In situ signal amplification for spatial transcriptomics using programmable DNA assemblies

Thesis by

Katsuya Lex Colón

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2025

(Defended March 21st, 2025)

Katsuya Lex Colón

ORCID: 0000-0002-7347-6128

ACKNOWLEDGMENTS

I would first like to express my gratitude to my advisor, Long Cai, for his unwavering support and invaluable expertise. Under his mentorship, I have grown immensely as a scientist. His guidance has challenged me in many ways, encouraging me to think critically and to embrace failure as an essential part of the scientific process. He taught me to approach experiments boldly, even when venturing into areas outside my expertise. Through this, I learned to draw meaningful conclusions and to think beyond conventional approaches or generally accepted perspectives in the field. Under his supervision, I learned how to truly think like a scientist. I am also deeply grateful to Lior Pachter for his unwavering support and guidance throughout my graduate studies. He provided much-needed encouragement during my most challenging moments, and he taught me a great deal about scRNA-seq analysis when I had no background in data analysis or programming. His dedication continues to inspire me. I would also like to extend my thanks to my committee members, Mikhail Shapiro, Rustem Ismagilov, and Matt Thomson for their support and guidance.

Much of the work I have done would not have been possible without the contributions of current and past members of the Cai lab. I would like to thank Chee-Huat Linus Eng for being my mentor. I value the discussions we shared and his significant contributions to the LANTERN project presented in Chapter II. Furthermore, I would also like to thank Chuqi Lu for her work and significant contributions to LANTERN in Chapter II. Additionally, I am grateful to Jina Yun and Saori Lobbia for their support with experiments. I would also like to thank Chris Cronin for his invaluable expertise in engineering to create various devices that have contributed significantly to all experiments. I cherish our discussions about music and science, and he has taught me much about the general principles of mechanical engineering to create devices of my own in the future.

Graduate school has been a challenging journey, and I could not have done it without the support of my friends and colleagues. I would like to thank John Thompson for being both a mentor and a friend. I am grateful for the patience he had for me during my first year of graduate school. I am also thankful for my friends Andrew Schacht, Mackenzie Strehle, Ariana Hardy, and Tess Ricciutti. I cherish the many evenings we spent hanging out, talking about science and life, drinking wine, exploring SoCal, finding new restaurants, playing chess (especially with Andrew), and much more. Their companionship kept me grounded, and they became like family to me. I would like to extend my heartfelt thanks to my dear friend, Terry Kim. Our friendship began on the very first day of graduate school and has continued to grow through the many rewarding and challenging experiences it has brought. Terry has aided me through every hardship and has celebrated alongside me during the most memorable moments of graduate school. I've always enjoyed our conversations about science, and his thoughtful advice over the years has been invaluable. I hold him in the highest regard and will always consider him a part of my family.

I would like to extend my deep gratitude to Carsten Tischbirek. Over the past two years, we have worked closely together, sharing ideas and accomplishing meaningful scientific work. His approach to research has broadened my perspective and shown me new ways of thinking about and conducting science. Carsten has also been a source of support during some of the most challenging times in graduate school, encouraging me to keep going when I considered stepping away. Outside of science, we've shared many memorable moments, enjoying good food and great wine, hiking the SoCal mountains, and going on many astrophotography trips. I look forward to collaborating further in science and continuing our friendship in the years ahead.

I would like to thank my family for shaping me into the person I am today. Being the first in my family to graduate from college and go as far as earning a doctorate would not have been possible without their unwavering support and encouragement to pursue my dreams. My father, Eduardo Colón, has been a profound source of inspiration. He has always reminded me that I can accomplish anything and emphasized the importance of academic success as a foundation for building a fulfilling career. I am deeply grateful for all that he has done for me. My mother, Ayano Donnelly, has also been an incredible source of support throughout my graduate school journey. I deeply admire her relentless drive to pursue her goals and her determination to excel in whatever she does. She has always been independent and a go-getter, qualities I greatly respect. Her bravery in embracing life and its challenges inspires me, and she has been a powerful role model, encouraging me to follow my dreams and live boldly. I would also like to extend my heartfelt thanks to my brothers, Jay Colón and Chase Donnelly. I like to think we share a bond closer than most siblings, and I truly cherish every moment we spend together. As the oldest sibling, I strive to lead by example and show them that they can take on any academic challenge. I hope to make them proud and inspire them to pursue their goals with unwavering determination and passion.

Finally, I am profoundly grateful to my wife and soulmate, Madeline Vera-Colón. She has been by my side through every challenge, the stressful moments, and the times I doubted myself. Her unwavering support has been invaluable. As a fellow scientist, she has also been an incredible source of insightful discussion and inspiration. I could not have accomplished this without her. She has always believed in my potential. I eagerly look forward to future collaborations where we can merge our expertise and passions. Additionally, I extend my heartfelt thanks to my cats, Myelin and Helix, for their emotional support.

ABSTRACT

Sequential Fluorescent *In Situ* Hybridization (seqFISH) has been an invaluable tool in imaging-based spatial transcriptomics, aiding researchers in elucidating spatially-resolved, gene expression patterns in intact tissues and cell culture models. However, methods that rely on smFISH, such as seqFISH, suffer from poor signal-to-noise ratio in certain tissue types or target RNA, require many fluorescently labeled RNA targeting probes which prohibits imaging of small RNA species, and exhibit poor sample throughput due to the need of high magnification objective or long exposure times. Herein, we develop solutions to these limitations by developing and utilizing a robust signal amplification strategy. While various amplification technologies exist, their limitations often hinder broad applicability. Moreover, we desire an amplification platform that is amenable to the denaturing wash conditions used in seqFISH. We will begin Chapter I by discussing the background, technical challenges, and utility of various *in situ* signal amplification technologies. Chapter II details the exploration and technical limitations of rolling circle amplification (RCA) and branched DNA (bDNA) assembly utilizing ssDNA padlock amplifier strands. Chapter III discusses the design and development of a novel amplification strategy called Signal amPlicAtion by Recursive Crosslinking (SPARC), which builds upon the knowledge gained from Chapter II. We highlight SPARC as a unique photochemical signal amplification method that iteratively deposits amplifier strands near the primary probe target for linear signal amplification. Then, the deposited amplifier strands act as a scaffold for branched DNA assembly, leading to an exponential signal amplification. Through each deposition and assembly step, amplifier strands are photo-crosslinked to the extracellular matrix, forming highly stable DNA nanostructures that can withstand harsh denaturing wash conditions. We demonstrate the utility of SPARC in amplifying signal of both single-molecule transcripts and proteins.

TABLE OF CONTENTS

NOMENCLATURE

bDNA	branched DNA
bp	base pair
cDNA	complementary DNA
CuAAC	Cu-assisted Azide Alkyne Cycloaddition
DBCO	Dibenzocyclooctyne
dUTP	Deoxyuridine triphosphate
FISH	Fluorescent *In Situ* Hybridization
HCR	Hybridization Chain Reaction
mRNA	messenger RNA
nts	nucleotides
RCA	Rolling Circle Amplification
rDNA	ribosomal DNA
rRNA	ribosomal RNA
RT	Reverse Transcription
rGRFT	Recombinant *Griffithsia* sp.
seqFISH	sequential FISH
smFISH	single-molecule FISH
ssDNA	single-stranded DNA

Chapter I

Brief history of *in situ* amplification technologies for spatial transcriptomics

1.1 Overview of imaging-based spatial transcriptomic methods

In situ hybridization (ISH) methods have become the standard approach for profiling various genes in intact cells while preserving their native spatial context. The earliest report of ISH was in 1969 by Gall. J. G. *et al.* and John. H.A. *et al.*, where they generated tritium-labeled rRNA probes to target rDNA in various tissues and cell culture models.[1,2] They demonstrated not only that stable RNA-DNA hybrids can form *in situ* but also that these probes exhibit high specificity. For instance, they discovered that rRNA probes generated from one species do not hybridize with rDNA targets from another species.[1,2] Moreover, Harrison. P.R. *et al.* demonstrated that tritium-labeled cDNA from 9S RNA in mouse reticulocyte can be used to detect globin mRNA *in situ* from erythroid cells in 1973.[3] Since their demonstrations, there has been numerous iterations on improving ISH methods, including novel ways for signal detection which includes fluorometric and colorimetric readouts.[4,5] However, most fluorometric and colorimetric approaches rely on the enzymatic deposition of small molecules, which can diffuse away, or the accumulation of antibodies on the target, both of which fail to provide high spatial resolution of the target within cells.

In 1998, Femino. A.M. *et al.* demonstrated for the first time that single-molecule transcripts can be detected via fluorescence microscopy by using 5 unique ssDNA probes, each conjugated to multiple fluorophores. These probes hybridize to the same RNA target, generating sufficient signal to visualize diffraction-limited spots.[6] This is defined today as single-molecule FISH (smFISH). In 2008, seminal work by Raj. A. *et al.* improved upon this method, showing that 48 unique single-dye-modified ssDNA probes targeting the same RNA can provide better signal quality using fluorescence microscopy.[7] Although this appeared to be a minor improvement, the utilization of numerous single-dye modified ssDNA probes provided several advantages compared to the multi-fluorophore

conjugated probes. For instance, having multi-fluorophore conjugated ssDNA can provide some variability in signal if 1) the synthesized oligonucleotide is heterogeneously modified and 2) large accumulation of multi-dye conjugated ssDNA probes to a target leads to self-quenching.[7] Finally, if the multi-dye conjugated probe binds non-specifically, it will be difficult to distinguish true signal from false signal considering there is already variability in signal intensity.[7] The work of Raj. A. *et al.* offered a more robust method for performing smFISH while also simplifying the synthesis of detection probes.

