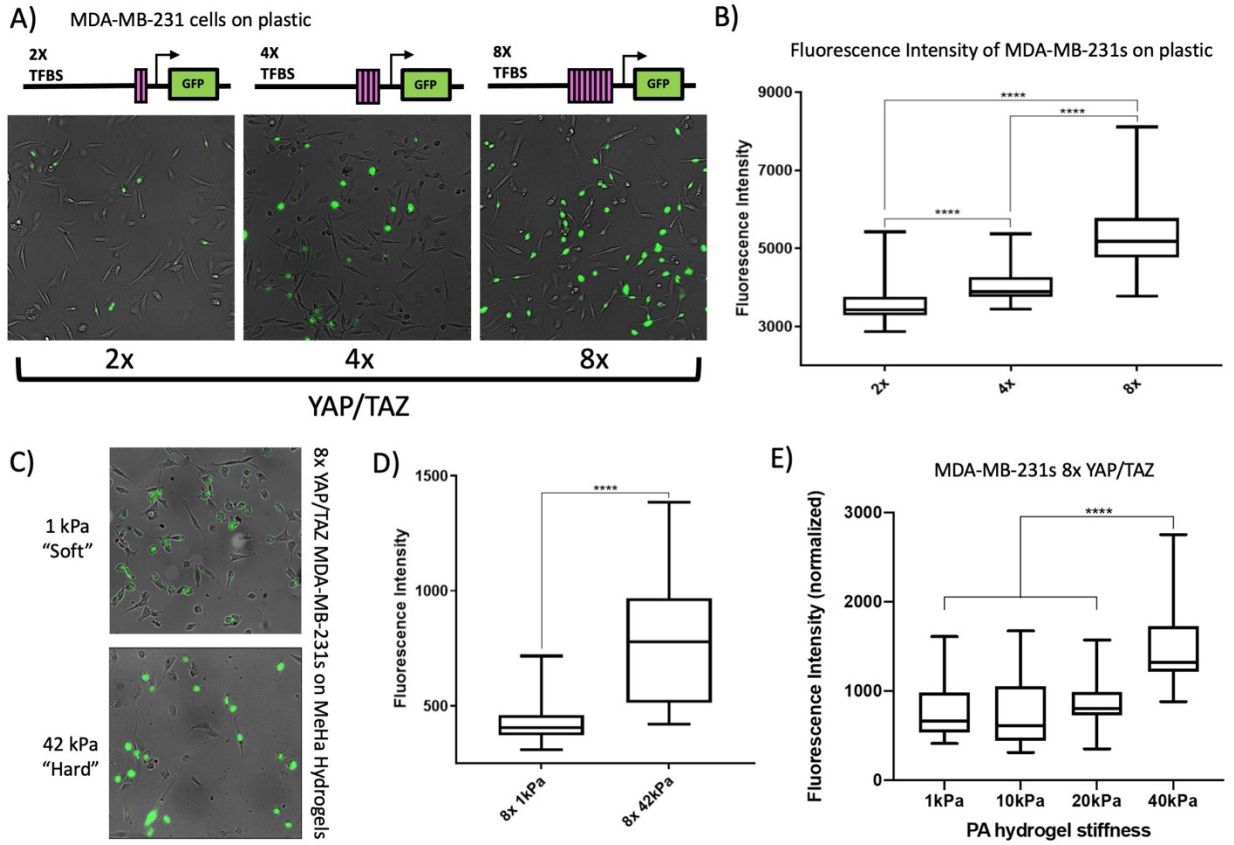


Figure 5.2 Validation of YAP/TAZ mechanosensor
A) Fluorescence images of MDA-MB-231 cells with 2x, 4x, and 8x TFBSs on a 6 well plate (plastic). B) Quantification of 2,4,8x YAP/TAZ sensors on stiff plastic. C) Fluorescence images of MDA-MB-231 cells with 8x YAP/TAZ 1 and 42 kPa MeHa hydrogel. D) Quantification of the 8x YAP/TAZ MDA-MB-231 cells on MeHA hydrogels of 1 and 42 kPA stiffness. E) MDA-MB-231 cells with 8x YAP/TAZ on 1, 10, 20, 40 kPa PA hydrogel stiffnesses. (n>100 for all images) Student t-test P<0.0001

