UNIVERSITY OF CALIFORNIA,
IRVINE

Profiling Multiomic Biomarkers using Particle Detection Counters and Spectral-FLIM Microscopy

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Tam Vu

Dissertation Committee:
Professor Weian Zhao, Chair
Professor Enrico Gratton
Professor Jered Haun

Oversight Committee:
Professor Andrej Luptak

2021

DEDICATION

To my amazing

family, friends, colleagues, and mentors

who always continue to inspire

and push me to follow for my dreams

"*Failure is only the opportunity to begin again, this time more intelligently.*"

- Henry Ford

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My graduate school experience has been one of the most exciting, challenging, and rewarding rollercoasters of emotions in my life. During this journey, I have been tremendously blessed with the privilege of working with and learning from some of the most amazing mentors and friends that I have ever met. First and foremost, I would like to thank Dr. Weian Zhao for always pushing me to the next level and reminding me to worry less and to not let fear hold me back. He was exactly the endlessly energetic mentor with the indomitable passion for science and entrepreneurship which I needed in my life to grow both as a scientist as well as an individual. I cannot really ask for a better graduate school lab to have been a part of. I would like to also thank Dr. Enrico Gratton who has been nothing short of amazing as a mentor to me, who has imbued within me a profound appreciation and respect for the field of Biophotonics. Being able to learn so much from the de facto world expert in Spectral-FLIM microscopy is a privilege and honor which I will cherish for the rest of my life. The many valuable lessons which I acquired from our weekly discussions, classes, and workshops have taught me how beautiful and vastly boundless the field of Biophotonics can be. Through Dr. Gratton, I have also gotten the fortunate opportunity to work with Dr. Alex Vallmitjana who is both a close friend and mentor to me as well as one of the most amazing programmers I have ever met.

I would like to thank Dr. Michelle Digman as well as all the Laboratory of Fluorescence Dynamics members who have always treated me with great kindness and made me feel at home. I wish I can work with you guys forever! To Dr. Jered Haun, Dr. Andrej Luptak, and Dr. Marian Waterman for serving on my committee and giving me valuable feedback, which has created

many exciting new directions for this project to embark on. Thank you to all Zhao lab members, both past and present who always treated me like a family member and gave me endless unconditional support as we shared our stories and challenges through many long nights and tough days. I would like to specially thank Dr. Jan Zimak who is always there whenever I needed him and was both a lab and life mentor that helped me survive the many brutal projects and deadlines that I encountered every month. I would like to thank Josh Gu who is an amazing graduate student with endless potential and talent. To Jaya Jayaraman who has been like a sister to me, always making sure I have snacks at hand and lending me an ear whenever I needed one. To Kieu La and Qi Xu who were always encouraging and believed in me and taught me how to be a more effective communicator. I would like to thank you all the professors, graduate students, and post-doc collaborators who have been a vital part of my journey. Without your help and guidance, this journey would have not been possible. I would like to thank you Dr. Alan Hauser for being an amazing mentor who is incredibly patient and understanding and will become an instrumental figure in my life for these next few years to come. Last but not least, I would like to thank my family who with words I cannot describe enough with to show how much I appreciate their help and support regardless of whatever situation I am in. I owe them everything and am truly blessed to be in their care!

CURRICULUM VITAE

Tam Minh Vu

EDUCATION

University of California, Irvine — Irvine, CA
Ph.D., Biomedical Engineering — 2017 - 2021

University of California, Irvine — Irvine, CA
M.S., Biomedical Engineering — 2017 – 2020

University of California, Davis — Davis, CA
B.S., Neurobiology, Physiology, and Behavior — 2006 - 2010

RESEARCH EXPERIENCE

Department of Pharmaceutical Sciences (Zhao Lab) — UCI

Graduate Student Researcher — 2017 - 2021

- Develop spatial multiomics profiling assays with Spectral-FLIM microscopy
- Develop ddPCR assays with Integrated Comprehensive Droplet Digital Detection (IC3D) platform for liquid biopsy and sepsis detection

Department of Immunology (Van de Water Lab) — UCD

Staff Research Associate II — 2014 - 2017

- Develop ELISA and Microarray ELISA assays for auto-immune antibodies panels
- Profile human serum and plasma with Luminex panels, western blots, and ELISAs

Department of Biomedical Engineering (Revzin Lab) — UCD

Staff Research Associate — 2010 - 2014

- Develop microfluidic antibody microarray assays for detection of Tuberculosis markers
- Develop Lab-on-a-Chip biosensors platform for cytokine detection

PUBLICATIONS

1. **Tam Vu**‡, Alexander Vallmitjana‡, Joshua Gu‡, Kieu La, Qi Xu, Jesus Flores, Jan Zimak, Jessica Shiu, Linzi Hosohama, Jie Wu, Christopher Douglas, Marian Waterman, Anand Ganesan, Per Niklas Hedde, Enrico Gratton*, and Weian Zhao*. *Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis*. BioRxiv 449468. [Preprint] (2021). doi: https://doi.org/10.1101/673665. ‡:co-first authors.

2. Per Niklas Hedde, Margaux Bouzin, Timothy J. Abram, Xiaoming Chen, Melody N. Toosky, **Tam Vu**, Yiyan Li, Weian Zhao and Enrico Gratton. *Rapid Isolation of Rare Targets from Large Fluid Volumes*. Scientific Reports. (2020); Vol. 10, Issue 1:12458.

3. Alexandra Ramirez-Celis, Elizabeth Edmiston, Joseph Schauer, **Tam Vu**, and Judy V.D. Water. *Peptides of neuron specific enolase as potential ASD biomarkers: from discovery to epitope mapping*. Brain Behav Immun. (2020); Vol. 84: 200-208.

4. Per Niklas Hedde, Tim Abram, **Tam Vu**, Weian Zhao, Enrico Gratton*. *Fluorescence lifetime detection with particle counting devices*. Biomedical Optics Express. (2019); Vol. 10, Issue 3: 1223-1233.

5. Timothy J. Abram,* Hemanth Cherukury, Chen-Yin Ou, **Tam Vu**, Michael Toledano, Yiyan Li, Jonathan T. Grunwald, Melody N. Toosky, Delia F. Tifrea, Anatoly Slepenkin, Jonathan Chong, Lingshun Kong, Domenica Vanessa Del Pozo, Kieu Thai La, Louai Labanieh, Jan Zimak, Byron Shen, Susan S. Huang, Enrico Gratton, Ellena M. Peterson, Weian Zhao*. *Rapid bacterial detection and antibiotic susceptibility testing in whole blood using one-step, high throughput blood digital PCR*. Lab on a Chip. (2019); Vol. 20, Issue 3: 477-489.

6. Chen-Yin Ou‡, **Tam Vu‡**, Jonathan T. Grunwald, Michael Toledano, Jan Zimak, Melody Toosky, Byron Shen, Jason A. Zell, Enrico Gratton, Timothy J. Abram*, Weian Zhao*. *An ultrasensitive test for profiling circulating tumor DNA using integrated comprehensive droplet digital detection*. Lab on Chip. (2019); Vol. 19: 993-1005. ‡: Co-first authors.

7. Elizabeth Edmiston, Karen L. Jones, **Tam Vu**, Paul Ashwood, Judy V.D. de Water. *Identification of the antigenic epitopes of maternal autoantibodies in autism spectrum disorders.* Brain Behav Immun. (2018); Vol. 69: 399-407.

8. Ying Liu, Ali Rahimian, Sergiy Krylyuk, **Tam Vu**, Bruno Crulhas, Gulnaz Stybayeva, Meruyert Imanbekova, Dong-Sik Shin, Albert Davydov, Alexander Revzin. *Nanowire aptasensors for*

electrochemical detection of cell-secreted cytokines. ACS Sensors. (2017); Vol 2, Issue 11: 1644-1652.

9. Pantea Gheibi, Shuxiong Zeng, Kyung Jin Son, **Tam Vu**, Ai-Hong Ma, Marc A Dall'Era, Stanley Alexander Yap, Ralph W de Vere White, Chong-Xian Pan, Alexander Revzin. *Microchamber cultures of bladder cancer: a platform for characterizing drug responsiveness and resistance in PDX and primary cancer cells*. Sci Reports. (2017); Vol 7, Issue 1: 12277.

10. Dong-Sik Shin, Zimple Matharu, Jungmok You, Christian Siltanen, **Tam Vu**, Vijay Krishna Raghunathan, Gulnaz Stybayeva, Ashley E Hill, Alexander Revzin. *Sensing Conductive Hydrogels for Rapid Detection of Cytokines in Blood*. Advanced Healthcare Materials. (2016); Vol. 5, Issue 6: 659-64.

11. **Tam Vu**, Ali Rahimian, Gulnaz Stybayeva, Yandong Gao, Timothy Kwa, Judy Van de Water, Alexander Revzin. *Reconfigurable microfluidic device with integrated antibody arrays for capture, multiplexed stimulation, and cytokine profiling of human monocytes*. Biomicrofluidics. (2015). Vol 9, Issue 4:044115.

12. Dong-Sik Shin, Jungmok You, Ali Rahimian, **Tam Vu**, Christian Siltanen, Arshia Ehsanipour, Gulnaz Stybayeva, Julie Sutcliffe, Alexander Revzin. *Photodegradable hydrogels for capture, detection, and release of live cells*. Angewandte Chemie Int Ed Engl. (2014). Vol 53, Issue 31: 8221-4.

13. Dipali Patel, Amranul Haque, Caroline N Jones, Nazgul Tuleouva, Elena Foster, **Tam Vu**, A Hari Reddi, Alexander Revzin. *Local control of hepatic phenotype with growth factor-encoded surfaces*. Integrative Biology. (2014). Vol 6: 44-52.

14. Arnold Chen‡, **Tam Vu‡**, Gulnaz Stybayeva, Tingrui Pan, Alexander Revzin. *Reconfigurable microfluidics combined with antibody microarrays for enhanced detection of T-cell secreted cytokines.* Biomicrofluidics. (2013). Vol 7, Issue 2:024105-9, 2013. ‡: co-first authors.

15. Sunny S Shah, Mihye Kim, Elena Foster, **Tam Vu**, Dipali Patel, Li-Jung Chen, Stanislav V Verkhoturov, Emile Schweikert, Giyoong Tae, Alexander Revzin. Electrochemical release of hepatocyte-on-hydrogel microstructures from ITO substrates. Anal Bioanal Chem. (2012). Vol 402, Issue 5: 1847-56.

PRESENTATIONS

1. "Spatial Profiling of mRNA Transcripts with Spectral and Fluorescence Lifetime Imaging and Microscopy". The Biophysical Society Virtual Annual Meeting, Virtual Meeting, February 2021.

2. "Gene expression by FLIM". Laboratory of Fluorescence Dynamics Workshop, Irvine, CA, October 2019.

3. "Ultra-sensitive Detection of Circulating Colorectal Cancer Genetic Biomarkers using Integrated Comprehensive Droplet Digital Detection". NanoEngineering for Medicine and Biology Conference, Los Angeles, CA, August 2018.

4. "Orthogonally Perfused Multi-Chamber Microfluidic Device for Blood Analysis" Biomedical Engineering Society Annual Meeting, Seattle, WA, September 2013.

PATENTS

Zhao W., Gratton, E., Zimak, J., **Vu, T.**, Vallmitjana, A., Gu, J., Hedde, P.N., Scipioni, L. 2020. Compositions and Methods for Spatial Profiling of Biological Materials Using Time-Resolved Luminescence Measurements. Publication No. WO/2021/102140. International Filing Date 19.11.2020

AWARDS/FELLOWSHIPS

Sigma Xi Honor Society, 2018 – 2021

NSF GRFP Fellowship, 2018 – 2021

NSF Student's Travel Award, 2018

NSF GRFP Honorable Mention, 2017

ABSTRACT OF THE DISSERTATION

Profiling Multi-Omic Biomarkers using Particle Counters and Spectral-FLIM Microscopy

By

Tam Vu

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2021

Professor Weian Zhao, Chair

Profiling multiomic biomarkers in bulk and *in situ* provides critical information which enables basic research and clinical applications. Unfortunately, most existing profiling methods are limited due to either low multiplexing, sensitivity, costs, or assay complexity. This thesis aims to develop two core technologies that address 1) bulk profiling issues with sensitivity and low throughput as well as 2) in situ profiling issues with low multiplexing capabilities, costs, and limited throughput. To address the first issue, this work introduces a novel liquid biopsy approach that utilizes a platform technology called Integrated Comprehensive Droplet Digital Detection (IC3D). This integrated approach combines microfluidic droplet partitioning technology, fluorescent multiplexed PCR chemistry, and our own unique and rapid particle counting technology to deliver ultrasensitive and ultrafast detection of colorectal cancer-specific genomic biomarkers from minimally processed blood samples.

To address the second issue, this work introduces a new spatial multi-omics technology termed Multi Omic Single-scan Assay with Integrated Combinatorial Analysis (MOSAICA) 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, and cost-effective spatial profiling of multi-omics biomarkers. By utilizing both time and intensity domains for labeling and imaging, this technology seeks to 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. Overall, these two technologies represent simple, versatile, and scalable tools which enable more rapid, sensitive, and/or multiplexed protein/transcriptomic analysis.

CHAPTER 1:
BACKGROUND AND INTRODUCTION

1.1 Introduction

Conventional tools for bulk and in situ profiling of mRNA and protein biomarkers include PCR, southern/western blots, fluorescence in situ hybridization (FISH) for DNA or RNA analysis, and/or immunohistochemistry (IHC) for protein detection. These assays are expensive and especially difficult to scale as large libraries of potential epitopes and/or genetic targets against different tissues are required to be screened and validated [1-6]. Moreover, these traditional approaches encounter issues with low throughput and multiplexing capabilities because of complicated/limited assay workflows and/or limited number of spectrally available channels [6-9]. For bulk RNA detection, more recent iterations such as the droplet digital PCR platforms attempts to address issues of low throughput and sensitivity issues with partitioning but remain inadequate for profiling ultra-rare targets due to low sensitivity or scanning speed.

As for in situ profiling, the more recent spatial transcriptomic methods that are based on sequential labeling, stripping, and imaging (e.g. seqFISH, MERFISH), branched amplification (e.g. SABER), or barcoded labeling with down-stream sequencing (e.g. Slide-seq) can achieve substantially greater multiplexing capabilities but encounter issues such as being time-consuming, laborious and costly to scale up and often are limited to few specialized laboratories [10-13]. Furthermore, for proteomic methods including Imaging Mass Cytometry (IMC) and CODEX where higher multiplexing can be achieved over conventional immunohistochemistry

methods, there remains a tradeoff with decreased sensitivity, throughput, accessibility and/or high cost [14-16].

This thesis aims to develop two core technologies to address bulk profiling issues with sensitivity and low throughput and in situ profiling issues with low multiplexing capabilities, costs, and limited throughput. To address the first issue, this work introduces a platform technology called the Integrated Comprehensive Droplet Digital Detection (IC3D) approach. This integrated approach combines microfluidic droplet partitioning technology, fluorescent multiplexed PCR chemistry, and our own unique and rapid particle counting technology to deliver ultrasensitive detection of genetic or mRNA biomarkers in plasma, serum, or minimally processed blood samples. When catered to detect CRC-specific genomic biomarkers from patients, this tool has the potential to serve as a liquid biopsy test for clinicians/oncologists who are interested in interrogating a panel of genetic markers from patients for routine monitoring of disease such as minimal residual disease and attaining critical info for treatment stratification. To address the second issue, this work introduces a new spatial multi-omics technology termed Multi Omic Single-scan Assay with Integrated Combinatorial Analysis (MOSAICA) 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, and cost-effective spatial profiling of multi-omics biomarkers. This method utilizes both time and intensity domains for labeling and imaging and seeks to 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.

1.2 Need for rapid and ultra-sensitive bulk profiling tools

Colorectal cancer (CRC) is the third deadliest cancer worldwide, causing an estimated 774,000 deaths in 2015 (WHO). A major contributing factor to this is the lack of available clinical methods that are effective at monitoring cancer recurrence at early stages when treatment is most effective. For example, even after initial treatment with a combination of targeted radiation, chemotherapy, and/or biologics, patients who achieved remission still face a 30 - 50% chance of relapsing. Monitoring for minimal residual disease (MRD) or cancer relapse is, thus, a vital component for effective clinical management of colorectal cancer progression by allowing physicians to stratify treatment successfully. Unfortunately, current conventional clinical methods (e.g. surgical biopsy, carcinoembryonic antigen test, histology, and radiological imaging) to monitor MRD suffer from poor sensitivity and/or specificity, tumor heterogeneity issues, and the requirement of repeated invasive sampling and have limited success in predicting MRD early. An alternative approach which can address these issues while providing rapid to sample results turnaround time will have considerable value in the oncology scene.

To address these issues, this thesis introduces a novel liquid biopsy approach that utilizes the IC3D platform technology. This integrated approach combines microfluidic droplet partitioning technology, fluorescent multiplexed PCR chemistry, and our own unique and rapid particle counting technology to deliver ultrasensitive detection of CRC-specific genomic biomarkers from minimally processed blood samples. Through this strategy, we demonstrated that the IC3D liquid biopsy system can detect oncogenic KRAS G12D mutant alleles against a background of wild-type genomes at a sensitivity of 0.00125-0.005% which is 50 to 1000× more

sensitive than existing commercial liquid biopsy ddPCR and qPCR platforms, respectively. Furthermore, we have also collaborated with the Laboratory of Fluorescence Dynamics (LFD) at UCI to adapt fluorescence lifetime imaging to this system in order to greatly improve our multiplexing capabilities to examine many more cancer mutations simultaneously. Lastly, we are currently leveraging this system in clinical studies to evaluate the performance of our system against current conventional and state-of-art diagnostic methods in predicting MRD from CRC patients. We aspire to one day provide clinicians with a powerful decision-making tool to monitor and treat MRD with unprecedented sensitivity for earlier stage intervention.

1.3 Need for in situ validation of biomarkers

Multiplexed mRNA profiling in the spatial context provides important new information enabling basic research and clinical applications. Unfortunately, most existing spatial transcriptomics methods are limited due to either low multiplexing or assay complexity. Here, we introduce MOSAICA as a tool that integrates *in situ* labeling of mRNA 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 transcriptomics. By integrating the time dimension with conventional spectrum-based measurements, MOSAICA enables direct, highly multiplexing of *in situ* spatial biomarker profiling in a single round of staining and imaging while providing error correction removal of background autofluorescence. We first demonstrate mRNA encoding using combinatorial spectral and lifetime labeling and target decoding and quantification using a

phasor-based image segmentation and machine learning clustering technique. We then showcase MOSAICA's multiplexing scalability in detecting 10-plex targets in fixed colon cancer cells using combinatorial labeling of only five fluorophores with facile error-correction and removal of autofluorescent moieties. MOSAICA's analysis is strongly correlated with sequencing data (Pearson's r = 0.9) 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 that have a high degree of tissue scattering and intrinsic autofluorescence, demonstrating the robustness of the approach. MOSAICA represents a simple, versatile, and scalable tool for targeted spatial transcriptomics analysis that can find broad utility in constructing human cell atlas, elucidating biological and disease processes in the spatial context, and serving as companion diagnostics for stratified patient care.

As summarized in Table 1.1, several key capabilities and advantages of the proposed technology for spatial omics compared to existing technologies are expected, including a) capable of 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) eliminating sample autofluorescence (by choosing probes with lifetimes different from autofluorescent moieties) and therefore improving detection sensitivity, signal-to-noise ratio (SNR), and detection efficiency as well as preclude the need for additional complicated, harsh, and time laborious processing steps, and c) highly multiplexing. As standard in situ fluorescence staining reagents and protocols which are compatible across different target molecular species are used, this platform strives to 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 diseases. 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, the proposed technology seeks to utilize 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). This proposed technology, once developed, is anticipated to be quickly and broadly adopted in the scientific community to address many unsolved biological questions at an affordable cost.

Table 1.1. Expected quantitative advances of the proposed Spectral-FLIM 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.

1.4 References

1. Kim, S.W., J. Roh, and C.S. Park, *Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips.* J Pathol Transl Med, 2016. **50**(6): p. 411-418.
2. Miller, R.T. *Avoiding pitfalls in diagnostic immunohistochemistry-important technical aspects that every pathologist should know*. in *Seminars in diagnostic pathology*. 2019.
3. Taylor, C. and R.M. Levenson, *Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II.* Histopathology, 2006. **49**(4): p. 411-424.
4. Jensen, E., *Technical review: In situ hybridization.* The Anatomical Record, 2014. **297**(8): p. 1349-1353.
5. Chu, Y.-H., et al. *In Situ Hybridization: Introduction to Techniques, Applications and Pitfalls in the Performance and Interpretation of Assays*. in *Seminars in diagnostic pathology*. 2019. Elsevier.
6. 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, 2017. **8**(59): p. 100863-100898.
7. Blom, S., et al., *Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis.* Sci Rep, 2017. **7**(1): p. 15580.
8. Kalra, J. and J. Baker, *Multiplex immunohistochemistry for mapping the tumor microenvironment*, in *Signal Transduction Immunohistochemistry*. 2017, Springer. p. 237-251.
9. 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, 2016. **113**(50): p. 14456-14461.
10. Wang, F., et al., *RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues.* J Mol Diagn, 2012. **14**(1): p. 22-9.
11. Chen, K.H., et al., *RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.* Science, 2015. **348**(6233): p. aaa6090.
12. Eng, C.L., et al., *Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH.* Nature, 2019. **568**(7751): p. 235-239.
13. Kishi, J.Y., et al., *SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues.* Nat Methods, 2019. **16**(6): p. 533-544.
14. Porta Siegel, T., et al., *Mass Spectrometry Imaging and Integration with Other Imaging Modalities for Greater Molecular Understanding of Biological Tissues.* Mol Imaging Biol, 2018. **20**(6): p. 888-901.
15. Ferguson, C.N., et al., *Mass spectrometry-based tissue imaging of small molecules.* Adv Exp Med Biol, 2014. **806**: p. 283-99.
16. Goltsev, Y., et al., *Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging.* Cell, 2018. **174**(4): p. 968-981.e15.

CHAPTER 2:
AN ULTRASENSITIVE TEST FOR PROFILING CIRCULATING TUMOR DNA USING INTEGRATED COMPREHENSIVE DROPLET DIGITAL DETECTION

Authors: Chen-Yin Ou[1]‡, **Tam Vu**[2,6]‡, Jonathan T. Grunwald[1], Michael Toledano[2], Jan Zimak[2], Melody Toosky[1], Byron Shen[1], Jason A. Zell[4,8], Enrico Gratton[6,9], Timothy J. Abram*[1] and Weian Zhao*[2,3,4,5,6]

Formal analysis: T. A., T. V.; methodology: T. A., W. Z., C.-Y. O., T. V.; writing – original draft: C.-Y. O., T. V., T. A., W. Z.; investigation: C.-Y. O., T. V., M. T., J. Z., M. T.; data curation: C.-Y. O., T. V., J. G., T. A.; conceptualization: W. Z., C.-Y. O., T. V., B. S., T. A.; resources: W. Z., J. Z., E. G.

Affiliations:

*Corresponding authors

[1]Velox Biosystems, 5 Mason, Suite 160, Irvine, CA 92618, USA

[2]Sue and Bill Gross Stem Cell Research Center, Sue & Bill Gross Hall CIRM Institute, University of California, Irvine, 845 Health Sciences Road, Suite 3027, Irvine, CA 92697, USA

[3]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA

[4]Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, CA 92697, USA

[5]Edwards Life Sciences Center for Advanced Cardiovascular Technology, University of California, Irvine, Irvine, CA 92697, USA

[6]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, USA

[7]Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92697, USA

[8]Division of Hematology/Oncology, University of California Irvine Medical Center, Orange, USA

[9]Laboratory for Fluorescence Dynamics, Department of Biomedical Engineering, University of California, Irvine, USA

† These authors contributed equally to this work.

2.1 Abstract

Current cancer detection systems lack the required sensitivity to reliably detect minimal residual disease (MRD) and recurrence at the earliest stages when treatment would be most effective. To address this issue, we present a novel liquid biopsy approach that utilizes an integrated comprehensive droplet digital detection (IC3D) digital PCR system which combines microfluidic droplet partitioning, fluorescent multiplex PCR chemistry, and our rapid 3D, large-volume droplet counting technology. The IC3D ddPCR assay can detect cancer-specific, ultra-rare genomic targets due to large sample input and high degree of partitioning. We first demonstrate our droplet digital PCR assay can robustly detect common cancer mutants including KRAS G12D spiked in wild-type genomic background or isolated from patient samples with 100% specificity. We then demonstrate that the IC3D ddPCR system can detect oncogenic KRAS G12D mutant alleles against a background of wild-type genomes at a sensitivity of 0.00125–0.005% with a false positive rate of 0% which is 50 to 1,000× more sensitive than existing commercial liquid biopsy ddPCR and qPCR platforms, respectively. In addition, our technology can uniquely enable detection of circulating tumor cells using their genetic markers without a pre-enrichment step, and analysis of total tumor DNA isolated from blood samples, which will increase clinical sensitivity and specificity, and minimize inter-assay variability.

Therefore, our technology holds the potential to provide clinicians with a powerful decision-making tool to monitor and treat MRD with unprecedented sensitivity for earlier stage intervention.

2.2 Introduction

At some point during their lifetimes, approximately 38.4% of men and women will be diagnosed with cancer [1]. Colorectal cancer (CRC), in particular, is the third most commonly diagnosed cancer, and the third most common cause of cancer-related deaths in the United States. In 2018, it is estimated that 140, 000 people were diagnosed with CRC, and 50, 000 people succumbed to the disease [2]. Despite major advances in treatment (*i.e.*, surgery, radiation therapy, chemotherapy, biologic/immune therapy) and resultant improved outcomes over time, colorectal cancer survivors with potentially curable (non-metastatic) disease still face a 10–50% risk of relapse. Relapse carries an extremely high mortality risk, as most relapses are not curable by current methods, except in cases of oligometastatic disease, where surgical resection of limited areas of tumor burden may result in cure. As such, monitoring for minimal residual disease (MRD) in CRC survivors has become a major investigational focus, as advances in MRD detection are expected to lead to early intervention and improved outcomes for this deadly disease. National Comprehensive Cancer Center Network (NCCN) Guideline-based surveillance monitoring with clinical exams, serial serum carcinoembryonic antigen testing, annual imaging assessments, and regular colonoscopic evaluations, however, lack the required sensitivity and/or specificity to reliably detect MRD at the earliest stages when treatment would be most effective [3].

Mounting evidence has recently demonstrated that “liquid biopsy” blood tests for cancer biomarkers, *i.e.*, circulating tumor cells (CTC), circulating cell-free tumor DNA (ctDNA), and exosomes, can serve as a non-invasive, cost effective and viable alternative for cancer detection,

surveillance monitoring of MRD, and drug response evaluation [4-5]. The use of liquid biopsy can 1) avoid invasive and repeated tumor sampling involved in conventional tissue biopsy or toxic radiation exposure involved in conventional imaging modalities, and 2) better recapitulate real-time genetic profiles in the tumor as a whole, which can be missed in traditional biopsies due to intra-tumor heterogeneity. Several liquid biopsy tests have been FDA-approved, including the CellSearch system for CTC and the PCR-based Cobas® EGFR mutation test v2 for non-small cell lung cancer [6]. Many other methodologies including various CTC isolation and enrichment methods, digital PCR (dPCR) and next-generation sequencing (NGS) for cell-free DNA (cfDNA) are being developed and validated in the pipeline by both academia and industry worldwide, as surveyed in recent reviews [7-8]. Early clinical validation of these liquid biopsy approaches has led to promising results as it can provide actionable information to potentially improve clinical outcome. For instance, Diehl *et al.* [9] demonstrated that PCR measurements of cfDNA aberrations (*e.g.*, APC, TP53 and KRAS) in the plasma could be used to reliably detect residual disease and monitor tumor dynamics in CRC patients who were undergoing surgery or chemotherapy. Furthermore, Tie *et al.* [10] demonstrated that sequencing of cfDNA after stage II CRC resection could identify patients at high risk for recurrence to help stratify treatment decisions.

Despite these early promises, detecting and monitoring cancer *via* liquid biopsy assays have yet to transform routine practice in clinical oncology. A major current limitation is the poor sensitivity of existing platforms to efficiently and effectively detect low-abundance circulating cancer biomarkers especially for early-stage cancers, MRD, and recurrence [11]. For instance, the abundance of cfDNA and CTCs in patients with early stage disease is known to be significantly

lower (10-fold) than those in patients with more advanced stages. Various reports have also demonstrated that CTCs can exist in low numbers in blood (<1–100 CTC per ml) [8, 12]. As a result, it was estimated that only 30–40% of stage I and II cancers could be detected using current methods. For cell surface protein capture approaches, the use of single or multiple surface protein markers (*e.g.*, EpCAM) in the CellSearch system and other methods are inadequate to capture and detect all CTCs. Alternatively, detection of CTCs using direct genetic analyses has been largely impractical due to the vast background of genomic, wild-type (WT) DNA contributed from white blood cells; therefore, an enrichment step is inevitably needed prior to their analysis. Indeed, the current assays for tumor DNA detection including qPCR (*e.g.*, PNA clamp-PCR and ARMS), digital PCR (*e.g.*, BEAMing and Bio-Rad droplet digital PCR (ddPCR)), and NGS have been confirmed to possess a sensitivity of only 0.01–1% [13]. The challenges with detection of CTCs have resulted in a shift in the field to the detection of cfDNA in plasma. However, for cfDNAs that exist in low total quantities (typically 10 s ng ml^{-1}) and are surrounded by excess WT DNA in plasma, these assays have been notably inadequate to detect these rare targets in early stage cancer patients. Furthermore, recent studies have suggested that CTCs, cfDNAs, and other circulating cancer markers including exosomes are distinct entities; therefore, each of them alone may not be adequate to reveal complete profiles of cancer [14-17]. Additionally, different isolation and sample processing kits and protocols for each of these cancer marker subtypes have created large variabilities and inconsistences between different settings, which represent a major roadblock to fully realize liquid biopsy's potential to cancer management in the clinic.

We reason the sensitivity and robustness of cancer liquid biopsy tests that use circulating tumor DNA as markers can be greatly improved using a dPCR platform that can accommodate 1)

much larger DNA sample input, 2) greater numbers of partitions, and 3) total DNA isolated from blood samples regardless of their origins in a single assay. dPCR formats permit absolute quantitation of target DNA with improved sensitivity and precision compared to qPCR assays. The compartmentalization of target DNA in droplets or microwells significantly increases the effective target concentration following PCR while reducing interference from background (WT genomic DNA), to permit absolute, digital ("1" or "0") quantification without the need for calibration curves. Despite these advances, existing dPCR assays still lack the required sensitivity for rare target DNA detection in the context of early stages of cancer, MRD, and recurrence due to their limited input sample content and number of partitions. For instance, Bio-Rad's QX-200 ddPCR assay can only handle 20 µL sample volumes (approximately 20 000 droplets) at a time with typically 0–66 ng of intact DNA input per single reaction [18-20]. Each run can thus accommodate only a small fraction of the total DNA isolated from several mls of blood samples, an amount typically required in order to detect rare targets effectively without subsampling issues [21-23].

We recently developed a platform technology called integrated comprehensive droplet digital detection (IC3D) that can selectively detect biomarkers from large volumes of biological samples such as bacteria in blood or circulating miRNA in plasma at single-cell or single-molecule sensitivity with a limit of detection (LOD) of single-digit number of targets per ml of measurement volume [24-25]. Our system integrates target-specific fluorescent chemistry, droplet microfluidics, and a high throughput 3D particle counting system. Unlike conventional particle-counting systems (*e.g.*, flow cytometry and 1D on-chip droplet counting in the Bio-Rad ddPCR system) that have low throughput and, therefore, can only accommodate a small sample volume or require multiple runs and much longer time to analyze greater sample volume, the IC3D

system can robustly detect fluorescent particles from ml volumes (tens of millions of droplets) at single-particle sensitivity within several minutes by collecting droplets in a cuvette and analyzing them in bulk using a high throughout 3D particle counter (Fig. 2.1). In this study, we evaluated IC3D as a novel ddPCR platform for ctDNA detection in cancer liquid biopsy. We demonstrated that the unique capabilities of rapid and large volume (ml) analysis of IC3D allow us to accommodate a significantly higher amount of DNA (20 μg ml^{-1}) and greater numbers of partitions (18 million reactions per ml), compared to commercial assays such as Bio-Rad ddPCR and qPCR. As a proof of concept, we demonstrated that our system can achieve a sensitivity improvement of at least 50- to 1,000-fold compared to Bio-Rad ddPCR and qPCR, respectively. In addition, we demonstrated cancer cells spiked in blood (modelling CTCs) can be directly analyzed with IC3D ddPCR using their genetic markers without the need for pre-enrichment. Therefore, unlike previous studies that have to profile cfDNA, CTCs, or other ctDNA carriers

separately, the IC3D system provides additional ability to analyze these markers together in a single assay from the same blood draw to greatly increase clinical sensitivity and specificity and reduce inter-assay variability.

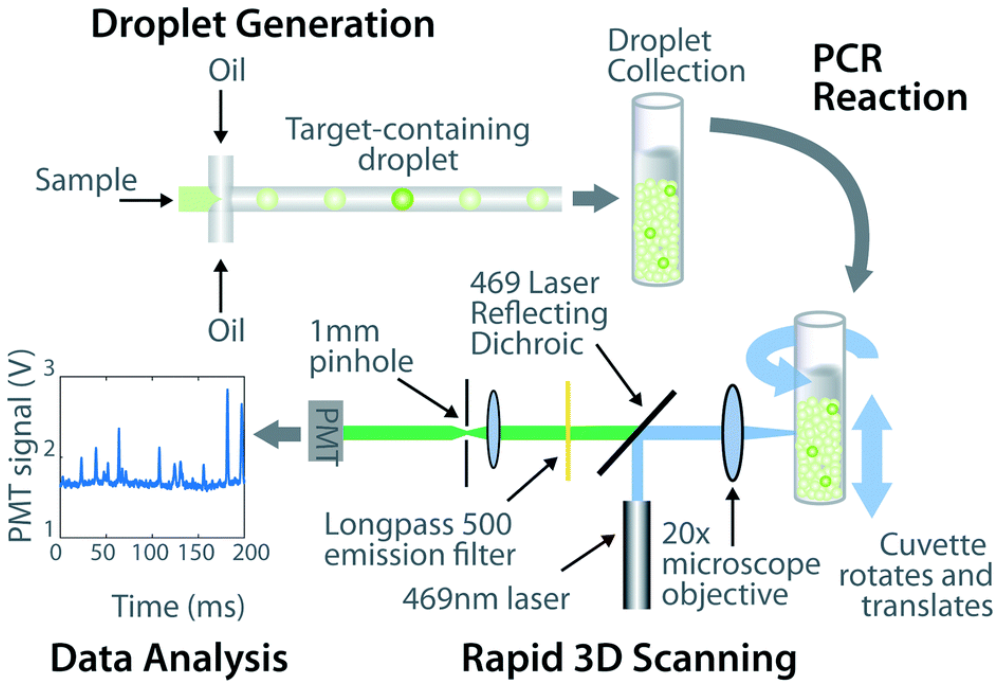


Fig. 2.1 IC3D ddPCR workflow diagram for a typical clinical sample. The IC3D technology involves partitioning a sample into millions of picoliter-sized droplets, thermocycling the droplets to amplify specific fluorescent signals, and detecting/quantifying droplets that are positive for one or more specific targets. Unlike commercial dPCR systems, the IC3D technology can analyze large sample volumes with great number of partitions, which can collectively improve sensitivity of ctDNA detection.