Once smFISH became a widely adopted method, scalability emerged as a challenge for detecting a larger number of unique transcripts or DNA loci. In traditional smFISH, one can profile multiple genes in a single imaging round if they assign certain fluorescent channels to a single target. However, to visualize additional genes, the existing signal must first be removed through methods such as photobleaching,[8] chemically cleaving fluorophores,[9] enzymatic digestion,[10] or destabilization of bound fluorescent probe using formamide,[11] followed by hybridizing a new set of probes. Therefore, the number of genes that can be visualized scales linearly as $F \times N$, where F is the number of distinguishable fluorophores, and N is the number of cycles involving the removal of existing signals and the addition of new ones, which can be time-consuming.[12] Furthermore, each target will require a unique detection probe which can be costly and difficult to design without off-target binding. To overcome this problem, a combinatorial smFISH-based method was developed by Lubeck *et al.* in 2014 that scales F^N by utilizing temporal barcodes that are assigned to each gene.[13] In practice, there is a sequential addition and removal of fluorescent signal targeting genes of interest. During each cycle, diffraction-limited spots appear in a specific sequence. Once decoded, this sequence reveals which gene corresponds to each spot. This method, termed sequential FISH (seqFISH), solved the scaling challenge and was further improved in 2016 by Shal. S. *et*

al. to enable error-detectable codes which utilizes checksum.[10] To summarize, seqFISH substantially reduced the number of hybridization cycles and the number of unique detection probes required to profile thousands of genes. Aside from seqFISH, Chen. K. H. *et al*. in 2015 developed a similar combinatorial smFISH approach termed multiplexed error-robust FISH (MERFISH), which utilizes a modified hamming code system.[8] Unlike seqFISH, MERFISH employs barcodes designed with error-detection and error-correction capabilities, ensuring that each barcode maintains a Hamming distance of 4 from all others. Moreover, they utilize hamming weights which ensures that each barcode will need to be called *N* times.[8] Both technologies have been able to demonstrate scalability by profiling 10,000 unique transcripts in intact cells. More recently, seqFISH was able to profile ~100,000 unique DNA loci along with ~18,000 unique introns.[14] The development of seqFISH served as a catalyst that jump-started the field of imaging-based spatial transcriptomics and genomics. This breakthrough led to the creation of numerous spatial technologies, enabling researchers to delve deeper into cellular communication by examining the transcriptional or genomic states of cells within their native spatial context.

1.2 Imaging-based vs sequencing-based spatial transcriptomics

The most notable technologies for spatial transcriptomics fall into two major categories: imaging-based or sequencing-based. While many sequencing-based platforms exist, Visium, STOmics, and Curio Seeker are the most widely commercialized technologies.[15] Sequencing-based platforms have garnered unprecedented attention due to several advantages: (1) ease of use, (2) lower costs, (3) ability to circumvent signal detection challenges faced by imaging-based platforms, (4) robustness across diverse sample types, (5) smaller data file sizes, (6) enables whole transcriptome profiling, and (7) compatibility with existing bioinformatics tools for preprocessing. These technologies generally utilize substrates with spatially defined areas containing poly-T capture

sequences or probe-specific capture sequences.[15–22] Tissues are placed on these substrates, allowing mRNA or probes to be captured through diffusion or electrophoretic transfer.[19] However, in most cases, mRNA can diffuse laterally from its original location, ranging from a few microns to hundreds, potentially mapping to a different cell and compromising spatial accuracy.[15–17,19,20,23] Among the various platforms, Visium HD and STOmics experience the least transcript diffusion during the transfer of probes or RNA onto their capture substrates.[15,20] However, the detection efficiency of these capture platforms is estimated to be well below 5%, significantly hindering the identification of biologically relevant genes.[15,16,20–23] Furthermore, these platforms do not provide subcellular information since the transcript can be anywhere in the spatially-defined block. Moreover, in some cases, the spatially defined blocks can encompass multiple cells which may be difficult to deconvolute. Typically, cell types are inferred by analyzing the abundance of specific marker genes enriched within the capture area.[16,17] Finally, sequencing-based platforms do not support thick tissue processing, resulting in the loss of information regarding cell-to-cell communication along the z-axis. Imaging-based spatial transcriptomics, unlike sequencing-based platforms, enables RNA detection and measurement *in situ*, avoiding issues related to lateral diffusion. Furthermore, imaging-based platforms support the analysis of thick tissues, offer true single-cell resolution, reveal the subcellular localization of biomolecules, and obtain detection efficiencies greater than 90%.[17,24]

1.3 Rise of *in situ* amplification technologies

As mentioned in the previous sections, smFISH is a powerful method but presents several limitations. These include a low signal-to-noise ratio in certain tissue types or for some target RNAs, reliance on multiple probes for sufficient signal (which can complicate the detection of small RNA species), and low sample throughput resulting from the need

for high magnification objectives or prolonged imaging times. To address these challenges, three distinct *in situ* amplification approaches were developed, which include Rolling Circle Amplification (RCA), Hybridization Chain Reaction (HCR), and branched DNA (bDNA). RCA was first introduced by Liu. D. *et al.* in 1996, which highlights the use of Klenow fragment of DNA Polymerase I to isothermally amplify small, circularized DNA templates.[25] Following that discovery, in 1998 Lizardi P.M. *et al.* introduced a seminal method for *in vitro* allele discrimination by using cyclizable ssDNA padlock probes and RCA with phi29 DNA polymerase to amplify a specific genomic locus.[26] Moreover, they demonstrated single-molecule counting by immobilizing wild-type and mutant DNA loci on a modified glass surface then performing RCA. Finally, they showed that RCA can be performed *in situ* in intact cells for the detection of a specific genomic locus.[26] This work inspired the development of numerous RCA-based methods for *in situ* detection of single biomolecules. One such method is the work of Ke. R. *et al.* in 2013, which enables RNA target detection by employing *in situ* RT and ssDNA padlocks designed to target specific cDNA products. These padlocks are ligated and amplified using RCA, followed by sequence-by-ligation to identify the targets.[27] Building on this approach, by Lee. J.K. *et al.* introduced Fluorescent *In Situ* SEQuencing (FISSEQ) in 2014, which employed *in situ* RT to generate cDNA and attach sequencing adapters to its target, followed by CircLigase to circularize the cDNA and performing RCA.[28] Once multiple copies of cDNA are generated *in situ*, sequencing primers are used to iteratively identify dinucleotide sequences which are then mapped to the transcriptome. Although this work had many technical limitations that significantly hindered detection efficiency of RNA molecules and was laborious to perform, FISSEQ demonstrated a proof-of-concept for whole transcriptome profiling in intact cells. The limiting step for the methods mentioned is the use of *in situ* RT, which is presumed to be highly inefficient.[28–30] To address this limitation, Deng R. *et al*. developed

a method using ssDNA padlocks that target RNA directly, along with the enzyme SplintR to ligate the RNA-DNA hybrid to form the circularized padlock enabling RCA.[29] Following this work, Wang. X. *et al*. introduced Spatially-resolved Transcript Amplicon Readout mapping (STARmap) in 2018.[30] STARmap employs a set of unique probes called Specific amplification of Nucleic Acids via Intramolecular Ligation (SNAIL). These SNAIL probes feature an inverted ssDNA padlock design combined with a proximity probe, enabling padlock ligation and providing a primer site for RCA initiation only when both probes are present. This probe pair design not only bypasses *in situ* RT but also ensures high specificity toward the target.[30] Although the two methods mentioned previously circumvents the need for *in situ* RT, neither enables unbiased, whole-transcriptome profiling. Despite numerous advancements in RCA-based methods, a common challenge is their low detection efficiency which can range from <1 – 20%, necessitating the development of more efficient *in situ* amplification strategies.[24,28,29,31] Additionally, some studies have reported significant variance in the number of genes detected and reads per cell, further emphasizing amplification biases.[30]

Aside from RCA, another commonly used *in situ* amplification approach is HCR which was first introduced in 2004 by Dirks. R. M and Pierce. N.A.[32] HCR utilizes metastable, fluorophore-labeled hairpin probe pairs, where each hairpin probe is designed to conceal a binding site from its complementary hairpin partner until triggered. In brief, HCR starts with an initiator probe binding to its target sequence. Upon addition of the hairpin probe pairs, one hairpin will bind to the initiator probe which leads to a conformational change that unmasks the binding site for the second hairpin probe to bind. Upon binding of the second probe, the first hairpin probe can bind which leads to a chain polymerization reaction that amplifies signal. In 2010, Choi. H.M.T *et al.* showcased the use of HCR for *in situ* amplification of mRNA targets in zebrafish.[33] While HCR is a unique,

non-enzymatic approach for *in situ* amplification, its main challenge lies in achieving effective multiplexing. Due to the inherent design of HCR forming dsDNA, it is not possible to encode barcode sequences that can be read out by *in situ* sequencing or hybridization-based approaches such as in RCA or bDNA, a limitation that is critical for high multiplexing experiments. Furthermore, the limited number of available HCR hairpin probe pairs presents a significant challenge for scaling to detect many genes, especially when individual hairpin probe pairs are uniquely assigned to each gene. However, one demonstration showed that the scaling issue can be addressed by coupling HCR with seqFISH.[10] Nonetheless, this approach required extended incubation times of HCR-based readout probes and DNase digestion steps to enable the detection of temporal barcodes, posing significant challenges. Finally, HCR can lead to off-target signal amplification unless split-initiator probes are used.[34] Due to the complexity and the need for numerous hairpin probes to achieve multiplexing, HCR has not been as widely adopted as RCA.