The PEST domain, a sequence rich in proline, glutamic acid, serine, and threonine, was used to act as a signal peptide for protein degradation to prevent oversaturation of the sensor quantification. While the PEST domain was useful in the MBA-MB-231 model system, it caused the fluorescence signal to degrade rapidly in the MSC model system. The fluorescence intensity of cells with and without the PEST domain were compared. The constructs in the MSCs required the removal of the PEST domain as the signal was diminished by the time imaging was performed (Figure 5.3). Based on FACS data, the 8x YAP/TAZ construct in MSCs without the PEST domain (Figure 5.3A) has higher GFP signal than the 8x YAP/TAZ contract in MSCs with PEST domain (Figure 5.3B). Based on quantification of fluorescence microscopy images, the 8x MRTFα construct in MSCs without PEST domain had a significantly higher fluorescence intensity than constructs with PEST domain. To study the fluctuations of the GFP signal in the MSCs with 8x MRTFα construct, a live cell imaging study was performed (Figure 5.4). The live cell study demonstrated that the no PEST construct had more fluctuations during the 23-hour time frame and had an overall increase in signal. The construct with the PEST domain remained significantly lower and had smaller and fewer fluctuations.

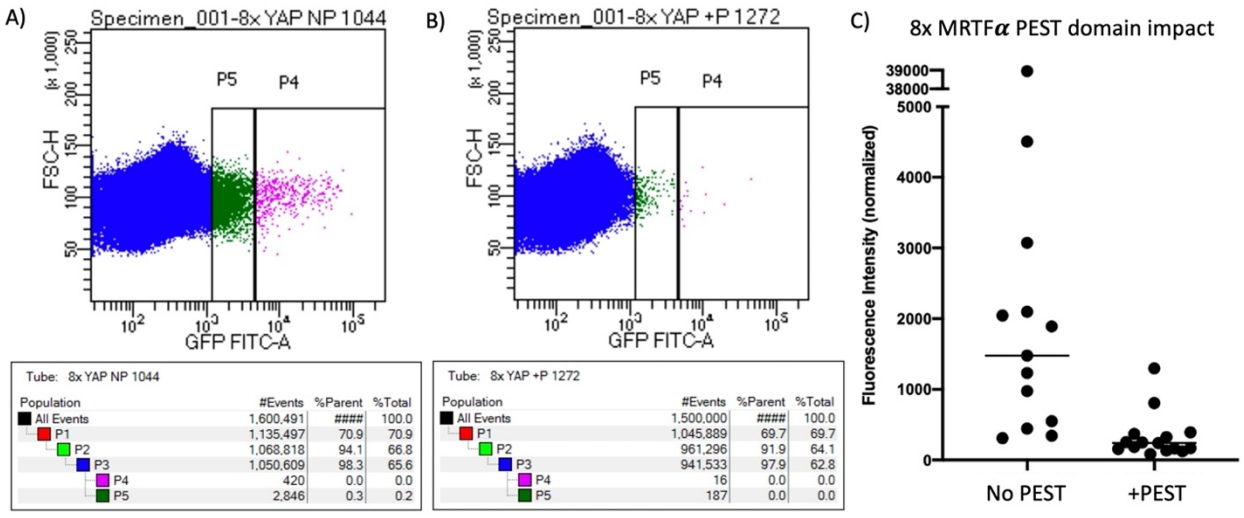


Figure 5.3 Impact of PEST domain on fluorescence intensity in MSCs
A) The 8x YAP/TAZ without PEST domain construct in MSCs FACs data B) The 8x YAP/TAZ with PEST domain construct in MSCs FACs data C) Fluorescence intensity quantification of 8x MRTFα constructs with and without PEST domain in MSCs.