2.3 Materials and methods

2.3.1 Materials

Phosphate-buffered saline (PBS, 10×) without Ca^{2+} and Mg^{2+} salts and bovine serum albumin (BSA) were purchased from Sigma-Aldrich (Saint Louis, MO). RPMI 1640 and DMEM with L-glutamine, were purchased from VWR (Radnor, PA). Fetal bovine serum (FBS), and penicillin/streptomycin (PS) were purchased from Invitrogen (Carlsbad, CA). All primers and probes used for this study were purchased from LGC Biosearch Technologies and Integrated DNA Technologies (IDT) and their sequences are listed in Table S2.1.† Synthetic and HPLC-purified KRAS G12D and BRAF V600E mutants were synthesized as double-stranded gBlocks® gene fragments and verified by IDT to guarantee the accuracy of the sequences (details regarding the synthetic gene fragments are listed in Table S2†). PerfeCTa qPCR ToughMix was purchased from Quanta BioSciences. Nuclease-free water, Rox reference dye (50×) and Countess cell counting chamber slides (for microscopic imaging of the droplets) were purchased from Thermo-Fisher Scientific (Waltham, MA). SU-8 2075 negative photoresist and developer were purchased from Microchem (Newton, MA). Silicon 4 in wafers were purchased from University Wafer (South Boston, MA). Photomasks were purchased from CAD/Art Services (Bandon, OR). Sylgard 184 polydimethylsiloxane (PDMS) and curing agent were purchased from Ellsworth Adhesives (Concord, CA). PFPE-based surfactant was prepared by Velox Biosystems with our proprietary formulation. HFE-7500 fluorocarbon oil was purchased from 3M (St. Paul, MN).

2.3.2 Primers, probes, and gBlocks preparation

Primers and probes from LGC Biosearch Technologies and IDT were centrifuged at 8000 rpm for 1 minute and reconstituted to a final concentration of 100 µM in TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0). Following IDT recommended procedures, KRAS c.35G>A (G12D) and BRAF c.1799T>A (V600E) gBlocks® gene fragments were reconstituted to a final concentration of 10–20 ng μl^{-1} in TE buffer. A NanoDrop 2000 spectrophotometer (Life Technologies) was used to assess the purity, quality, and concentrations of gene fragments. Primers, probes, and gene fragments were stored in DNA LoBind tubes at −20 °C prior to experiments.

2.3.3 Cell-lines and DNA sample preparation

The human colon adenocarcinoma cell line, LS174T (CL-188), and human T-cell lymphoblast cell line, Jurkat E6-1 (TIB-152), were obtained from the American Type Culture Collection (ATCC). LS174T cells were cultured in DMEM media while Jurkat cells were cultured in RPMI 1640 media, both supplemented with L-glutamine, 10% FBS, and 1% PS. All cells were cultured at 37 °C with 5% atmospheric CO_2 in a humidified incubator. Prior to ddPCR or blood spiking experiments, cells were cultured in flasks until sub-confluence, trypsinized (if adherent), resuspended in PBS, and then subjected to either genomic DNA (gDNA) extraction protocols or to whole blood titration experiments. For gDNA extraction, QIAamp Blood DNA miniprep kit (Qiagen) was used according to manufacturer's instructions, and a NanoDrop 2000 spectrophotometer (Life Technologies) was used to assess the purity and quality of these purified gDNA samples. Only DNA samples with absorbance ratios of $A_{260/280}$ and $A_{260/230}$ greater

than 1.8 were used for this study.

2.3.4 Healthy donor samples

De-identified and healthy donor blood samples used in this study were obtained with informed consent from donors and approval from the Institutional Review Board (IRB 2012-9023) *via* the Blood Donor Program at the UCI Institute for Clinical and Translational Science (ICTS). Samples were collected in lavender top (K2EDTA) vacutainer tubes using venipuncture under sterile conditions and processed within several hours for downstream experiments.

2.3.5 CRC patient samples

A clinical pilot study was completed to validate the ability of our system to detect and quantify SNP cancer mutations in patient plasma. IRB-exempted and de-identified plasma samples from patients diagnosed with CRC (n = 7, average age = 59, stages I to III) were purchased from BioIVT (Westbury, NY). These samples were pre-determined to be positive for KRAS G12D by genetic sequencing by the vendor. As negative controls, non-screened plasma samples from healthy patients were obtained from the aforementioned ICTS donor program. In brief, cfDNA was extracted from 1 ml of patient plasma samples using a QIAamp Circulating Nucleic Acid kit (Qiagen, Cat#: 55114), following manufacturer's protocols. Purified cfDNA was eluted with Qiagen AE buffer and stored at −20 °C in DNA LoBind tubes prior to ddPCR assays. For ddPCR assays, purified cfDNA samples were mixed with a PCR master mix containing 0.5 µM primers and 0.25 µM probe specific against KRAS G12D (Table S2.1†) before droplet encapsulation using a flow-focusing microfluidic device. Droplet PCR conditions for this set of experiments were 95

°C for 3 min (1 cycle), 95 °C for 20 s and 60 °C for 20 s (40 cycles), and 4 °C (hold) using a C1000 Touch thermal cycler (Bio-Rad). After thermocycling, droplets were imaged on a fluorescent microscope as described below and median droplet fluorescence intensity was measured. Droplets were defined to be positive for KRAS G12D if they had a minimum signal-to-noise ratio of 3.0 compared to background non-fluorescent droplets.

2.3.6 IC3D digital droplet system

The IC3D platform technology has been utilized for a number of different applications including the detection of bacterial species-specific targets in blood using DNAzyme sensors and detection of circulating miRNA in plasma using EXPAR probes [24-25]. Without modifying the scanning instrumentation, we have adapted this platform into a droplet digital PCR assay by employing and optimizing several technical parameters including droplet size, DNA loading content, and the use of propynyl-dC (pdC) and propynyl-dU (pdU) modified dual-labelled PCR probes in order to detect ultra-rare circulating tumor DNA targets with high sensitivity and SNP-specific selectivity in this present study.

2.3.7 Microfluidic chip design and fabrication

We employed a custom microfluidic chip design with the architectures and operating mechanism described in Fig. 2.2. The fabrication for this four-nozzle droplet generation platform was based on a modified version of an established soft lithography procedure [24]. Microchannel architectures were designed using AutoCAD (Autodesk, San Rafael, CA, USA) and sent to CAD/ART services (Bandon, OR, USA) who provided high-resolution photomasks. SU-8 master molds were

produced with the recommended protocol from Micro-Chem. After wafer fabrication, PDMS was mixed in a 1 : 10 ratio of curing agent to elastomer and subsequently degassed and poured onto the SU-8-on-Si wafer master. PDMS was cured in an oven for 4 hours at 95 °C. Once cured, PDMS layers were peeled off and punched with a 1 mm diameter biopsy punch (Kay Industries Co., Tokyo, Japan) to create inlets and outlets. To bond PDMS to glass, both surfaces were treated with oxygen plasma in a PDC-32G Harrick Plasma cleaner (Harrick Plasma, Ithaca, NY) for 30 s and bound together. Finally, the droplet generation devices were post-baked in a convective oven at 95 °C for at least 24 hours.

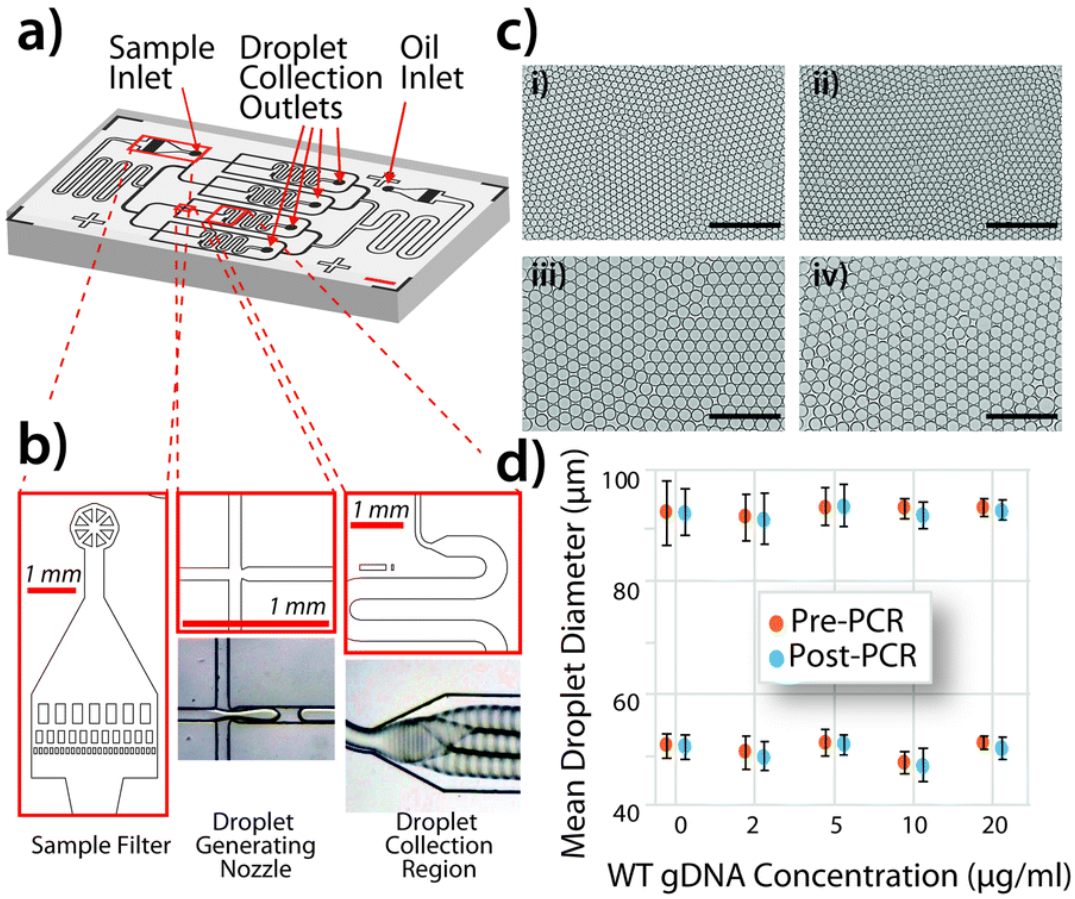


Fig. 2.2 High-throughput droplet generation and thermocycling stability. a) Schematic of custom droplet generation chip with inlets and outlets identified. b) Close-up view of three important features including: integrated sample and oil filter, droplet generating nozzle array (flow-focusing principle), and droplet collection region. c) Brightfield microscope images of a 2D monolayer of droplets in a hemocytometer, i) 50 µm droplets pre-PCR, ii) 50 µm droplets post-PCR, iii) 90 µm droplets pre-PCR, iv) 90 µm droplets post-PCR. Scale bar = 500 µm. d) Quantitative measurements of individual droplet diameters demonstrate robust droplet stability before and after thermocycling across a range of gDNA concentrations from 0–20 µg ml−1.

2.3.8 Droplet generation

Samples were emulsified with the use of 3M™ Novec™ HFE-7500 Engineered Fluid (fluorinated oil) containing 2% PFPE surfactant. For the disperse phase, aqueous solutions containing PCR reagents and DNA sample were loaded into syringes on top of 200 µl of HFE-7500 fluorocarbon oil without surfactant to displace sample volume and ensure full encapsulation of the aqueous layer. A second syringe was loaded with fluorinated oil with 2% PFPE surfactant to

serve as the continuous phase. Both syringes were mounted onto PHD Ultra™ syringe pumps (Harvard Apparatus) and connected to the respective aqueous and oil inlets on the microfluidic chip. Droplet generation was monitored with optical microscopes (Celestron® LCD digital microscope) to ensure consistent droplet production. Droplets were prepared for each sample using separate devices to avoid contamination and were collected in sterile collection tubes.

2.3.9 Microscope evaluation of droplet PCR assay performance

Microscope images of a 2D monolayer of droplets were recorded to evaluate droplet generation uniformity, droplet thermal stability, and PCR efficiency. Approximately 10 µl of droplets were dispensed onto a disposable hemocytometer (Invitrogen™ Countess cell counting chamber slides) prior to analysis. Quantitative measurements for each of these evaluations included average droplet diameter and % CV before and after thermocycling (using brightfield imaging at 10× magnification) and signal-to-noise ratios computed from median fluorescence signal intensity of droplets from a positive control sample (using fluorescence imaging (FITC) at 10× magnification).

2.3.10 Sample preparation for droplet partitioning experiment

To demonstrate the effects of rainfall on assay performance as higher concentrations of wild-type gDNA are loaded, 300 µl of PCR reaction mixture containing (PerfeCTa qPCR ToughMix reagent (1×), NP-40, forward and reverse primers (1 µM each), mutant probe (0.25 µM)), and 225,000 copies of synthetic KRAS G12D mutant were mixed with 0.1, 0.5, 1, 2.5, 5 or 10 µg of gDNA isolated from Jurkat cells. Additionally, a blank sample (which does not contain wild-type

Jurkat gDNA and KRAS G12D mutant) was also prepared as a non-template control. Fluorescent signal suppression ("rainfall") was assessed *via* quantitative intensity measurements of 2-dimensional droplet microscopy images (described above).

2.3.11 Sample preparation for multiplex droplet digital PCR study

To demonstrate the capacity of multiplex droplet digital PCR to detect mutant alleles (MT) in the presence of wild-type genome backgrounds (WT) using 90 µm droplets, 100 µl of PCR reaction mixture (PerfeCTa qPCR ToughMix reagent (1×), NP-40, forward and reverse primers (1 µM each), wild-type probe (0.25 µM), mutant probe (0.25 µM)), containing 75,000 copies of synthetic targets (or 0 as negative control) were mixed with 250 ng of Jurkat gDNA (equivalent to 75,757 haploid genome copies) to reach an mutant allele frequency of ~50% (or 0% for negative control). For calculating gene copy number in gDNA, haploid copy number dilutions were calculated based on the molecular weight of one normal haploid female genome equaling 3.3 pg [26].

2.3.12 Sample preparation for synthetic KRAS G12D study

For our test trial to assess the analytical sensitivity of our IC3D platform using 90 µm droplets, 850 µl of PCR reaction mixture (PerfeCTa qPCR ToughMix reagent (1×), NP-40, forward and reverse primers (1 µM each), mutant probe (0.25 µM)) containing different amounts of synthetic KRAS G12D targets (0, 68, 680, 3,400, 6,800, 34,000, 637,500 copies) were mixed with 2.125 µg of Jurkat gDNA (equivalent to 643,939 haploid genome copies) to reach mutant allele frequencies of 0, 0.01, 0.1, 0.5, 1, 5, and 50%, respectively.

2.3.13 Sample preparation for LS174T spiked blood experiment

To demonstrate the potential of our IC3D platform to detect CTCs in whole blood without pre-enrichment, we used the human colorectal adenocarcinoma model cell line, LS174T, to assess the analytical sensitivity of our assay. We spiked varying numbers of LS174T cells into 1 ml whole blood sample aliquots obtained from a healthy donor. Serial dilutions of LS174T prepared at 400k, 100k, 25k, 6.4k, 1.6k, 400, and 100 cells were spiked into each respective blood sample to establish a titration series that spans a broad dynamic range of sensitivity (Table S2.3†). To determine the wild-type leukocyte background of each sample, a set of three 1 ml blood samples from the same donor were lysed with an ACK lysis buffer kit (Thermo-Fisher). Red blood cells were fractionated and removed during the process while white blood cells were fully retained to allow precise and accurate leukocyte quantification on a hemocytometer. In short, 1 ml of EDTA-treated whole blood was diluted with ACK lysis buffer at a ratio of 1 : 10 and allowed to incubate at RT for 5 minutes for each replicate. The remaining leukocytes were centrifuged at 300 × *g* for 7 min at RT, washed with 4 °C cold PBS, centrifuged at 300 × *g* for 7 min at 4 °C, and resuspended with 4 °C cold PBS. Following ACK lysis, leukocytes were mixed with trypan-blue and loaded into a Hausser Scientific 3100 hemocytometer (Thermo-Fisher). White blood cells were counted and averaged from 4 quadrants for each replicate and were determined to have an expected background concentration of 3.95×10^6 cells per ml. A QIAamp Blood DNA miniprep kit (Qiagen) was then used, according to manufacturer's instructions, to purify and isolate all gDNA from the LS174T-spiked whole blood samples. This kit contains the proteolytic agent, proteinase K, which lyses all cells in the blood samples and allows us to perform downstream PCR-based molecular

detection without CTC loss. Purified gDNA samples were eluted with Qiagen AE buffer and stored at −20 °C in DNA LoBind tubes prior to experiments. Lastly, these samples were analyzed on a NanoDrop and determined to have absorbance ratios of $A_{260/280}$ and $A_{260/230}$ greater than 1.8. For IC3D experiments, each PCR condition utilized 6 µg of gDNA as template and had a final reaction volume of 300 µl. The PCR reactions contained PerfeCTa qPCR ToughMix reagent (1×), NP-40, forward and reverse primers (1 µM each), and mutant probe (0.5 µM). Due to the high concentration of background gDNA, samples were encapsulated in smaller 50 µm droplets in order to increase the fraction of binary droplets.

2.3.14 Droplet PCR

Prior to thermocycling, droplet samples were transferred to thin-walled PCR tubes, each containing 30 µL Novec™ 7500 oil with surfactant and 70 µL of droplets. For droplet PCR, the following PCR protocol was used: (1) 3 min at 95 °C (initial denaturation), (2) 20 s at 95 °C (denaturation), (3) 60 s at 58.1 °C, (4) repeat steps 2 & 3 for 40 cycles, and (5) an infinite 12 °C hold. Following thermocycling, samples were assessed *via* fluorescence microscopy (described above) and 3D fluorescent scanning (described below).

2.3.15 Droplet detection using 3D particle counter

The droplet scanning instrument used in this study consisted of a bench-top, horizontal-geometry confocal microscope. In this embodiment, a visible-wavelength excitation laser beam (λ = 469 nm with ∼2.2 mW typical output power) was focused by an under-filled low numerical aperture 20× objective (MV-20×, Newport) inside a cylindrical cuvette (Abbott) containing the

droplet sample. Slow (5 mm s^{-1}) vertical translation and fast (200 rpm) rotation of the cuvette transported the target fluorescent droplets across the Gaussian-shaped excitation volume. The emitted fluorescence signal was collected by the same objective, transmitted through a dichroic filter and collimated onto the sensitive area of a photomultiplier tube (H9305-04 PMT, Hamamatsu) for the time-trace acquisition.

Droplet samples for IC3D scanning were manually transferred from the individual PCR tubes of each sample into thoroughly cleaned, borosilicate glass cuvettes containing 1.5 ml of 3M™ Novec™ 7500 engineered fluid (fluorinated oil) with 2% surfactant and scanned for 2 minutes. Raw fluorescent time trace data collected from the droplet scanning instrument was analyzed using SimFCS (Laboratory for Fluorescence Dynamics, Irvine, CA, USA: http://www.lfd.uci.edu). To achieve high specificity while counting the number of detection events, the fluorescence time trace data was fit to a pre-determined Gaussian profile of fixed standard deviation and amplitude. Hits were only counted if the chi-square value of the shape-fit was statistically significant ($X^2 < 0.003$) and the peak amplitude was between a user-defined minimum and maximum threshold. Parameters were determined based on the highest concentrated sample in each of the following experiments to correctly identify hits with consistent widths and amplitudes and applied to remaining data (synthetic KRAS G12D study: SDV = 24, minimum amplitude threshold = 510 mV, maximum amplitude threshold = 3000 mV, chi-square threshold = 0.003, LS174T cell-spiking study: SDV = 16, minimum amplitude threshold = 200 mV, maximum amplitude threshold = 5000 mV, chi-square threshold = 0.01) (an SDV of 24 corresponds to a Gaussian peak of width equal to 145 data points, which at a sampling frequency of 64 kHz is approximately 2.26 milliseconds). Though the shape-fitting algorithm utilized in the

IC3D system can accurately differentiate positive, fluorescent droplets from other non-specific fluctuations in the recorded fluorescence signal, it does not exclude the possibility of false positives due to rare PCR false positive amplification.

2.3.16 qPCR for KRAS G12D mutant detection

Per 20 µl reaction volume, different amount of synthetic KRAS G12D targets (0, 16, 80, 160, 800, 1,700, 15,000 copies) and 50 ng of Jurkat gDNA were added to a PCR reaction mixture that included the following components: PerfeCTa PCR ToughMix reagent (1×), NP-40, forward and reverse primers (1 µM each), mutant probe (0.25 µM), wild-type probe (0.25 µM) and 1× Rox. Each sample condition was run in quadruplicate (20 µl in each well) with 7900 HT fast real-time PCR system (Applied Biosystems) using the same PCR thermal cycling condition as droplet PCR.

2.3.17 Assessing analytical sensitivity of Bio-Rad ddPCR QX-200 platform with LS174T spiked whole blood samples

To compare the performance of IC3D with the leading commercial ddPCR platform, we utilized the Bio-Rad QX-200 digital PCR system (Bio-Rad Laboratories, Hercules, CA, USA) to assess the analytical sensitivity of this commercial ddPCR system in detecting CTCs in whole blood without pre-enrichment, using samples of LS174T spiked in whole blood. In brief, ddPCR reaction master mixes were made containing 10 µl of 2× ddPCR Supermix, 1 µl of 20× human mutant KRAS p.G12D (c.35G>A) ddPCR FAM probe (dHsaCP2500596), 1 µl of 20× human KRAS wild-type for p.G12D (c.35G>A) ddPCR HEX probe (dHsaCP2000002), 7 µl of nuclease-free water, and 1 µl of purified gDNA for each reaction. LS174T spiked whole blood samples were prepared in duplicates

and at a concentration of 66 ng μl^{-1} based on manufacturer's recommendations for loading intact genomes.

Furthermore, according to Bio-Rad's calculations and estimates of 1 haploid genome weighing 3.3 pg, 66 ng of DNA should contain 20,000 haploid genome equivalents (or gene copies if homozygous) [21,26]. Each reaction mixture was then loaded into DG8 cartridges along with droplet generation oil to partition the samples using the QX-200 droplet generator. Thermal cycling conditions for this set of experiments were 95 °C for 10 min (1 cycle), 94 °C for 30 s and 55 °C for 1 min (40 cycles), 98 °C for 10 min (1 cycle) and 4 °C (hold). The QX-200 droplet reader was then used to assess droplets as positive or negative based on fluorescence amplitude and a set threshold. Data analysis was performed as recommended using the QuantaSoft Software version 1.7.4 and their Rare Mutation Best Practices Guidelines. A threshold of 2500 was used for both mutant and WT channels to determine the fractional abundance of KRAS G12D for each LS174T spiked whole blood sample.

2.3.18 Validating KRAS G12D mutant allele frequency of LS174T with Bio-Rad QX-200 ddPCR

Upon contacting ATCC, we discovered that there has been no previous validation study conducted by the vendor to determine the mutant allele frequency (AF) of KRASG12D for the clonal cell line, LS174T. Based on existing independent reports and to our best knowledge, LS174T has been confirmed to express KRAS G12D but at variable frequencies depending on lab settings [27-29]. This prompted us to use the Bio-Rad QX-200 Digital PCR system (Bio-Rad Laboratories, Hercules, CA, USA) in order to accurately and precisely determine the percent mutant zygosity of our cell source.

In brief, ddPCR reaction master mixes were made containing 10 µl of 2× ddPCR Supermix, 1 µl of 20× human mutant KRAS p.G12D (c.35G>A) ddPCR FAM probe (dHsaCP2500596), 1 µl of 20× human KRAS wild-type for p.G12D (c.35G>A) ddPCR HEX probe (dHsaCP2000002), 7 µl of nuclease-free water, and 1 µl of cell line gDNA for each reaction. LS174T or Jurkat (negative control) gDNA was prepared at a concentration of 12 ng μl^{-1} and added to PCR reaction mixtures in duplicates. As previously mentioned, with 1 haploid genome weighing 3.3 pg, 12 ng of DNA was expected to contain 3636 haploid genome equivalents (or gene copies if homozygous). The same droplet generation, PCR, and scanning conditions used on the Bio-Rad QX-200 platform described earlier were used for this set of experiments. For mutant zygosity confirmation, data analysis was performed as recommended by the manufacturer using the QuantaSoft software version 1.7.4. Threshold was determined by comparing ddPCR results between Jurkat, LS174T, and a no template control (nuclease-free water).

2.4 Results and discussion

2.4.1 High-throughput droplet generation

The ability of the IC3D system to efficiently analyze large sample volumes is enabled by the combination of high-throughput droplet generation and rapid, three-dimensional scanning. While other design strategies have been utilized to demonstrate ultrahigh-throughput droplet generation [30-31], the flow-focusing 4-nozzle droplet generator used in this study was chosen due to its ability to robustly generate precise droplets that maintain stability during thermocycling and its simple, single-layer fabrication and assembly (Fig. 2.2a). In order to

improve device robustness by lowering the risk of nozzle clogging, a series of micro-posts with variable spacing were integrated into the aqueous and oil inlets to serve as an in-line filter (Fig. 2.2b). By altering the channel geometry at the nozzle region, devices were designed to produce droplets of various diameters. In this study, devices were fabricated to produce droplets with average diameters of 50 and 90 μm (Fig. 2.2c).

Flow rates for the aqueous and oil streams to produce stable 90 μm droplets were 50 μl min^{-1} and 75 μl min^{-1}, respectively and 15 μl min^{-1} and 80 μl min^{-1}, respectively for the production of stable 50 μm droplets. These flow rates correspond to droplet generation times of 66 min and 20 min for a 1 ml aqueous sample for 50 μm and 90 μm droplets, respectively. Droplet stability was assessed before and after 40 cycles of thermocycling by measuring droplet diameters in a 2-dimensional monolayer on a hemocytometer. Over a range of background gDNA content from 0–20 μg ml^{-1}, average diameters for the 50 μm before and after thermocycling were 50.11 μm (% CV = 3.73%) and 49.98 μm (% CV = 4.31%), respectively, while the 90 μm droplets averaged 91.56 μm (% CV = 3.60) before thermocycling and 91.62 μm (% CV = 3.28%) after thermocycling (Fig. 2.2) (data from 20 unique fields of view with approximately 900–1000 droplets per field for 50 μm droplets and 200–300 droplets per field for 90 μm droplets).

2.4.2 Droplet PCR assay

Somatic point mutations within tumoral DNA are highly specific biomarkers that can distinguish cancer cells from normal cells. These biomarkers have been shown to be applicable for cancer diagnosis, prognosis, selection of rational combination therapies, and monitoring of patients, but have yet to become routine assays employed in clinical oncology due to the unmet

need for a more highly sensitive strategy that can discriminate tumor-specific mutations in a large excess of non-mutated DNA from normal cells. Nevertheless, the sensitivity of traditional approaches to SNP detection (techniques such as Sanger sequencing or TaqMan PCR) possess only a sensitivity of 1–10% [32-33], thereby suffering from false negatives and limited reproducibility. A combination of conventional target-specific TaqMan PCR and a droplet microfluidic system allows the partitioning of bulk PCR solution into a massive number of picoliter droplets in which each partition contains one or no target molecules. Partitioning of target mutant DNA away from highly homologous wild-type DNA permits absolute digital quantification of mutant targets with single-molecule sensitivity, reduces interference from background DNA, and, therefore greatly increases sensitivity over conventional approaches.

We first demonstrated the capacity of our system for multiplexed droplet digital PCR detection of different targets by detecting mutant alleles in the presence of wild-type genomic backgrounds (Fig. 2.3). Primers and probes used for different SNP mutation detection (KRAS G12D and BRAF V600E, genetic markers among the most prevalent for CRCs and many other types of cancer) were designed and optimized to have the same annealing temperature for droplet digital PCR. Gene fragments carrying KRAS G12D or BRAF V600E were synthesized and sequence-verified by IDT, whereas background genomic DNAs harbouring wild-type KRAS and BRAF alleles were isolated and purified from Jurkat cells. In a 100 µl PCR volume, 250 ng of Jurkat gDNA was mixed with 75,000 mutant copies to reach approximately an equal molar ratio of wild-type to mutant copies. Using the droplet generation approach described previously, mutant copies as well as background DNA were compartmentalized in 90 µm droplets together with FAM probe specific for mutant targets and CAL560 probes specific for corresponding wild-type targets.

Droplets were then thermocycled and the fluorescence signal of droplets was measured. The amplification of mutant DNA resulted in a green-fluorescent droplet while the amplification of wild-type DNA resulted in a red-fluorescent droplet. As shown in Fig. 2.3, in the wild-type only control, only red-fluorescent droplets but not green-fluorescent droplets were observed. For samples containing both wild-type and mutant genes, the fluorescent droplets were either red (due to amplification of a wild-type gene), green (due to amplification of a mutant gene) or yellow (due to amplification of both a mutant and wild-type gene). This result suggests that we could specifically discriminate mutant alleles from wild-type alleles *via* a droplet digital PCR assay. While multiplex detection and analysis is possible with the IC3D system, the following studies demonstrate IC3D performance on SNP detection, using single-color mutation detection assays targeting KRAS G12D, one of the key and most prevalent cancer mutations associated in many different cancers for the remaining experiments [34].

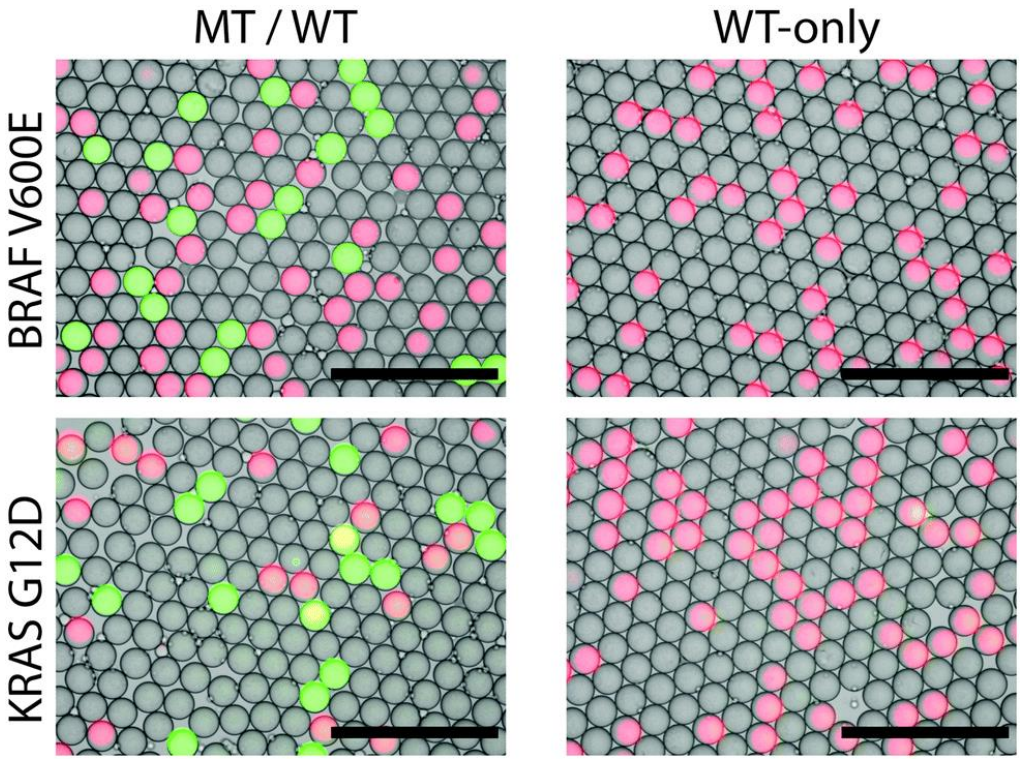


Fig. 2.3 Fluorescent microscope images of droplets with different target mutations. RGB multichannel images composed of brightfield droplet image (grey), MT-specific probe (FAM fluorophore) (green), and WT-specific probe (CAL560 fluorophore) (red). Scale bar = 500 µm.

2.4.3 Analysis of KRAS G12D CRC patient samples

Furthermore, a pilot study using KRAS G12D mutant-positive (pre-determined by sequencing) CRC patient plasma samples (*n* = 7) and healthy donor samples (*n* = 5) demonstrated that our droplet PCR assay targeting KRAS G12D can accurately identify these cancer cases with 100% clinical sensitivity and specificity (Fig. 2.4).

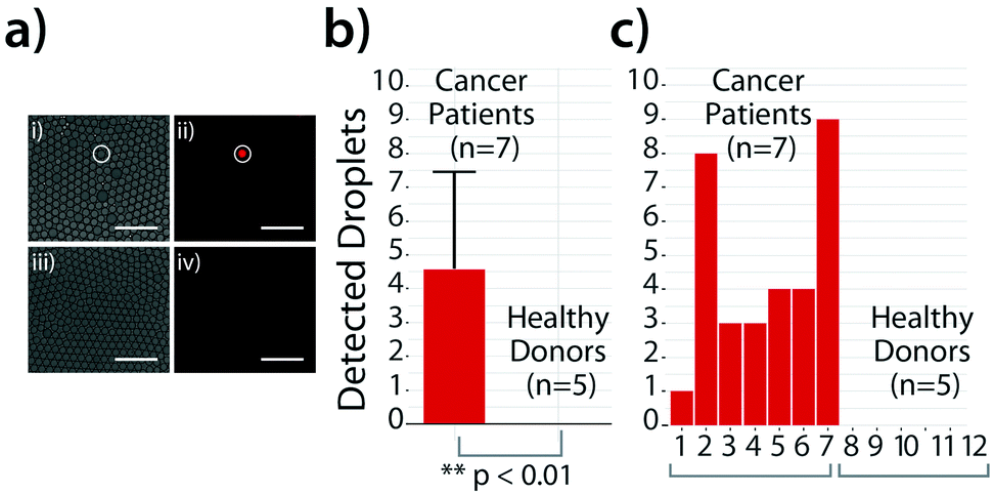


Fig. 2.4 Analysis of KRAS G12+ CRC patient plasma samples using our droplet PCR assay. a) Representative fluorescent microscope images of droplets following PCR from a stage IIA cancer patient (i – BF, ii – fluorescent) and a control healthy patient (iii – BF, iv – fluorescent) (scale bar = 500 µm). Positive droplets were defined as those with a signal-to-noise ratio (SNR) greater than 3.0 and then enumerated in group (b) or as individual patient (c). Y axis = number of detected droplets in 100 µl final volume.