In addition to HCR and RCA, the final class of amplification is bDNA. This approach was first conceived in 1989 by Horn T. and Urdea M. S., who chemically synthesized bDNA structures by attaching DNA sequences to multiple sites on modified nucleic acid scaffolds.[35] Signal amplification is achieved by incorporating multiple binding sites on the branched structures, allowing secondary readout probes (fluorescent or colorimetric) to bind and enhance signal. They also showcased its ability for signal amplification in nucleic acid quantification assays in 1997.[36] However, hybridization-based approach for *in situ* bDNA assembly became prevalent in 2000 when it was demonstrated by Antao V.P. *et al.*[37] Instead of chemically synthesizing bDNA structures, they leveraged the hybridization properties of DNA and employed a rational design approach to assemble large structures directly in intact cells. Since then, there has been many iterations of hybridization-based bDNA assembly, one of which is RNAscope developed by Wang. F*. et al.* in 2012.[38]

RNAscope employs a clever split-probe design called z-probes, which are RNA-targeting probes engineered to enhance specificity. The assembly of the bDNA structure occurs only when two z-probes are in proximity, reducing non-specific signal amplification. This assembled tree-like structure, with multiple binding sites for secondary probes, enables substantial signal amplification for *in situ* detection of single RNA molecules. This design is essentially the same, in principle, to the split-HCR design proposed by Choi. H. M. T. *et al.* in 2018.[34] Aside from RNAscope, other methods like SABER-FISH, developed by Kishi J. Y. *et al.,* employ a similar approach but omitted the use of z-probes.[30] Instead, SABER-FISH utilizes multiple branching cycles for non-linear signal amplification, achieving up to a 500-fold increase in signal. The limitation to most of these methods is the lack of multiplexing where they can only profile a hand-full of genes. The work of Xia. C. *et al.* in 2019, demonstrated how bDNA can be integrated with MERFISH to profile over 100 genes effectively.[39] Moreover, in the same year came ClampFISH by Rouhanifard *et al.,* which is a unique method to generate highly stable structures that are assembled using click-enabled ssDNA padlocks.[40] Since ClampFISH employs ssDNA padlocks, the assembly of bDNA structures is expected to be highly specific. Additionally, due to each amplifier strand being locked in place by the padlock scheme, the risk of bDNA structures disassembling after denaturing or harsh conditions is unlikely. Finally, in its latest conception, a company named NanoString used *in vitro* assembled bDNA structures as readouts for direct amplified detection of single-molecule transcripts.[41] By employing a similar encoding scheme as MERFISH and seqFISH, they were able to profile >900 unique transcripts in intact tissue samples. More recently, the same company was able to scale up their method to profile the whole protein coding transcriptome (18933 genes), a feat that has been a challenge for over a decade.[42] However, their detection efficiency is similar to scRNA-seq which is around 5-10% and their amplification factor is presumably

less than 50-fold. Therefore, many of the low abundance transcripts will be missed which may be important for discovery-driven experiments.

1.4 Why develop new amplification technologies?

As discussed in the previous section, many derivatives of RCA, HCR, and bDNA amplification have been developed for *in situ* single-molecule detection. However, there are a few limitations to each method. RCA, for instance, generally suffers from poor detection efficiency (approx. 10%), requires padlock probes with perfect blunt ends matching to its target that can be difficult and costly to synthesize, and demonstrates poor spatial resolution and sub-pixel localization due to amplicons generally being > 200 nm and amorphous.[24,27–31] Despite these limitations, RCA remains a widely adopted method due to its ease of use and affordability. Moreover, RCA demonstrates effective performance in thick tissue samples, handling depths ranging from 50 to 200 microns.[43,44] Industries such as 10x Genomics have adopted RCA as their main amplification platform for *in situ* detection of up to 5000 genes, but their detection efficiency is estimated to be less than 5%.[45–47] Considering detection efficiency drops as more genes are profiled, due to decoding or amplification inefficiencies, the likelihood that RCA based approaches for whole transcriptome profiling is unlikely. Finally, since the amplicons are larger than 200 nm, they occupy significantly more space within the cell compared to those produced by other amplification methods. This can restrict the number of amplicons generated *in situ* and make it challenging to resolve nearby amplicons when profiling many genes. To circumvent the issue of space limitation, Shahar. A. *et al*. utilized expansion microscopy to better resolve spots and theoretically increase the number of amplicons that can be generated.[44] In their study, they reported an estimated detection efficiency of 40%, which is significantly higher than other RCA-based methods. However, this estimate was derived from profiling only four genes, which may not accurately reflect the true efficiency as the

performance of RCA could vary depending on the number of targets being amplified. Additionally, their detection efficiency was compared against HCR with a presumed detection efficiency of 70%. Finally, expansion microscopy is not a practical solution for RCA as it significantly increases imaging time and increases the complexity of sample preparation.[48]

Apart from RCA, HCR suffers from the lack of hairpin probe pairs. With only 10 pairs currently available, the method is restricted to measuring a maximum of 10 targets.[49] Even when HCR is combined with the current seqFISH barcoding strategy that employs pseudocolors,[48,50] the DNAse digestion step which was used in previous HCR-seqFISH will lead to the digestion of primary probes prohibiting downstream temporal barcodes to be read.[10] In theory, the original seqFISH scheme combined with spectral imaging could enable the imaging of thousands of genes using 4-5 barcoding rounds. However, in practical applications, this method is limited by high optical density, making it impractical, and it also requires extended imaging times with the DNAse digestion steps. Additionally, HCR will have difficulty with multiplexed protein profiling without spectral imaging, as the use of DNase to remove channel-specific signals would also digest other antibody barcodes that have not yet been imaged. Unless HCR signal can be removed without the removal of primary probe sequences, multiplexing can become a difficult task. Furthermore, HCR can lead to off-target polymerization and produce non-specific signal unless split initiators are used.[34] Finally, the size of HCR spots has been reported to be approximately 1 micron in some cases which can complicate sub-pixel localization and make resolving multiple spots challenging.[51] The main benefit of HCR, in comparison to other methods, is the ability to amplify signal in whole mount samples and sections greater than 500 microns.[34,49,52–54] Moreover, HCR is significantly easier to perform than other methods and is presumably more cost-effective, as it relies on unmodified ssDNA.

Unlike RCA and HCR, bDNA-based signal amplification offers higher detection efficiency than RCA and has the potential for deep tissue applications, similar to HCR, with the added advantage of high multiplexing capabilities. Additionally, bDNA will have reduced variability in amplification compared to HCR and RCA, as its inherent design enables a pre-defined and consistent amplification factor. There are two general approaches to bDNA-based signal amplification. The first involves generating bDNA-based reporters *in vitro* with a predefined amplification factor, a method predominantly utilized by NanoString Technologies.[41,42] However, there are two major limitations to this approach. One problem is limited diffusion and probe penetration as the generated reporters are ~20 nm (nearly double the size of antibodies), which limits its utility in thick tissue samples. A study by Sundah N. R. *et al.* demonstrated that molecular reporters approximately 40 nm in size exhibit poor subcellular diffusion.[55] This study underscores the significance of reporter size and probe penetration, making these critical factors to consider when designing bDNA-based reporters. With this in mind, bDNA-based reporters used by NanoString Technologies have only been demonstrated in ~5-micron thick tissue samples. Second, due to the size constraints, there is an upper limit for signal amplification before the reporter becomes impractically large. To address these limitations, bDNA can be assembled *in situ*, potentially bypassing the diffusion challenge and enabling researchers to achieve amplification factors exceeding 100-fold.[40,51,56–58] Previously mentioned methods, such as SABER-FISH, ClampFISH, and MERFISH-bDNA, implemented this approach. However, SABER-FISH does not generate highly stable structures, making it susceptible to disassembly during multiple rounds of harsh washing steps to remove bound fluorophores, as required in multiplexed FISH protocols like seqFISH. Alternatively, multiple rounds of reductive cleavage could be performed to remove fluorophores, like in MERFISH, but this approach in some cases can lead to

sample degradation. Currently, SABER-FISH has not demonstrated its ability for highly multiplexed profiling of genes or proteins. Moreover, SABER-FISH can only be performed in thin tissue sections due to the amplifier strands being greater than 500 nts.[57,58] To circumvent these limitations, ClampFISH was developed to generate highly stable bDNA structures *in situ*, by leveraging click-chemistry enabled ssDNA padlocks.[40,56] ClampFISH employs amplifier strands shorter than 100 nts, which theoretically allows for improved diffusion. However, ClampFISH has only been demonstrated to profile up to 10 genes.[56] While there is not any inherent reason to why it cannot scale to thousands of genes when coupled with seqFISH, ClampFISH amplifiers are fairly expensive and relies on reagents that can degrade quickly unless stored under inert atmospheric conditions. Moreover, the click-chemistry-enabled ligations used in ClampFISH may lack specificity, as ligations can occur independently of off-target hybridizations. As a result, the generated amplicons may contribute to non-specific signals.

In designing a new amplification strategy, we prioritized specificity, cost-effectiveness, high multiplexing capabilities, compatibility with thick tissue samples, and the ability to amplify both single transcript and protein molecules. Moreover, we desired a method that does not involve specialized sample preparation protocols such as hydrogel embedding or expansion. We also aimed to develop a method capable of scaling to the whole transcriptome while achieving detection efficiencies significantly greater than 5% and surpassing other single-cell profiling technologies. With these goals in mind, we began by exploring the technical limitations of RCA to assess its capacity for profiling many genes with high detection efficiency. Next, we pursued a modified version of ClampFISH, termed Linked AmplificatioN Tethered by Exponential RadiaNce (LANTERN). Unlike ClampFISH, LANTERN leverages ligases to enhance the signal specificity and ligation efficiency of padlock probes, eliminates the need for degradable reagents, and reduces amplifier costs.

Using LANTERN, we were able to achieve high amplification factors and profile up to 3000 genes *in situ*. Although this method revealed several unforeseen limitations, the insights gained from its design paved the way for the development of a novel amplification scheme called Signal amPlificAtion by Recursive Crosslinking (SPARC). This innovative approach incorporates features such as iterative deposition of amplifiers using a toehold-mediated strand displacement scheme, *in situ* bDNA assembly, and photochemistry. In the following chapters, I will further discuss the exploration of RCA and LANTERN, and the subsequent development of SPARC. We demonstrate the utility of SPARC for amplified detection of single-molecule transcripts and proteins.

1.5 References

1. John, H. A., Birnstiel, M. L. & Jones, K. W. RNA-DNA Hybrids at the Cytological Level. *Nature* **223**, 582–587 (1969).
2. Gall, J. G. & Pardue, M. L. Formation and detection of rna-dna hybrid molecules in cytological preparations*. *Proc. Natl. Acad. Sci.* **63**, 378–383 (1969).
3. Harrison, P. r., Conkie, D., Paul, J. & Jones, K. Localisation of cellular globin messenger RNA by in situ hybridisation to complementary DNA. *FEBS Lett.* **32**, 109–112 (1973).
4. Rudkin, G. T. & Stollar, B. D. High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence. *Nature* **265**, 472–473 (1977).
5. Langer-Safer, P. R., Levine, M. & Ward, D. C. Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc. Natl. Acad. Sci.* **79**, 4381–4385 (1982).
6. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* **280**, 585–590 (1998).

7. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
8. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, (2015).
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.* **113**, 11046–11051 (2016).
10. 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).
11. Shah, S. *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* **174**, 363-376.e16 (2018).
12. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
13. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
14. Takei, Y. *et al.* High-resolution spatial multi-omics reveals cell-type specific nuclear compartments. 2023.05.07.539762 Preprint at https://doi.org/10.1101/2023.05.07.539762 (2023).
15. Oliveira, M. F. *et al.* Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. 2024.06.04.597233 Preprint at https://doi.org/10.1101/2024.06.04.597233 (2024).

16. Du, M. R. M. *et al.* Spotlight on 10x Visium: a multi-sample protocol comparison of spatial technologies. 2024.03.13.584910 Preprint at https://doi.org/10.1101/2024.03.13.584910 (2024).

17. Tian, L., Chen, F. & Macosko, E. Z. The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* **41**, 773–782 (2023).

18. Janesick, A. *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).

19. Borm, L. E. *et al.* Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat. Biotechnol.* **41**, 222–231 (2023).

20. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* (2022) doi:10.1016/j.cell.2022.04.003.

21. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* 1–7 (2020) doi:10.1038/s41587-020-0739-1.

22. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

23. Wirth, J. *et al.* Spatial transcriptomics using multiplexed deterministic barcoding in tissue. *Nat. Commun.* **14**, 1523 (2023).

24. Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017).

25. Liu, D., Daubendiek, S. L., Zillman, M. A., Ryan, K. & Kool, E. T. Rolling Circle DNA Synthesis: Small Circular Oligonucleotides as Efficient Templates for DNA Polymerases. *J. Am. Chem. Soc.* **118**, 1587–1594 (1996).

26. Lizardi, P. M. *et al.* Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19**, 225–232 (1998).

27. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

28. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014).

29. Deng, R., Zhang, K., Sun, Y., Ren, X. & Li, J. Highly specific imaging of mRNA in single cells by target RNA-initiated rolling circle amplification. *Chem. Sci.* **8**, 3668–3675 (2017).

30. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).

31. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).

32. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proc. Natl. Acad. Sci.* **101**, 15275–15278 (2004).

33. Choi, H. M. T. *et al.* Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat. Biotechnol.* **28**, 1208–1212 (2010).

34. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

35. Horn, T. & Urdea, M. S. Forks and combs and DNA: the synthesis of branched oligodeoxyribonucleotides. *Nucleic Acids Res.* **17**, 6959–6967 (1989).

36. Horn, T., Chang, C.-A. & Urdea, M. S. Chemical synthesis and characterization of branched oligodeoxyribonucleotides (bDNA) for use as signal amplifiers in nucleic acid quantification assays. *Nucleic Acids Res.* **25**, 4842–4849 (1997).

37. Antao, V. P., Player, A. N. & Kolberg, J. A. In Situ Hybridization Using the bDNA Technology. in *Techniques in Quantification and Localization of Gene Expression* (ed.

Patterson, B. K.) 81–93 (Birkhäuser, Boston, MA, 2000). doi:10.1007/978-1-4612-1342-0_6.

38. Wang, F. *et al.* RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).

39. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci.* **116**, 19490–19499 (2019).

40. Rouhanifard, S. H. *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry–based amplification. *Nat. Biotechnol.* **37**, 84–89 (2019).

41. He, S. *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).

42. Khafizov, R. *et al.* Sub-cellular Imaging of the Entire Protein-Coding Human Transcriptome (18933-plex) on FFPE Tissue Using Spatial Molecular Imaging. 2024.11.27.625536 Preprint at https://doi.org/10.1101/2024.11.27.625536 (2024).

43. Sui, X. *et al.* Scalable spatial single-cell transcriptomics and translatomics in 3D thick tissue blocks. 2024.08.05.606553 Preprint at https://doi.org/10.1101/2024.08.05.606553 (2024).

44. Alon, S. *et al.* Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, (2021).

45. Xenium Prime 5K Gene Expression Workflow, Analysis & Data Highlights - Official 10x Genomics Support. *10x Genomics* https://www.10xgenomics.com/support/in-situ-gene-expression/documentation/steps/assay/xenium-prime-5k-gene-expression-workflow-analysis-data-highlights.

46. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* 1–10 (2020) doi:10.1038/s41587-020-0465-8.

47. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**, 311–322 (2008).

48. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235 (2019).

49. Schulte, S. J., Fornace, M. E., Hall, J. K., Shin, G. J. & Pierce, N. A. HCR spectral imaging: 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent samples. *Development* **151**, dev202307 (2024).

50. Eng, C.-H. L., Shah, S., Thomassie, J. & Cai, L. Profiling the transcriptome by RNA SPOTs. *Nat. Methods* **14**, 1153–1155 (2017).

51. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* **9**, 7721 (2019).

52. Wang, Y. *et al.* EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. *Cell* **0**, (2021).

53. Shah, S. *et al.* Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* **143**, 2862–2867 (2016).

54. Kumar, V. *et al.* Optimization and evaluation of fluorescence in situ hybridization chain reaction in cleared fresh-frozen brain tissues. *Brain Struct. Funct.* **226**, 481–499 (2021).

55. Sundah, N. R. *et al.* Barcoded DNA nanostructures for the multiplexed profiling of subcellular protein distribution. *Nat. Biomed. Eng.* **3**, 684–694 (2019).

56. Dardani, I. *et al.* ClampFISH 2.0 enables rapid, scalable amplified RNA detection in situ. *Nat. Methods* **19**, 1403–1410 (2022).

57. Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533 (2019).

58. Attar, S. *et al.* Efficient and highly amplified imaging of nucleic acid targets in cellular and histopathological samples with pSABER. *Nat. Methods* 1–10 (2024) doi:10.1038/s41592-024-02512-2.

Chapter II

Exploration of RCA and branched DNA assembly using ssDNA padlocks for amplified smFISH and seqFISH

This work was a collaborative project conducted with Chee-Huat Linus Eng and Chuqi Lu. Contribution to figures that was not primarily obtained by me is outlined in the figure caption.

2.1 Abstract

Highly multiplexed spatial transcriptomics has emerged as a powerful tool for elucidating gene expression patterns in cells while maintaining their native spatial context. Imaging-based spatial transcriptomic methods that rely on smFISH offer high detection efficiency of biomolecules, provide single-cell spatial coordinates, and reveal the 3D locations of biomolecules at the subcellular level. Although a powerful technique, it often suffers from a poor signal-to-noise ratio, which depends on tissue type and gene targets, as well as low sample throughput caused by long exposure times or use of high magnification objectives. These limitations could be addressed with a robust signal amplification strategy that significantly improves signal-to-noise ratio in suboptimal samples, reduces imaging duration by shortening exposure times or enabling the use of lower magnification objectives, and requires fewer probes to detect genes of interest. In this study, we evaluated two widely used *in situ* signal amplification strategies, RCA and bDNA, for multiplexed RNA profiling using seqFISH. Although RCA is one of the most common amplification strategies, we found significant limitations with this method regarding its ability to profile many targets while maintaining high detection efficiency. Furthermore, we identified sub-cellular localization inaccuracies and some inefficiencies amplifying protein targets. To address these limitations, our investigation of bDNA based strategies using a ligation-enabled padlock system demonstrated its ability to profile 150 genes with significantly higher detection efficiency compared to RCA and effectively amplify dense targets. Additionally, the method was able to profile up to 3000 genes with high correlation with RNA-seq. However, we observed a decline in detection efficiency as the number of targets increased. Herein, we will discuss design constraints with our padlock system that led to poor sub-pixel colocalization and amplifier binding events.

2.2 Introduction

In recent years, spatial transcriptomics has become a standard method for studying the spatial organization of gene expression within tissues, offering valuable insights into cellular communication and interactions. While imaging-based spatial transcriptomic methods that rely on smFISH have gained substantial popularity, they still present several limitations. These challenges include a low signal-to-noise ratio in suboptimal samples, the need for many probes to target specific genes, and limited sample throughput due to long exposure times and the requirement for high-magnification objectives during imaging. To overcome these challenges, many have focused on developing robust *in situ* amplification technologies. These methods include various forms of rolling circle amplification (RCA), including RollFISH,[1] HybISS,[2] STARmap,[3,4] FISSEQ,[5,6] and BaristaSeq,[7] among others. Alternatively, there has been much development in hybridization chain reaction (HCR) technology such as third-generation HCR[8] and HCR spectral imaging.[9] Another class of amplification methods, known as branched DNA (bDNA), has gained significant momentum over the years, inspiring the development of techniques like RNAScope,[10] SABER,[11–13] ClampFISH,[14,15] and SMI probes.[16,17] However, many amplification methods face various challenges, including poor probe penetration in thick tissue samples, non-specific signals, low detection efficiency, high reagent costs, and limited scalability for whole-transcriptome profiling with high detection efficiency. The specific combination of these challenges varies across different methods and platforms.

While many methods exist in imaging-based spatial transcriptomics, seqFISH has emerged as a powerful tool for profiling thousands of genes in intact tissues and cell culture systems.[29,32,33] However, the method relies on smFISH, and we wanted to further improve seqFISH for robust detection and quantification of single-molecule transcripts

across various sample types. Furthermore, we wished to the reduce the number of probes required for accurate detection of biomolecules. In this report, we explored two amplification methods for seqFISH which includes RCA and a modified ClampFISH method called, Linked AmplificatioN Tethered by Exponential RadiaNce (LANTERN). Both methods are highly amenable to seqFISH as they can withstand stringent DNA denaturing wash conditions to remove bound fluorescent readouts. This study highlights significant limitations and inefficiencies associated with RCA, which are often overlooked in the existing literature. We also observed that RCA-based seqFISH has poor detection efficiency. Unlike RCA, LANTERN demonstrated significant potential for highly multiplexed RNA profiling, with detection efficiencies far exceeding those reported by most methods. Additionally, LANTERN showed a strong correlation with bulk RNA-seq measurements in NIH 3T3 cells and E9.5 mouse embryos. However, we identified design constraints with LANTERN that limited its capability to achieve whole transcriptome-level profiling. Regardless, LANTERN demonstrates its ability to profile more than 1000 genes in NIH 3T3 cells.