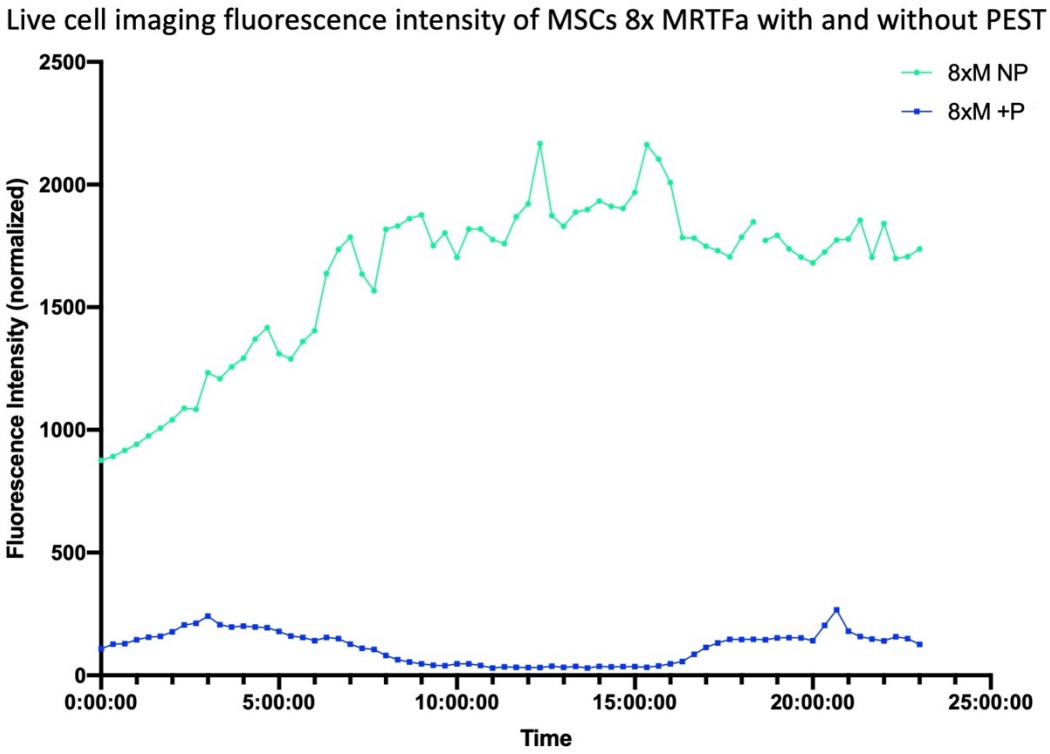


Figure 5.4 MSC 8x MRTFα live cell study on GFP signal fluctuation
The green line was the GFP signal of the 8x MRTFα without PEST and the blue line was the GFP signal of the 8x MRTFα with PEST. The x-axis is the time in hours.

Mechanosensors with 2x, 4x, and 8x of YAP/TAZ, MRTFα, and TWIST were engineered into the MBA-MB-231 model system using nucleofection. For the MSC model, the constructs will be put into lentiviral vectors. If we encounter expression or response heterogeneity, we will purify sensors or create clonal sensor cell lines using fluorescence-activated cell sorter (FACS). Further bioinformatic analyses can be performed to identify transcriptional stiffness response elements that are uniquely and specifically responsive to stiffness but not to nonspecific factors (e.g., hypoxia and inflammation). If our stiffness sensitive promoters do not provide sufficient coverage of biologically relevant stiffness (Young's modulus of 0.1 to >40 kPa), we will change TFBS consensus sequence or

choose other mechanotransductive transcription factors that inherently sense different range of stiffness, e.g. TWIST1 which is sensitive between 0.5 and 2 kPa [21]. Moreover, we will also consider other clonal cell lines to harness different mechanotransduction mechanisms to target a distinct stiffness.

5.3 Methods

Cell engineering

The plasmids, as seen in Supplementary Figure 1, were cloned using Gibson assemblies. After the CAG promoter are a puromycin resistant gene, TEAD binding sites, UNS promoter, TATA box, mNeonGreen, PEST domain, and zeocin resistant gene. These plasmids were nucleofected into the cell lines. After transfection, cells that successfully integrated the plasmids were selected by adding puromycin and zeocin antibiotics into the cell culture media.