2.4.4 Demonstration of partitioning effect on assay sensitivity

Unlike commercially available dPCR systems, the IC3D technology can uniquely analyze large sample volumes, resulting in two key advantages: (1) the ability to load more volume (or DNA content) from a clinical sample per assay, and (2) the ability to achieve a higher degree of sample partitioning. To demonstrate the relationship between detection sensitivity and the degree of sample partitioning, an experiment was performed using increasing amounts of genomic wild-type gDNA and measuring the total fraction of droplets that reached peak signal amplification. The increasing concentration of wild-type gDNA simulates the effect of generating fewer partitions for the same sample. Theoretical calculations for estimating the percent of "rainfall" (defined as the number of intermediate-intensity droplets divided by the total number of positive droplets) were based on partitioning statistics for a Poisson distribution as described

in Milbury *et al.* [35] This theory estimates the probability of *k* targets in one droplet given an average "loading" of λ targets per droplet (λ_w = WT loading, λ_m = MT loading) as:

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Therefore, the rainfall fraction can be determined as the number of droplets containing both MT and WT copies divided by the total number of droplets containing MT copies, or N_{Dual}/N_{MT}:

$$\text{Rainfall fraction} = \frac{\left(1-e^{-\lambda_w}\right)\cdot\left(1-e^{-\lambda_m}\right)}{\left(1-e^{-\lambda_m}\right)}$$

In theory, for digital PCR to be truly digital, a sample must be partitioned into only empty droplets and single-copy droplets (*i.e.* 0's and 1's, respectively) to achieve single-copy amplification. As seen in Fig. 2.5, as the amount of WT gDNA background increases (and consequently the fraction of truly binary droplets decreases), the fraction of droplets at maximal PCR amplification decreases accordingly. This phenomenon, commonly referred to as "rainfall", exists when the binary partitioning condition is not met. In these scenarios, absolute quantification can result in inaccurate measurements. Additionally, as in the case of SNP detection where wild-type and mutant sequences are amplified with the same primer sequence, incomplete partitioning can yield droplets containing both mutant and wild-type copies which results in compromised detectable signal.

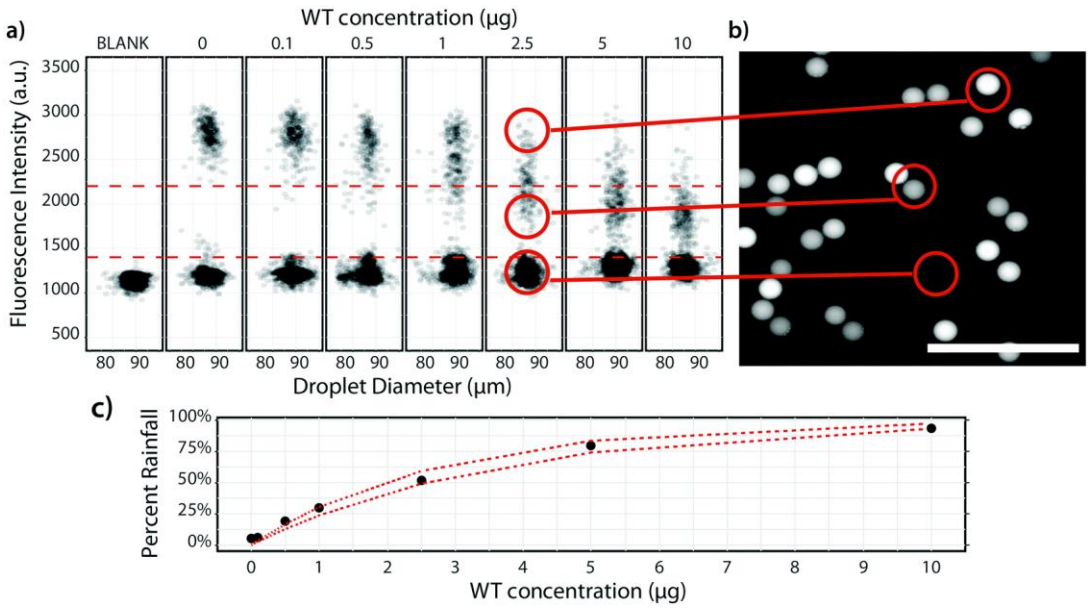


Fig. 2.5 Effect of number of partitions on assay performance. a) Calculated percent rainfall (intermediate droplet fraction/total positive droplet fraction) from droplet intensity data in panel b. b. Dashed red lines refer to the lower and upper boundaries of the "intermediate-intensity" zone, where droplets are distinguished from empty ("negative") droplets and fully amplified droplets. b) Raw fluorescence intensity data collected for partitioning experiment. c) Agreement with partitioning theory. Solid circles represent calculated percent rainfall from panel a. Dotted red lines refers to upper and lower 95% confidence interval for theoretical percent rainfall given MT and WT copy numbers and an average droplet diameter of 90 µm. With its ability to accommodate very large sample volumes, the IC3D system can efficiently generate and analyze significantly more partitions than competing dPCR systems. Though different strategies exist for increasing the throughput of droplet detection in the context of ddPCR assays including large 2-dimensional arrays [36] and optofluidic devices [37-38], the simplicity of the 3-dimensional droplet analysis technique presented here allows IC3D to scale easily to accommodate large sample volumes while maintaining exceptional sensitivity. In applications that require high concentrations of gDNA to be analyzed, the ability to interrogate significantly more partitions results in a higher fraction of truly binary droplets, which has the potential to greatly improve assay sensitivity.

2.4.5 IC3D ddPCR detection of synthetic KRAS G12D mutant in Jurkat gDNA

Plasma cell-free DNA (cfDNA) as a liquid biopsy assay has been shown to be a valuable surrogate specimen for detecting tumor-specific aberrations. However, detection of tumor-derived cfDNA has proven to be challenging in clinics, because tumor derived cfDNA often represents a small fraction (<1%) of total cfDNA [32]. We aimed to explore the clinical utility of detecting cfDNA isolated from plasma using IC3D. As a proof of concept, various copies of synthetic KRAS G12D mutant fragments were spiked into Jurkat gDNA (serving as the background of normal DNA), resulting in a wide range of mutant allele frequencies from 0.01% to 5%. In order to analyze these samples on the IC3D system, they were first encapsulated into 90 µm droplets (ensuring a high fraction of truly binary droplets), thermocycled to amplify the fluorescence signal from droplets containing the KRAS G12D target, and transferred to a cuvette for rapid 3-dimensional fluorescent scanning. Following scanning, the fluorescent time trace data was fit to a pre-determined shape fitting algorithm corresponding to the signal from true positive droplets and events were enumerated (Fig. 2.6). As WT-containing droplets exhibit a dimly fluorescent signal in between the intensity of empty droplets and MT-containing droplets, stringent shape fitting algorithm parameters were selected to successfully differentiate positive droplets containing mutant targets (Fig. 2.6C) from other non-specific fluctuations in the recorded fluorescence signal (Fig. 2.6B).

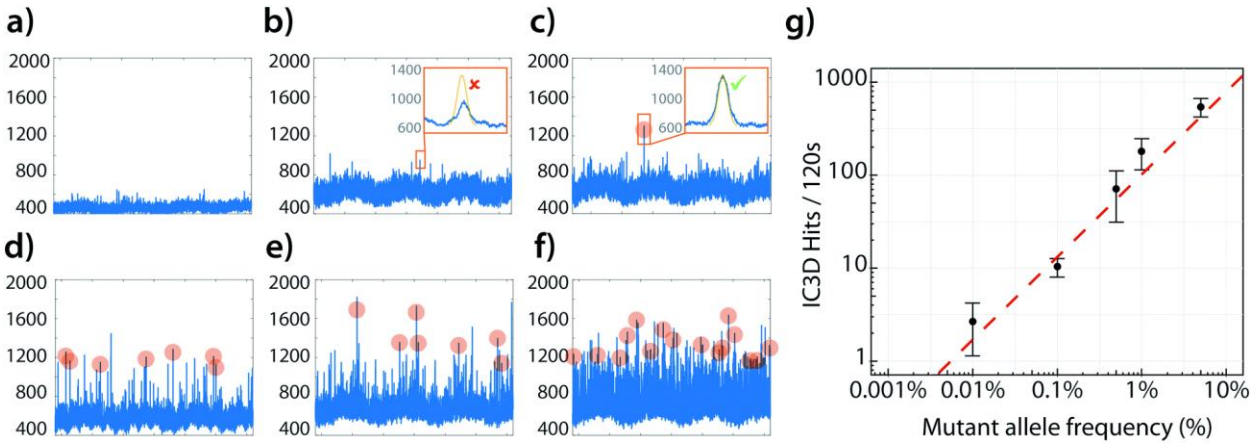


Fig. 2.6 Detection of synthetic KRAS G12D mutant in Jurkat gDNA by IC3D. Panels a–f: raw fluorescent time trace data (y-axis = PMT signal (mV), x-axis = 5 seconds acquisition time): a) blank, b) 0% AF, c) 0.1% AF, d) 0.5% AF, e) 1% AF, f) 5% AF. Orange circles denote confirmed positive droplet. g) Concentration curve of detected positive droplets versus % AF. Points represent combined average of three replicates (error bars denote ± standard deviation, n = 3).

We demonstrated that the full range of allele frequencies from 0.01% to 5% can be consistently detected with the IC3D system across three independent replicates (Fig. 2.6). Furthermore, the aggregate response of all 3 replicates demonstrated a strong linear relationship between the sample concentration (0.01–5% AF) and the number of IC3D hits (adjusted R^2 = 0.9723). With a false positive rate of 0%, the IC3D system was able to detect all 3 replicates of the 0.01% AF sample. This experiment effectively demonstrated clinical feasibility of using IC3D for detecting cfDNA isolated from plasma.

For comparison, a conventional real-time PCR assay was performed for detection of synthetic KRAS G12D mutant allele frequencies of 0% (negative control), 0.1%, 0.5%, 1%, 5%, 10% and 50% in the presence of the same amount of background WT Jurkat gDNA (50 ng/20 µl). The KRAS G12D and WT targets were amplified with common primers, but differentiated by separate fluorescence-labelled BHQplus probes. The FAM-labelled probe targeted the KRAS G12D allele,

while the CAL Fluor Orange560-labeled probe (VIC alternative probe) targeted the WT allele. In four replications per real-time PCR reaction, VIC fluorescence was detected in all samples with C_T (cycle threshold) values ranging between 25.1–25.7, indicating that a consistent level of WT background was present in all tested conditions. With regard to mutant detection, no FAM fluorescence was observed for the negative samples, but successful detection was observed in samples with 50% and 10% mutant allele frequencies (Fig. S2.2). However, no FAM fluorescence was detected for samples with 1%, 0.5%, and 0.1% mutant allele frequencies. For the 5% allele frequency sample, weak fluorescence was detected in two replications only, whereas no fluorescence was detected in the other two replications (Fig. S2.2). Hence, stable amplification was only observed for 10% mutant allele frequency samples or higher, indicating the lowest limit of detection by conventional real-time PCR for KRAS G12D detection with these study conditions is between 5% and 10% mutant allele frequency. By comparison, the IC3D system demonstrated a 500-fold increase in sensitivity by detecting an allele frequency of 0.01% in an equivalent assay.

2.4.6 IC3D ddPCR detection of spiked LS174T cells in whole blood

We next evaluated the potential of IC3D ddPCR in analyzing total DNA isolated from blood specimens. Because the amount of cfDNA is negligible compared to DNA isolated from white blood cells, for simplicity, our model used healthy donor blood samples spiked with cancer cells in this set of experiments. Therefore, this study further allowed us to evaluate whether IC3D ddPCR can directly detect CTCs using genetic markers without a pre-enrichment step, which was largely impossible before due to the interference of WT gDNA background.

In this experiment, cancer cells harboring the KRAS G12D mutation were spiked into aliquots of unprocessed, whole blood. To accommodate this large genomic background, we utilized smaller droplets with an average diameter of 50 µm (compared to 90 µm in previous experiments) to achieve efficient partitioning of samples with a high background of gDNA (approx. 20 µg ml^{-1}), resulting in the detection of between 100–400 cancer cells per ml of whole blood in this study (Fig. 2.7) (false positive rate = 0%). By loading significantly more gDNA than demonstrated in the previous synthetic KRAS G12D detection study, an equivalent allele frequency between 0.00125–0.005% was detected in this experiment. Furthermore, increasing number of cells corresponded with an increase in the number of detected IC3D hit events and agreed well with results from the synthetic KRAS G12D detection study.

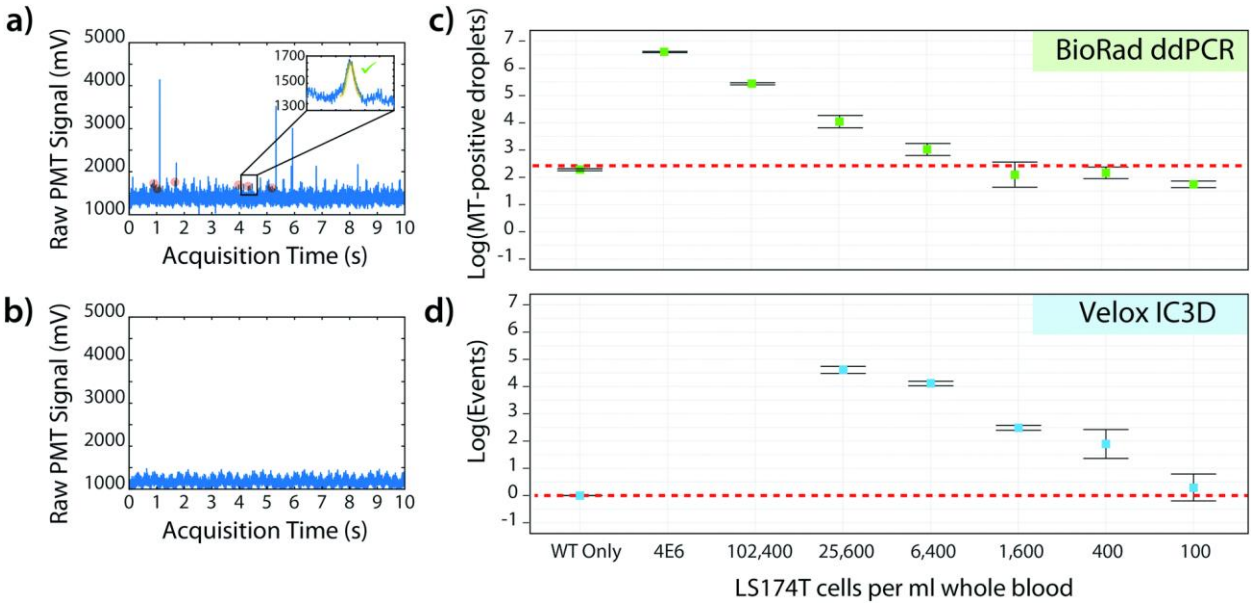


Fig. 2.7 Detection of spiked LS174T in whole blood using IC3D. Panels a and b represent raw fluorescent time trace data recorded on IC3D for a positive sample (1600 cells per ml) and negative sample, respectively. The inset in panel a is of a representative positive event where the shape-fitting algorithm criteria was met and displayed over an x-axis of 6 ms. Panels c and d demonstrate the sensitivity of the Bio-Rad ddPCR and IC3D systems for the detection of KRAS G12D mutations as a genetic approach for quantifying CTCs in whole blood extractions. c) Bio-Rad ddPCR results; y-axis = log reported Poisson MT-positive droplets, red dashed line = false positive rate based on events detected in WT-only negative control, error bars = relative error bar (for symmetrical display on log-scale), defined as ±0.434*stDev/y across 2 replicate wells. d) IC3D results; y-axis = log events (per 120 s), error bars = relative error (described above) across 3 independent sample replicates.

2.4.7 Comparison to Bio-Rad ddPCR

In order to determine the detection limit of the leading commercial ddPCR system using the same approach for KRAS G12D genetic detection of CTCs, a similar experiment was performed with samples from the same cell spiking dilution from above using the Bio-Rad QX-200 ddPCR system. Like many other conventional particle counting systems, droplets in the Bio-Rad ddPCR system are counted on a 1D, flow-based, on-chip system that operates at ~100 s of particles per s with a relatively small number of parallel reactions (20 µL sample volume or 20,000

droplets) [21, 39, 40]. Based on published guidelines for the Bio-Rad system, the maximum number of intact genomic DNA copies per 20 µl reaction volume is approximately 20,000 copies (or 66 ng per 20 µl reaction volume), above which the PCR reaction is inhibited [21, 41]. We have found that the sensitivity of the Bio-Rad system is approximately 0.08 to 0.33% (*i.e.*, 16 to 66 targets in 20,000 copies of genomic DNA) (Fig. 2.7c), which agrees with published studies from multiple groups concerning the detection of KRAS G12D [42-45].

2.4.8 Data analysis and statistics

In order to assess the analytical performance of the IC3D and Bio-Rad ddPCR systems in terms of quantitation limit, we employed a simple definition of the analytical cutoff as the lowest measured concentration where the lower relative error bar does not cross the average false positive rate. Similar to the analytical cutoff used in the Bio-Rad's Rare Mutation Best Practice Guidelines, this metric establishes a straight-forward way of empirically determining the limit of detection of each assay based on actual tested mutant positive and negative controls without contingency on theoretical extrapolation. For both systems, we defined the false positive rate as the average number of positive detection events (*i.e.* droplets) for negative control samples that contain an equivalent amount of WT but without the addition of LS174T cells.

Since the Bio-Rad's QuantaSoft software analysis tool allows for user-defined thresholding to define the acceptance criteria for positive events, we set a conservative threshold at 2,500, validated by the strong agreement of higher mutant fraction samples with expected mutant fractions (Fig. S2.1†) [MF = 0.33–4.76%, confidence level > 96.9%]. Applying this same threshold to wells containing the wild type only negative control samples resulted in a false

positive rate equivalent to a mutant fraction of approximately 0.07%, which agrees well with reported false positive rates exceeding 0.02% for KRAS G12D [42-45]. Due to this high false positive rate, the limit of detection for detecting CTCs spiked in whole blood on the Bio-Rad ddPCR system is 6,400 cells per ml (approximately 0.08% in allele frequency). Conversely, the high specificity of the IC3D detection and analysis platform facilitated by a robust shape-fitting algorithm resulted in a false positive rate of 0%, indicating a limit of detection of less than 400 cells per ml (approximately 0.005% in allele frequency). As a result, 2/3 replicates of 100 cells per ml (0.00125% allele frequency) and 3/3 replicates of 400 cells per ml (0.005% allele frequency) were in a detectable range above the negative control line. This suggests that our IC3D system can be used as a powerful tool to detect ctDNA from total DNA isolated from blood specimens with an allele frequency sensitivity at least 50-fold higher than the Bio-Rad ddPCR system per run. For additional context, the performance and throughput of the IC3D system as demonstrated in this study are compared to two commercial ddPCR systems in Table S2.6.† Therefore, we emphasize that droplet counting throughput is a critical performance metric which IC3D outperforms existing digital PCR systems.

2.5 Conclusions

We demonstrated that the IC3D platform can greatly improve the sensitivity of ctDNA tests (*i.e.* at least 50 to 1000× more sensitive than current dPCR and qPCR assays, respectively) due to larger sample input, and greater numbers of partitions. The IC3D platform uniquely enables analysis of total tumor DNA isolated from blood samples regardless of their origins (*i.e.* cfDNA, CTC DNA, or exosomal DNA), which will increase clinical sensitivity and specificity,

and minimize biases of individual biomarkers and inter-assay variability due to pre-analytical preparations of individual markers. In fact, this is one of the first demonstrations that CTCs can be directly detected and profiled using their genetic markers without a pre-enrichment step, therefore eliminating the technical issues related to the efficiency of CTC purification and enrichment [46,47]. We expect our assay sensitivity can be further improved through genomic DNA digestion (more input sample loading), smaller droplets (more partitioning, as demonstrated by Pekin *et al.*, 2011 [26]) and the use of competitive probes such as PNA and XNA (higher SNR). We acknowledge that the Bio-Rad ddPCR and other systems can analyse large amount of DNA to achieve similar sensitivity if they run many parallel assays for a single sample (*e.g.* Bio-Rad ddPCR's 96 well-plate format), however, such implementation will incur significantly increased assay time and cost. For instance, the Bio-Rad ddPCR system would need to run ~50 reactions (half of a 96 well-plate) to analyze 1 ml of blood in 2 hours, which would only require a single run using the IC3D system in several minutes.

The inability of current technologies to sufficiently identify target ctDNA from a vast excess of wild-type counterparts especially in early-stage (stages 0, I and II) cancers has limited the use of liquid biopsy for early detection and screening, prognosis, and treatment stratification. With this unprecedented high sensitivity, our IC3D assay has the potential to change how we detect, treat, and monitor cancer patients at an early stage when interventions are most effective. We are currently evaluating the prognostic implications of this IC3D ctDNA assay for MRD and recurrence detection in samples from early stage patients with CRC in an ongoing clinical study. In this study, we will have primary tumors genetically profiled and then run a targeted pre-defined tumor genetic panel using IC3D to guide personalized cancer monitoring

and treatment. We emphasize that the IC3D technology intends to offer new capabilities (*i.e.*, unprecedented high sensitivity) and to complement, rather than replace, existing technologies including NGS that can analyze broader genetic profiles. Finally, IC3D ddPCR assays can be applied to other areas where large sample input is required for rare target detection including HIV reservoir analysis, prenatal screening, and study of transplanted cells *ex vivo* in cell therapy.

2.6 Acknowledgements

This study is supported by NIH/NIAID (1 R01 AI117061), UCI Applied Innovation's Proof of Product (POP) Grants, and the National Cancer Institute of the National Institutes of Health under award number P30CA062203 (The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health). T. V. is supported by NSF GRFP (DGE-1839285).

Supplementary Table 2.1. List of primers and probes used in this paper

Oligo Name	Sequence	Vendor	Used in Figure(s)
Multiplex droplet PCR, IC 3D ddPCR detection of KRAS G12D mutant in Jurkat gDNA and LS174T-Spiked WB			
KRAS_forward primer	5'-GGCCTGCTGAAAATGACTGAA-3'	LGC	Figures 5,6,7
KRAS_reverse primer	5'-GCTGTATCGTCAAGGCACTCTT-3'	LGC	Figures 5,6,7
KRAS Wild-type Probe	5'CAL560/pdUGGAGpdCpdUGGpdUGGpdCGpdU/3'BHQ-plus	LGC	Figures 5,6,7
KRAS Mutant Probe	5'FAM/pdUGGAGpdCpdUGApdUGGpdCGpdUA/3'BHQ-plus	LGC	Figures 5,6,7
BRAF_forward primer	5'-ACCTCAGATATATTTCTTCATG-3'	LGC	Figure 3
BRAF_reverse primer	5'- CCAGACAACTGTTCAAAC-3'	LGC	Figure 3
BRAF Wild-type Probe	5'CAL560/pdUpdCGAGApdUpdUpdUpdCApdCpdUGpdUAGpdC/3'BHQ-plus	LGC	Figure 3
BRAF Mutant Probe	5'FAM/pdUpdCGAGApdUpdUpdUpdCpdUpdCpdUGpdUAGpdC/3'BHQ-plus	LGC	Figure 3
Analysis of KRAS G12D in CRC patient samples			
KRAS_forward primer	5'- ATTATAAGGCCTGCTGAAAATGAC-3'	IDT	Figure 4
KRAS_reverse primer	5'- TGTATCGTCAAGGCACTCT-3'	IDT	Figure 4
KRAS Mutant Probe	5'ALEXA647/ TTGGAGCTGATGGCGCCGAC/3'BHQ-2	IDT	Figure 4

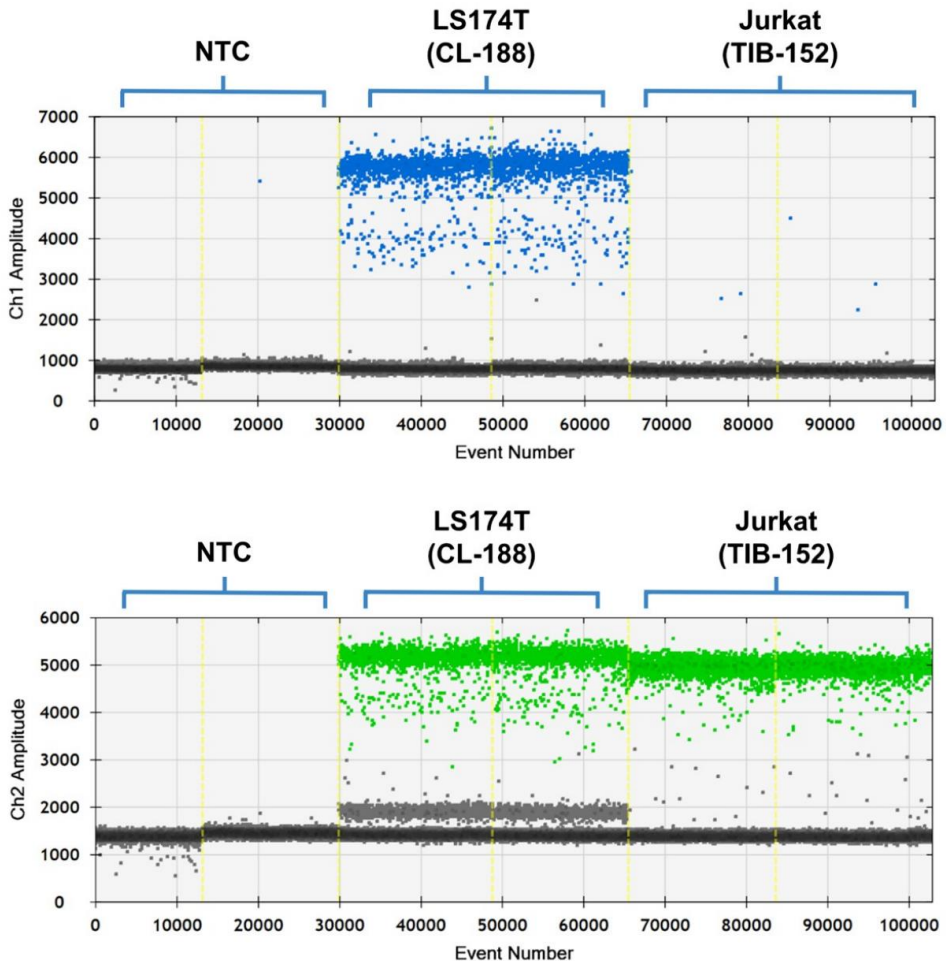
*pdC: propynyl-dC; pdU: propynyl-dU

Supplementary Table 2.2. Summary of synthetic KRAS G12D and BRAF V600E gBlocks® gene fragments

Synthetic KRAS G12D gBlocks® gene fragment	
Molecular Weight	335304.6
fmoles/ng	2.98
Sequence Length (bp)	543
Sequence content (5' -> 3'):	
TGACATACTCCCAAGGAAAGTAAAGTTCCCATATTAATGGTTACATATAACTTGAAACCCAAGGTACATTTCAGATAACTTAACTTTCAGCA TAATTATCTTGTAATAAGTACTCATGAAAATGGTCAGAGAAACCTTTATCTGTATCAAAGAATGGTCCTGCACCAGTAATATGCATATTAAA ACAAGATTTACCTCTATTGTTGGATCATATTCGTCCACAAAATGATTCTGAATTAGCTGTATCGTCAAGGCACTCTTGCCTACGCCATCAGC TCCAACTACCACAAGTTTATATTCAGTCATTTTCAGCAGGCCTTATAATAAAAATAATGAAAATGTGACTATATTAGAACATGTCACACATA AGGTTAATACACTATCAAATACTCCACCAGTACCTTTTAATACAAACTCACCTTTATATGAAAAATTATTTCAAAATACCTTACAAAATTCAA TCATGAAAATTCCAGTTGACTGCAGACGTGTATCGTAATGAACTGTACTTCATTTACAAACTCCTCCATCGACGCTTAAGAA	

Synthetic BRAF V600E gBlocks® gene fragment	
Molecular Weight	131462.3
fmoles/ng	7.61
Sequence Length (bp)	213
Sequence content (5' -> 3'):	
AGGGCCAAAAATTTAATCAGTGGAAAAATAGCCTCAATTCTTACCATCCACAAAATGGATCCAGACAACTGTTCAAACTGATGGGACCCAC TCCATCGAGATTTCTCTGTAGCTAGACCAAAATCACCTATTTTTACTGTGAGGTCTTCATGAAGAAATATATCTGAGGTGTAGTAAGTAAA GGAAAACAGTAGATCTCATTTTCCTATCAGA	

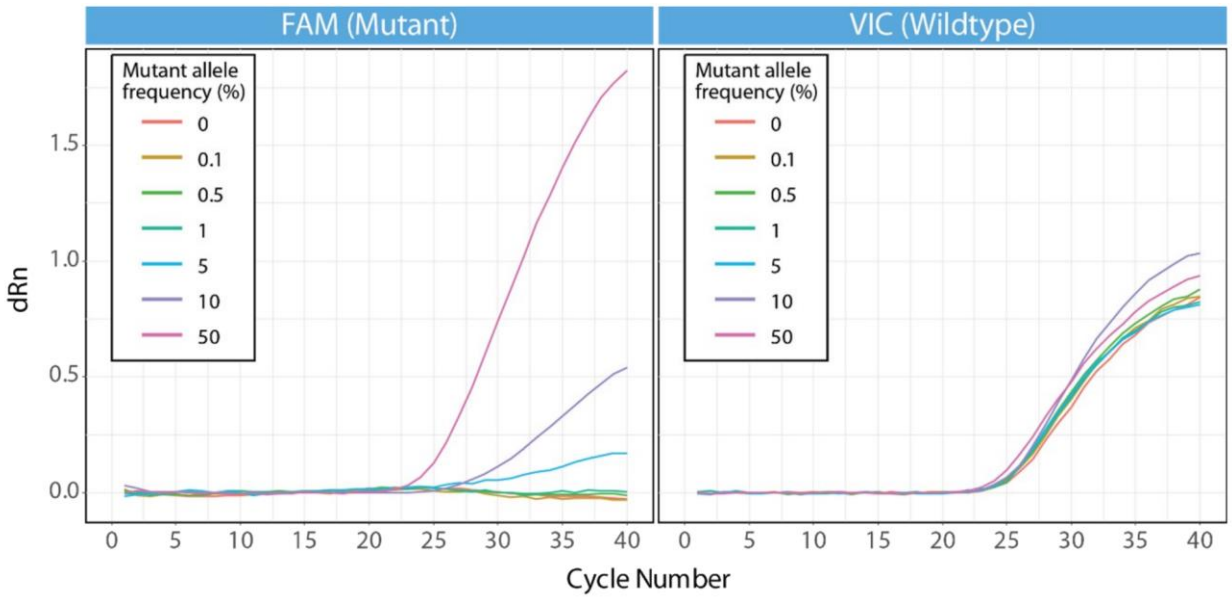
Supplementary Figure 2.1. Determining KRASG12D mutant allele frequency of LS174T via Bio-Rad QX-200 ddPCR



Sample Type	gDNA Loaded per well (ng)	KRASG12D Mutant (CH1 - FAM)			Wildtype (CH2 - VIC)			Measured Mutant Fraction (%)	Measured WT Fraction (%)
		Positive Droplets	Negative Droplets	Copies Detected	Positive Droplets	Negative Droplets	Copies Detected		
Blank	0	1	30,022	1.4	0	30,023	0	0	0
LS174T	12	2,505	33,042	3,446	5,184	32,163	3,352	50.69	49.31
Jurkat	12	7	37,340	9	6,768	7,426	7,040	0.12	99.88

*n = 2 for all sample types. Blank = no template control (nuclease-free water).

Supplementary Figure 2.2. Amplification plots for qPCR data (parallel to synthetic KRAS G12D detection experiment)



Supplementary Table 2.3. Blood spiking conditions with LS174T cells and theoretical mutant and wild-type copies calculations

# of LS174T Cells Spiked In 1 ml WB	# of Leukocytes in 1 mL WB	Theoretical # of KRASG12D copies	Theoretical # of Wild-Type copies	Theoretical KRASG12D AF (%)
409,600	3.95×10^6	415,252	8,309,522	4.76
102,400	3.95×10^6	103,813	8,002,380	1.28
25,600	3.95×10^6	25,953	7,925,595	0.327
6,400	3.95×10^6	6,488	7,906,398	0.082
1,600	3.95×10^6	1,622	7,901,600	0.021
400	3.95×10^6	405	7,900,399	0.005
100	3.95×10^6	101	7,900,100	0.00125
0 (WT)	3.95×10^6	0	7,900,000	0

* An expected background of 3.95 x 106 leukocytes cells per ml were determined with a hemocytometer. 50.69% mutant zygosity was used to calculate the theoretical # of KRASG12D copies contribution from LS174T.