2.3 Results

2.3.1 Probe designs and quality checks

We investigated primary probe designs with different ligation configurations for RCA and LANTERN. The two main approaches were (1) padlock probes modified with an alkyne on the 5' end and an azide on the 3' end to enable click ligation at the DNA/RNA template, and (2) padlock probes with 5' phosphorylated, perfectly

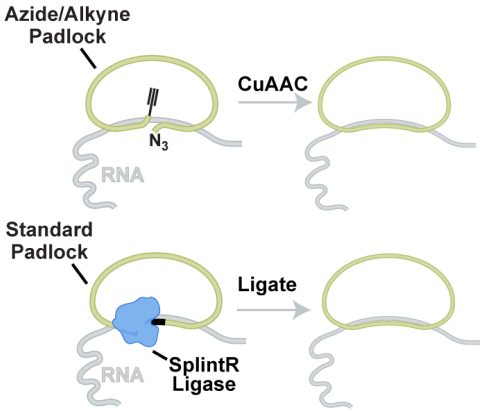


Figure 2.1: Schematic of azide/alkyne padlock probe and standard phosphorylated padlock probe for chemical or enzymatic ligation at the DNA/RNA hybrid template, respectively.

blunted ends complementary to the RNA target for SplintR ligation at the DNA/RNA template (**Figure 2.1**). Each configuration ensures circularization of primary probes, stabilizing them to their target. The SplintR ligated padlock is compatible with both LANTERN and RCA, whereas the click-ligated padlock is exclusively compatible with LANTERN. Each ligation method offers distinct advantages and limitations. In practice, probes for click-based ligation are easier to generate compared to standard padlocks, which require blunt ends that perfectly match their target without any gaps. More importantly, the ligation efficiency is much higher using click chemistry compared to ligases.[14,15] Furthermore, the reagents used for Cu-catalyzed Azide-Alkyne Cycloaddition (CuAAC) can diffuse through tissue more readily than ligase enzymes. However, CuAAC reagents can also cause RNA degradation if reaction conditions are not properly optimized. The benefit of standard padlocks using SplintR ligation is high specificity. Ligation events occur only when there is significant alignment with the target and can even discriminate single-nucleotide differences at the nick site under certain conditions.[18,19] Moreover, compared to traditional approaches that utilizes *in situ* RT to enable ligation of padlocks against cDNA, direct RNA ligation approaches offer greater detection efficiency.[18] The drawback to this approach is the synthesis of these probes. To obtain perfectly blunted ends matching to its target, primers used for probe generation will need to be cleaved and the final probe will need to be purified by PAGE (**see Methods**). Furthermore, the SplintR ligase will have limited diffusion and ligation efficiency compared to the click-based method.

Each probe design described was synthesized in-house (**see Methods**). Probes containing azide and alkyne functional groups were validated using in-gel fluorescence by clicking on dye molecules (**Figure 2.2A**). We reasoned that the 5' alkyne is present in 100% of our probes because it is introduced during RT using an alkyne-modified primer

during probe generation (**see Methods**). To determine if the addition of the 3' azido-nucleotide with terminal deoxynucleotidyl transferase occurs at near-quantitative yields, a mass shift assay was performed by appending a DBCO-modified oligonucleotide (25 nts). Our results indicate that greater than 90% of our generated probes contain 3' azides (**Figure 2.2B**). Moreover, the absence of smearing on the gel and supporting evidence from previous literature suggest that a single azide was successfully added.[15,20] For our standard padlock probes with blunted ends complementary to its target, we designed probes containing a BspQI restriction site on the 3' end and a USER cleavage site on the 5' end of our oligonucleotides (**Figure 2.2C**). Upon digest we observed 3 bands as expected, corresponding to no cleavage, single digest, and complete digest. The completely cleaved probes were purified in-house with PAGE at moderate yields (**Figure 2.2D**). Upon the successful generation of these probes, we utilized complex opool mixtures with these modifications for most multiplexed assays.

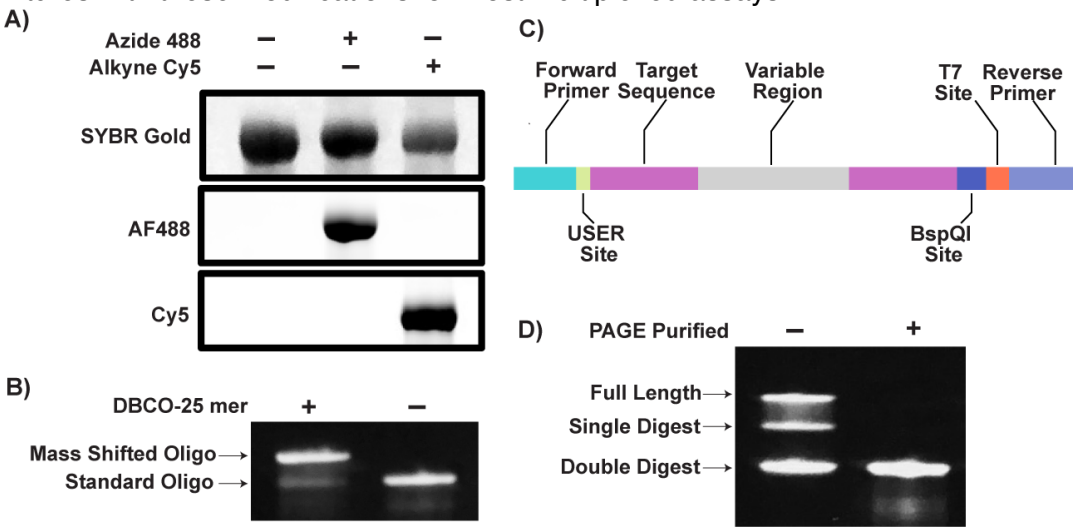


Figure 2.2: A) In-gel fluorescence assay to verify the presence of azido and alkynyl functional groups on generated padlock probes. B) Mass shift assay confirming near quantitative addition of azido-nucleotides with TdT. C) Probe design for the generation of padlock probes with blunted ends matching the target sequence. D) Digestion and PAGE purification leads to moderate yields of standard padlock probes with high purity.

2.3.2 RCA demonstrates high correlation with RNA-seq and smFISH measurements but has poor detection efficiency

RCA is commonly employed in many spatial transcriptomic technologies. We sought to couple RCA with seqFISH for amplified detection of numerous genes in NIH 3T3 cells. To achieve this, we utilized padlock probes and SplintR for direct ligation of the DNA/RNA template, followed by RCA with aminoallyl-dUTPs to enable amplicon crosslinking. This approach, commonly employed in methods like FISSEQ[5,6] and Starmap[3,4,21], stabilizes the amplicons by crosslinking them to the extracellular matrix. Additionally, it condenses the amplicons by linking amino groups on the DNA to neighboring sites, ensuring spatial retention and structural integrity (**Figure 2.3A**). We utilized 4 barcoding rounds with 6 pseudocolors encoded across channels (216 codewords) to profile 150 unique RNA transcripts, following a similar scheme to RNA SPOTs.[22] RCA-seqFISH showed significant amplification and sufficient generation of amplicons *in situ.* We observed ~300 counts per cell with a false discovery rate less than 5%, demonstrating minimal noise from the data (**Figure 2.3B**). Furthermore, there was high correlation with bulk RNA-seq for the 150 genes profiled and against 40 genes measured by standard smFISH (**Figure 2.3C-D**). However, the detection efficiency was ~12%, which was considerably low (**Figure 2.3D**). Nevertheless, this value aligns well with previous estimates in the literature.[18,23] The underlying cause of inefficient amplification remains unclear. However, it may stem from inefficient RCA initiation (due to inaccessibility of enzyme) or reagent depletion. It could be that hydrogel-embedding and protein digestion may improve the accessibility of the polymerase enzyme and enhance amplicon generation. Although a platform that does not require such extensive sample preparation will be most ideal. Alternatively, the ligation from the SplintR enzyme could also be inefficient or that the padlock probe is too large for efficient amplification.

Regardless, these findings suggest that RCA-based seqFISH provides highly amplified signals with excellent specificity and minimal noise. However, the ability to profile thousands of genes or the entire transcriptome at moderate to high detection efficiency remains a challenge and may not be the most suitable method.