Cell culture

Triple negative breast cancer MDA-MB-231 cells were cultured in Leibovitz L-15 medium (L-15; Corning) containing 10% Fetal Bovine Serum (FBS; Seradigm) with 1% Penicillin and Streptomycin (GenClone) and incubated at 37 °C. Each transfection was performed using 2 μg of DNA (1 μg of the engineered plasmid, 0.33 μg of Cas9 plasmid, and 0.66 μg of gRNA plasmid. The cells were selected with puromycin (Invitrogen) and zeocin (Alfa Aesar) at a concentration of 2 μg/mL each and increased to 20 μg/mL.

FACS

Sorting was performed through the UCI Institute for Immunology Flow Core Facility using the BD FACSAria Fusion Sorter with a 70 μm nozzle. After 2 weeks of antibiotic selection, 100,000-1,000,000 cells were collected for sorting. The cells were sorted based

on their expression of mNeonGreen gated at 10^3 au. Data was analyzed using FCS Express 7.

Imaging

Images were acquired with an inverted Nikon Ti-1000E using a 10x plan apochromat oil objective with a numerical aperture of 0.45. Samples were illuminated with a Spectra-X (Lumencor) LED light source at the 395 nm, 470 nm, and/or 640 nm excitation wavelengths. Images were acquired with an Andor Zyla 4.2 sCMOS camera. Live cell imaging was acquired with the samples inside a stage top incubator (WKSM series) to maintain the temperature at 37°C and 5% CO2 during the time study.

Image analysis

The images obtained from the Nikon microscope were then analyzed in CellProfiler to detect the puncta and quantifying the intensity of each while also obtaining the background intensity. The signal-to-noise ratio (SNR) was calculated with the values of fluorescence intensity from each puncta and the background.