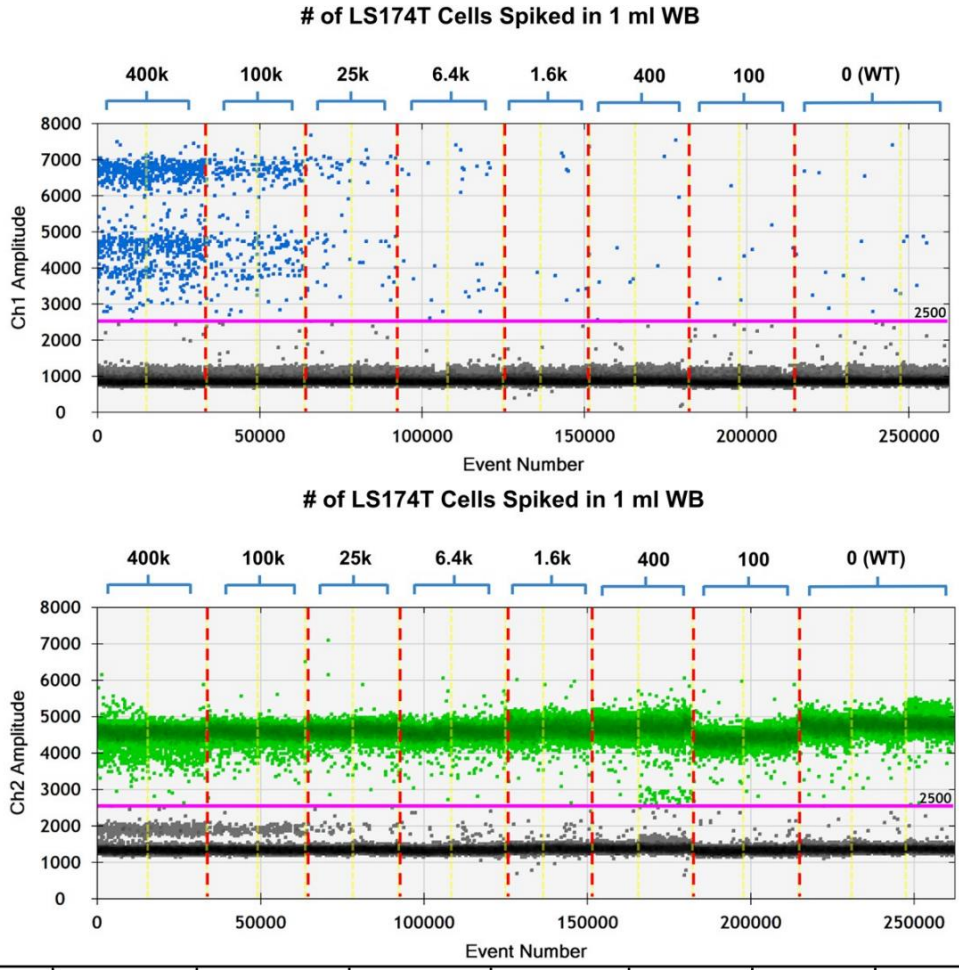
Supplementary Table 2.4. IC3D hit counts for synthetic KRAS G12D experiment

Replicate	AF %	Hits / 120s
A	BLANK	0
A	0.00000%	0
A	0.01000%	1
A	0.10000%	9
A	0.50000%	82
A	1.00000%	257
A	5.00000%	427
B	0.00000%	0
B	0.01000%	3
B	0.10000%	13
B	0.50000%	27
B	1.00000%	146
B	5.00000%	536
C	0.00000%	0
C	0.01000%	4
C	0.10000%	9
C	0.50000%	104
C	1.00000%	139
C	5.00000%	668

Supplementary Table 2.5. IC3D hit counts for LS174T cell-spiking experiment

Replicate	# Spiked Cells/ ml whole blood	AF %	Hits / 120s
A	0	0.00000%	0
A	100	0.00125%	3
A	400	0.00500%	1
A	1600	0.02100%	9
A	6400	0.08200%	48
A	25600	0.32700%	67
B	0	0.00000%	0
B	100	0.00125%	1
B	400	0.00500%	3
B	1600	0.02100%	14
B	6400	0.08200%	66
B	25600	0.32700%	123
C	0	0.00000%	0
C	100	0.00125%	0
C	400	0.00500%	16
C	1600	0.02100%	13
C	6400	0.08200%	70
C	25600	0.32700%	114

Supplementary Figure 2.3. Bio-RAD QX-200 ddPCR for LS174T cell-spiking experiment



Expected Mutant Fraction (%)	# of LS174T Cells Spiked In 1 ml WB	KRASG12D (CH1 - FAM)	Wild-Type (CH2 - VIC)	Measured Mutant Fraction (%)	Standard Deviation	*t*-value	*p*-value	Confidence Level (%)
		Copies Detected	Copies Detected					
4.76	409,600	1480	27,840	5.05	0.0044	980	< 0.01	< 99.99
1.28	102,400	458	30,760	1.47	0.0682	19.4	< 0.01	< 99.9
0.33	25,600	114	30,480	0.37	0.1996	1.44	0.031	96.9
0.08	6,400	41.2	30,360	0.14	0.0723	0.869	0.09	91
0.02	1,600	16.2	31,220	0.05	0.0551	0.299	0.295	70.5
0.005	400	17.4	31,980	0.05	0.0297	0.473	0.207	79.3
0.0013	100	11.4	28,560	0.04	0.0108	2.52	0.008	99.2
0	0 (WT)	35.4	51,080	0.07	0.0013	**reference**	**reference**	**reference**
0	0 (NF H2O)	5.8	0	100.00	0	464	< 0.01	100

*n = 3 for WT and NF H20 controls; n = 2 for all other samples

Supplementary Table 2.6. Droplet throughput comparison of Velox prototype IC3D system to commercial ddPCR systems

		Velox IC3D (3-D) *	RainDance (1-D)	BioRad (1-D)
	Analytical Sensitivity	0.00125 – 0.005% (present study)	0.0022% [a]	0.08 – 0.33% (present study)
Sample	WT concentration limit per reaction	> 400 ng/ 20 µl (20 µg/ml)	1500 ng/ 25 µl (60 µg/ml) [b]	66 ng/ 20 µl [f] (3.3 µg/ml)
	Final dilution volume (clinical sample = 20 µg total)	1 ml	0.33 ml	6.06 ml
	Number of samples	1 x 1 ml	14 x 25 µl	304 x 20 µl
	Total number of droplets (partitions)	15.3 million	70 million [c]	6.1 million [g]
Throughput	Droplet Generation	66 min (1 ml/ 66 min)	60 min 2 x (8-wells/30 min) [d]	180 min 4x (96-wells/45 min) [h]
	Droplet scanning	2 min (2 min/ sample)	4 hours 2 x (2 hrs/8 wells in "fast mode") [e]	380 min 304 x (120 min/96 wells) [i]
	Total Time**	1hr, 8 min	5 hours	9hr, 20min

a) Reported empirical LOD of 0.0022% AF for KRAS G12D (3.5 E7 WT copies) [C. A. Milbury, Q. Zhong, J. Lin, M. Williams, J. Olson, D. R. Link and B. Hutchison, Biomol Detect Quantif, 2014, 1, 8–22.]
b) Recommended DNA loading up to 10% target occupancy (RainDrop® assay guidelines), which for a 25 µl reaction (5E6 droplets) is approximately 500,000 DNA molecules, which is the equivalent of approximately 1.5 µg gDNA
c) Approximately 5E6 droplets per 25 µl reaction (RainDrop® assay guidelines)
d) Approximately 30 minutes to generate droplets from 1 chip (8 wells) (ThunderBolts™ System/RainDance Source Operator's Manual)
e) Droplet scanning time for 8 wells = 2 hours (fast mode); 4 hours (standard mode) (RainDrop® Sense Operator's manual)
f) "Recommendations for Optimal Results" (ddPCR™ Supermix for Probes): "The concentration of intact human genomic DNA should be ≤ 66 ng per 20 µl reaction"

g) Approximately 20,000 droplets per sample (“QX100/QX200 Workflow / Droplet Generation”): “Droplet Digital™ PCR Applications Guide.” Bulletin 6407 Ver B. Bio-Rad Laboratories, Inc.)
h) Approximately 45 minutes to generate droplets for 96 wells (“1.3 Installation and General Operation”, pg. 4): “Automated Droplet Generator.” Instruction Manual. Catalog #1864101. BioRad © 2018)
i) Approximately 120 minutes to read/analyze droplets from one 96-well plate [S. H. Te, E. Y. Chen and K. Y.-H. Gin, Appl. Environ. Microbiol., 2015, 81, 5203.], [E. Mazaika and J. Homsy, Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.], 2014, 82, 7.24.1-7.24.13.]
* IC3D performance and throughput estimates are solely based on this proof-of-concept study and do not represent limitations of the technology. ** Droplet thermocycling and droplet transfer steps not included

2.7 References

1. Cancer Statistics, https://www.cancer.gov/about-cancer/understanding/statistics, (accessed December 21, 2018).
2. R. L. Siegel , K. D. Miller and A. Jemal , *Ca-Cancer J. Clin.*, 2018, **68** , 7 —30.
3. NCCN Clinical Practice Guidelines in Oncology, Colon Cancer, Vers 4.2018, (accessed December 19, 2018).
4. E. Crowley , F. Di Nicolantonio , F. Loupakis and A. Bardelli , *Nat. Rev. Clin. Oncol.*, 2013, **10** , 472 —484.
5. M. Ignatiadis and S.-J. Dawson , *Ann. Oncol.*, 2014, **25** , 2304 —2313.
6. E. Heitzer , I. S. Haque , C. E. S. Roberts and M. R. Speicher , *Nat. Rev. Genet.*, 2019, **20** , 71 —88.
7. M. G. Krebs , R. L. Metcalf , L. Carter , G. Brady , F. H. Blackhall and C. Dive , *Nat. Rev. Clin. Oncol.*, 2014, **11** , 129 —144.
8. C. Alix-Panabieres and K. Pantel , *Cancer Discovery*, 2016, **6** , 479 —491.
9. F. Diehl , K. Schmidt , M. A. Choti , K. Romans , S. Goodman , M. Li , K. Thornton , N. Agrawal , L. Sokoll , S. A. Szabo , K. W. Kinzler , B. Vogelstein and L. A. J. Diaz , *Nat. Med.*, 2008, **14** , 985 —990.
10. J. Tie , Y. Wang , C. Tomasetti , L. Li , S. Springer , I. Kinde , N. Silliman , M. Tacey , H.-L. Wong , M. Christie , S. Kosmider , I. Skinner , R. Wong , M. Steel , B. Tran , J. Desai , I. Jones , A. Haydon , T. Hayes , T. J. Price , R. L. Strausberg , L. A. Diaz Jr , N. Papadopoulos , K. W. Kinzler , B. Vogelstein and P. Gibbs , *Sci. Transl. Med.*, 2016, **8** , 346ra92.
11. S. Webb *Nat. Biotechnol.*, 2016, **34** , 1090.
12. H. Schwarzenbach , D. S. B. Hoon and K. Pantel , *Nat. Rev. Cancer*, 2011, **11** , 426 —437.
13. G. Siravegna , S. Marsoni , S. Siena and A. Bardelli , *Nat. Rev. Clin. Oncol.*, 2017, **14** , 531.
14. C. Bettegowda , M. Sausen , R. J. Leary , I. Kinde , Y. Wang , N. Agrawal , B. R. Bartlett , H. Wang , B. Luber , R. M. Alani , E. S. Antonarakis , N. S. Azad , A. Bardelli , H. Brem , J. L. Cameron , C. C. Lee , L. A. Fecher , G. L. Gallia , P. Gibbs , D. Le , R. L. Giuntoli , M. Goggins , M. D. Hogarty , M. Holdhoff , S.-M. Hong , Y. Jiao , H. H. Juhl , J. J. Kim , G. Siravegna , D. A. Laheru , C. Lauricella , M. Lim , E. J. Lipson , S. K. N. Marie , G. J. Netto , K. S. Oliner , A. Olivi , L. Olsson , G. J. Riggins , A. Sartore-Bianchi , K. Schmidt , L.-M. Shih , S. M. Oba-Shinjo , S. Siena , D. Theodorescu , J. Tie , T. T. Harkins , S. Veronese , T.-L. Wang , J. D. Weingart , C. L. Wolfgang , L. D. Wood , D. Xing , R. H. Hruban , J. Wu , P. J. Allen , C. M. Schmidt , M. A. Choti , V. E. Velculescu , K. W. Kinzler , B. Vogelstein , N. Papadopoulos and L. A. J. Diaz , *Sci. Transl. Med.*, 2014, **6** , 224ra24.
15. R. Deng , K. Zhang and J. Li , *Acc. Chem. Res.*, 2017, **50** , 1059 —1068.
16. R. Deng , K. Zhang , L. Wang , X. Ren , Y. Sun and J. Li , *Chem*, 2018, **4** , 1373 —1386.
17. X. Tang , R. Deng , Y. Sun , X. Ren , M. Zhou and J. Li , *Anal. Chem.*, 2018, **90** , 10001 —10008.
18. D. George , J. Czech , B. John , M. Yu and L. J. Jennings , *Chimerism*, 2013, **4** , 102 —108.
19. B. Zhou , M. S. Haney , X. Zhu , R. Pattni , A. Abyzov and A. E. Urban , *Methods Mol. Biol.*, 2018, **1768** , 173 —190.

20. Y. Cao , M. R. Raith and J. F. Griffith , *Water Res.*, 2015, **70** , 337 —349.
21. B. J. Hindson , K. D. Ness , D. A. Masquelier , P. Belgrader , N. J. Heredia , A. J. Makarewicz , I. J. Bright , M. Y. Lucero , A. L. Hiddessen , T. C. Legler , T. K. Kitano , M. R. Hodel , J. F. Petersen , P. W. Wyatt , E. R. Steenblock , P. H. Shah , L. J. Bousse , C. B. Troup , J. C. Mellen , D. K. Wittmann , N. G. Erndt , T. H. Cauley , R. T. Koehler , A. P. So , S. Dube , K. A. Rose , L. Montesclaros , S. Wang , D. P. Stumbo , S. P. Hodges , S. Romine , F. P. Milanovich , H. E. White , J. F. Regan , G. A. Karlin-Neumann , C. M. Hindson , S. Saxonov and B. W. Colston , *Anal. Chem.*, 2011, **83** , 8604 —8610.
22. Y. Ono , A. Sugitani , H. Karasaki , M. Ogata , R. Nozaki , J. Sasajima , T. Yokochi , S. Asahara , K. Koizumi , K. Ando , K. Hironaka , T. Daito and Y. Mizukami , *Mol. Oncol.*, 2017, **11** , 1448 —1458.
23. A. S. Basu *SLAS Technol.*, 2017, **22** , 369 —386.
24. K. Zhang , D.-K. Kang , M. M. Ali , L. Liu , L. Labanieh , M. Lu , H. Riazifar , T. N. Nguyen , J. A. Zell , M. A. Digman , E. Gratton , J. Li and W. Zhao , *Lab Chip*, 2015, **15** , 4217 —4226.
25. D.-K. Kang , M. M. Ali , K. Zhang , S. S. Huang , E. Peterson , M. A. Digman , E. Gratton and W. Zhao , *Nat. Commun.*, 2014, **5** , 5427.
26. D. Pekin , Y. Skhiri , J.-C. Baret , D. Le Corre , L. Mazutis , C. B. Salem , F. Millot , A. El Harrak , J. B. Hutchison , J. W. Larson , D. R. Link , P. Laurent-Puig , A. D. Griffiths and V. Taly , *Lab Chip*, 2011, **11** , 2156 —2166.
27. A. How Kit , N. Mazaleyrat , A. Daunay , H. M. Nielsen , B. Terris and J. Tost , *Hum. Mutat.*, 2013, **34** , 1568 —1580.
28. J. Soh , N. Okumura , W. W. Lockwood , H. Yamamoto , H. Shigematsu , W. Zhang , R. Chari , D. S. Shames , X. Tang , C. MacAulay , M. Varella-Garcia , T. Vooder , I. I. Wistuba , S. Lam , R. Brekken , S. Toyooka , J. D. Minna , W. L. Lam and A. F. Gazdar , *PLoS One*, 2009, **4** , e7464.
29. D. Sefrioui , F. Mauger , L. Leclere , L. Beaussire , F. Di Fiore , J.-F. Deleuze , N. Sarafan-Vasseur and J. Tost , *Clin. Chim. Acta*, 2017, **465** , 1 —4.
30. E. Amstad , M. Chemama , M. Eggersdorfer , L. R. Arriaga , M. P. Brenner and D. A. Weitz , *Lab Chip*, 2016, **16** , 4163 —4172.
31. H.-H. Jeong , V. R. Yelleswarapu , S. Yadavali , D. Issadore and D. Lee , *Lab Chip*, 2015, **15** , 4387 —4392.
32. L. A. J. Diaz and A. Bardelli , *J. Clin. Oncol.*, 2014, **32** , 579 —586.
33. F. Riva , O. I. Dronov , D. I. Khomenko , F. Huguet , C. Louvet , P. Mariani , M.-H. Stern , O. Lantz , C. Proudhon , J.-Y. Pierga and F.-C. Bidard , *Mol. Oncol.*, 2016, **10** , 481 —493.
34. K. M. Haigis *Trends Cancer*, 2017, **3** , 686 —697.
35. C. A. Milbury , Q. Zhong , J. Lin , M. Williams , J. Olson , D. R. Link and B. Hutchison , *Biomol. Detect. Quantif.*, 2014, **1** , 8 —22.
36. A. C. Hatch , J. S. Fisher , A. R. Tovar , A. T. Hsieh , R. Lin , S. L. Pentoney , D. L. Yang and A. P. Lee , *Lab Chip*, 2011, **11** , 3838 —3845.
37. V. R. Yelleswarapu , H.-H. Jeong , S. Yadavali and D. Issadore , *Lab Chip*, 2017, **17** , 1083 —1094.
38. M. Kim , M. Pan , Y. Gai , S. Pang , C. Han , C. Yang and S. K. Y. Tang , *Lab Chip*, 2015, **15** , 1417 —1423.

39. L. B. Pinheiro , V. A. Coleman , C. M. Hindson , J. Herrmann , B. J. Hindson , S. Bhat and K. R. Emslie , *Anal. Chem.*, 2012, **84** , 1003 —1011.
40. C. M. Hindson , J. R. Chevillet , H. A. Briggs , E. N. Gallichotte , I. K. Ruf , B. J. Hindson , R. L. Vessella and M. Tewari , *Nat. Methods*, 2013, **10** , 1003 —1005.
41. Q. Kang , B. Parkin , M. D. Giraldez and M. Tewari , *BioTechniques*, 2016, **60** , 175 —176. , 178, 180 passim.
42. A. S. Whale , A. S. Devonshire , G. Karlin-Neumann , J. Regan , L. Javier , S. Cowen , A. Fernandez-Gonzalez , G. M. Jones , N. Redshaw , J. Beck , A. W. Berger , V. Combaret , N. Dahl Kjersgaard , L. Davis , F. Fina , T. Forshew , R. Fredslund Andersen , S. Galbiati , Á. González Hernández , C. A. Haynes , F. Janku , R. Lacave , J. Lee , V. Mistry , A. Pender , A. Pradines , C. Proudhon , L. H. Saal , E. Stieglitz , B. Ulrich , C. A. Foy , H. Parkes , S. Tzonev and J. F. Huggett , *Anal. Chem.*, 2017, **89** , 1724 —1733.
43. L. Dong , S. Wang , B. Fu and J. Wang , *Sci. Rep.*, 2018, **8** , 9650.
44. C. Demuth , K.-L. G. Spindler , J. S. Johansen , N. Pallisgaard , D. Nielsen , E. Hogdall , B. Vittrup and B. S. Sorensen , *Transl. Oncol.*, 2018, **11** , 1220 —1224.
45. J. A. Denis , A. Patroni , E. Guillerm , D. Pepin , N. Benali-Furet , J. Wechsler , G. Manceau , M. Bernard , F. Coulet , A. K. Larsen , M. Karoui and J.-M. Lacorte , *Mol. Oncol.*, 2016, **10** , 1221 —1231.
46. L. M. Millner , M. W. Linder and R. Valdes Jr , *Ann. Clin. Lab. Sci.*, 2013, **43** , 295 —304.
47. M. T. Gabriel , L. R. Calleja , A. Chalopin , B. Ory and D. Heymann , *Clin. Chem.*, 2016, **62** , 571 —581.

CHAPTER 3:
VALIDATION OF PROBE LABELING STRATEGY

Authors: Qi Xu[1,3], **Tam Vu**[1,3] Josh Gu[1], Jan Zimak[1], Enrico Gratton*[3,4], and Weian Zhao*[1,2,3]

Formal analysis: T. V., Q. X., J. G., J. Z., E. G., W.Z.; methodology: T. V., Q. X., J. G., J. Z., E. G., W.Z.; writing –Q. X., T. V., J. G.; investigation: T. V., Q. X., J. G., J. Z., E. G., W.Z.; data curation: T. V., Q. X., J. G.; conceptualization: T. V., Q. X., J. G., J. Z., E. G., W.Z.; resources: W. Z., E. G.

Affiliations:

*Corresponding authors

[1]Sue and Bill Gross Stem Cell Research Center, Sue & Bill Gross Hall CIRM Institute, University of California, Irvine, 845 Health Sciences Road, Suite 3027, Irvine, CA 92697, USA

[2]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA

[3]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, USA

[4]Laboratory for Fluorescence Dynamics, Department of Biomedical Engineering, University of California, Irvine, USA

3.1 Background Information

3.1.1. Probe Labeling Strategy

In situ hybridization was first invented by Gall and Purdue in 1969 using radioisotopes and was then modified for easier use with any standard commercial fluorophores by Femino et al. in 1998 [1-2]. Termed single molecule FISH (smFISH), smFISH decorates each mRNA transcript with multiple fluorescent oligonucleotide probes and allows users to visualize and detect labeled transcripts under a fluorescent microscope. For this approach, using many probes (10 – 50 per transcript) was required to attain a sufficient signal-to-noise ratios to differentiate specific decorated signals from the nonspecific signals which bound everywhere (Figure 3.1A) [3]. Currently, there are many variants of smFISH (Figure 3.1).

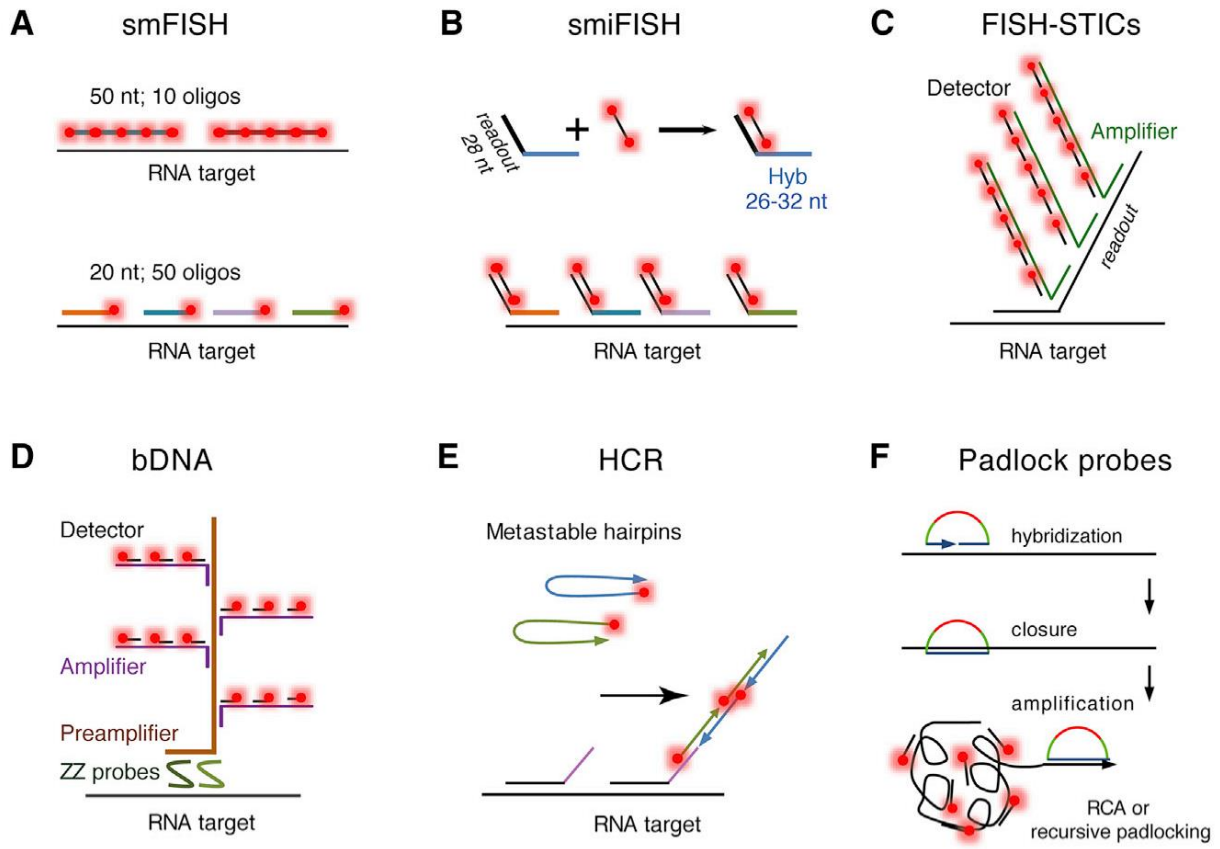


Figure 3.1. Variants of single molecule FISH methods. (A) Conventional single molecule FISH (smFISH). (B) Single molecule indirect FISH. Nonfluorescent primary probes are first conjugated to mRNA and then subsequently bound to fluorescent secondary probes (C) FISH with Sequential Tethered and Intertwined 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 [3].

Of particular interest are the indirect labeling schemes where nonfluorescent primary probes are first conjugated to mRNA which are then subsequently bound to fluorescent secondary probes (Figure 3.1B). Since unconjugated oligonucleotide probes can be easily programmed to target any complementary regions on the RNA transcript, building an inexpensive library of probes towards many different genes become highly scalable and cost-

effective. The more expensive secondary conjugated fluorophores 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 3.1B–1E). As a preliminary study, this thesis utilizes the labeling scheme depicted in Figure 3.1B to demonstrate how combinatorial labeling with a common panel of fluorophores can be integrated with spectra-FLIM microscopy to achieve greater multiplexed detection and error-correction. However, this approach is not restricted to only this labeling technique and should be compatible with any of the aforementioned labeling techniques, opening many exciting avenues for future work on this project. To examine the effects of different FISH hybridization and probe design parameters on RNA-FISH signal, a model cell line, mNeon Green HEK293T, was employed.

3.1.2 mNeonGreen HEK293T cells

A cell line with stable expression is required to validate the labeling and detection of transcripts. For this cell line, the expectation is: 1) it needs to have consistent expression of transcript; 2) the expression level needs to be easy to confirm under a fluorescent microscope, e.g. a fluorescent protein. For these reasons, a stable mNeonGreen expression in HEK293T cells was engineered. MNeonGreen is a monomeric yellow-green fluorescent protein, which was first published in 2013. It is derived from the cephalochordate *Branchiostoma lanceolatum*. Its excellent performance has been well validated under various microscope systems. The excitation wavelength of mNeonGreen is 506 nm, while its emission wavelength is 517 nm [4]. Any cell with mNeonGreen protein inside will show green fluorescence. In addition to visualizing the cells, in our case, the expression of mNeonGreen protein provides a great option for FISH probe design.

3.1.4 Construction of mNeonGreen HEK293T cells

To ensure the consistent expression of mNeonGreen protein, Human embryonic kidney (HEK293T) cells were transfected with an engineered mNeonGreen plasmid (Fig. 3.1A). First a plasmid with mNeonGreen sequence (GenBank: KC295282.1) was constructed. The plasmid was then transfected into HEK293T cells with FuGENE HD Transfection Reagent. One day later, the cells were checked under a fluorescence microscope and exhibited heterogeneity in mNeonGreen expression. Not every cell has mNeonGreen protein inside. Therefore, three days after transfection, the cells were selected with puromycin and Zeocin. One day after selection, the cells were checked again. This time all the cells show the green fluorescence. They are identified as positive cells, while the normal HEK293T cells are negative ones. In the following experiments, both positive and negative cells were mixed to form a cell mixture called mixed mNeonGreen HEK293T cells. And the ratio of two types of cells is around 1:1. Therefore, both positive results and negative controls are available in the same region of interest (ROI) (Fig. 3.2B,C).

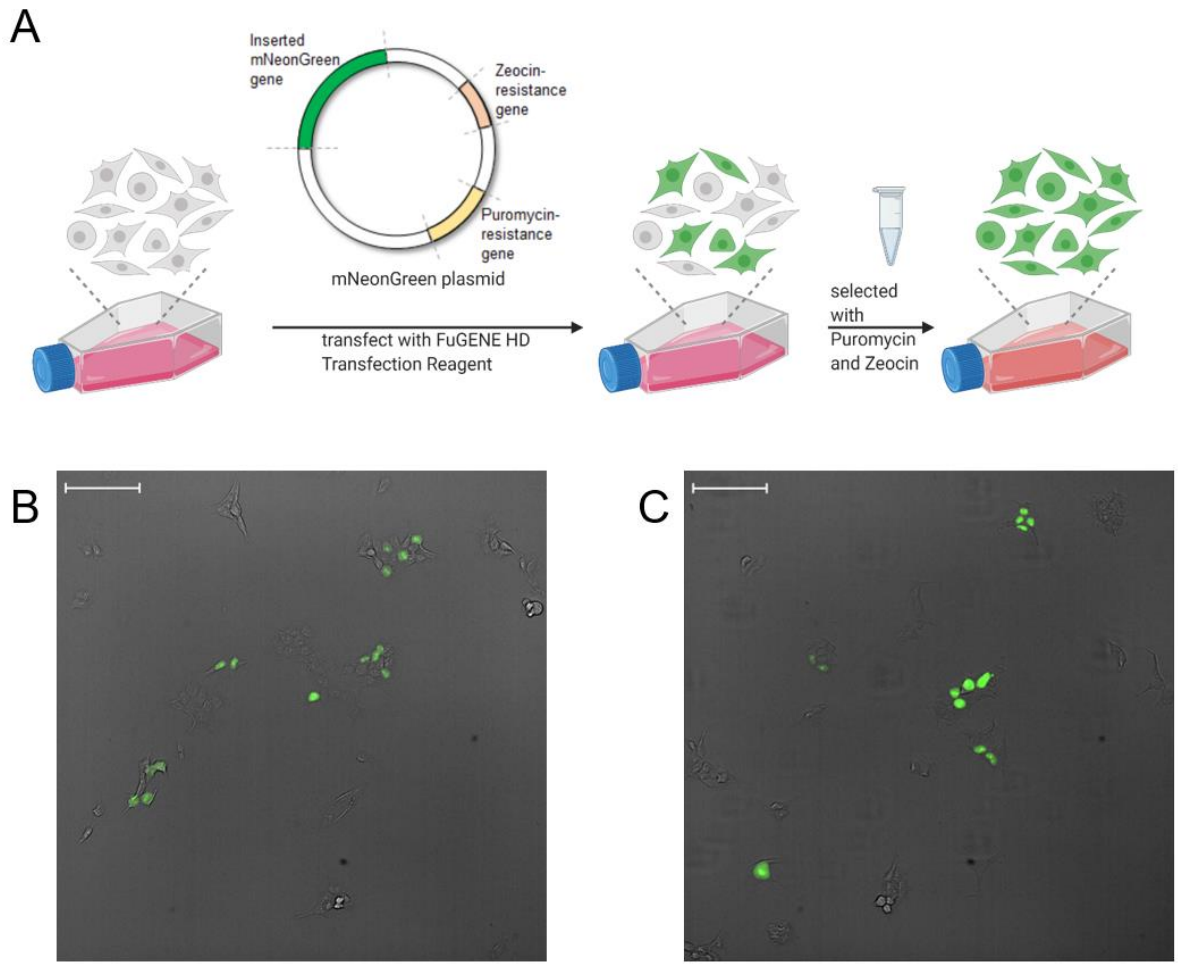


Figure 3.2 mNeonGreen HEK293T cells. A) Construction of mNeonGreen HEK293T cells. B&C) Mixed mNeonGreen cells under the microscope. The images in bright field and FITC channel were merged. Scale bar = 10 µm.

3.2.1 Probe design

The indirect binding method mentioned above was also applied into the following RNA-FISH design. The primary probes consist of a target sequence (complimentary to a part of target mRNA), a binding site for secondary probes and a TTT linker in between to allow the flexibility of the probe. The number of each component in a primary probe can be adjusted as needed. Alexa 647 was chosen as the fluorophore for secondary probes. In a set of primary probes for a certain

mRNA, each target sequence is unique, targeting the different regions of the mRNA, while the binding site for secondary probes is the same.

All target sequences were designed via the Stellaris® Probe Designer version 4.2, targeting the coding sequence of mNeonGreen [5]. Previous research [6,7] have shown that a narrow range of the melting temperature (Tm) and GC content of probes allows high hybridization efficiency. Specifically, Tm and GC% determines when the denaturation (unbinding) occurs between the target and probes. Therefore, if these two metrics are in a narrow range, the possibility of denaturation would be lower. For this reason, any sequence which deviates from the main range of the Tm and GC% was removed. Additionally, any sequences including four or more consecutive bases of the same kind (e.g. GGGG) were dropped to eliminate the presence of quadruplet structure. To further ensure the specificity of probes, an online BLAST query was also run on each sequence against the human transcriptome. After these selections, 14 binding sequences with mNeonGreen mRNA were designed. Each sequence has 27 nucleotides (nt).

For the design of a binding site for secondary probes, the design of readout sequences was borrowed from the Zhuang Lab [8]. These 20-nt sequences were designed randomly, with the possibility of 50% G, 25% A and 25% T per base. The three-letter readout sequences are artificial, so it is less likely to bind randomly on the sample. This helped to decrease their nonspecific binding. Sequences including four or more consecutive G bases (e.g. GGGG) were removed. And these readout sequences have been BLAST against human transcriptome.

Finally, each probe includes a 27-nt target sequence, a TTT linker and a 20-nt readout sequence. It is:

target sequence (27-nt) —TTT— readout sequence (20-nt)

The specificity of these 14 probes was checked against human genome again via BLAST. The range of GC content was 40-50%.

The secondary probe includes a 20-nt sequence, which is the reverse complement of the readout sequences in primary probes, and one Alexa 647 fluorescence molecule:

/5Alex647N/ readout sequence* (20-nt)

* means the reverse complement.

Later, based on the results from parameter optimization, these probe designs were improved. Currently each primary probe includes a 27-nt target sequence, two TTT linkers and two 20-nt readout sequences. And each secondary probe is attached with two Alexa 647 fluorescence molecules (schematic shown in Fig. 3.3).

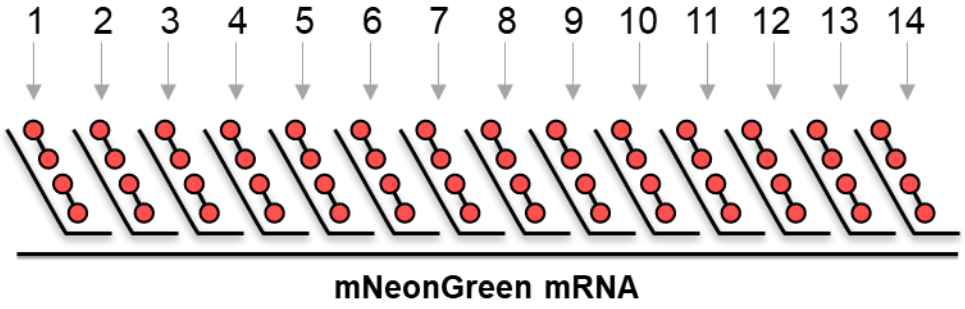


Figure 3.3 The schematic of probe binding. Red dots represent Alexa 647 fluorescence molecules.

3.2.2 Validation

Although the expression of mNeonGreen protein can be confirmed under the microscope, the hybridization efficiency of probes in mixed mNeonGreen HEK293T cells requires optimization. Therefore, a validation experiment was conducted to visualize the location of mNeonGreen

transcript to assess the hybridization efficiency. The images were based on the optimized experiment parameters (Fig. 3.4A). Figure 3.4B shows the detection of mNeonGreen mRNA. The signal from Alexa 647 was specifically located around the HEK293T cells expressing mNeonGreen protein, indicating the accuracy of probe design and the specificity of target sequences (individual images in each channel shown in Fig. 3.4C & D & E).