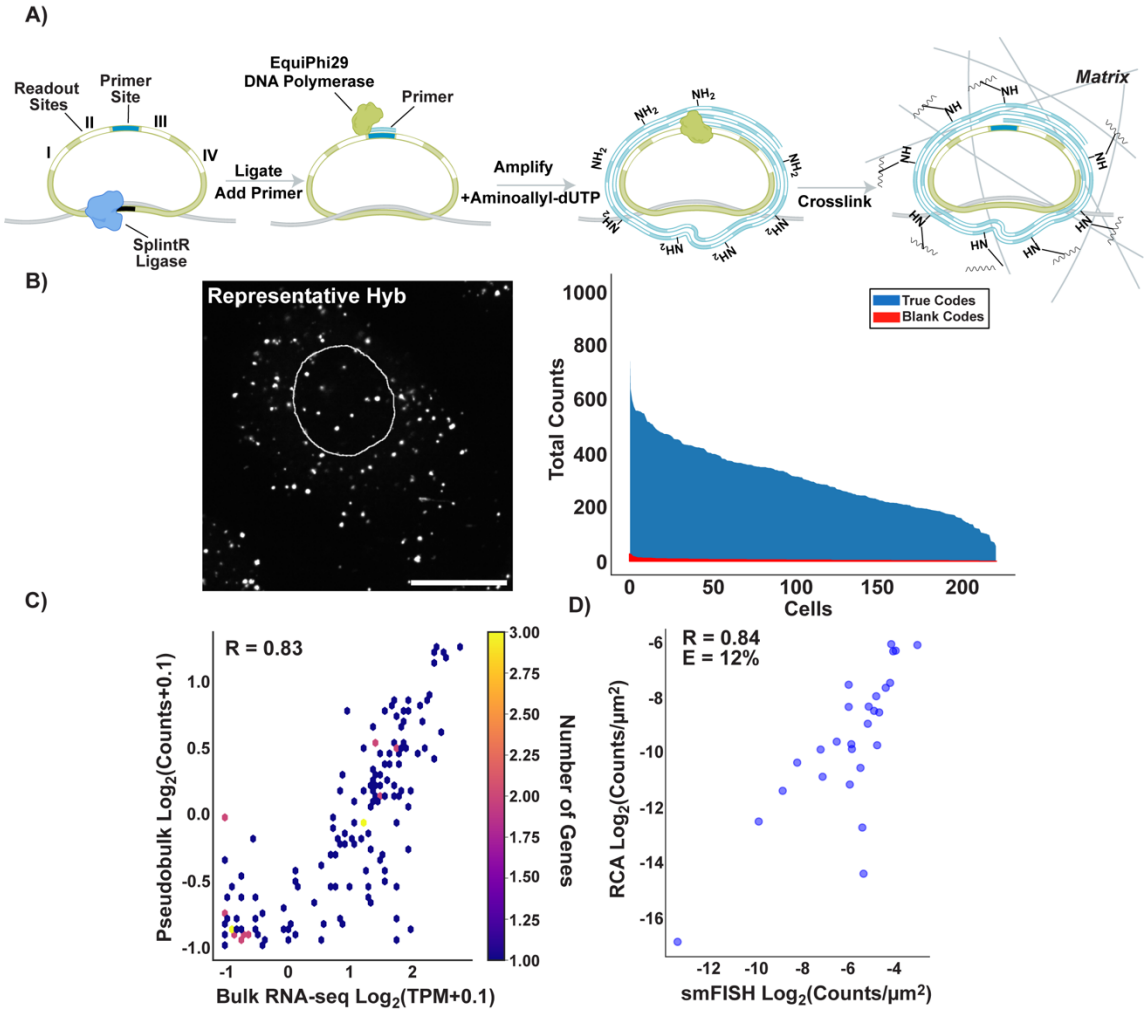


Figure 2.3: A) Schematic of RCA-based seqFISH using 4 barcoding rounds and 6 pseudocolors. B) Representative hybridization round (left, nucleus represented by white outline) and counts per cell distribution (right). Scale bar = 20 microns. C) Correlation between pseudobulked counts from RCA-based seqFISH and bulk RNA-seq measurement from NIH 3T3 cells. D) Correlation between RCA-based seqFISH and standard smFISH measurement of 40 unique genes in NIH 3T3 cells. Detection efficiencies were estimated from the slope of a linear regression fitted to the unlogged data.

2.3.3 RCA exhibits poor colocalization with smFISH and inefficient signal amplification of dense targets

Although numerous literatures exist for *in situ* RCA, there is limited data on the accuracy of subcellular localization of RNA targets. This is especially important if detecting exogenous RNA or DNA barcodes like in MEMOIR, where multiple spots targeting different elements need to colocalize.[24] Similarly, this is essential when barcodes are split across two probes, requiring the presence of both probes to decode the target accurately. Additionally, split barcoding reduces probe size, thereby enhancing amplification efficiency. Therefore, we designed smFISH padlock probes targeting Eef2 transcripts in NIH 3T3 cells and two distinguishable RCA probes to assess colocalization (**Figure 2.4A**). To our surprise, we observed that most amplicons do not colocalize with smFISH (~50%). Moreover, the two different RCA amplicons generated from the same transcript had worse colocalization compared to smFISH (~10%, **Figure 2.4B**). This may be due to both

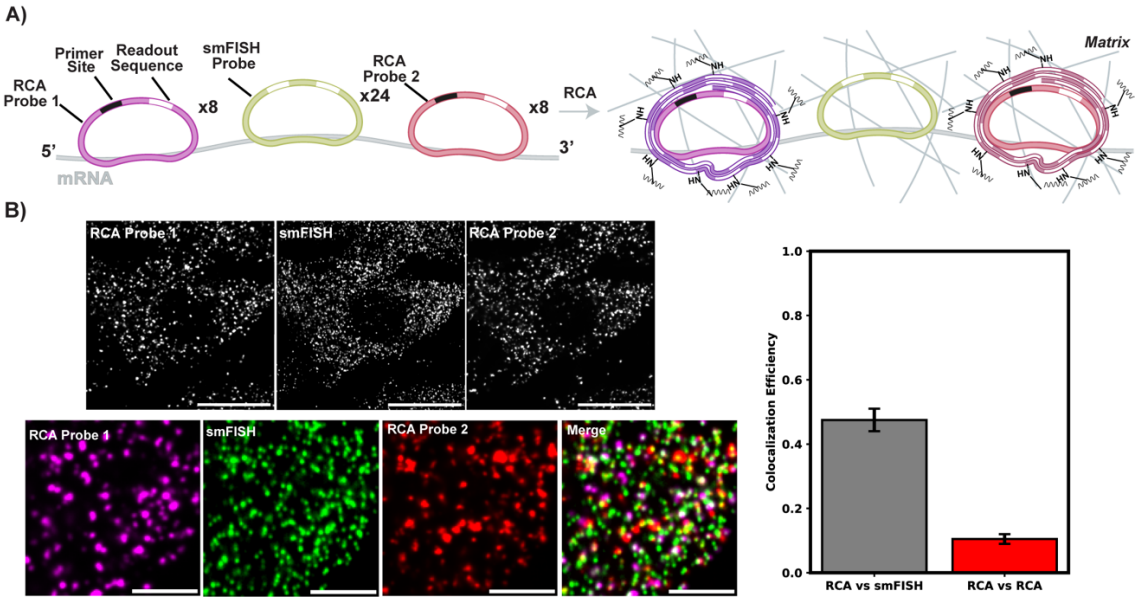


Figure 2.4: A) Schematic of 2 unique sets of RCA probes and 1 set of smFISH probes for assessing colocalization. B) Representative cell depicting spot density across the 3 probe sets (top left, scale bar = 20 microns) and zoomed in images of spots (bottom left, scale bar = 5 microns). Quantification shows poor colocalization efficiency across the three probes set. Additionally, RCA spots have greater colocalization efficiency against smFISH spots compared to the two RCA amplicons (right panel). Error bar represents 95% C.I. using the 2 different amplicons as reference for colocalization efficiency.

amplicons moving away from the transcript. Additionally, the result could be an indication of steric hindrance or local reagent depletion which is affecting amplification on the same transcript.

In addition to smFISH colocalization, we aimed to evaluate the amplification efficiency of RCA against protein targets, such as Lamin proteins, and glycans with terminal manα(1,2)man. As discussed previously, RCA amplicons often diffuse away from their target. However, we hypothesized that anchoring these amplicons to an antibody or lectin would mitigate diffusion, thereby supporting the amplicon displacement theory. Tethered amplicons should recapitulate the expected spatial pattern of the nuclear lamina (stained by Lamin-B1) and the cell membrane (stained by rGRFT). A loss of the expected pattern would indicate that amplicons are still detaching from their targets. Moreover, given the historically poor detection efficiency of RCA against RNA targets, it remains uncertain whether proteins or glycans can be uniformly amplified using this approach. To investigate this, we targeted nuclear Lamins and manα(1,2)man residues using Lamin-B1 antibody and rGRFT lectin conjugated with ssDNA (**Figure 2.5A**). When we amplified these affinity reagents, the signal appeared as punctates with heterogenous labeling. However, as expected, the spatial patterning was preserved (**Figure 2.5B**). This indicates that if amplicons generated from RCA are tethered to some biomolecule, the subcellular localization is conserved. However, the formation of punctate signals highlights inherent inefficiencies with RCA, likely stemming from reagent depletion or inefficient initiation of RCA *in situ*.

In summary, our findings indicate that RCA amplicons can become dislodged from their target transcripts, compromising subcellular localization accuracy, as shown by their low colocalization with smFISH signals. Furthermore, amplification of two different sets of

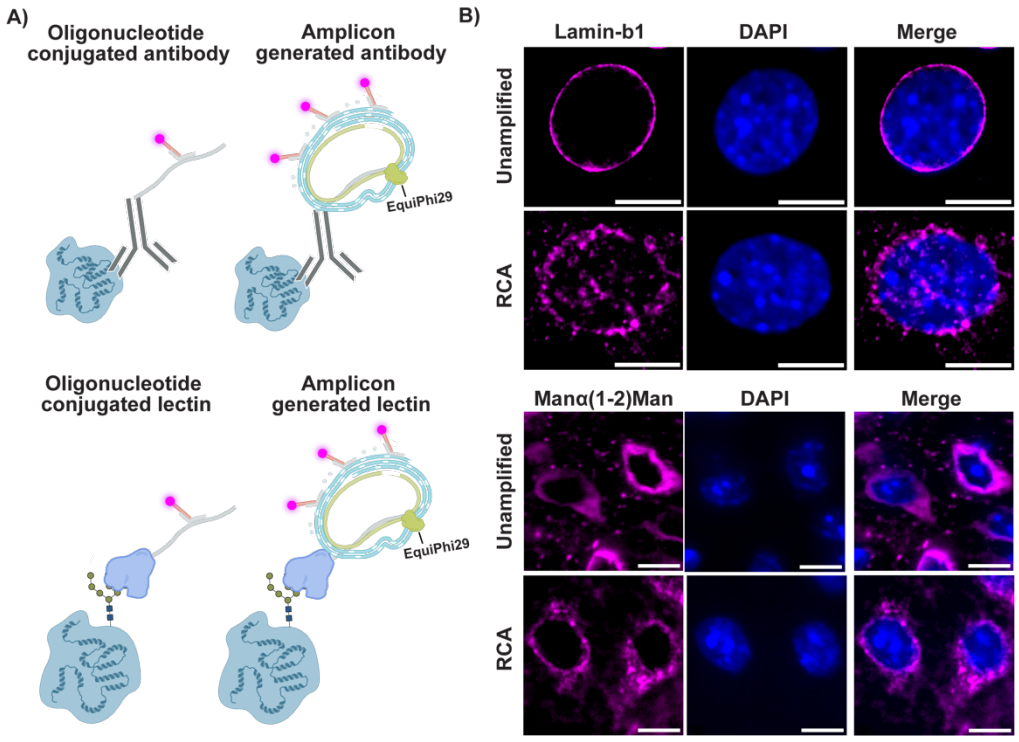


Figure 2.5: A) Schematic of antibody and lectin RCA. B) Representative images depicting standard labeling and RCA of the nuclear lamina (top, scale bar = 10 microns). Representative images of standard labeling and RCA of manα(1,2)man (bottom, scale bar = 10 microns).