5.4 Supplementary information

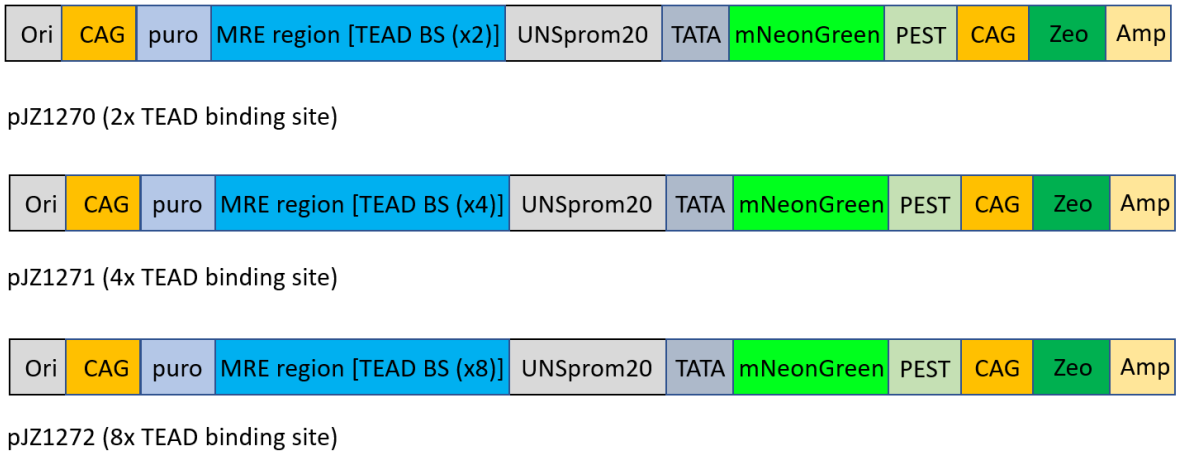


Figure 5.5 S1 YAP/TAZ plasmids altering TFBS 2x-8x

5.5 References

1. Nagelkerke, A., et al., *The mechanical microenvironment in cancer: How physics affects tumours.* Seminars in Cancer Biology, 2015. **35**: p. 62-70.
2. Fattet, L. and J. Yang, *Molecular and Cellular Mechanobiology of Cancer*, in *Molecular and Cellular Mechanobiology*, S. Chien, A.J. Engler, and P.Y. Wang, Editors. 2016, Springer New York: New York, NY. p. 277-290.
3. Maman, S. and I.P. Witz, *A history of exploring cancer in context.* Angiogenesis, 2018. **1501**: p. 1980s.
4. Cox, T.R., et al., *LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis.* Cancer research, 2013: p. canres. 2233.2012.
5. Cukierman, E. and D.E. Bassi, *Physico-mechanical aspects of extracellular matrix influences on tumorigenic behaviors.* Seminars in Cancer Biology, 2010. **20**(3): p. 139-145.
6. Przybyla, L., J.M. Muncie, and V.M. Weaver, *Mechanical Control of Epithelial-to-Mesenchymal Transitions in Development and Cancer.* Annual Review of Cell and Developmental Biology, 2016. **32**(1): p. 527-554.
7. Iskratsch, T., H. Wolfenson, and M.P. Sheetz, *Appreciating force and shape — the rise of mechanotransduction in cell biology.* Nature Reviews Molecular Cell Biology, 2014. **15**: p. 825.
8. Wickramaratne, D., et al., *Fine Needle Elastography (FNE) device for biomechanically determining local variations of tissue mechanical properties.* Journal of biomechanics, 2015. **48**(1): p. 81-88.
9. Keating, M., et al., *Spatial distributions of pericellular stiffness in natural extracellular matrices are dependent on cell-mediated proteolysis and contractility.* Acta Biomaterialia, 2017. **57**: p. 304-312.

10. Roca-Cusachs, P., V. Conte, and X. Trepat, *Quantifying forces in cell biology.* Nature Cell Biology, 2017. **19**: p. 742.
11. Kumar, A., et al., *Talin tension sensor reveals novel features of focal adhesion force transmission and mechanosensitivity.* The Journal of Cell Biology, 2016. **213**(3): p. 371-383.
12. Grashoff, C., et al., *Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics.* Nature, 2010. **466**: p. 263.
13. Chang, C.-W. and S. Kumar, *Vinculin tension distributions of individual stress fibers within cell-matrix adhesions.* Journal of Cell Science, 2013.
14. Oldach, L. and J. Zhang, *Genetically Encoded Fluorescent Biosensors for Live-Cell Visualization of Protein Phosphorylation.* Chemistry & Biology, 2014. **21**(2): p. 186-197.
15. Dupont, S., et al., *Role of YAP/TAZ in mechanotransduction.* Nature, 2011. **474**: p. 179.
16. Pushkarsky, I., et al., *Elastomeric sensor surfaces for high-throughput single-cell force cytometry.* Nature Biomedical Engineering, 2018. **2**(2): p. 124.
17. Duportet, X., et al., A platform for rapid prototyping of synthetic gene networks in mammalian cells. Nucleic Acids Research, 2014. 42(21): p. 13440-13451.
18. Guye, P., et al., Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Research, 2013. 41(16): p. e156-e156.
19. Liu, L., et al., Mechanoresponsive stem cells to target cancer metastases through biophysical cues. Science translational medicine, 2017. 9(400): p. eaan2966.
20. Engler, A.J., et al., Matrix Elasticity Directs Stem Cell Lineage Specification. Cell, 2006. 126(4): p. 677-689.
21. Wei, S.C., et al., Matrix stiffness drives epithelial–mesenchymal transition and tumour metastasis through a TWIST1–G3BP2 mechanotransduction pathway. Nature Cell Biology, 2015. 17: p. 678.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

Contribution: J.G. designed, prototyped, and optimized the sideSPIM chamber.