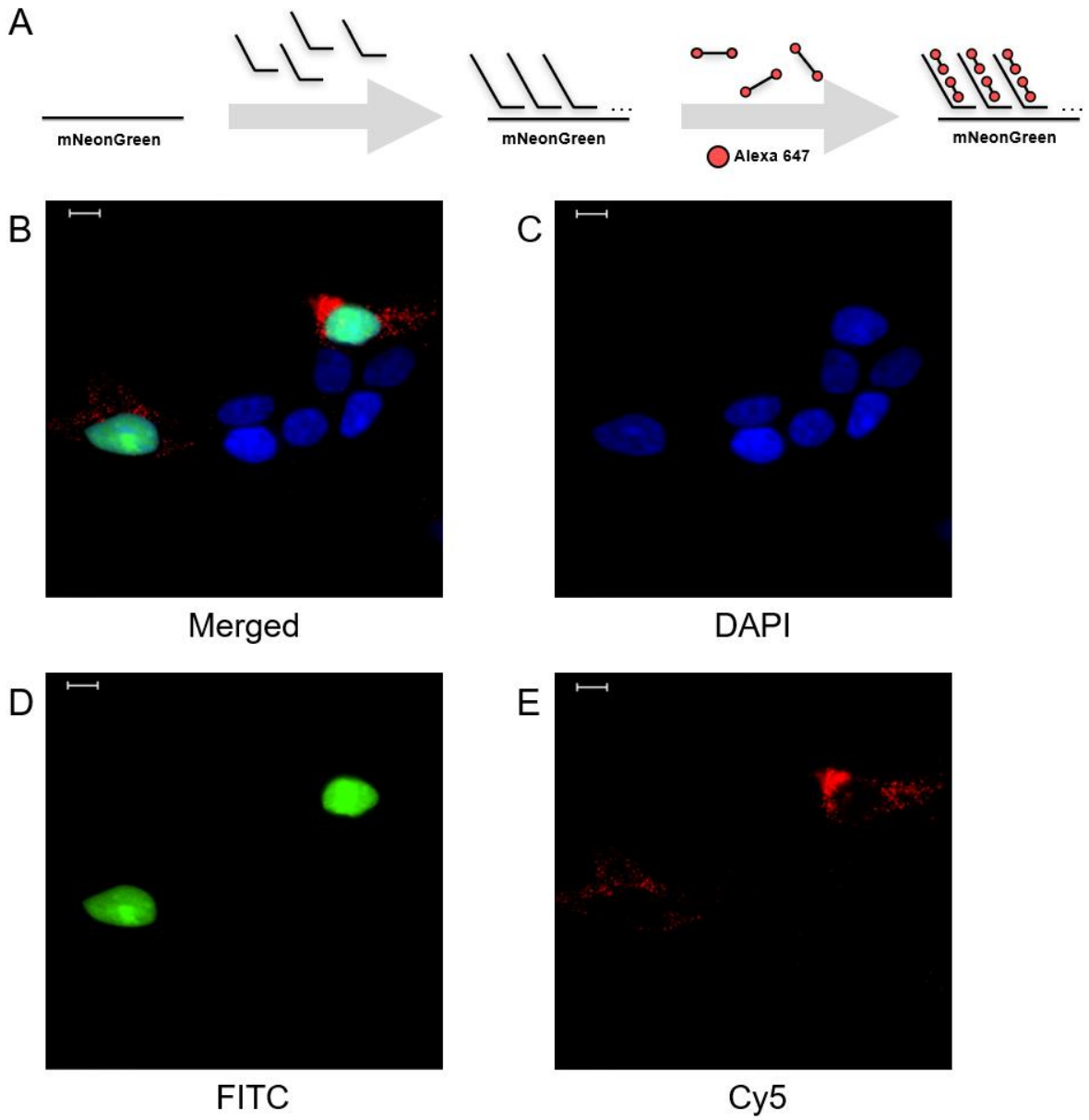


Figure 3.4 Validation of hybridization efficiency in mixed mNeonGreen HEK293T cells. A) The schematic showing the workflow of RNA-FISH. B) Merged image showing the detection of mNeonGreen transcript. Four channels were shown, bright field, DAPI (blue), FITC (green), Cy5 (red). Scale bar = 10 μm.

In this experiment, RNA-FISH was performed on three types of control. In the scramble control (Fig. 3.5A), the target sequences of the primary probes were designed for a mouse gene, while their readout sequence was still complementary to the secondary probe labeled with Alexa

647. This control indicates any nonspecific primary probe binding. The negative control (Fig. 3.5B) is the condition with no primary probes adding, indicating any nonspecific binding of the secondary probe alone. The blank control (Fig. 3.5C) is the condition without any probes added. Very few red signals came from scramble control (Fig. 3.5D), while there was no red signal in negative control (Fig. 3.5E) and blank control (Fig. 3.5F).

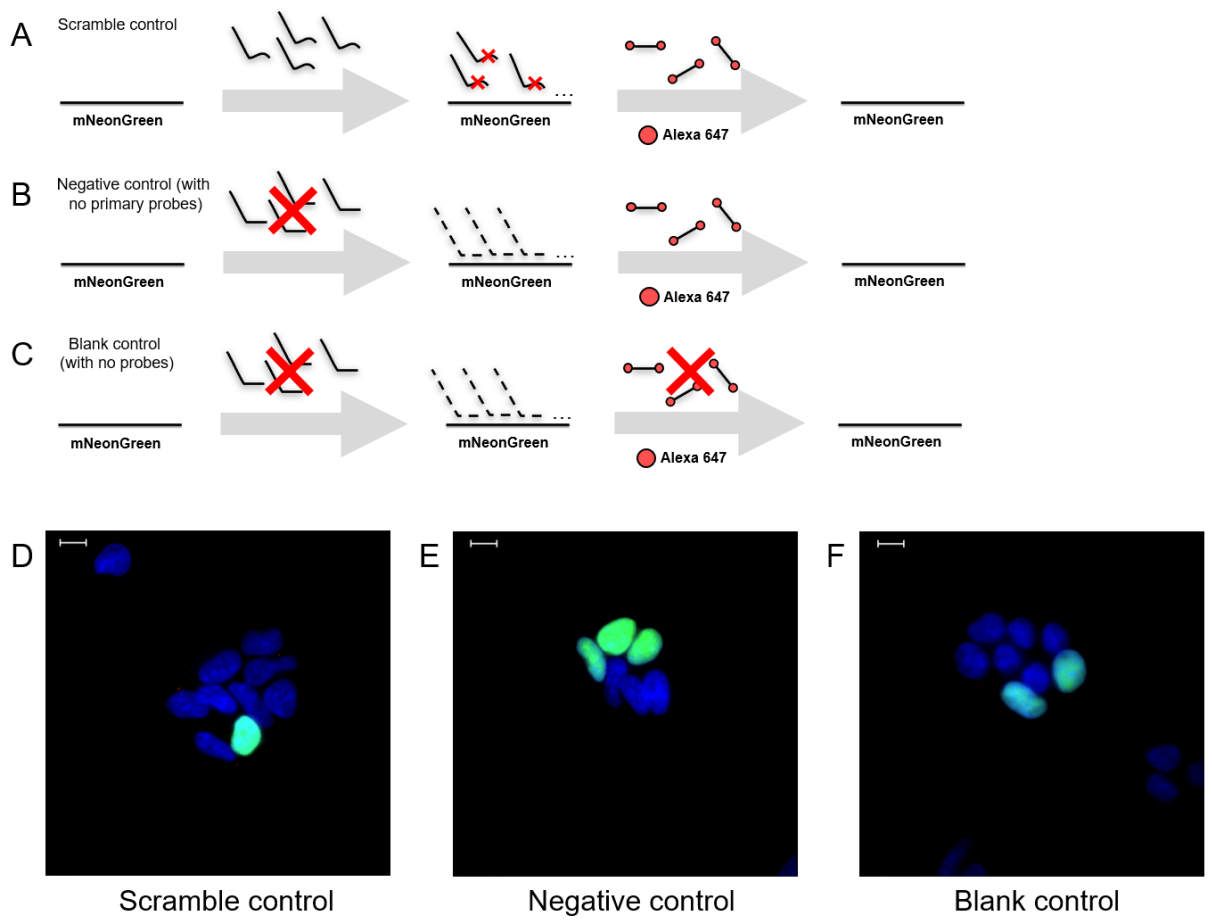


Figure 3.5 Scramble control, negative control, and blank control for validation. A) The schematic of scramble control. B) The schematic of negative control. C) The schematic of blank control. D) Merged image of scramble control. Three channels were shown, DAPI (blue), FITC (green), Cy5 (red). Scale bar = 10 µm. E) Merged image of negative control. Three channels were shown, DAPI (blue), FITC (green), Cy5 (red). Scale bar = 10 µm. F) Merged image of blank control. Three channels were shown, DAPI (blue), FITC (green), Cy5 (red). Scale bar = 10 µm.

Based on the results above, the hybridization efficiency and specificity of mNeonGreen probes was validated in mixed mNeonGreen HEK293T cells. The strong specific signal from the probe set for the mNeonGreen transcript indicates the high labeling efficiency of this RNAFISH method for the detection of mNeonGreen mRNA transcripts.

3.2.4 Detection of multiplex target

To test the multiplexing capability of this RNA-FISH platform, a probe set for the UBC transcript was designed (Fig. 3.6A). UBC (ubiquitin C, NM_021009.7) is a housekeeping gene in the human genome and is one of the recommended genes by RNAscope® to be used as positive control for medium or high expression. The 14 target sequences for UBC mRNA were designed in the same way as the target sequences which were used for mNeonGreen mRNA. Sequences with four or more consecutive G bases or deviating from the main range of the Tm and GC% were removed. Its assigned fluorophore is Atto 565, using a different readout sequence and its reverse complement.

Figure 3.6B showed the detection of multiplex transcripts in mixed mNeonGreen HEK293T cells. As shown, UBC was expressed in all the HEK293T cells (Fig. 3.6E), while only the cells showing green fluorescence expressed the mNeonGreen transcript (Fig. 3.6F). According to the individual images in each channel (Fig. 3.6C & D & E & F), there was no interference between the two detection channels. Therefore, with more detection channels equipped, it is possible to further improve the multiplexing capability.

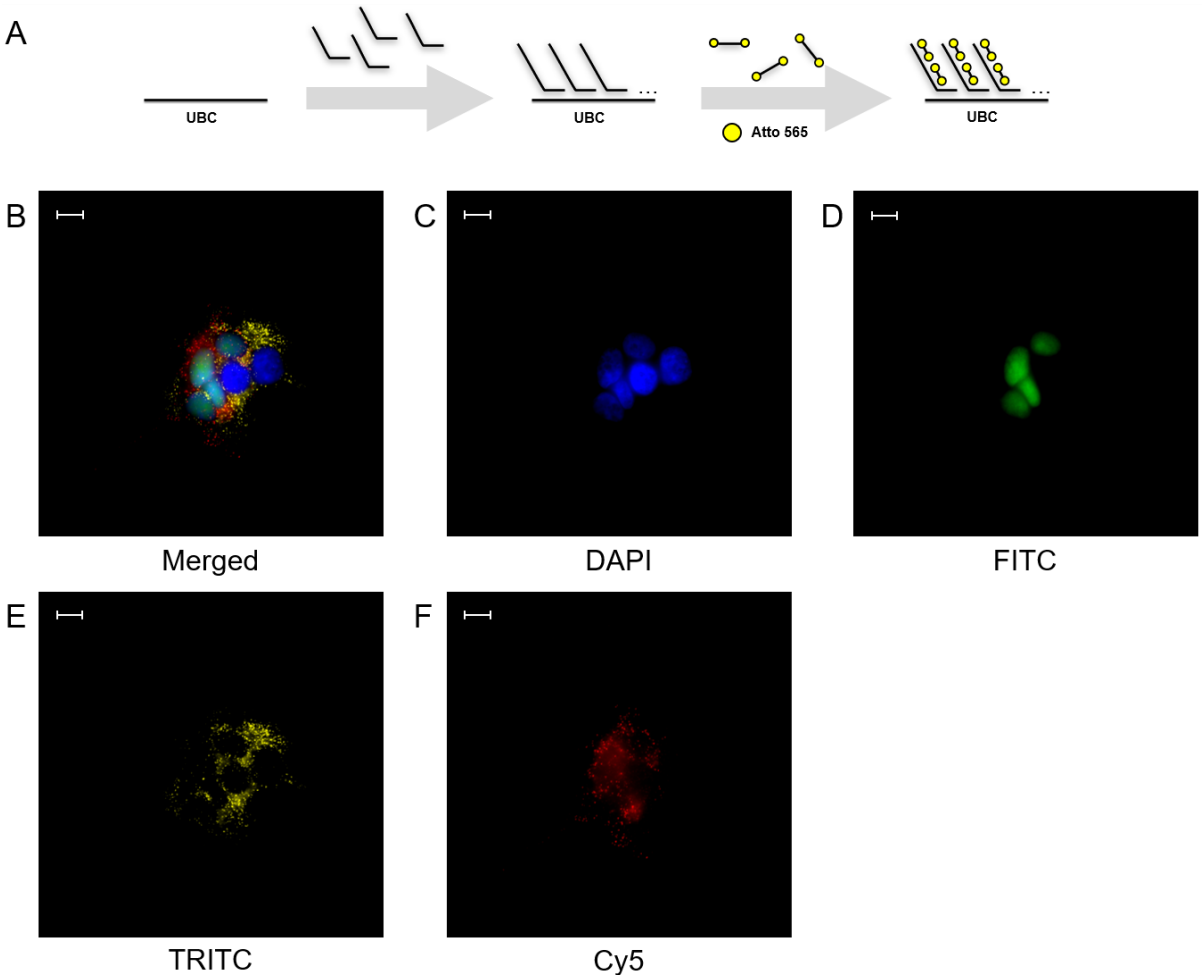


Figure 3.6 Detection of multiplex transcripts in mixed mNeonGreen HEK293T cells. A) The schematic showing the workflow of RNA-FISH for UBC transcript. B) Merged image showing the detection of both UBC and mNeonGreen transcripts. Four channels were shown, DAPI (blue), FITC (green), TRITC (yellow) and Cy5 (red). Scale bar = 10 µm. C) All the HEK293T cells in DAPI channels. Scale bar = 10 µm. D) mNeonGreen cells in FITC channel. Scale bar = 10 µm. E) Detection of UBC transcripts in TRITC channel. Scale bar = 10 µm. F) Detection of UBC transcripts in mNeonGreen channel. Scale bar = 10 µm.

3.2.5 Parameter optimization

Using the mNeonGreen HEK293T cell model and the probe set for mNeonGreen transcript, the effects of different probe design and hybridization parameters were tested. The results were analyzed to indicate the optimal conditions. Here, noise refers to the fluorescence

dots from the cells not expressing the mNeonGreen protein or the baseline levels of nonspecific binding.

3.2.6 Primary probe structure

Previous research from Yin Lab [9] has shown, as the number of readout sequences on primary probes increased, the signal intensity also improved. However, it remained unknown whether the increase would reach a maximum limitation when primary probes were continuously extended. To test this hypothesis, three sets of primary probes were designed, which only differed in the number of readout sequences (schematics shown in Fig. 3.7A). On each primary probe in set 1, there was only one readout sequence. In set 2, the number was doubled. This number was further increased to four in set 3.

The results showed that set 2, primary probes with two readout sequences, provided the strongest signal, 949 relative fluorescent intensity (RFI) (±213), despite the high variation (Fig. 3.7B). The SNR in set 2 was highest as well, reaching 3.52 (±0.73) (Fig. 3.7C). The intensity of average noise in each condition was similar, suggesting that noise mainly came from the nonspecific binding of secondary probes. Based on the result, primary probes with two readout sequences were chosen for use in the following tests.

The high variation in intensity in set 2 and 3 resulted from the nonspecific binding of several secondary probes. A reason why the ones with the four readouts sequences failed to further increase the signal might be the structure. The probes were too long (119 nucleotides) to efficiently diffuse and bind with the target transcript in a limited time. Also, since each probe has a binding region of only 27 nucleotides out of 119 nucleotides bound, it was less difficult to wash off with repetitive washing after the primary probe hybridization.

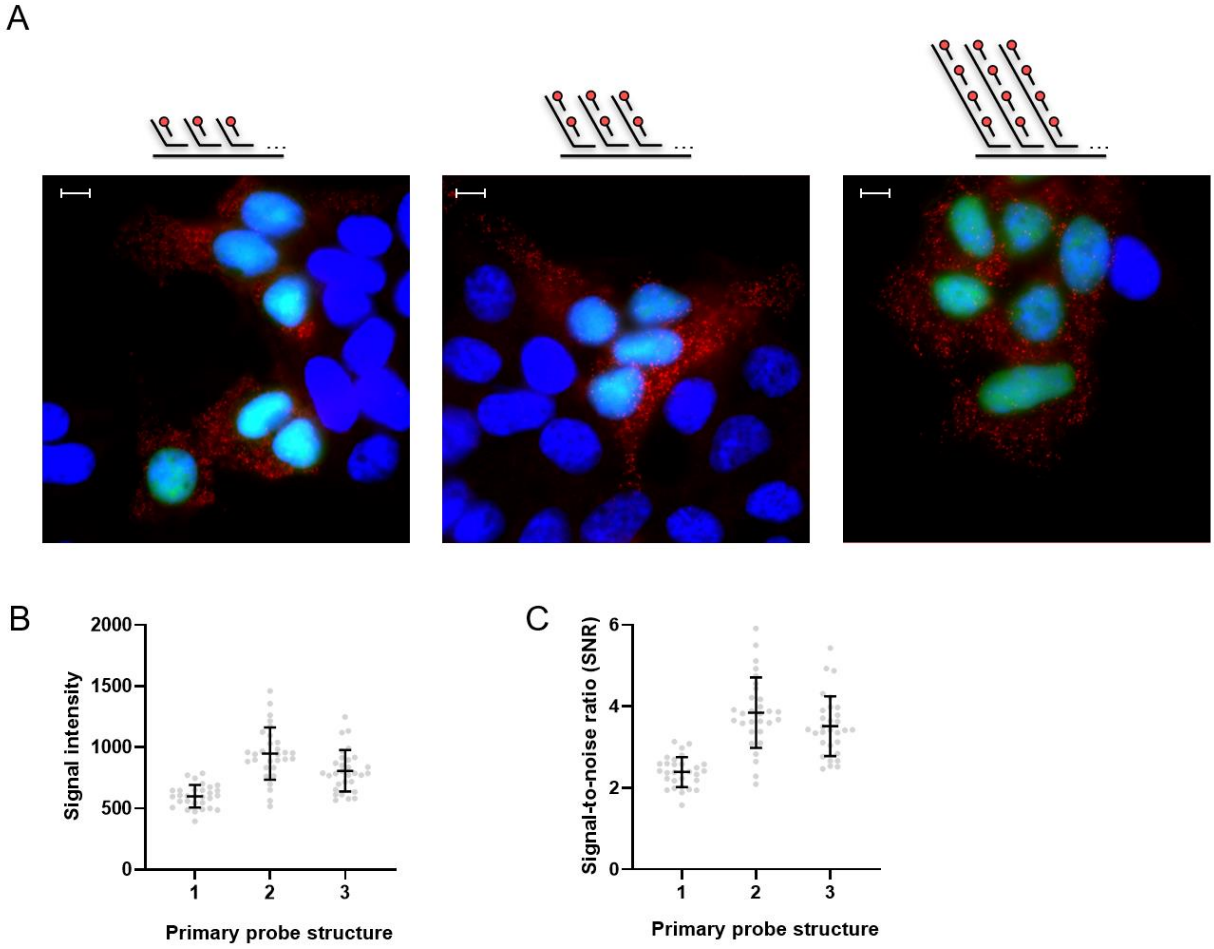


Figure 3.7 Influence of primary probe structure on the detection of mNeonGreen transcripts. A) (from left to right) Merged images showed detection of mNeonGreen transcript using primary probes with one, two, and four readout sequences, respectively. The schematics of binding were shown above the figures. Three channels were shown, DAPI (blue), FITC (green), and Cy5 (red). Scale bar = 10 μm. B) Scatter plot showing the signal intensity on three conditions. n=30 puncta. C) Scatter plot showing the signal-to-noise ratio (SNR) on three conditions. n=30 puncta.

3.2.7 Fluorophore probe structure

Based on the results on the effects of primary probe structure on transcript detection, secondary probe structures were assessed as well. Secondary probes with different structures were tested. Secondary probe 1 was the reverse complement of the readout sequence, attached with one Alexa 647 molecule, while secondary probe 2 has the same DNA sequence but with one Alexa 647 molecule at each end (schematics shown in Fig. 3.8A).

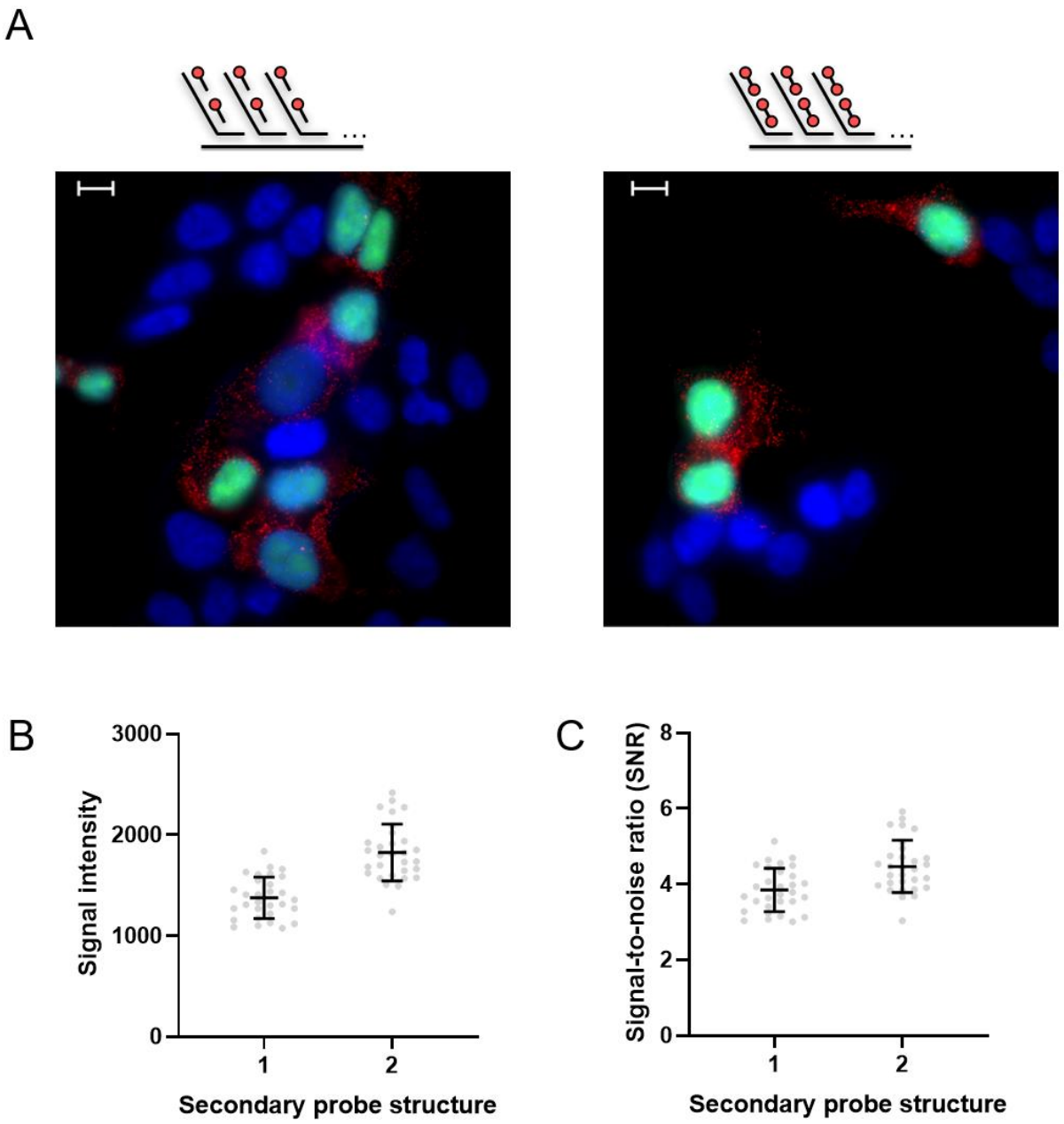


Figure 3.8 Influence of secondary probe structure on the detection of mNeonGreen transcripts. A) (from left to right) Merged images showed detection of mNeonGreen transcript using secondary probes with one, and two readout sequences, respectively. The schematics of binding were shown above the figures. Three channels were shown, DAPI (blue), FITC (green), and Cy5 (red). Scale bar = 10 μm. B) Scatter plot showing the signal intensity on three conditions. n=30 puncta. C) Scatter plot showing the signal-to-noise ratio (SNR) on three conditions. n=30 puncta.

The signal intensity from the condition using secondary probe 2 was 1826 RFI (±281) (Fig. 3.8B), significantly higher than the one using secondary probe 1. However, the two-fluorophore structures also led to higher noise. This was consistent with the conclusion from the tests of primary probe structure that noise should primarily come from the nonspecific binding of secondary probes. The difference in SNR between secondary probe 1 and 2 was smaller, but secondary probe 2 still showed high SNR (Fig. 3.8C). Therefore, to maximize the difference in the signal intensity, secondary probe 2 was preferred in these tests.

3.2.8 Blocking buffer

FISH is based on the specific binding between the mRNA targets and DNA probes. However, nonspecific binding occurs, affecting overall SNR of transcript detection. To lower the noise, a blocking buffer can be exploited before (optional) and along the incubation with probes. In general, an effective blocking buffer should [3]: 1) bind with the surface firmly, stronger than nonspecific binding, but still weaker than the specific binding; 2) not impede the specific binding and the fluorescence signal (no cross-reaction, no depression effect); 3) not bring any undissolved large particles or contamination like bacteria; 4) be not sensitive to temperature change (avoid denaturation); 5) perform consistently across various batches. With the application of a blocking buffer, the interference from the background should decrease, while the signal intensity remains the same in order to improve hybridization efficiency and thus improve the SNR.

Currently, most blocking buffers that have been applied in FISH are mixed solutions of irrelevant nucleotide fragments, like yeast tRNA and herring sperm DNA, along with many other compounds. The DNA fragments were chemically and physically hydrolyzed into short base pairs to nonspecifically bind everywhere. Their application before the incubation with probes can pre-

bind with the sample surface. However, since these fragments do not contain a readout region to the secondary fluorescent probe, their binding with the target transcript should deter any subsequent specific labeling with the primary probes.

In addition to irrelevant nucleotide fragments, many other reagents might also help to improve the performance of FISH. For example, dextran sulfate was widely used [6-9], to accelerate the hybridization of probes with nucleic acid. Alternatively, ethylene carbonate might help the diffusion of probes [7]. BSA, bovine serum albumin, has often been used as a blocker of nonspecific sites.

Based on these chemistry information, three blocking buffers were designed and tested. Blocking buffer #3 (BB #3) was the simplest one, with only dextran sulfate and ethylene carbonate. Blocking buffer 2 (BB #2) has one more component, herring sperm DNA fragments, while blocking buffer #1 (BB #1) had one more component, BSA, and doubled the concentration of each component in BB #2. Condition 1, 2 and 3 refers to the conditions in which BB #1, #2 and #3 were used, respectively.

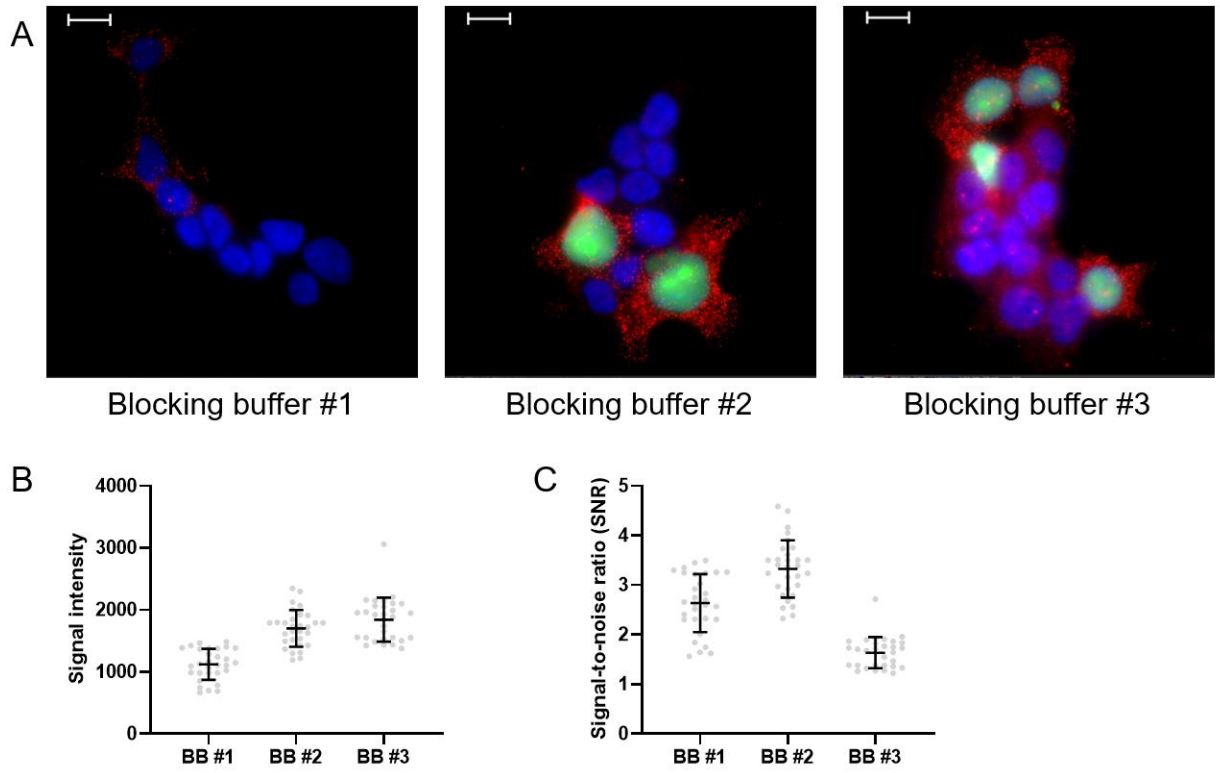


Figure 3.9 Influence of different blocking buffer on the detection of mNeonGreen transcripts. A) (from left to right) Merged images showed detection of mNeonGreen transcript with different blocking buffers used. Three channels were shown, DAPI (blue), FITC (green), and Cy5 (red). Scale bar = 10 µm. B) Scatter plot showing the signal intensity on three conditions. BB represents a blocking buffer. n=30 puncta. C) Scatter plot showing the signal-to-noise ratio (SNR) on three conditions. n=30 puncta.

In the condition 3, although the average signal intensity was highest (Fig. 3.9B), the noise was strongest as well, leaving a red background in the right image (Fig. 3.9A) and a lowest SNR (Fig. 3.9C). Some nonspecific binding even occurred on glass surface. Therefore, BB #3 was not a good choice of blocking buffer. Instead, with relatively high signal and normal noise, condition 2 showed the best SNR. In condition 1, the signal from mNeonGreen transcript was much lower but still detectable. However, the protein was hard to observe under the microscope. This suggested this blocking buffer might have impeded the hybridization as well as damaged the

fluorescent proteins expressed in these cells. So, in the following experiment, blocking buffer #2 was used.

3.2.9 Primary probe concentration

The concentration of primary probes varies depending on the RNA-FISH platforms [6-9]. To determine the best working concentration, five conditions were tested, with 100 nM, 50 nM, 20 nM, 5 nM, and 1 nM primary probes. Here, the term concentration referred to the concentration of each primary probe, instead of the overall one. Based on these experiments, it appears that the higher the concentration of primary probes was used, the higher the noise, especially in the condition with 100 nM of each probe (Fig. 3.10A). As seen in this image, several large clusters of puncta might have resulted from the aggregation of primary probes, which later bound with the secondary probes. In the conditions with 100 nM and 50 nM primary probes, the range of signal intensity was similar. This might result from the fluorescence saturation of each puncta. Based on the analysis on signal intensity (Fig. 3.10B) and SNR (Fig. 3.10C), 5 nM was found to be the best working concentration per probe.

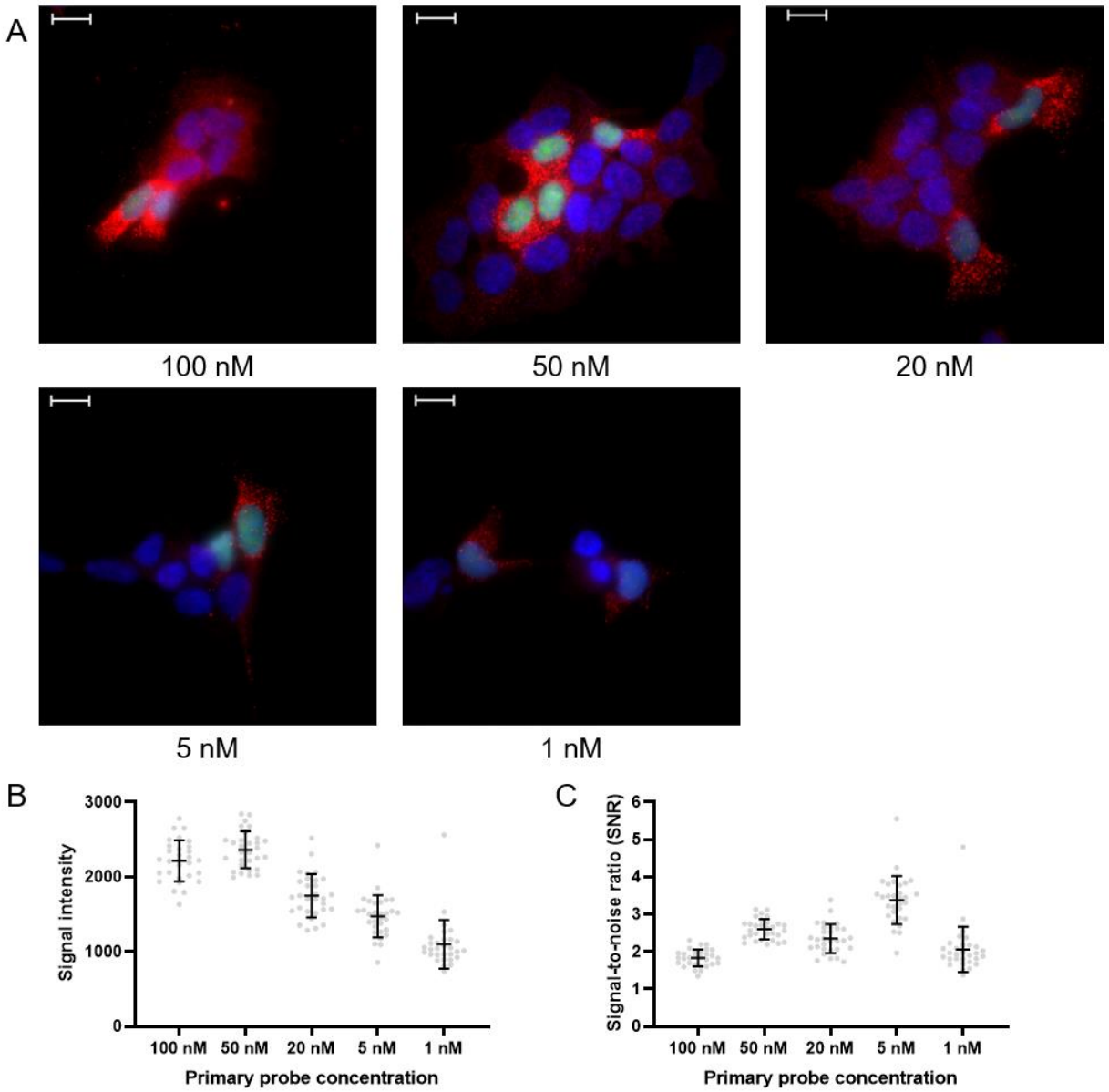


Figure 3.10 Influence of primary probe concentration on the detection of mNeonGreen transcript. A) (from left to right, top to bottom) Merged images showed detection of mNeonGreen transcript using 100, 50, 20, 5, and 1 nM primary probes, respectively. Three channels were shown, DAPI (blue), FITC (green), and Cy5 (red). Scale bar = 10 µm. B) Scatter plot showing the signal intensity on three conditions. n=30 puncta. C) Scatter plot showing the signal-to-noise ratio (SNR) on three conditions. n=30 puncta.