probes on the same transcript is poor. These issues can potentially be alleviated if the amplicons are tethered to some substrate. We also observed that RCA struggles with amplification against dense targets, as evidenced by the formation of punctate signals where a homogeneous and uniformly labeled spatial patterning was expected. Despite these challenges, RCA remains highly specific, as demonstrated by our seqFISH experiments and its utility for cell typing in numerous reports.[3,4,21,25–27] Notably, 10x Genomics recently released the Xenium 5k Prime platform, which shows improved colocalization across different RCA amplicons. This improvement may be attributed to their workflow, where the RNA molecule itself serves as the primer for RCA initiation, effectively tethering the amplicons in place.[28]

2.3.4 LANTERN shows high amplification factor, colocalizes with smFISH, and can amplify dense targets

Considering the shortcomings of RCA, we wanted to develop an amplification method that has high detection efficiency, compatible with seqFISH, and has diffraction-limited spots for improved sub-pixel localization. Building upon ClampFISH,[14,15] we investigated a bDNA system that utilizes ligase-enabled padlocks. We transitioned away from CuAAC-based ligations due to inconsistencies in amplification and the reliance on degradable reagents, such as ascorbate, which complicate automated amplification processes. This shift was particularly important for fluidic systems designed to operate continuously for extended durations, often exceeding 18 hours. Moreover, we wanted a ligation system that is highly specific towards its target where blunt-end ligation is highly dependent on amplifier hybridization to its correct target. Click chemistry, on the other hand, can theoretically have higher mismatches and still ligate.

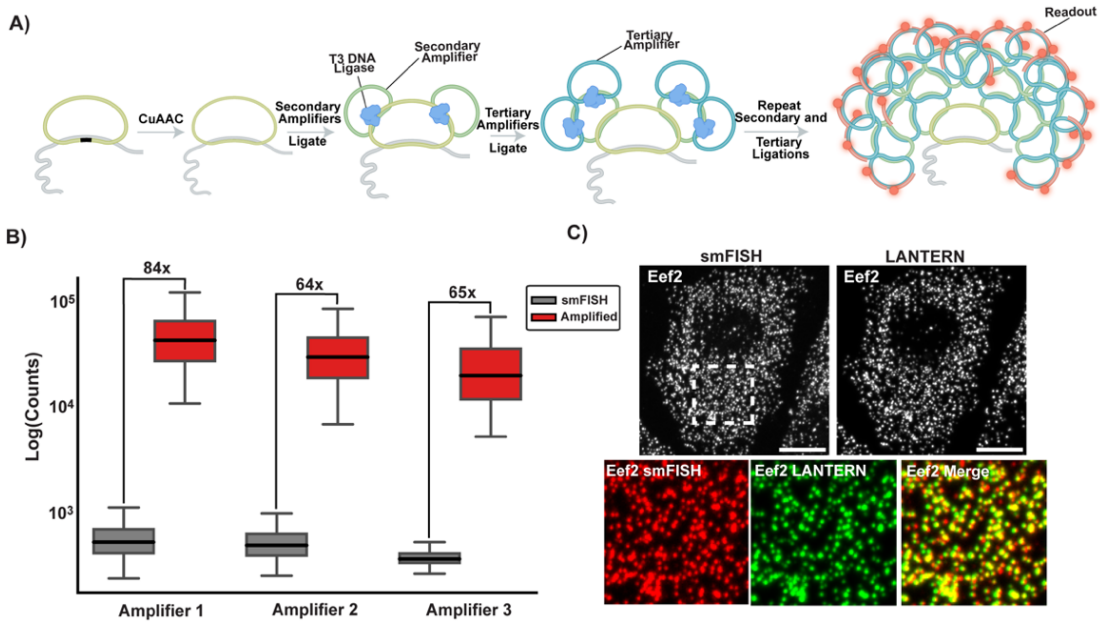


Figure 2.6: A) Schematic of LANTERN. B) Representative amplification fold was performed on colocalizing spots across conditions. Horizontal line corresponds to median. (Experiment performed by Chee-Huat Eng, analysis performed by Katsuya Colón) C) LANTERN shows high colocalization with smFISH (inset) and similar number of spots. Scale bar = 10 microns.

In brief, LANTERN utilizes ssDNA padlock primary probes containing an azide and alkyne on the terminal ends for ligation at the DNA/RNA hybrid template. After click

ligation, secondary amplifiers are hybridized onto to the probe containing 2 tertiary amplifier binding sites. Once secondary amplifiers are ligated, tertiary amplifiers are hybridized and ligated similarly as the secondary amplifiers. This process is repeated to gain an exponential boost in signal (**Figure 2.6A**). Using this method, we were able to obtain greater than 60-fold signal amplification *in situ* (**Figure 2.6B**). Additionally, we observed significantly more amplified spots than RCA and nearly all spots colocalize with smFISH (**Figure 2.6C**). These results show that LANTERN is a viable signal amplification strategy that can be coupled to seqFISH.

Considering RCA showed some inefficiencies in amplifying dense targets such as nuclear Lamin proteins or glycans with terminal manα(1,2)man, we wanted to determine if LANTERN can amplify highly abundant species such as 18s rRNA (Rn18s). While amplifying Rn18s, we simultaneously targeted Eef2 using a single probe. The inclusion of Eef2 serves as a control to identify potential amplification inefficiencies, as the absence of its signal indicates suboptimal amplification. To our surprise, we observed whole cell labeling (lack of punctates) when targeting Rn18s and signal from a single probe bound to Eef2 (**Figure 2.7**). Our results demonstrate that our amplification protocol does not suffer from reagent depletion or other enzymatic inefficiencies. Additionally, unlike RCA, the results confirm that amplification efficiency is not compromised by high probe density.

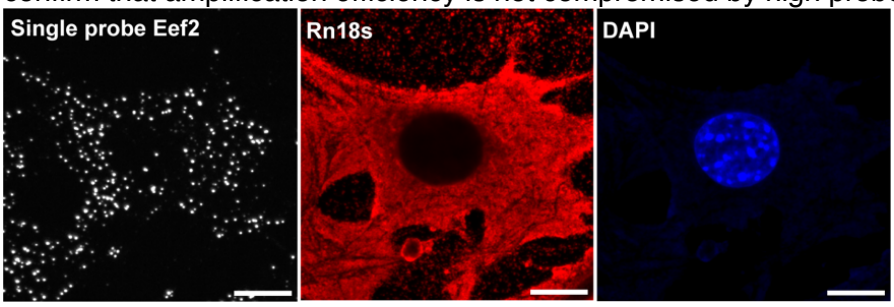


Figure 2.7: LANTERN shows whole cell amplification when targeting Rn18s and single probe Eef2 amplification. These results indicate that LANTERN can amplify many targets. Scale bar = 10 microns.

2.3.5 150-plex LANTERN shows high detection efficiency and correlation with RNA-seq in both NIH3T3 cells and E9.5 mouse embryo

To perform high-multiplexing experiments using combinatorial barcoding, we needed to design and identify a large number of usable amplifiers. To achieve this, we generated 180 amplifier pairs (secondary and tertiary amplifier), assigning two amplifier pairs per probe to target 90 unique genes. We assessed the quality of each amplifier based on its ability to generate a bright signal. By measuring the colocalization between amplifiers, we could quickly identify those that failed to produce a bright signal by observing missing branches or a lack of spots in either branch. Interestingly, we observed that certain readout probes exhibited off-target binding (**Supplementary Figure 2.1**). It remains unclear whether this is due to off-target ligations or off-target readout probe binding. Nonetheless, we identified 60 amplifier pairs that demonstrated sufficient amplification signals and low off-target readout binding (**Supplementary Table 2.1**). It should be noted that there are limitations to this screening process. The amplification screen does not include colocalizing smFISH probes, making it difficult to determine the

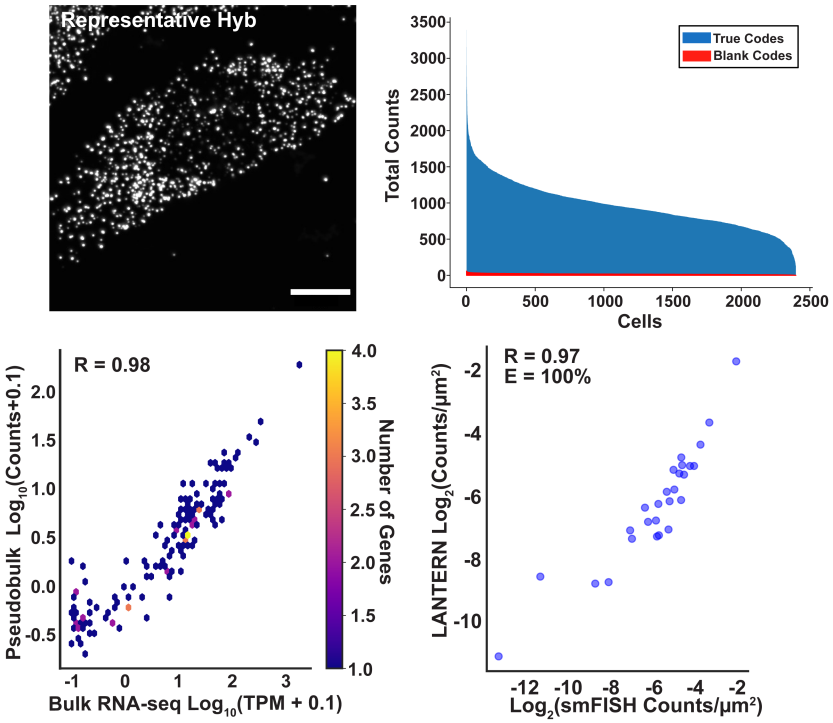


Figure 2.8: Summary of 150-plex LANTERN experiment in NIH 3T3 showing representative hybridization round (top left), total counts per cell (top right), correlation with bulk RNA-seq (bottom left), and correlation and detection efficiency compared to smFISH (bottom right). Detection efficiencies were estimated from the slope of a linear regression fitted to the unlogged data. Scale bar = 10 microns.

actual amplification efficiency without normalization to a standard signal when comparing amplifiers. Additionally, we cannot assess noise levels without colocalization to a ground truth signal.