6.1 Conclusion

With the field of spatialomics becoming a hot topic in the last few years, the potential for the MOSAICA platform to help elucidate biological questions in a spatial context is promising. MOSAICA can provide a holistic view of the tissue microenvironment at a low-cost and single round of labeling and imaging. The goal of this thesis was to develop a multi-omics spatial technology that could fill in the gap in the field and be able to detect RNA and protein targets in the range of 6-50 plex within a single round of labeling and imaging. Using combinatorial labeling to multiplex higher with fewer fluorophores and error correction. We validated our MOSAICA platform in our mNeonGreen HEK293T model system and benchmarked MOSAICA with RNAscope and LGC Stellaris. We then did a 10-plex demonstration in colon cancer SW480 cells with 5 fluorophores and correlated with puncta/cell count with sequencing data, obtaining a Pearson r of 0.96. A panel was then demonstrated in human melanoma FFPE skin tissue with a Pearson r of 0.97. We also demonstrated multi-omics detection with a simultaneous co-detection of RNA and protein in colon cancer SW480 cells. By utilizing the spectral and lifetime data along with combinatorial labeling, MOSAICA harnesses the power of lifetime to provide a new approach to the field of spatialomics.

Another platform that was developed for my thesis was the low-cost tissue imager that was built using 3D printed parts, acrylic boards, and commercial parts easily available for under $9,000 total. The tissue imager obtained similar image resolution as our

$400,000 Nikon epi-fluorescence microscope and image an entire skin biopsy tissue sample in a single field-of-view with 3-4 plex immunofluorescence staining.

As both the MOSAICA platform and tissue imager were recently developed, there are still aspects that can be improved or further development that can be done. To increase imaging throughput, we have investigated several possibilities such as two-photon microscopy, light-sheet microscopy (SPIM), and sequential staining and imaging. Regardless, our hope is that these new technologies can be easily adopted and benefit scientists or clinicians to expand the scientific frontier.

6.2 Future directions

There are many exiting directions for this project and the spatialomics field that are currently being developed not only for this project, but also the academic and industry setting as well. With all the groups working towards increasing multiplexing abilities, increasing imaging throughput, lowering cost, improving detection efficiency, improving image analysis methods and throughput, detecting DNA, RNA, and protein simultaneously, and more, there is great promise for the spatialomics field.

To tackle the issue of imaging throughput, both two-photon and light-sheet microscopy methods were tested for the MOSAICA platform. Selective plane illumination microscopy (SPIM) is a fast and gentle imaging technique that combines the speed of widefield imaging with quality optical sectioning and low photobleaching as only a thin section of the sample is illuminated at any given time (Figure 6.1A). Previous chambers designed for the sideSPIM by Dr. Per Niklas Hedde from the LFD at UCI used acrylic for the backbone of the chamber and coverslip glass glued with silicone to the bottom and side. This design was used for live cells embedded in collagen. However, for culturing

cells in an incubator, fixing the cells, running the MOSAICA FISH labeling protocol, and imaging, this design had issues such as toxicity from the acrylic, leaking wells, collagen embedding prevented hybridization of probes, and a material was needed to raise the sample from the bottom coverslip glass to avoid the distortion of the light sheet. To overcome these issues, the acrylic backbone was replaced with a PDMS material formed in a mold that was then bonded to the coverslips in a plasma chamber (Figure 6.1B). There were 2 different prototypes for the well size (large Figure 6.1B left and small Figure 6.1B right). The mold was made using a laser cut acrylic. Pre-cured 0.5mm layer of PEG (BIO-133 RI=1.33) was then cured to the bottom of the wells to seal the gaps and raise the sample from the bottom. After coating the wells with fibronectin, colon cancer SW480 cells were plated and fixed 2 days later. Preliminary data was generated with detection of POLR2A and UBC using the MOSAICA platform (Figure 6.1C).