4.2.10 Incubation time of primary probes

Longer incubation time with primary probes is recommended by many research groups [6-9], to ensure the binding maximum and thus achieve the highest signal. To confirm this idea, mixed mNeonGreen cells were incubated with primary probes for 4, 8 and 16 hours (Fig. 3.11A).

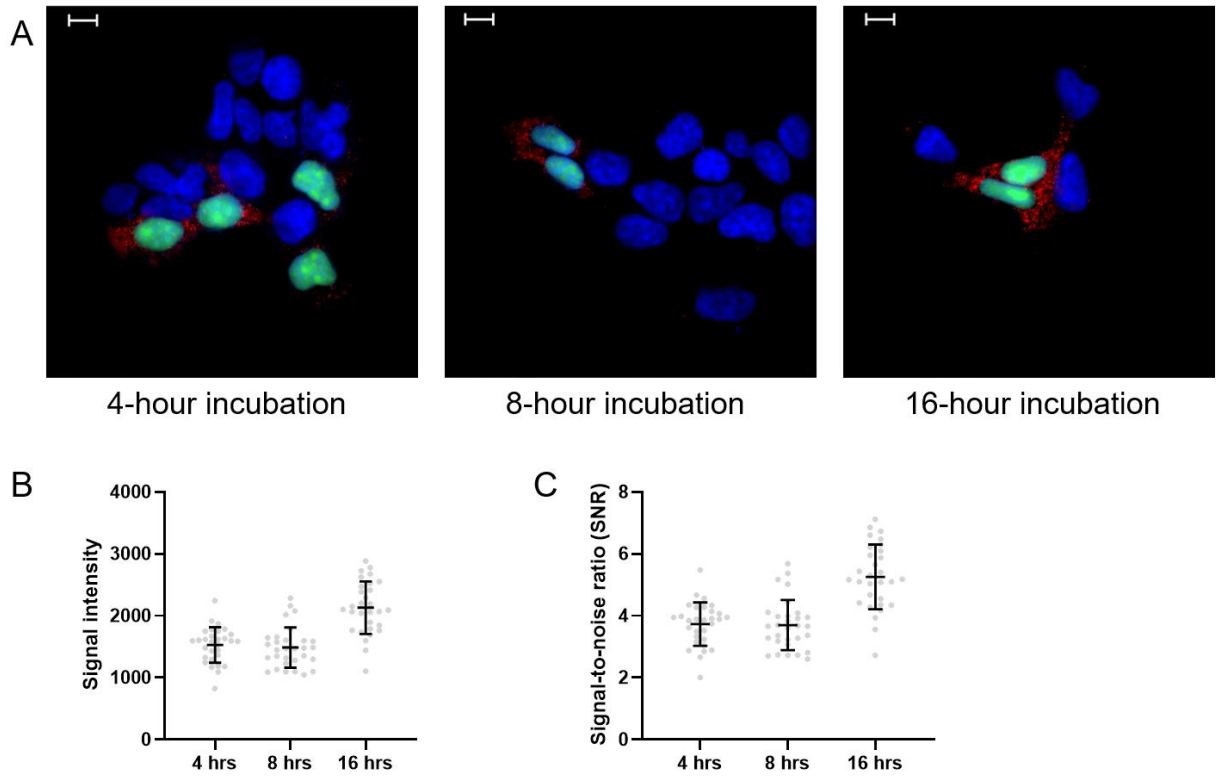


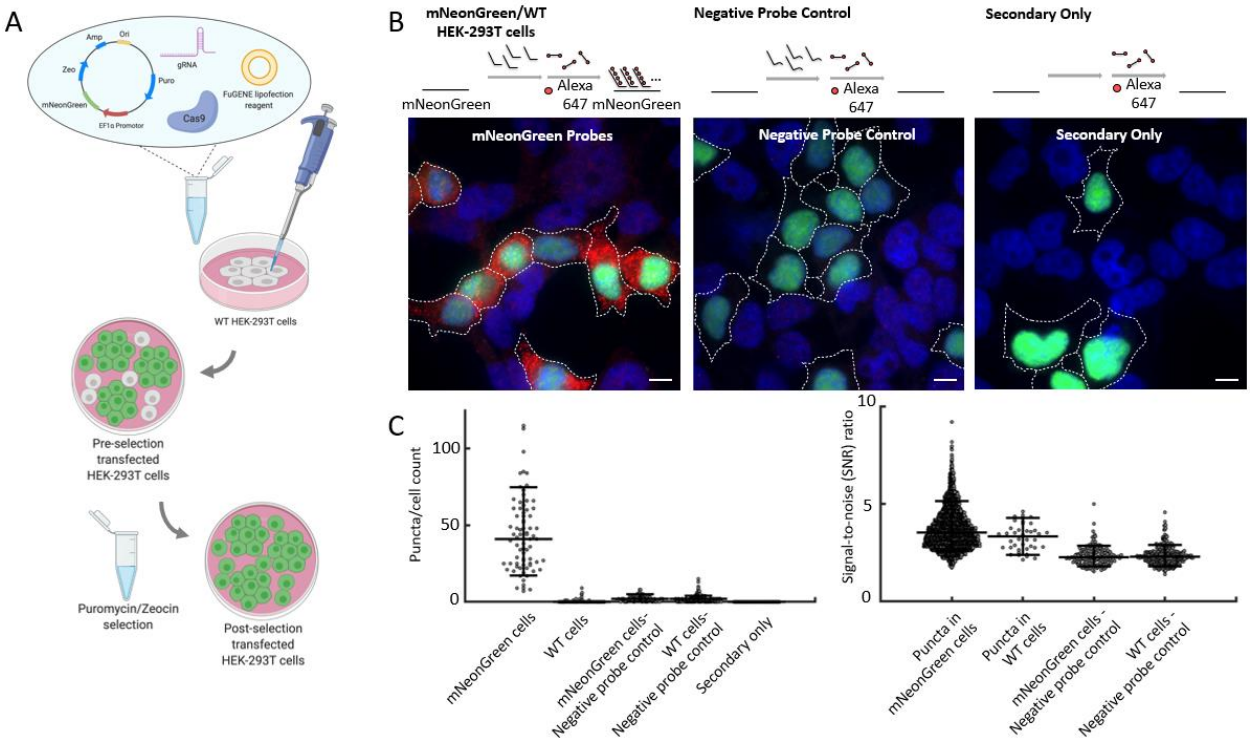
Figure 0.11 Influence of incubation time on the detection of mNeonGreen transcript. A) (from left to right) Merged images showed detection of mNeonGreen transcript with 4, 8, and 16-hour incubation of primary probes, respectively. Three channels were shown, DAPI (blue), FITC (green), and Cy5 (red). Scale bar = 10 µm. B) Scatter plot showing the signal intensity on three conditions. n=30 puncta. C) Scatter plot showing the signal-to-noise ratio (SNR) on three conditions. n=30 puncta.

Despite the great variations in signal intensity, the 16-hour incubation condition had the strongest signal (Fig. 3.11B) and highest SNR (Fig. 3.11C). However, the ones with 8-hour incubation failed to show much higher intensity than those with 4-hour incubation. The uneven

concentration of primary probes might account for this as well as the great variations in intensity. The different detection efficiency was clear in the mixed mNeonGreen cells with 4-hour incubation (Fig. 3.11A). Even in the same field of view, the number of fluorescent dots around each mNeonGreen cell differed significantly, suggesting the uneven concentration of primary probes in.

Thus, overall, we first investigated the specificity of our labeling condition using a simple cell mixture model comprising wild-type HEK293T-X cells and HEK293T-X cells engineered with mNeonGreen (Fig. S3.12A) by detecting mNeonGreen mRNA as the target gene. Since only fluorescent mNeonGreen positive cells can express the corresponding mRNA transcripts, this cell mixture model provides a straightforward tool to assess specificity and nonspecific binding. Using a Nikon epifluorescence microscope to image the samples following staining with primary and secondary probes, we detected on average 43.5 puncta per mNeonGreen positive cell (n = 76 cells) and 0.25 puncta per wild-type cell (n = 164) (Fig. S3.12,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 towards dopachrome tautomerase (DCT), 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 (Fig. S3.13B,C). We next optimized labeling efficiency by testing the number of primary probes and incubation times of primary probes and secondary probes. We determined our optimal condition to comprise a minimal of least 12 primary probes for each target mRNA (Fig. S3.13A,B) (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 hours for primary probe hybridization and 1 hour for secondary probe hybridization, respectively (Fig. S3.13C,D), which were used in subsequent experiments.



Supplementary Figure 3.12. 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).

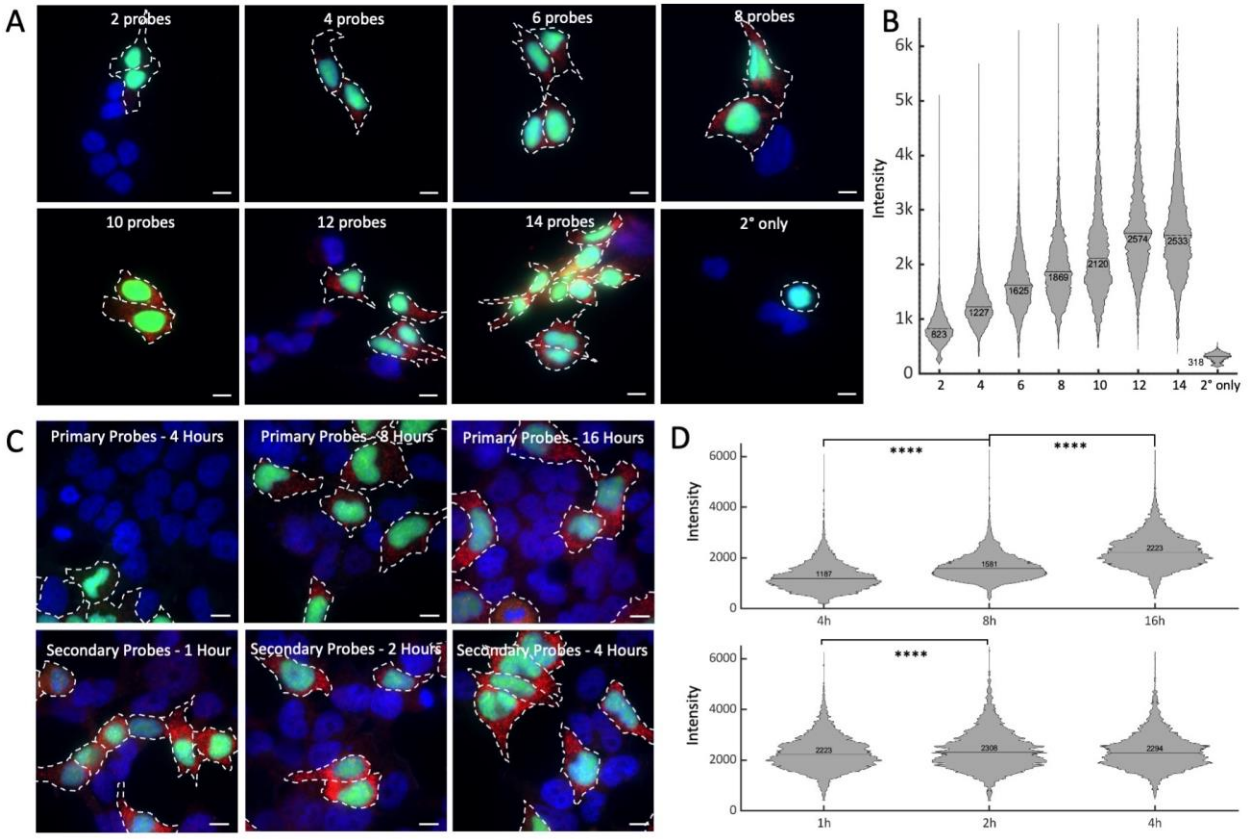


Figure 3.13. Optimization parameters of in situ hybridization conditions. A) Representative images of mNeonGreen cells with different numbers of primary probes. Conditions included 2, 4, 6, 8, 10, 12, and 14 mNeonGreen primary probes to the HEK293T-X mNeonGreen and WT model. The concentration of each 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 ****.

3.3 References

1. Pardue, M.L. and Gall, J.G., Molecular hybridization of radioactive DNA to the DNA of cytological preparations. PNAS, 1969. **64**(2): p. 600-604.
2. Femino, A.M., et al., Visualization of single RNA transcritps in situ. Science, 1988. **280**(5363): p. 585-90.
3. Pichon, et al., A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. Mol Cell, 2018. **71**(3): p. 468-480.
4. “mNeonGreen.” FPbase, www.fpbase.org/protein/mneongreen/.
5. “Synthetic Construct Monomeric Green Fluorescent Protein (MNeonGreen) g - Nucleotide - NCBI.” National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/nuccore/KC295282.1.
6. Chen, Kok Hao, et al. "Spatially resolved, highly multiplexed RNA profiling in single cells." Science 348.6233 (2015): aaa6090.
7. Eng, Chee-Huat Linus, et al. "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+." Nature 568.7751 (2019): 235-239.
8. Moffitt, Jeffrey R., et al. "High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization." Proceedings of the National Academy of Sciences 113.39 (2016): 11046-11051.
9. Kishi, Jocelyn Y., et al. "SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues." Nature methods 16.6 (2019): 533.

CHAPTER 4:
SPECTRAL-FLIM MICROSCOPY AND INSTRUMENTATION

Authors: **Tam Vu**[1,3] Alex Vallmitjana [3,4], Jan Zimak[1], Per Niklas Hedde [3,4], Enrico Gratton*[3,4], and Weian Zhao*[1,2,3]

Formal analysis: T. V., A. V., J. Z., P. H., E. G., W.Z.; methodology: T. V., A. V., J. Z., P. H., E. G., W.Z.; writing – T. V., A. V.; investigation: T. V., A. V., J. Z., P. H., E. G., W.Z.; data curation: T. V., A. V.; conceptualization: T. V., A. V., J. Z., P. H., E. G., W.Z.; resources: W. Z., E. G.

Affiliations:

*Corresponding authors

[1]Sue and Bill Gross Stem Cell Research Center, Sue & Bill Gross Hall CIRM Institute, University of California, Irvine, 845 Health Sciences Road, Suite 3027, Irvine, CA 92697, USA

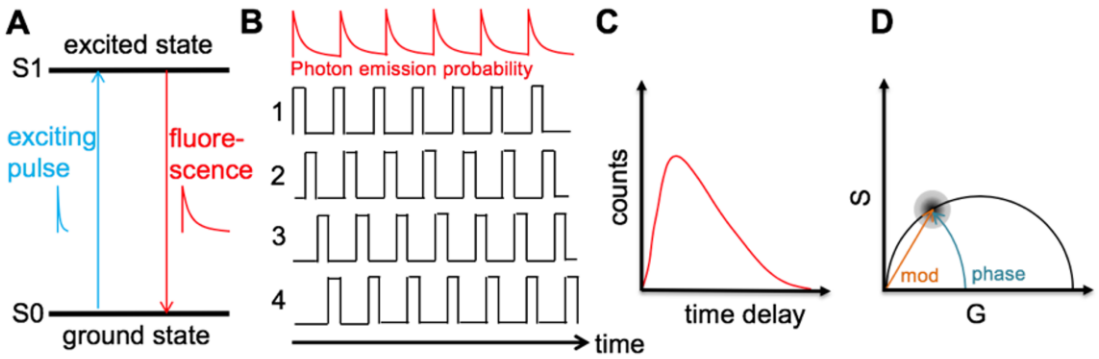
[2]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA

[3]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, USA

[4]Laboratory for Fluorescence Dynamics, Department of Biomedical Engineering, University of California, Irvine, USA

4.1 Spectral and FLIM Microscopy

Fluorescence imaging is the visualization of fluorescent dyes or proteins as labels for molecular processes or structures. It enables a wide range of experimental observations including the location and dynamics of gene expression, protein expression and molecular interactions in cells and tissues. The most common way that biologists use fluorescence to study cells and tissues is to measure light intensity produced by fluorophores or fluorescent reporters. However, intensity is only one dimension of fluorescence. Intrinsic to each fluorophore, is also its fluorescence lifetime which is a measure of how long a fluorophore remains on average in its excited state before returning to the ground state by emitting a fluorescence photon (Figure 4.1). This parameter may change with the probe's conformational state and has often been exploited by researchers to probe a fluorophore's surrounding molecular environment for its composition, such as ion concentration, pH, lipophilicity or viscosity [1-2].

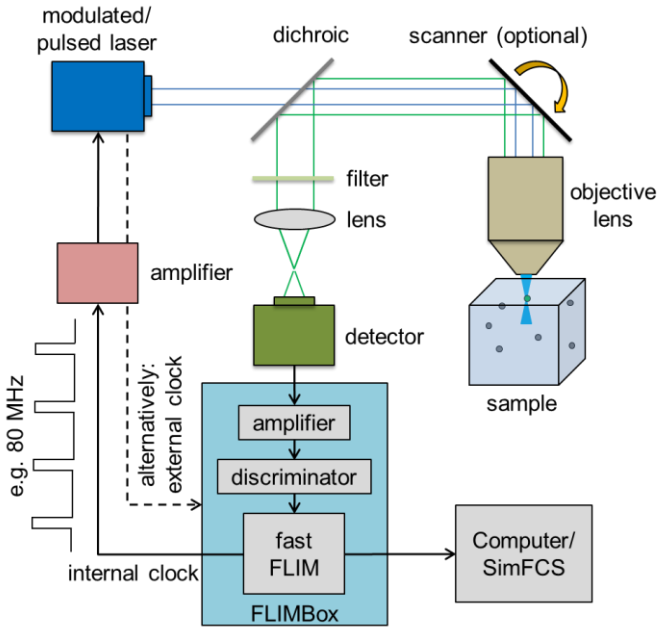


Hedde et al. Biomed Opt Express. 2019. [3]

Figure 4.1. Principle of fluorescence lifetime detection. (A) Fluorescence mechanism. (B) The distribution of time delays can be measured by repeatedly exciting and detecting fluorescence photons in multiple time bins. (C) From the number of photons detected in each time bin, the characteristic lifetime decay can be determined. (D) This data can be quantified on in the 2D phasor space by plotting the lifetime phase and modulation of the measured signal on a polar plot.

To measure a fluorophore's lifetime, specialized software and hardware are required [3]. A typical FLIM setup is shown in Figure 4.2. 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. From 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 (Figure 4.2). The phasor method draws from the digital frequency domain hardware and software that permits using all the photons detected from a sample [4]. 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.



Hedde et al. Biomed Opt Express. 2019 [3].

Figure 4.2. A typical optical setup for fluorescence lifetime measurement and analysis. A light source such as a laser excites the samples that are stained with probes and the emission is collected by the detector(s) which can be photomultipliers or cameras. Individual photon counts are registered by the lifetime imaging electronic device (FLIMBox) which in turn uses the clock from the light source or an internal clock to establish the time of arrival of the photons with respect to the excitation. If a scanner is used, it also provides a trigger signal to synchronize the spatial origin of each photon.

Fluorescence lifetime imaging microscopy (FLIM) is an especially powerful method that can supplement intensity-based measurements with a whole new dimension of data. Fluorophores which have the same excitation and emission spectra but different lifetimes can be differentiated with FLIM, opening up a new avenue for multiplexing capabilities [5]. Table 4.1 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 to be able to successfully image and differentiate them in spatial multiomics has yet to be done.

Table 4.1. List of Commonly Used Fluorophores Differing in Spectra or FLIM Characteristics

	Exciation Max	Dichroic (ideal)	Emmision Max	Extinction Coefficient	Quantum Yield	Brightness	Lifetime
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 ns
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 ns
ATTO 425	439	462	485	45,000	0.9	40500	3.6 ns
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 ns
BODIPY FL-X	504	507	510	85000	N/A	N/A	N/A
Alexa 488	495	507	519	73,000	0.92	67160	4 ns
FITC	495	507	519	N/A	N/A	N/A	4.1 ns
BODIPY FL	505	509	513	80000	N/A	N/A	5.7
ATTO 488	500	510	520	90,000	0.8	72000	4.1 ns
ATTO 495	498	512	526	80,000	0.2	16000	1.0 ns
Oregon Green 488	498	512	526	76000	N/A	N/A	4.1 ns
ATTO 514	511	521.5	532	115,000	0.85	97750	3.9
ATTO 520	517	527.5	538	110,000	0.9	99000	3.6
Cy3	512	531	550	136,000	0.15	20400	0.3 ns
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 ns
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 ns
ATTO 542	542	552	562	120,000	0.93	111600	3.7 ns
BODIPY TMR-X	542	558	574	60000	N/A	N/A	6 ns
Alexa 555	555	560	565	155,000	0.1	15500	0.3 ns
Dylight 549	553	561	569	N/A	0.27	N/A	0.65 ns
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 ns
ATTO 565	564	577	590	120000	0.9	108000	4.0 ns
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
BODIPY 581/591	584	588	592	136000	N/A	N/A	N/A
ATTO Rho12	577	588.5	600	120,000	0.8	96000	4
Alexa 568	578	590.5	603	88,000	0.69	60720	3.6 ns
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 ns
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 ns
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 ns
ATTO 647N	646	655	664	150,000	0.65	97500	3.5 ns
ATTO 647	647	657	667	120,000	0.2	24000	2.4 ns
Alexa 647	650	659	668	270,000	0.33	89100	1.04 ns
CY5	650	660	670	250,000	0.3	75000	1 ns
ATTO 665	662	671	680	160000	0.6	96000	2.9
ATTO 655	663	671.5	680	125,000	0.3	37500	1.8 ns
Alexa 660	663	676.5	690	132,000	0.37	48840	1.2 ns
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 ns
Alexa 750	749	762	775	290,000	0.12	34800	0.66 ns
Cy7	750	769	788	199,000	0.3	59700	0.8 ns
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 easily distinguishable on a FLIM phasor plot (Figure 4.3). Through intensity measurements, these moieties are inseparable from each other or from a labeled fluorophore but through FLIM measurements, these moieties can be easily corrected or removed out from analysis via the phasor approach [6].

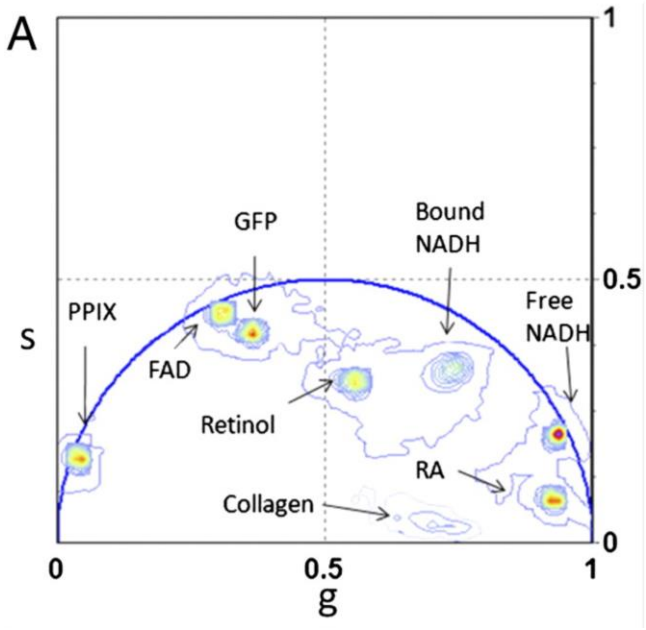


Figure 4.3. A representative phasor plot of multiple autofluorescent moieties. Cells and tissues have autofluorescence artifacts with unique characteristic lifetime signatures. These signatures occupy distinct positions on the phasor plot which can be removed from analysis. Stringari et al. PNAS. 2011[6].

4.2 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 to 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 [7-8]. 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 the supplemental material). 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 [9]. 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 5 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. A

typical image, on the order of 10^6 pixels, obtained with this method provides 10^6 points in this 5D space [10]. 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.

Figure 4.4 depicts the spectral/lifetime analogy for fluorescence microscopy using the phasor approach. As an example, it uses a hypothetical experiment where 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. 4.4A). In each pixel, we accumulate enough photons to build a spectral histogram and a lifetime histogram (Fig. 4.4B). 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 [11], 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. 4.4C). 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. 4.4D). Again, note that in this example in Figure 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.

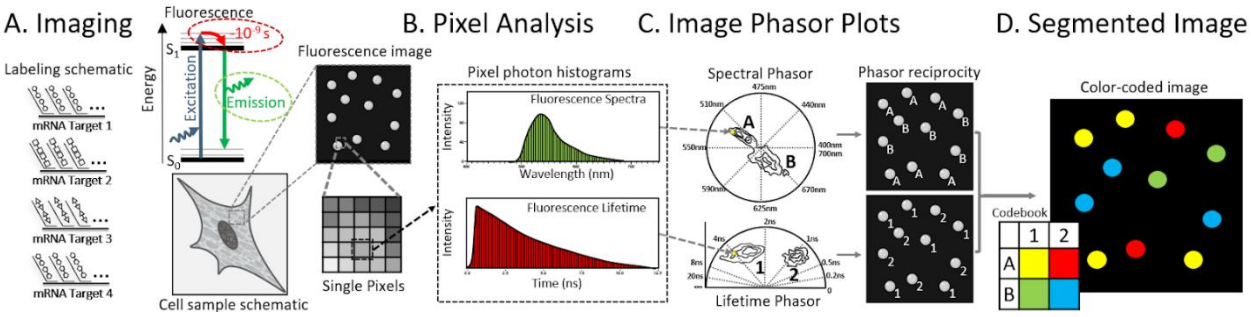


Figure 4.4. Imaging and phasor analysis with Spectral-FLIM. A) As an example, four different probes are used to target four different genes. The fluorescence is collected using the spectral-FLIM instrument to form images where each pixel carries information of the spectra and lifetime. B) At each pixel we compute the photon distribution in the spectral and temporal dimension. The lifetime phasor transform maps these distributions in each pixel to a position on the phasor space. C) The lifetime phasor plots are built by transforming every pixel in the image to 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.

4.3 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 4.5 shows a real case with such configuration, both for spectra and for lifetime. In Figure 4.5A, the cartoon represents the case of using two probes with distinct spectra. When imaging this sample, we can use two spectral channels, Figure 4.5B/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. 4.5D) and the overall counts of each combination in the field of view can be provided (Fig. 4.5E).

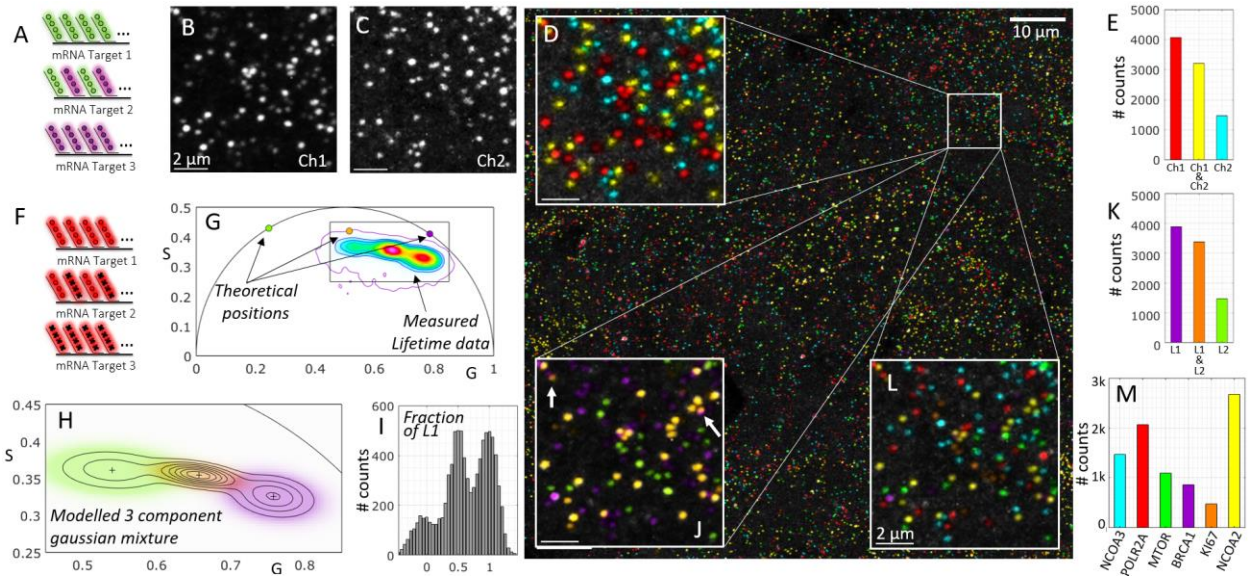


Figure 4.5. Working example of combinatorial labelling of three mRNA targets with two probes. A) Three different target genes are tagged using two probes with different spectra. Targets 1 and 3 are tagged each with one probe, Target 2 is tagged with both simultaneously. B,C) The fluorescence is collected in the two expected spectral channels for the known emission of the two probes. D) 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.

Similarly, we show a case in which the targets are now labeled with two probes that have similar spectra but different lifetimes (Fig. 4.5F). 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 4.5G 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 [11], and we use this machine learning technique to infer the phasor locations of the probe combinations (Fig. 4.5H). 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 3 clusters (Fig. 4.5J) and obtain the counts of the three lifetime components (Fig. 4.5K). Additionally, we apply our lifetime multicomponent analysis technique [12] 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, ½ and 1] (Fig. 4.5I).

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 Figure 4.5J, 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 Figure 4.5, 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. 4.5L,M). The specific 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:

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

4.4 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. 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 gene. The whole pipeline is depicted in Figure 4.6.

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[47]. We used an initial experiment tagging house-keeping 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 label to it by a priori knowing the expected positions of each combination of probes depending on the spectra and lifetime of the probes. 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 in between the segmented volumes. This yields a division of the imaged volume into a number of polyhedra where each face is exactly in 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.

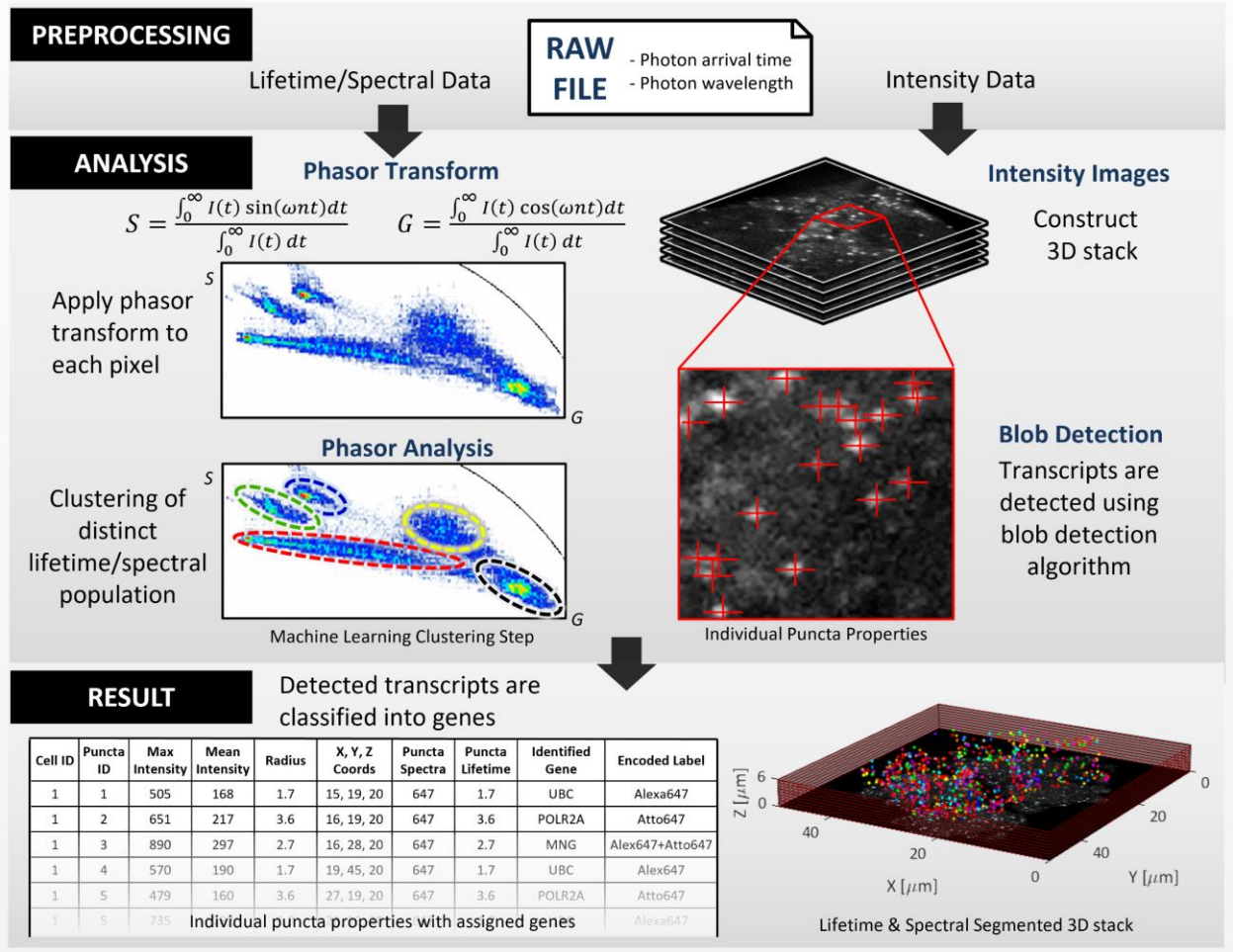


Figure 4.6. Automated pipeline of the processing and analysis. Raw data consists of a list of detected photons with their times of arrivals. Using the acquisition parameters, dwell time, number of pixels, number of repetitions per image etc. the image stacks are reconstructed. Knowing the laser frequency, 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.

4.5 References

1. Chacko, J.V. and Eliceiri, K.W., Autofluorescence lifetime imaging of cellular metabolism: Sensitivity toward cell density, pH, intracellular, and intercellular heterogeneity. Cytometry A, 2019. **95**(1): p. 56-69.
2. Lin, H.J.L, et al. Fluorescence lifetime-resolved pH imaging of living cells. Cytometry A, 2003. **52**(2): p. 77-89.
3. Hedde, P.N., et al. Fluorescence lifetime detection with particle counting devices. Biomed Opt Express, 2019. **10**(3): p. 1223–1233.
4. Colyer, R.A., C. Lee, and E. Gratton, A novel fluorescence lifetime imaging system that optimizes photon efficiency. Microscopy research and technique, 2008. **71**(3): p. 201-213.
5. Sardar, P., et al, Molecular probes for fluorescence lifetime imaging. Bioconjug Chem, 2015. **26**(6):963-74.
6. Stringari, C., et al., Phasor approach to fluorescence lifetime microscopy distinguishes different metabolic states of germ cells in a live tissue. PNAS, 2011. **108**(33): p. 13582-13587.
7 Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The phasor approach to fluorescence lifetime imaging analysis. *Biophys J* **94**, L14-16, doi:10.1529/biophysj.107.120154 (2008).
8 Ranjit, S., Malacrida, L., Jameson, D. M. & Gratton, E. Fit-free analysis of fluorescence lifetime imaging data using the phasor approach. *Nature protocols* **13**, 1979-2004 (2018).
9 Fereidouni, F., Bader, A. N. & Gerritsen, H. C. Spectral phasor analysis allows rapid and reliable unmixing of fluorescence microscopy spectral images. *Optics express* **20**, 12729-12741 (2012).
10 Scipioni, L., Rossetta, A., Tedeschi, G. & Gratton, E. Phasor S-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging. *Nature Methods* **18**, 542-550 (2021).
11 Vallmitjana, A., Torrado, B. & Gratton, E. Phasor-based image segmentation: machine learning clustering techniques. *Biomedical Optics Express* **12**, 3410-3422, doi:10.1364/BOE.422766 (2021).
12 Vallmitjana, A. *et al.* Resolution of 4 components in the same pixel in FLIM images using the phasor approach. *Methods and Applications in Fluorescence* **8**, 035001 (2020).

CHAPTER 5:
SPECTRAL-FLIM WORKFLOW AND POTENTIAL APPLICATIONS

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

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. consulted on the manuscript. E.G. and W.Z designed and directed the project.

Affiliations:

1. Department of Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

2. Laboratory for Fluorescence Dynamics, University of California, Irvine, Irvine, California 92697, United States

3. Department of Biological Chemistry, University of California, Irvine, Irvine, California 92697, United States

4. CIRM Stem Cell Research Biotechnology Training Program at California State University, Long Beach, Long Beach, California 90840, United States

5. Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, Irvine, California 92697, 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. School of Medicine, University of California, Irvine, Irvine, California, 92617, United States

9. Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, California 92697, United States

10. Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California 92697, United States

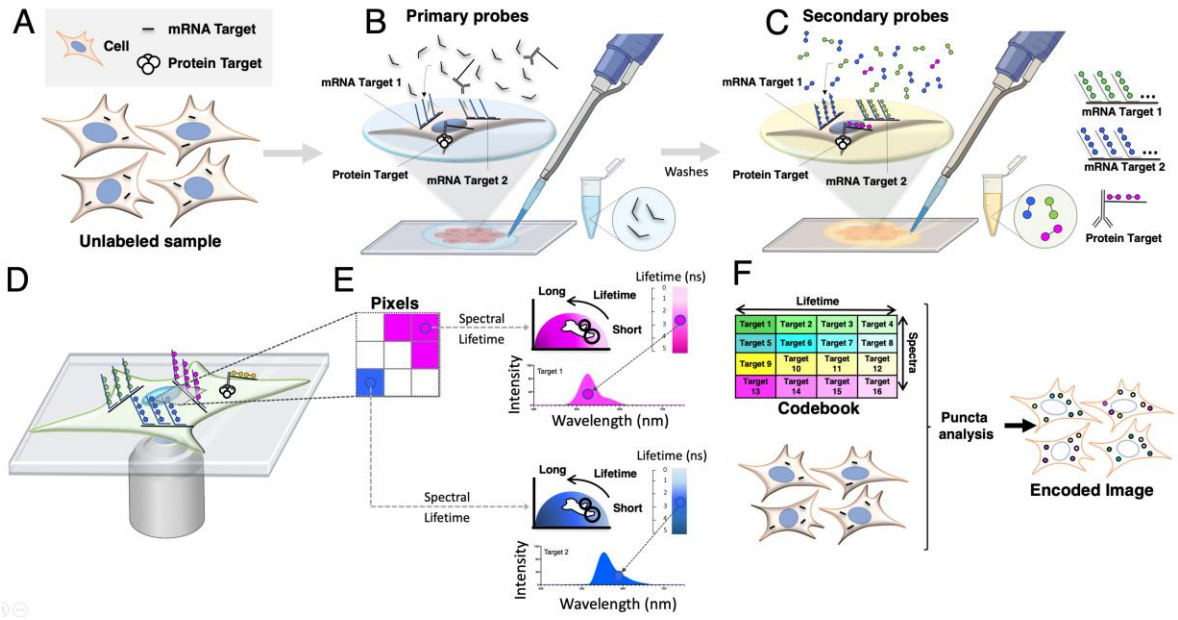
11. Beckman Laser Institute & Medical Clinic, University of California, Irvine, Irvine, California 92697, United States

† These authors contributed equally to this work.

* Corresponding Authors

Chapter 5.1 Probe design and labeling and detection strategy

To exploit the many benefits that Spectral-FLIM microscopy provides, the following probe labeling and detection strategy is proposed (Figure 5.1).



Artwork by Josh Gu

Figure 5.1. **Schematic of the Spectral-FLIM technology for spatial multi-omics analysis**. (A) Sample(s) comprise cells and target molecules. (B) Primary probes are added to the sample to bind to targets (e.g. nucleic acids, proteins). (C) Secondary label probes are added to bind to the primary labels through a "readout" domain to label each target molecule with a unique combinatorial fluorescence spectrum and lifetime encoded signature. (D) Labeled targets are imaged under a microscope (in this case, a SPIM) to interrogate both spectrum and lifetime using phasor analysis (E). (F) Labeled targets eliciting the encoded spectrum and lifetime signatures are automatically identified and quantified in a highly multiplexed fashion in 3D tissues using a codebook.

Starting with a fixed sample, primary probes designed to bind specifically to mRNA targets at a complementary target region were added. In parallel, for protein detection, primary antibodies conjugated with oligonucleotides were also added (Figure 5.1A-B). Both sets of

primary probes were designed to contain an adaptor region called a readout region which can hybridize to complementary secondary fluorescent probes. For each gene and protein biomarker, a different combination of readout regions is utilized to encode each biomarker with a unique fluorescent signature. Following incubation with fluorescent oligonucleotides, each type of biomarker will exhibit a distinct fluorescent signature (Fig. 5.1C). To detect these signatures simultaneously without numerous rounds of stripping and re-hybridization of probes or image registration, a microscope with spectral-FLIM capabilities was used (Figure 5.1D). Since each image contains both spectral and lifetime information, each pixel can be replotted onto a 2D phasor plot to decode and identify the fluorophore species present at each point (Figure 5.1E). Using a pre-designed codebook, each mRNA puncta or protein to generate an encoded image that provided the target gene xyz location and abundance (Fig. 5.1F).

To rapidly design oligo FISH probes for each gene, the validated python platform, OligoMiner, is used (Figure 5.2A) [1]. Briefly, after entering the accession number for the target gene, the program will obtain the input sequence. The blockParse script will screen the input sequence and output a file with candidate probes while allowing users to maintain consistent and customized length, GC, melting temp, 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 script to generate a file of unique candidates only. The primary probes comprise complementary sequence of typically 27-30 nucleotides (nt) and are designed mostly within the coding sequence (CDS), which has fewer variation than the untranslated region (UTR) [2]. Furthermore, primary probe “read-out” domains and secondary probes (typically 15 to 20 nt long) will be 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 will simply use those that have been previously validated [3]. 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 will be obtained through vendors such as IDT, Sigma, or Genscript. Lastly, to further ensure the specificity of probes, an online BLAST query was also run on each sequence against the human transcriptome (Figure 5.2B). After these selections, Sigma Oligoevaluator is implemented to screen for secondary structures and any potential primer dimers (Figure 5C).

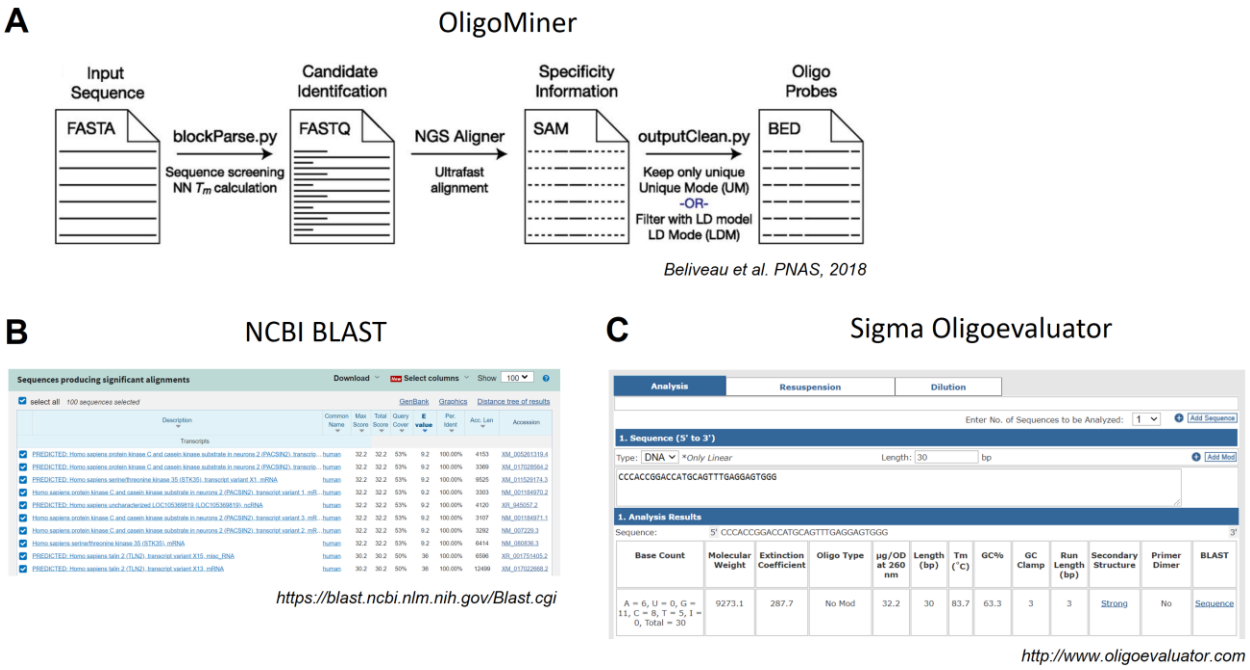


Figure 5.2. Probe Design Workflow. (A) OligoMiner is a validated probe design pipeline which can rapidly screen and generate tens to hundreds of probes per gene target following specific chosen parameters. (B) NCBI Blast is utilized to screen for probes which target only one specific transcript. (C) Sigma Oligoevalutor calculates any potential secondary structure or primer dimer.

5.2 Implementing Spectral-FLIM microscopy for detection and analysis

While this approach is designed to eventually work with many commercial (e.g. Leica FALCON, PicoQuant rapidFLIM, ISS FastFLIM) and custom microscope platforms, this technology was first validated and demonstrated on the ISS Alba STED FLIM microscope. The Alba-STED system is a confocal microscope equipped with a pulsed white laser system where the excitation wavelength(s) can be tailored to any combination of fluorescent probes. Single photons are detected with ultra-sensitive avalanche photodiode detectors and lifetimes are measured with the SPARTAN 6 FPGA-based 4-channels ultra-high speed (640 MHz) electronics. Imaging is achieved by fast beam scanning with galvo mirrors and 3D stacks of images can be acquired with a z-piezo stage. Spectral and lifetime data were analyzed using the phasor approach with simFCS software developed by Dr. Enrico Gratton. The phasor space is extremely useful to visually resolve different, heterogeneous spectrum and lifetime populations with each population representing a target molecule of interest.

For validation, fluorophores used for probe labeling were first prepared as a calibration solution to measure each fluorophore's expected spectral and lifetime properties. Samples containing cells expressing a particular mRNA target (e.g. mNEONGreen HEK293T) were then labeled accordingly and measured for its fluorescent and autofluorescence signatures and compared against nonlabeled samples and the corresponding fluorophore calibration files. For these same measurements, the corresponding spectral measurements alone taken under similar conditions will be compared against the FLIM measurements to assess the advantages conferred by FLIM.

5.3 Tissue Validation with Spectral-FLIM

Ideally, this technology should work with all tissues of origins (mouse, human, drosophila, etc.) and preservation matrices (OCT, FFPE, fresh frozen, etc). To validate the general applicability of this approach to these various types of samples, a panel of positive (e.g. housekeeping genes such as POLR2A) and negative control (e.g. genes from soil bacteria, DapB) markers will need to be assessed using this platform vs. a more validated gold-standard platform such as RNAScope. For each experiment, multiple replicates and matching positive and negative control tissues will be run for enough statistical rigor. Each experiment will also be repeated enough times to ensure reproducibility. In addition, unlabeled tissue samples will be imaged to determine the amount of autofluorescence inherent in each sample and then compared against a labeled sample.

5.4 Combinatorial labeling for greater multiplexing capabilities

Combinatorial labeling scales by (nCr) and can greatly confer greater multiplexing capabilities with each incremental increase in the number of fluorophores. For each fluorophore added to the panel, its spectral and lifetime characteristics will be assessed as a calibration solution in PBS or DI water as well as in cells and tissues to determine the expected lifetime shift. Its combination with other fluorophores will also be measured to determine if any unexpected interaction will occur between the neighboring fluorophores. To validate large panels of genes, it becomes crucially important to work with cells or tissues which have been pre-sequenced to determine if the correct ratio of genes is being detected as well as the expected variance. Preferably, the samples that will be worked with will have both bulk sequencing and single cell sequencing data as a benchmark comparison. Prior to detecting a larger panel of genes such as a 100-plex panel, smaller panels will be scaled up gradually, e.g. 6 to 16 to 32 to 100.

Moreover, several combinatorial-based labelling approaches (Figure 5.3) is proposed. First, transcripts can be labeled with one type of fluorophores (ALEXA 647 or ATTO 647) or a blend of fluorophores (ALEXA 647 and ATTO 647) with similar spectra but different lifetimes to elicit distinct fluorescence signatures (Figure 5.3A). Second, they can also be labeled with different combinations of or 3 to scale by (nCr) as long as each fluorophroe has a distinct spectral or lifetime signature (Figure 5.3B). Third, FLIM-FRET interactions resulting from two adjacent donor and acceptor can be elicited to create unique fluorescence signatures (acceptor lifetimes are proportionally decreased in response to donor quenching ability) (Figure 5.3C). Lastly, FRET lifetimes decay in proportion to the 6th power of the Forster radius, and the distance between the fluorophores can be modulated to program different FRET-dependent lifetimes combinations (Figure 5.3D). Uniquely, this molecular programming approach, using nucleic acids to direct FRET behavior, allows sub-5nm precision to resolve different lifetimes. Collectively, these approaches can occur synergistically to generate different fluorescent barcodes for greater multiplexing capabilities. For each proposed approach, validation will occur as previously described. Panels of positive and negative control genes and cell lines will be required and tested against a benchmark test such as RNAScope or sequencing to confirm.

Figure 5.3. Combinatorial labeling schemes. (A) Transcripts can be labeled with one type of fluorophores (Alexa 647 or Atto 647) or a blend of fluorophores (Alexa 647 and Atto 647) to elicit distinct fluorescence signatures. (B) Transcripts can be labeled with different combinations of fluorophores to increase the number of unique signatures combinatorically. (C) Different FRET pairs can be used to label transcripts to create fluorescent labels which differs in spectra and/or lifetime. (D) Varying the distance between FRET pairs will elicit changes in both spectral and lifetime to further enable greater multiplexing.

5.5 Validating Protein and mRNA co-detection

Simultaneous detection of protein and transcript levels will reveal the genotypic and phenotypic heterogeneity and provide enriched information for fundamental biology and disease diagnosis at many length scales (sub-cellular to whole tissue sections). Few existing commercial and academic technologies can profile both RNA and protein within the same sample either due to limitations of detection and/or sample processing. A key challenge is to harmonize upstream protocols to maintain the viability of both the target RNA sequence and target protein structure, while permeabilizing the sample enough such that targets can be detected [4-5]. For target retrieval, several approaches will be needed to test based on manufacturer's recommendations

and recent literature to determine if antigen structure of the expressed proteins can still be recognized through multiplexed immunofluorescence after RNA transcripts are labelled.

While RNA and protein detection will be imaged simultaneously using the fluorescence spectral and lifetime imaging approach, RNA and protein staining will be tested simultaneously as well as sequentially in both orders. First, target retrieval (RNA and protein) are crucial steps and will be performed using existing mild approaches described previously [6]. Protein staining reactions will then be performed (dye-conjugated antibodies) which will require a blocking and binding step. Following this, RNA staining will be performed as described previously. Each condition will have matching tissue and cell control replicates where only one assay is performed to compare and determine if any confounding effects occurred between the two assays.

5.5 References

1. Beliveau, B.J., et al., *OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes.* Proceedings of the National Academy of Sciences, 2018. **115**(10): p. E2183-E2192.
2. Eng, C.L., et al., *Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH.* Nature, 2019. **568**(7751): p. 235-239.
3. Xu, Q., et al., *Design of 240,000 orthogonal 25mer DNA barcode probes.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2289-94.
4. Grabinski, T.M., et al., *A method for combining RNAscope in situ hybridization with immunohistochemistry in thick free-floating brain sections and primary neuronal cultures.* PLoS One, 2015. **10**(3).
5. Kalra, J. and J. Baker, *Multiplex immunohistochemistry for mapping the tumor microenvironment*, in *Signal Transduction Immunohistochemistry*. 2017, Springer. p. 237-251.
6. Wang, F., et al., *RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues.* J Mol Diagn, 2012. **14**(1): p. 22-9.

CHAPTER 6:
MULTIPLEXED DETECTION WITH MOSAICA

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

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. consulted on the manuscript. E.G. and W.Z designed and directed the project.

Affiliations:

1. Department of Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

2. Laboratory for Fluorescence Dynamics, University of California, Irvine, Irvine, California 92697, United States

3. Department of Biological Chemistry, University of California, Irvine, Irvine, California 92697, United States

4. CIRM Stem Cell Research Biotechnology Training Program at California State University, Long Beach, Long Beach, California 90840, United States

5. Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, Irvine, California 92697, 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. School of Medicine, University of California, Irvine, Irvine, California, 92617, United States

9. Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, California 92697, United States

10. Department of Microbiology and Molecular Genetics, 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 Authors

6.1 MOSAICA

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 new 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[7-11]. 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 (scRNAseq) 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]. Spatial transcriptomic methods that are based on sequential labeling, stripping, and imaging (e.g. seqFISH, MERFISH), branched amplification (e.g. RNAscope™, SABER), or in situ sequencing (e.g. GeoMx™, Slide-seq) are often too complicated, error-prone, time-consuming, laborious and costly to scale up, limiting their broad usage [18-23]. Furthermore, repeated processing of the same sample can damage tissue structural integrity and target molecules and are often not feasible for precious samples such as clinical biopsies.

In this work we report a new spatial-omics 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. 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 [24]. By utilizing both time and intensity domains for labeling and

imaging, we were able to discriminate a large repertoire of spectral and lifetime components simultaneously within the same sample to enable increased multiplexing capabilities with standard optical systems.

In this study, we described 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 panel in SW480 colorectal 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-correction, and autofluorescence removal in highly scattering and autofluorescent clinical melanoma FFPE tissues, demonstrating its potential use in tissue for cancer diagnosis and prognosis.

6.2 10-plex Combinatorial Panel in SW480 Cells

We 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 [25], which will allow us to study tumorigenesis in the spatial context and potentially identify cancer stem cell populations in colorectal cancer in future. Here,

we selected this model as a validation platform to demonstrate the multiplexing scalability and error correction 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 each gene: BRCA1, BRCA2, CENPF, CKAP5, KI67, MTOR, NCOA1, NCOA2, and NCOA3. These genes were chosen due to their housekeeping status or involvement in tumorigenesis in colorectal cancer. By encoding each gene with a distinct combination of two fluorophores, we generated a codebook of 10 labelling combinations from only five fluorophores following Equation 1: $\binom{5}{2} = 10$ (Fig. 6.1A). 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. 6.1A right). Matching numbers and concentrations of primary and secondary probes that were used in the 10-plex panel were used in this sample.

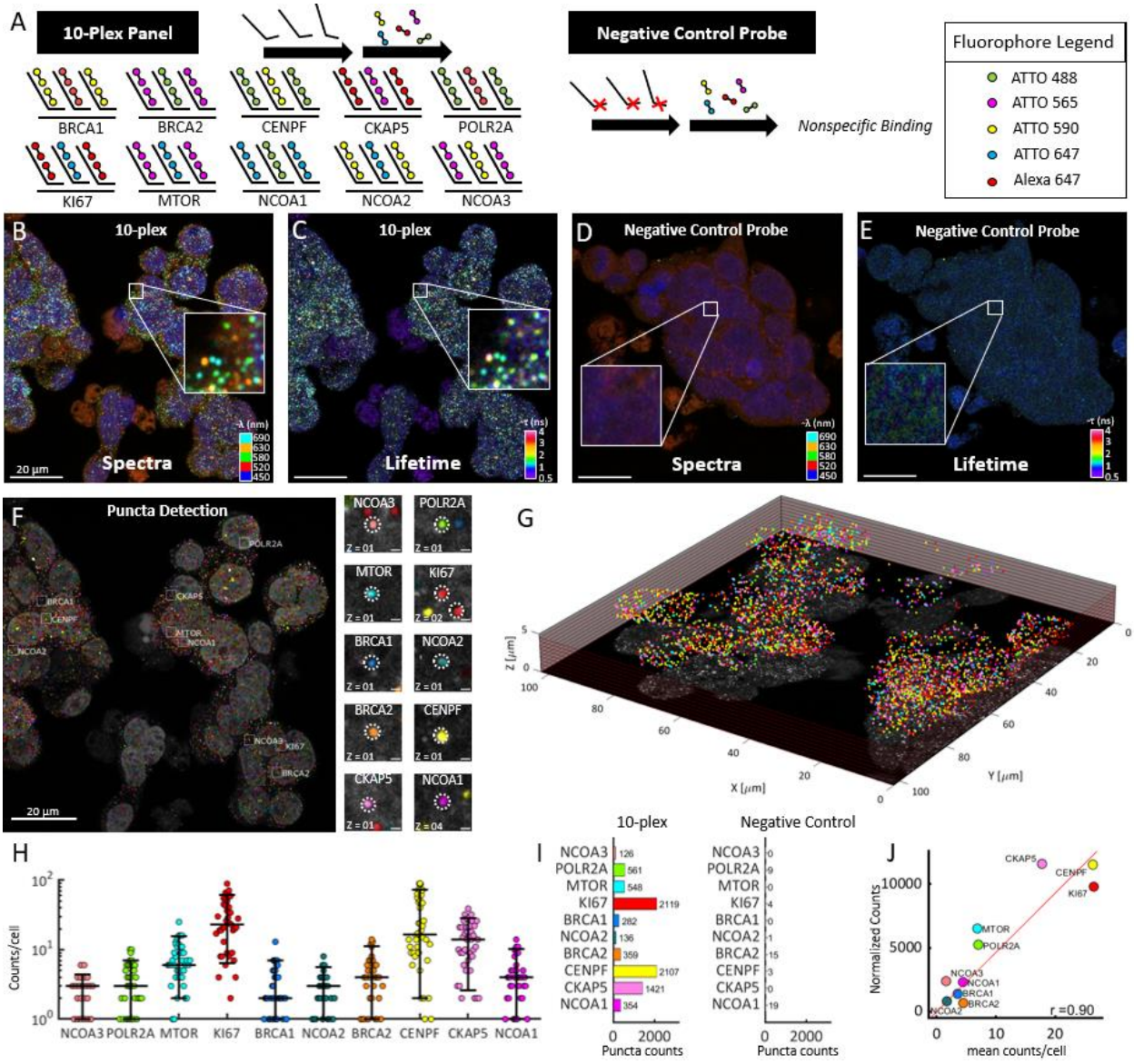


Figure 6.1. Simultaneous 10-plex detection of genes in colorectal cancer SW480 cells in a single round of labeling and imaging. A) 10 different types of gene transcripts are labeled with primary probes followed by respective and complementary fluorescent secondary probes. Each gene is labeled with a combination of 2 different 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). Scale bar 20 µm in main image and 1 µm in insets. G) 3D representation

of the field of view for the 10-plex sample. H) Number of puncta detected for each gene expression target in each cell for the labeled 10-plex sample (N=80 cells). I) Overall puncta count of each transcript in the 10-plex sample (left) and negative control probe sample (right). J) Correlation of detected puncta (mRNA puncta count) vs. RNA-bulk sequencing (normalized counts) is shown for each target with Pearson r of 0.9.

Figure 6.1B depicts a spectral image overlay (five fluorescent channels including DAPI) of the labeled 10-plex SW480 sample. Additionally, in the same measurement, the orthogonal lifetime information attained by interrogating each pixel for their lifetime components. These pixels were phasor-transformed and pseudo-colored based on their projected phasor coordinates on the universal circle, creating the image depicted in Figure 6.1C. 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, Figure 6.1D/E depict the merged composite spectral and lifetime images of the corresponding negative control probe sample. Figure 6.1F 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. With 40 Z-stack images, we generated a 3D spatial distribution of the field of view to visualize the spatial analysis in a 3D context (Fig. 6.1G).

MOSAICA employs an error-correction strategy by gating for specific and pre-encoded fluorophore combinations and rejecting any fluorescent signature which do not meet these criteria. Specifically, in addition to the population of pixels that represents the decodified puncta as mentioned above, there was a population of pixels of various shapes and sizes which did not fit the criteria. Simulations to look into overlapping and inconsistent signals were run (Fig 6.2A,B). Plotted in Figure 6.2C are the total detected puncta in this experiment, separated into

populations depending on our classification criteria. Of the total detected puncta (n = 13,521), a considerable fraction n = 4,488 (33%), were observed emitting fluorescence only in a single channel and was indicative of nonspecific binding events and/or autofluorescence moieties. As previously characterized by several groups, nonspecific binding events is an inherent issue with smFISH which arises from the stochastic binding of cDNA probes towards cellular components such as proteins, lipids, or nonspecific regions of RNA and follow a random distribution [26].

When combined with native autofluorescence moieties which can also exists as diffraction-limited structures that emits a strong fluorescent signal in any single channel, this group represents a confounding problem for standard intensity-based measurements and analysis because they share similar SNR and intensities to real labeled puncta and cannot be differentiated. This is the main benefit of implementing the combinatorial encoded criteria which rejects this large amount (around 33%) of stochastic and nonspecific binding labeling events as well as any event eliciting a lifetime signature that deviated from the utilized commercial 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 = 880. 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. We attained an interval for the fraction of lost puncta due to optical crowding ranging from 2.1% to 6.6% which accounts for the 880 puncta or 6% of the total detected puncta. We name this group the overlapped in Figure 6.2C,D.

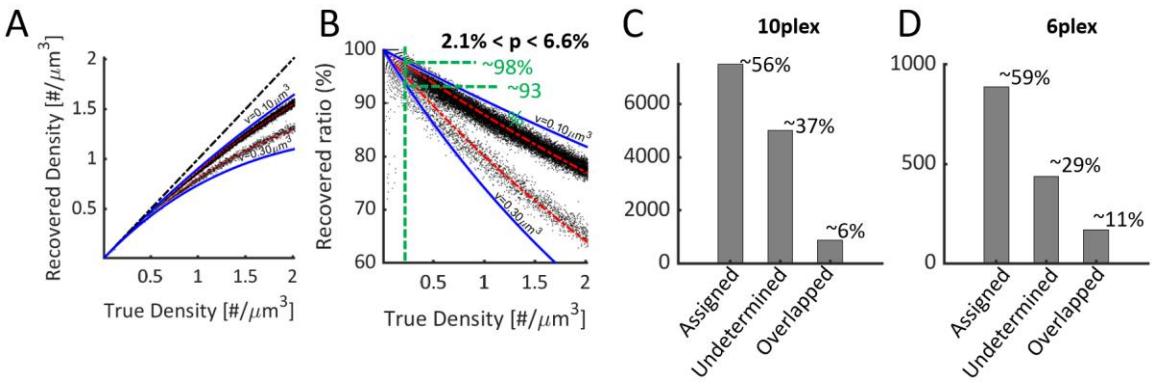


Figure 6.2. Overlapping and inconsistent signal simulations. Simulations are run at different densities to generate 3D image stacks and puncta are detected using our image processing pipeline. **A)** Recovered density by our system as a function of the true density for the limit considered puncta volumes (blue) and simulation trends (top trend for 1plex bottom trend for 10-plex). **B)** Percentage of recovered puncta for the 1plex (top trend, depicting ratio of underestimated due to overlap) and 10-plex (bottom trend, depicting ratio of inconsistent signal due to overlap), again between the limit considered puncta volumes (blue). **C,D)** Number of puncta assigned to a particular gene, undetermined puncta and overlapped puncta for the 10-plex experiment and 6-plex experiment.

The number of puncta detected for each target gene in each cell for the labeled 10-plex sample was plotted (Fig. 6.1H). Figure 6.1I plots the total number of detected puncta for the labeled 10-plex sample 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. To validate these puncta count, we compared them to matching RNA-seq data from the same cell type. Shown in Figure 6.1J 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 gene. When fitting with a straight line we obtained a Pearson correlation of r = 0.904, indicating a significant positive association between the two methods. The number of counts per cell was obtained by simply dividing the total number of detected genes into the number of cells

obtained by 3D segmentation of the individual nuclei stained with DAPI. The individual cell edges were estimated by growing the edges of the nuclei until convergence. A total of 80 cells were analyzed for this experiment.

6.3 6-plex Combinatorial Panel in Human FFPE Skin Tissue

We next investigated whether MOSAICA can provide multiplexed mRNA detection and phasor-based error-correction 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 [27-29]. 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 each gene to exhibit a unique fluorescent signature (Fig. 6.3A left).

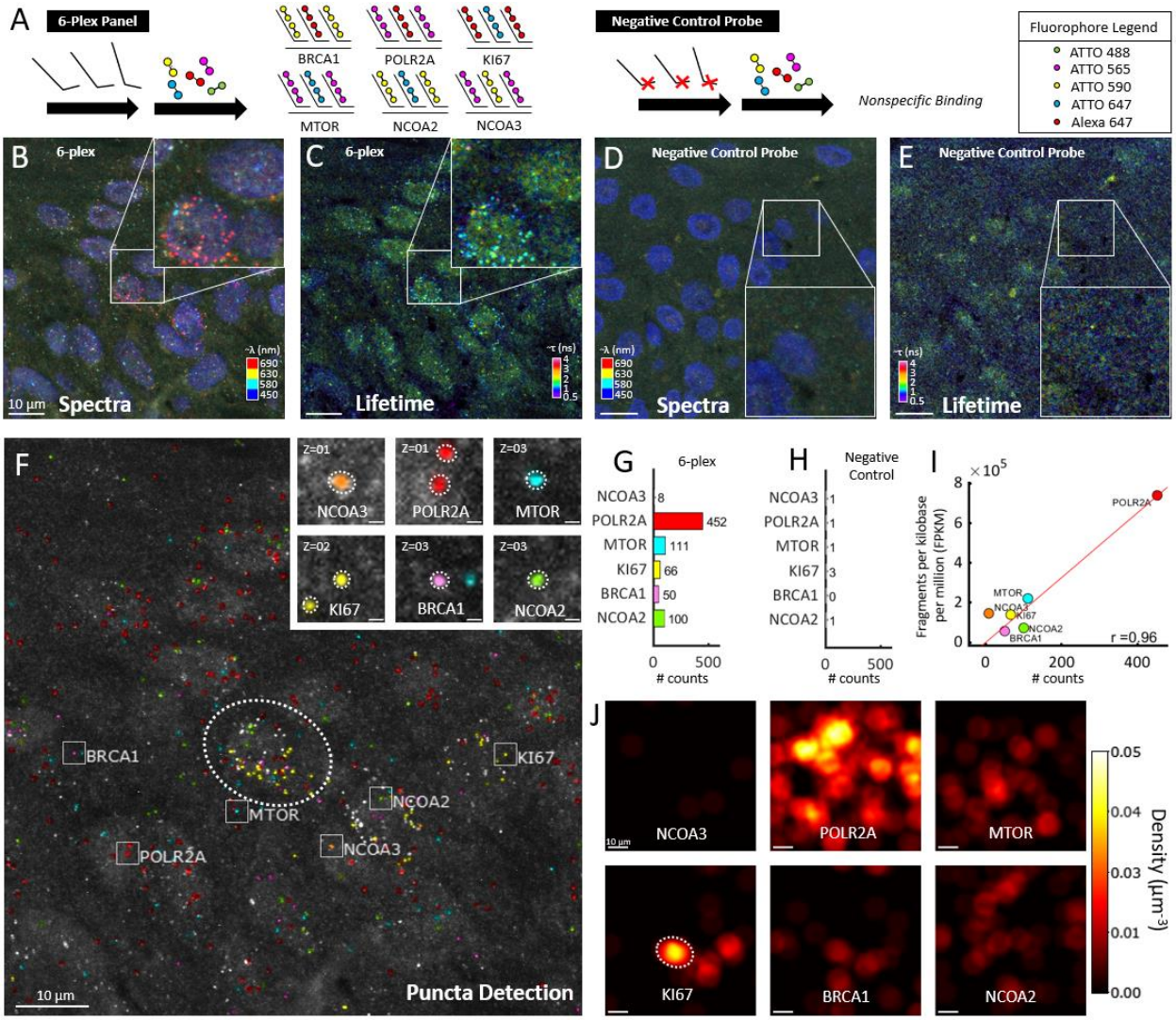


Figure 6.3. Multiplexed mRNA detection in epidermis region of human skin melanoma FFPE tissue. A) 6 different types of gene transcripts were labeled with primary probes followed by respective and complementary fluorescent secondary probes. Each gene was labeled with a combination of 2 different fluorophores for 10 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 10-plex sample (3 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 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). Scale bar is 10 µm in the main image and 1 µm in insets. G) Overall puncta count of each transcript in the 6-plex sample. H) Puncta count for the negative control probe sample. I) Correlation of detected puncta (mRNA puncta count) vs. bulk sequencing (fragments per kilobase per million

(FPKM)) 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.