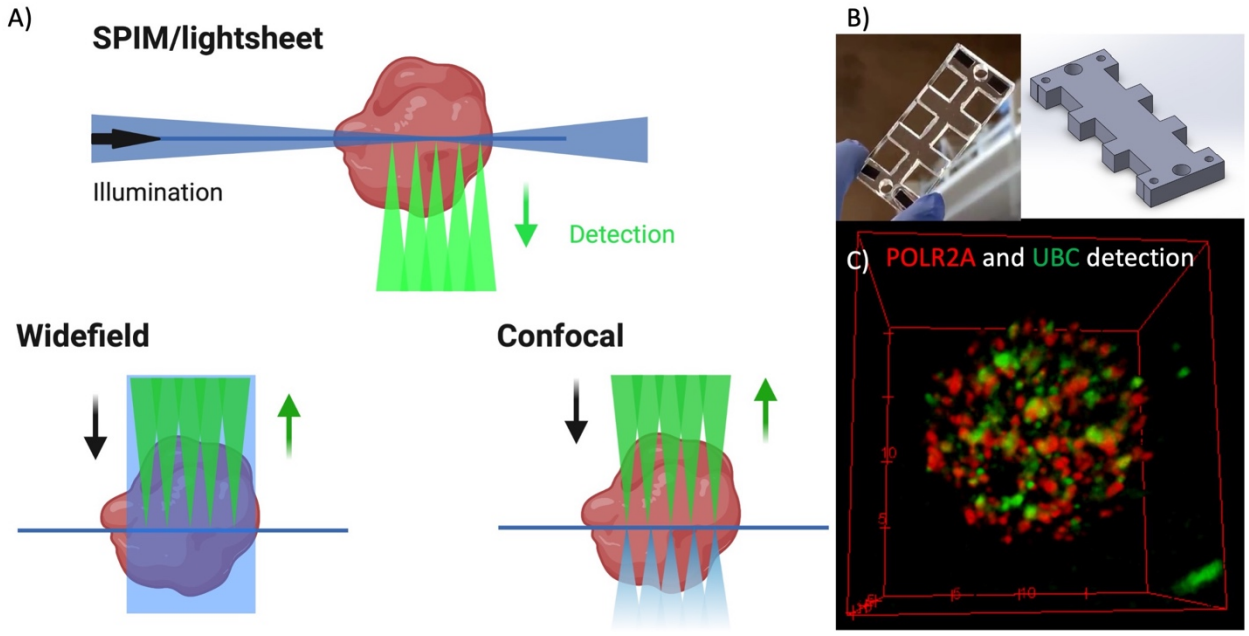


Figure 6.1 Selective plane illumination microscopy application for MOSAICA

With sideSPIM image acquisition being 50-100x faster than confocal, this was a potential method to increase imaging throughput. A 5 μm Z-stack image with 1 μm

spacing for 100 μm by 100 μm field-of-view was acquired in seconds. However, a challenge with this method was the custom sample chamber required and the image resolution. The numerical aperture (NA) of the excitation lens and 60x detection objective was 0.16 and 1.1 respectively. Increasing the NA of the detection objective and further modifying the custom chamber are possible ways to improve the image resolution.

As for protein detection, a large panel of over 30 primary antibodies have been validated and can be conjugated to oligos to bypass the species limit and multiplex the protein detection capabilities of MOSAICA. Using commercial kits such as SoluLINK has produced promising preliminary data for the antibody oligo conjugation. Once the antibody oligo conjugations are successfully validated and amplification is added for transcript detection, multiomics panels can be expanded on the MOSAICA platform.

To further expand the multiplexing abilities of MOSAICA, some issues that would need to overcome include further fluorophore discovery, optical crowding, and reducing non-specific binding. A possible solution would be to have each adopt the sequential method but have 6-16 plex in each round instead of 3-plex. This could significantly increase the speed of the sequential process while preventing optical crowding. However, the goal would be to develop an instrument that could automate the labeling/hybridization process, imaging, and analysis in a single device. This would not only eliminate a lot of the human error in the assay process, but also ensure that the sample remains at the same location for sequential labeling/imaging to avoid the difficulties of image registration. There is potential for MOSAICA to be a disruption in the spatialomics field even though there are numerous technologies being developed. By harnessing the power of fluorescence lifetime, MOSAICA leverages it to multiplex higher

in a single round and clean up the image. Compared to sequencing data, spatial analysis provides another layer of information, spatial location and clustering. Spatialomics tools like MOSAICA can be adopted easily help provide scientist and pathologists a holistic view of their sample.