Figure 6.3B 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 Figure 6.3C. Figure 6.3D/E depict the merged composite spectral and lifetime images of the corresponding negative probe sample also in the epidermis region. Figure 6.3F 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. Using this approach, we observed that a population of puncta consisting of nonspecific, autofluorescent, or unknown sample artifacts were rejected from analysis, (n = 438) or 29% of total detected clusters (Fig. 6.2D). In addition to this group, we observed MOSAICA rejecting a small group of puncta that emitted fluorescence in multiple spectral channels (n = 168). Since, this fraction (11%) exceeded the optical crowding range (2.1% to 6.6%) our simulations and model permits, we attribute this discrepancy to autofluorescence moieties which had fluorescent signatures that spans a broad emission spectrum and elicits a multispectral response. 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 [26,30]. With MOSAICA, these contaminating artifacts can be accounted for with the integration of spectral, lifetime, and shape-fitting algorithms.

Figure 6.3G/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. We observed MOSAICA platform results with the negative control sample obtaining counts of less than five per thousand. 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. Shown in Figure 6.3I is a scatter plot of mRNA puncta count for each cell plotted against fragments per kilobase per million (FPKM). We observed a Pearson correlation of r = 0.959 for this 6-plex sample, indicating a significant positive association between the two methods and the capability for MOSAICA to capture a wide range of differentially expressing markers. Lastly, as shown in Figure 6.3J, 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. 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 [31].

6.4 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 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 genes.

We generated the simulated images in a cubic space of 10x10x10 µm, discretized as an image stack of 33 images of 1000 x 1000 pixels (yielding a voxel resolution of 100x100x300 nm). This volume was generated containing increasing densities of transcripts, ranging from a single transcript of each gene (66 transcripts) up to 2,000 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 (Figure 6.1) where the genes are tagged with combinations of 2 out of 5 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 Figure 6.2).

6.5 Detection efficiency of Spectral-FLIM vs. LGC Stellaris vs. RNASCOPE

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 (Fig. 7.4). Using 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 ($p = 7.8x10^{-4}$ and $p = 3.4x10^{-4}$), 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 broad dynamic range of expression levels from single digit to hundreds of copies per cell.

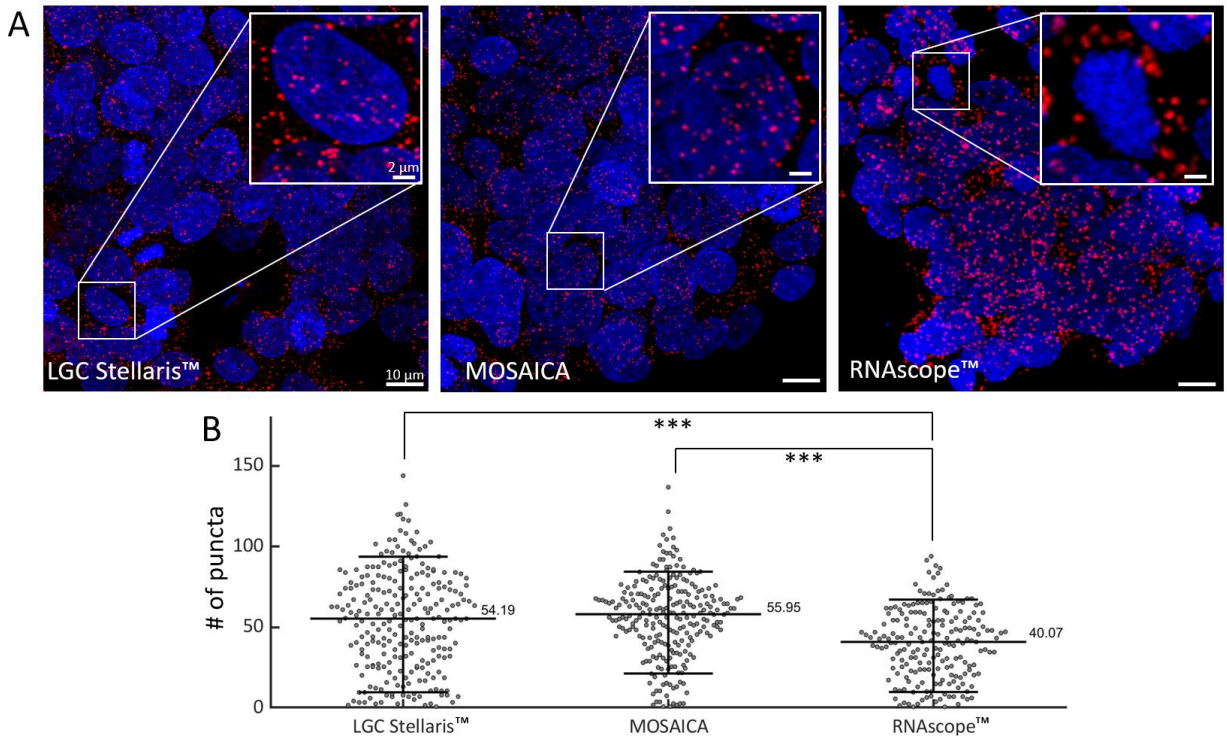


Figure 6.4. Benchmarking MOSAICA against RNAscope™ and LGC Stellaris™. A) POLR2A gene expression on colorectal cancer SW480 cells following RNAscope™, LGC Stellaris™ and MOSAICA protocols. Scale bars are 10µm. B) Puncta counts per cell volume between three platforms. MOSAICA exhibited comparable puncta 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 = 3.4×10^{-4} (LGC Stellaris™ vs RNAscope™ and p = 7.8×10^{-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.

6.6 Preliminary FLIM-FRET with Spectral-FLIM FISH

In addition to using different combinations of fluorophores that exhibit distinct spectral and lifetime signatures, preliminary FLIM-FRET experiments where different FRET pairs and their distances were be modulated to elicit changes in both spectral and lifetime were conducted (Figure 6.5). Using the same pair of fluorophores, this method demonstrates that distance

modulation can be a potential way to increasing multiplexing capabilities. As a complementary donor and acceptor pair are positioned closer together, the lifetime of the donor decreases with increasing quenching. This trajectory can be depicted in Figure 6.5A. To utilize this mechanism to increase multiplexing capabilities, mNeonGreen transcripts were targeted in 4 different conditions: 1) labeled with donor only and no acceptor, 2) labeled with FRET pairs separated by a distance of 25 bp or 7.5 nM, and 3) labeled with FRET pairs separated by a distance of 12 bp or 3.6 nM (Figure 6.5B). As expected, the sample labeled only with donor, had the closest lifetime to the donor in solution group (Figure 6.5C-D). The slight shift to the right appears to be the donor fluorophore blending in with the autofluorescence of the sample. As distance is decreased between the pair, from 25 bp to 12 bp, the donor exhibited quenching towards the right side of the phasor plot. This is a preliminary demonstration that this technique can be feasible and potentially scalable with more distances and FRET pairs. However, since not all regions are fully double-stranded, there remains many limitations for using this approach as a molecular ruler. Future work entails conjugated fluorophores at certain sites on the secondary probe to allow the conjoined strands to be double stranded everywhere.

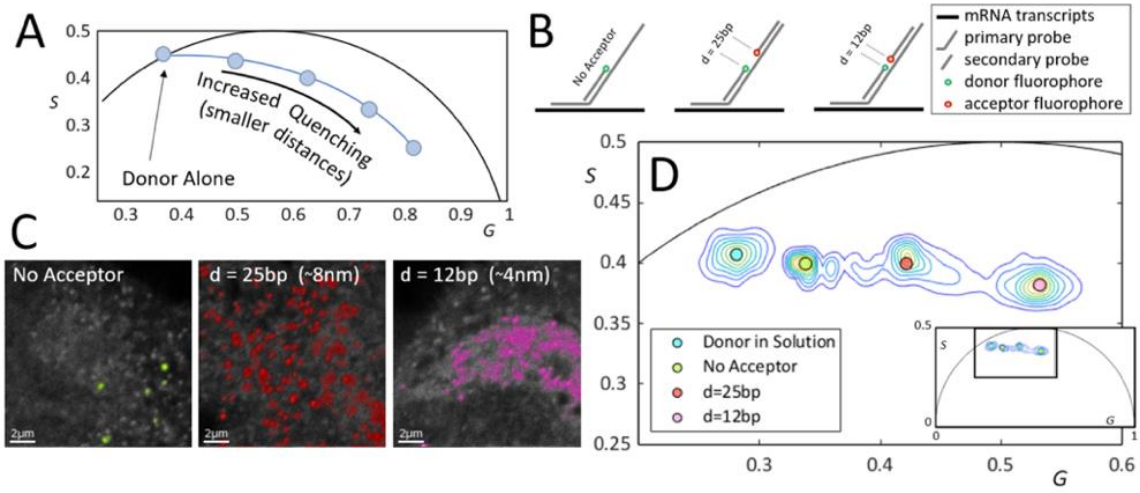


Figure 6.5. Preliminary lifetime measurements with FRET can be used for fluorescence barcoding/decoding. (A) The theoretical behavior on the lifetime phasor as one reduces the distance between the FRET probes. (B) Tagging mRNA transcripts with FRET probe pairs at different distances. (C) Real images of transcripts tagged with only the donor, and the probe pair at two different distances. (D) The phasor plot resolves the different cases of the images in panel C.

6.7 Detecting mRNA transcripts in highly scattering and auto-fluorescent tissues

As a preliminary demonstration to determine if this technology can be applicable to highly scattering and auto fluorescent tissues from different origins and prepared with different preservation methods, ubiquitin C (UBC) transcripts from mouse skin tissue preserved in optimum cutting temperature (OCT) medium and mouse colon tissue preserved in formalin fixed paraffin embedded (FFPE) medium were labeled with ALEXA 647 and detected with spectral-FLIM (Figure 6.6 A and C). A negative control sample labeled with no primary probes but with secondary fluorescent probes were ran in parallel (Figure 6.6 B and D). With spectral-FLIM, background tissue autofluorescence could be effectively subtracted to reveal distinct puncta which directly matched the encoded fluorescent signature.

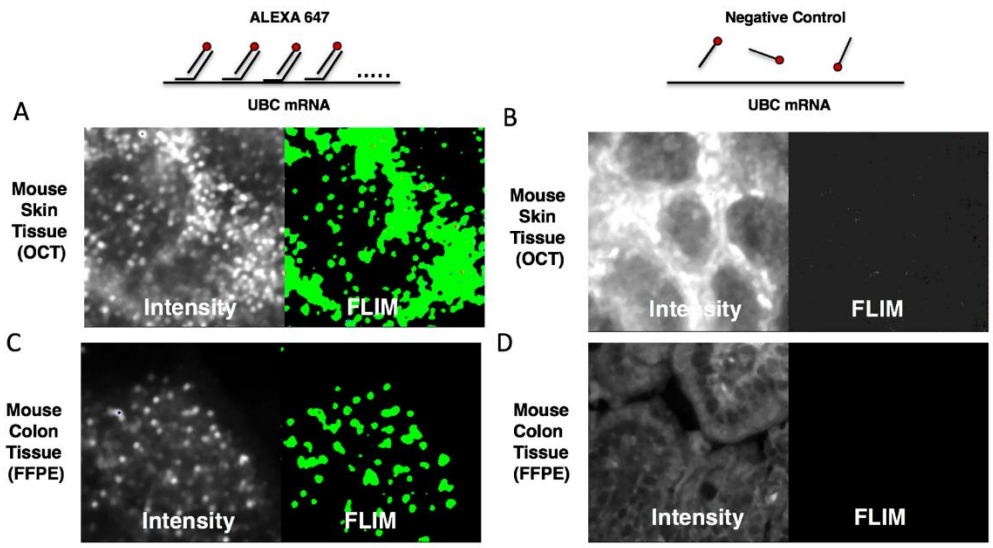


Figure 6.6. Detecting mRNA transcripts in highly scattering and auto-fluorescent tissues. (A) Mouse skin tissues preserved in OCT labeled with probes targeting UBC mRNA with Alexa 647. (B) A control sample labeled with secondary fluorescent probes only. (C) Similarly, mouse colon tissues preserved in FFPE exhibited clear and distinct UBC transcripts with spectral-FLIM. (D) A control sample with only secondary fluorescent probes.

Furthermore, POLR2A transcripts from human skin tissues preserved in formalin fixed paraffin embedded (FFPE) medium were detected and labeled with ALEXA 647 (Figure 6.7). Using standard intensity-based fluorescence microscopy, we could not differentiate between labeled puncta from autofluorescent moieties with similar SNR (Figure 6.7A - top). However, with spectral-FLIM, background tissue artifacts (red circles) could be effectively subtracted out to reveal distinct puncta (green circles) which directly matched the encoded fluorescent signature (Figure 6.67 - bottom). Probes containing a scrambled sequences targeting no mRNA region were also added to serve as a negative control to visualize the intrinsic autofluorescent moieties of each sample (Figure 6.7B).

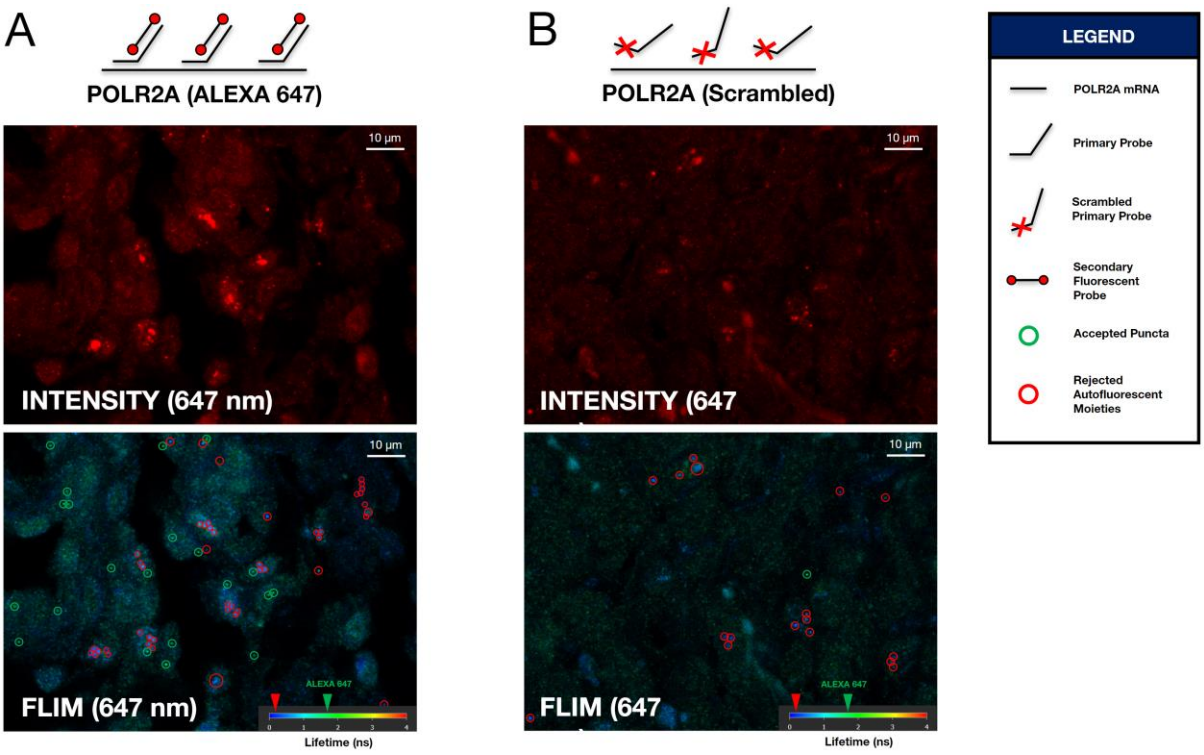


Figure 6.7. Error correction with Spectral-FLIM in highly scattering and autofluorescent tissues. (A) Human FFPE skin sections were labeled with probes targeting POLR2A with ALEXA 647. FLIM effectively discriminates the labeled puncta (green circles) against autofluorescent moieties (red circles) with similar SNR. (B) A scrambled control non-complementary towards POLR2A served as a negative control to highlight the diverse autofluorescent moieties which can be present in highly autofluorescent tissues.

6.8 Co-detection of mRNA and Protein in cells

Figure 6.8 shows an example of simultaneous detection of protein and RNA transcripts in the same cells using spectral-FLIM. To harmonize the FISH and IHC protocol, protein labeling was first conducted with RNAse inhibitors added to ensure that no confounding IHC reagent might degrade the mRNA prior to transcript labeling, e.g. serum or ascites derived antibodies. RNA labeling was then performed after IHC and then this sample was collectively imaged in one round.

For this experiment, SW480 cells were again used and labeled for the protein markers tubulin (ALEXA 488) and vimentin (TRITC) as well as for the mRNA markers POLR2A (ALEXA 647) and MTOR (ATTO 647) (Figure 6.8A-F). Figure 6.8H plots the number of puncta count found for the mRNA targets while Figure 6.8G depicts the merged images. These preliminary experiments demonstrate that this technology can work for spatial multiomic detection.

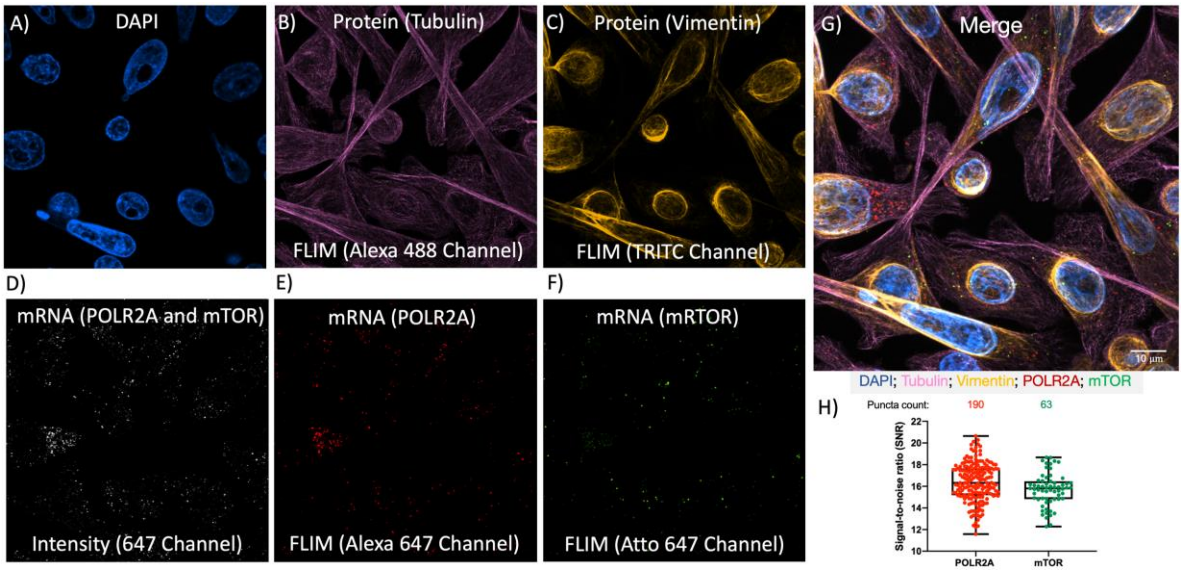


Figure 6.8. Simultaneous 4-plex co-detection of protein and mRNA in colon cancer SW480 cells using hyperspectral-FLIM. (A) The proteins Tubulin and Vimentin were labeled with TUBB4A mouse and VIM rabbit primary antibody, respectively. Secondary antibodies goat anti-mouse Alexa 488 and donkey anti-rabbit TRITC were then used respectively. The mRNA targets POLR2A and mTOR labeled with target specific primary probes were hybridized, then secondary probes with Alexa 647 and Atto 647 were hybridized to the primary probes, respectively. (B) Using spectral-FLIM, the two targets within the 647 nm spectral channel were separated. Signal-to-Noise and puncta count analysis was performed for the mRNA targets.

6.9 Methods

6.9.1 Primary Probe Design

A set of primary probes were designed for 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 CDS region. However, if that region does not provide over 40 probes, the exons from the CDS and UTR 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 genes in the panel to eliminate off-target hits. For this experiment, each gene had the maximum number of probes that could be designed with our pipeline and requirements. The final primary probe design included 2 assigned readout sequences of secondary probe with a “TTT” connector in between, another connector, then one of the probes specific for the gene. The primary probes were ordered from Sigma Aldrich and pooled together for each gene.

6.9.2 Secondary probe design

Secondary probe structures were based on the design from the Zhuang group [8]. 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 nt 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 [32]. Probes were then checked with BLAST to identify and eliminate sequences with contiguous homology regions longer than 11 nt to the human transcriptome. From the readout sequences satisfying the above requirements, 16 were selected.

6.9.3 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 adenocarcinoma (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). The cells were plated into

8-well chambers and then fixed. The 8-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℃ with 5% CO_2.

6.9.4 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.

6.9.5 Preparation of fixed cells in cell chambers

When the cells reached 70% confluency, cells were fixed for 30 minutes 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 minutes and washed with PBS 3 times. 0.5% Triton X-100 (T8787-100ML; Sigma-Aldrich) in PBS was incubated in each well for 5 minutes 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℃.

6.9.6 Preparation of FFPE tissues

The University of California Irvine 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 hour. For antigen unmasking, slides were deparaffinized, rehydrated then followed by target retrieval (with citrate buffer).

6.9.7 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℃ for 8 minutes and then at 37℃ for 7 minutes. Following this pre-block step, primary probes with 5nM of each probe in blocking buffer were added to the samples and incubated at 60℃ for 30 minutes and then overnight at 37℃.

6.9.8 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 (EC)) is used for 3 washes and incubated in 60℃ for 5 minutes each time. Blocking buffer is added and incubated at room temperature for 5 minutes. The sample is then incubated in a solution with 5 nM of the secondary probes in blocking buffer at 37℃ for an hour. The sample is washed with 2x SSCT twice before using wash buffer to wash 3 times and incubated in 42℃ for 5 minutes 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.

6.9.9 Codetection of protein and 9RNA

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℃. 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 hour. 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.

6.9.10 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.

6.9.11 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.

6.9.12 Microscopy Imaging

Our samples can be imaged with any instrument provided that it has spectral and lifetime acquisition capabilities. Our measurements were taken on two separate instruments; a generic FLIM instrument is depicted in Supplementary Figure 1. 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 (78MHz) 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, 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, 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 4K resolution with 6.5 µm pixels.

6.9.13 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:

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

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 depends on the point spread function of the instrument. Using the detected number of counts n=13.5k and the estimated cellular volume 68 kμm^3, both obtained from the two image stacks, 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 between 0.1 and 0.3 μm^3 we obtained an estimated percentage of overlapping puncta in the interval [2, 6.6]%, which accounts for the number of puncta detected in more than 2 channels in the 10-plex experiment (6%). See Figure 6.2.

6.9.14 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 gene. For the human skin melanoma FFPE tissue, the patient sample did not have corresponding sequencing data. RNA sequencing data was obtained 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 (FPKM). The average of the 5 patient samples for each transcript were used for correlation graphs with Spectral-FILM puncta count.

6.9.15 Statistical Analysis

When comparing distributions of puncta counts, signal-to-noise ratios, and intensity values, Student t-tests were performed for 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.

Phasor Transform Expressions

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

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

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

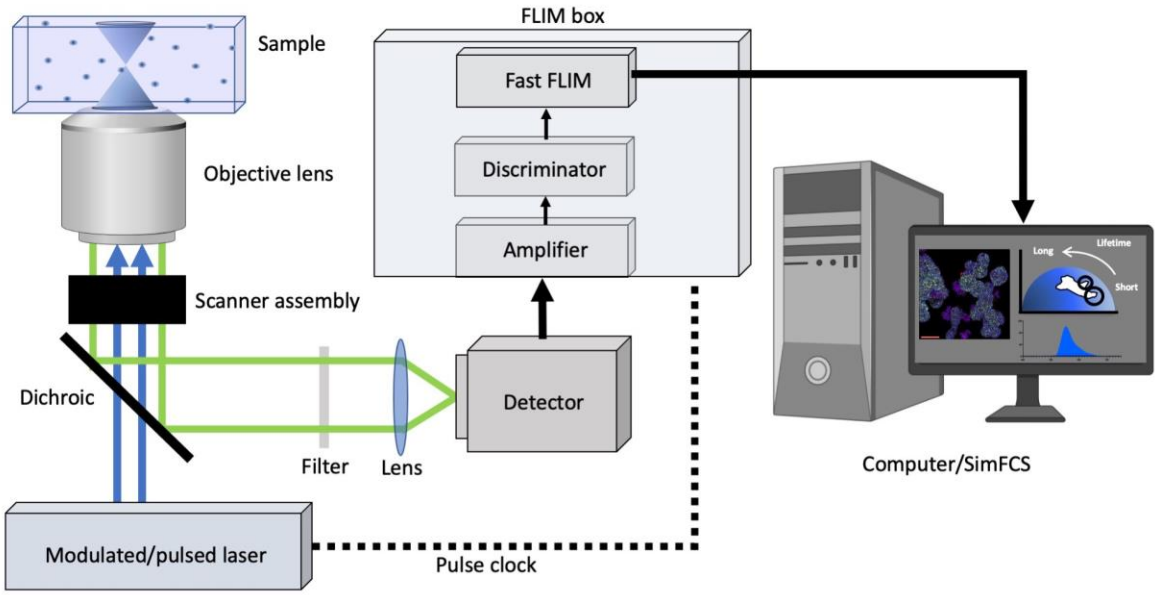
Where T is the period between excitation pulses (or modulation period), and $\omega = \frac{2\pi}{T}$ the pulsation frequency such that the period of trigonometric functions matches the excitation period T.

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

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

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

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



Supplementary Figure 6.1. 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).

Table S6.1. 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 S6.2. List of Genes Used and their Assigned Fluorophore Combination

Gene	Fluorophore 1	Fluorophore 2	No. of Probes per Gene
NCOA3	ATTO 565	ATTO 590	80
POLR2A	ATTO 488	ALEXA 647	80
MTOR	ATTO 565	ATTO 647	80
MKI67	ATTO 647	ALEXA 647	80
BRCA1	ATTO 590	ALEXA 647	80
NCOA2	ATTO 590	ATTO 647	80
BRCA2	ATTO 488	ATTO 565	80
CENPF	ATTO 488	ATTO 590	80
CKAP5	ATTO 565	ALEXA 647	80
NCOA1	ATTO 488	ATTO 647	80
MNEONG	ALEXA 647	N/A	14

6.9 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. *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 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).

13 Kalra, J. & Baker, J. in *Signal Transduction Immunohistochemistry* 237-251 (Springer, 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, doi:10.1073/pnas.1617699113 (2016).

15 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).

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 Diagn* **14**, 22-29, doi:10.1016/j.jmoldx.2011.08.002 (2012).
19 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).
20 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).
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, doi:10.1038/s41592-019-0404-0 (2019).
22 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
23 Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* **38**, 586-599 (2020).
24 Becker, W. Fluorescence lifetime imaging–techniques and applications. *Journal of microscopy* **247**, 119-136 (2012).
25. Chen, G. T. *et al.* Disrupting ß-catenin dependent Wnt signaling activates an invasive gene program predictive of colon cancer progression. *bioRxiv*, 667030 (2019).
26. 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).
27. 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, doi:10.1200/jco.2005.02.0735 (2005).
28. 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).
29. Allam, M., Cai, S. & Coskun, A. F. Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. *NPJ precision oncology* **4**, 1-14 (2020).
30. Trcek, T., Lionnet, T., Shroff, H. & Lehmann, R. mRNA quantification using single-molecule FISH in Drosophila embryos. *Nat Protoc* **12**, 1326-1348, doi:10.1038/nprot.2017.030 (2017).
31. Khodaeiani, E. *et al.* Immunohistochemical evaluation of p53 and Ki67 expression in skin epithelial tumors. *Indian J Dermatol* **58**, 181-187, doi:10.4103/0019-5154.110824 (2013).
32. 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. *Proceedings of the National Academy of Sciences* **116**, 19490-19499, doi:10.1073/pnas.1912459116 (2019).

CHAPTER 7:
CONCLUSION AND FUTURE WORK

7.1 Summary and Conclusion

MOSAICA provides a new solution that fills a major gap in the spatialomics field by simultaneously achieving both simplicity and multiplexing by enabling direct *in situ* spatial profiling of many biomarkers in a single round of staining and imaging. This solution contrasts existing approaches such as conventional direct labeling approaches (e.g., LGC Stellaris™, RNAscope™, etc.) which can only analyze 3 - 4 targets due to limited spectral channels as well as recent spatial transcriptomic approaches (e.g., seqFISH) which can provide greater multiplexing capabilities but at the expense of many iterations of sample re-labeling, imaging, indexing, and error-prone image registration. MOASICA accomplishes this by uniquely integrating the lifetime dimension with the conventional spectral dimension, employing combinatorial fluorescence spectral and lifetime target encoding, and exploiting machine learning- and phasor-based deconvolution algorithms to enable high-plex analysis and error-correction. This enables MOSAICA to simultaneously access both spectral and lifetime dimensions to achieve higher plex detection without sacrificing assay throughput as well as correcting for autofluorescent moieties and stochastic nonspecific binding artifacts, which are inherent challenges associated with existing intensity-based measurements.

Compared to existing sequential hybridization and imaging approaches, MOSAICA significantly reduces the number of hybridization and imaging rounds required to profile larger multiplexed panels of RNA biomarkers. By imaging every fluorophore and labeled transcript simultaneously rather than sequentially, the need to relabel probes or reimaging samples is

precluded which is often the bottleneck for these complex and lengthy assays. For a 10-plex panel, this drastically shortens assay time by at least of factor of 4 – 5x and scales up even more as the panel becomes more multiplexed, making these assays more practical, effective, and affordable for routine laboratory or clinical usage. This is particularly important in clinical settings, where biopsy samples are limited in quantity and the need to avoid repetitive stripping, rehybridization, and image registration to preclude damage to the tissue structure integrity and target molecules is highly critical. In terms of cost, compared to existing commercial platforms where assay ranges from several hundred dollars to thousand dollars per batch of samples, MOSAICA utilizes inexpensive DNA primary probes which can be purchased in batch or as microarrays for a minimal price. Fluorescently conjugated secondary probes can also be used and shared as a common set among many different genes, scaling down costs to several dollars per assay or 10 – 20x less than commercial vendors. Compared to indirect spatial transcriptomic methods that analyze barcoded regions of interest (ROIs) using downstream sequencing (e.g., GeoMx® Digital Spatial Profiler, 10x Genomics Visium), the direct, imaging-based MOSAICA would be advantageous for targeted mRNA profiling with higher spatial resolution (subcellular features or single molecules), simpler workflow, lower cost, and potentially higher throughput (number of tissue sections analyzed per day) with a camera-based system (below). Moreover, the MOSAICA platform utilizes standard fluorescent probes and fluorescence microscopy (several commercial instruments capable of acquiring both spectral and lifetime information are available, e.g., Leica SP8 FALCON, PicoQuant rapidFLIM, ISS FastFLIM, etc., and already exist in many shared core facilities in academia and industry). Fluorescence imaging remains the most

familiar and widely used technique in biological research and its minimal requirements of MOSAICA will permit quick and broad adoption in the scientific community.

As a complementary tool to existing spatial-omics methods, MOSAICA broadly enables scientists and clinicians to better study biology, health, and disease and to develop precision diagnostics and treatments. Gene expression is heterogeneous and many different cell identities/states exist in a particular tissue. To fully characterize cells *in situ,* we need to be able to assess multiplexed panels of genes within the same cell, which can be readily addressed by MOSAICA. As such, MOSAICA can help accelerate spatiotemporal mapping experiments to construct 3D tissue cell atlas maps. Additionally, MOSAICA represents a powerful tool for targeted *in situ* validation of RNA sequencing data. scRNAseq returns cell identities in the form of "differentially expressed gene lists" that "define" cell types. However, the clustering process is subjective, variable and error-prone. Multiplex spatial transcriptomics using MOSAICA can validate whether a pattern of gene expression really defines a cell type or conflates multiple cell types. Furthermore, we are currently developing MOSAICA to serve as a clinical companion diagnostic tool for stratified care. Of particular interest is the spatial profiling of the organization and interactions between tumor cells, immune cells and stromal components in tumor tissues that can inform cancer diagnosis, prognosis, and patient stratification [1-2]. With MOSAICA's ability to analyze numerous markers in a single round, we believe this feature will be particularly attractive for clinicians or researchers working with precious patient derived materials.

Further ongoing work for this platform includes improving multiplexing by scaling up the number of fluorophores and their combinations using our combinatorial-based labeling approach. For instance, our combinatorial encoding results and simulations predict that with

different combinations of 10 fluorophores using triple labeling strategy, MOSAICA can barcode and detect 120 mRNA targets. Our immediate next step is to scale our multiplexing capability to about 60-plex, which will bridge the gap between conventional FISH and sequential- and sequencing-based methods and can be rapidly employed to numerous applications in targeted spatial transcriptomic analysis in basic and translational research. In the future, we will further expand our codebook by implementing a Förster resonance energy transfer (FRET)-based barcoding method [3] by using different FRET fluorophore pairs and tuning the distance between fluorophore donor and acceptor. We can utilize the FRET phenomena as an additional error correction mechanism at the nanometer level to resolve multiple transcripts in the same voxel. Moreover, we will improve imaging throughput with our recently developed camera-based light-sheet imaging [4] and hyperspectral imaging [5]. Indeed, MOSAICA is versatile and can be further integrated with other imaging modalities such as super-resolution, expansion microscope and multi-photon imaging techniques [6-8] to increase subcellular resolution and enable imaging large highly scattering tissues.

Additionally, we will further develop our image analysis software with a user interface which enables classification and visualization of single-cell phenotypes, spatial organization and neighborhood relationship. Our current software platform's puncta detection and puncta classifier algorithm can also be improved using modern image processing techniques such as convolutional neural networks with clinical training sets to optimize true and false positive detection of biomarkers. We anticipate on benchmarking these sets with experimental data with stripping methods to determine fidelity of labeled targets. Lastly, we will further improve protein

multiplexing in our multi-omics analysis using antibody-DNA conjugates where we can conduct combinatorial labeling and barcoding to scale up multiplexing as we do for the mRNA detection.

7.2 References

1. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615-620 (2020).
2. 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).
3. Dagher, M., Kleinman, M., Ng, A. & Juncker, D. Ensemble multicolour FRET model enables barcoding at extreme FRET levels. *Nature nanotechnology* **13**, 925-932 (2018).
4. Hedde, P. N., Malacrida, L., Ahrar, S., Siryaporn, A. & Gratton, E. sideSPIM–selective plane illumination based on a conventional inverted microscope. *Biomedical Optics Express* **8**, 3918-3937 (2017).
5. Scipioni, L., Rossetta, A., Tedeschi, G. & Gratton, E. Phasor S-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging. *Nature Methods* **18**, 542-550 (2021).
6. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. *Nature methods* **15**, 173 (2018).
7. Crosignani, V. *et al.* Deep tissue fluorescence imaging and in vivo biological applications. *Journal of biomedical optics* **17**, 116023 (2012).
8. Crosignani, V., Dvornikov, A. S. & Gratton, E. Enhancement of imaging depth in turbid media using a wide area detector. *J Biophotonics* **4**, 592-599, doi:10.1002/jbio.201100001 (2011).