With LANTERN showcasing significant improvement over RCA, we performed a 150-plex RNA profiling experiment in NIH 3T3 cells. Using the same encoding scheme used in RCA-seqFISH, we obtained ~100% detection efficiency with an FDR less than 5% (**Figure 2.8**). LANTERN shows significantly higher correlation and detection efficiency compared to RCA-seqFISH, making it a promising method for highly multiplexed profiling of RNA species. Using the same probes, we evaluated LANTERN in E9.5 mouse embryos to assess its effectiveness in tissue samples. As expected, LANTERN successfully amplified and detected targets, with expression levels aligning closely with those obtained from bulk RNA-seq analysis (**Figure 2.9**). With these demonstrations, we show that LANTERN achieves high detection efficiency and performs effectively in both tissues and cultured cells. The observed expression levels align closely with bulk RNA-seq and standard smFISH measurements, underscoring its reliability and versatility.

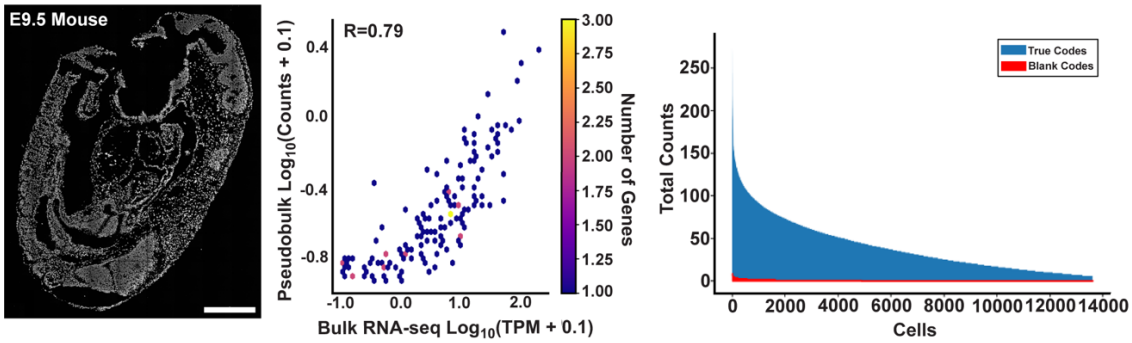


Figure 2.9: Summary of LANTERN applied to E9.5 mouse embryo (left panel) showing high correlation with RNA-seq (middle panel) and counts per cell distribution (right panel). Scale bar = 200 microns. (Experiment performed by Chee-Huat Eng, analysis performed by Katsuya Colón).

2.3.6 The detection efficiency of LANTERN declines with increasing gene targets

While we have demonstrated LANTERN in a smaller-scale 150-plex experiment, our ultimate goal is to scale the method to near-transcriptome levels, comparable to the capabilities of seqFISH+.[29] In doing so, we performed a 1,000 and 3,000-plex RNA profiling experiment in NIH 3T3 cells using seqFISH with 4 barcoding rounds and 15 pseudocolors (3375 codewords). To our surprise, we observed a significant drop in detection efficiency when profiling 1,000 genes compared to our 150 genes experiment, decreasing from 100% to 30% (**Figure 2.10A**). Scaling further to 3,000 genes resulted in a similar drop in detection efficiency, from 30% to 10% (**Figure 2.10B**). Additionally, correlations with RNA-seq and smFISH also decreased as the number of profiled genes increased (**Figure 2.10A-B**), likely due to the under detection of genes. Furthermore, we observed consistent counts per cell across both the 1,000-plex and 3,000-plex

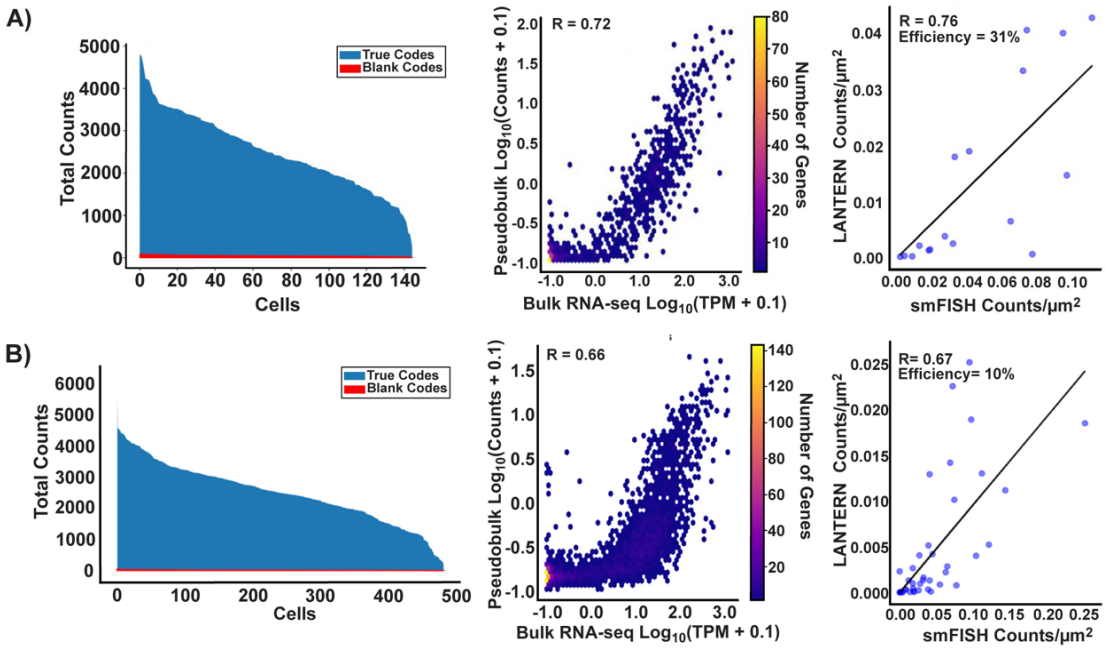


Figure 2.10: A) Summary of the 1,000-plex RNA profiling experiment in NIH 3T3 cells, displaying the counts per cell distribution (left), correlation with bulk RNA-seq (middle), and correlation with standard smFISH (right). (Experiment and analysis performed by Chuqi Lu) B) Summary of the 3000-plex RNA profiling experiment in NIH 3T3 cells, displaying the counts per cell distribution (left), correlation with bulk RNA-seq (middle), and correlation with standard smFISH (right). Detection efficiencies were estimated from the slope of a linear regression fitted to the unlogged data.

experiments, with counts remaining at approximately 2,500 per cell, despite the increase in the number of genes profiled. Regardless, the correlations with RNA-seq and smFISH remain reasonable and are comparable to those observed in other studies comparing smFISH and RNA-seq data.[29,30]

We further investigated the potential reasons for the reduced detection efficiency observed as more genes were profiled. Our findings suggest that the decline in detection efficiency stems from poor decoding, likely caused by undetectable branches that produce incomplete barcodes. It is unlikely that poor branching is caused by insufficient ligase enzymes or reagents, as the Rn18s experiment demonstrated efficient amplification of dense targets, as well as the successful detection of Eef2 using a single probe. To better understand this, we examined the completion rate of our bDNA assembly by hybridizing a single padlock probe onto Eef2 with five branching sites and performed LANTERN. We observed that many spots have incomplete branching with ~30% having complete barcodes (**Figure 2.11A**). This issue likely stems from missed binding events. In an exponentially growing bDNA system, if one amplifier fails to bind while another successfully hybridizes, the bound branch will grow much faster, creating significant variability in branch growth. Moreover, if binding probabilities vary among amplifiers, these disparities could further amplify the observed variability. Consistent with this explanation, imaging reveals substantial variation in signal intensities across branches, illustrating this heterogeneity (**Figure 2.11A**). In the 5 binding site data, we also observe that ~90% of spots have at least 2 amplifiers that colocalize. This outcome is intuitive given the likelihood of at least any two amplifiers binding is relatively high and the colocalization across many more branches will decrease when we require more successful binding events. This issue can theoretically be mitigated by increasing the number of binding sites during the initial phase of amplifier binding. In simulations, we modeled binding events as

a binomial process (p = 0.75), revealing that the coefficient of variation in branch growth decreases as the initial number of binding sites increases (**Figure 2.11B**). This suggests that employing more probes per target can increase completion rates by reducing variability in branch growth and improving the likelihood of successful binding events. Indeed, our observations indicate that using around 6–12 probes on a single Eef2 target (data not shown) leads to complete barcodes and reduced variability in branch signal. In our 1k- and 3k-plex experiments we used 12 probes per gene (150-plex had 16 probes). It could be that generating probes from a highly complex oligopool mixture can lead to probe dropouts. When coupled with poor binding efficiency of padlock probes, it could be that these experiments had less than 6 probes bound per target which can lead to incomplete barcodes due to high variability in branch growth.

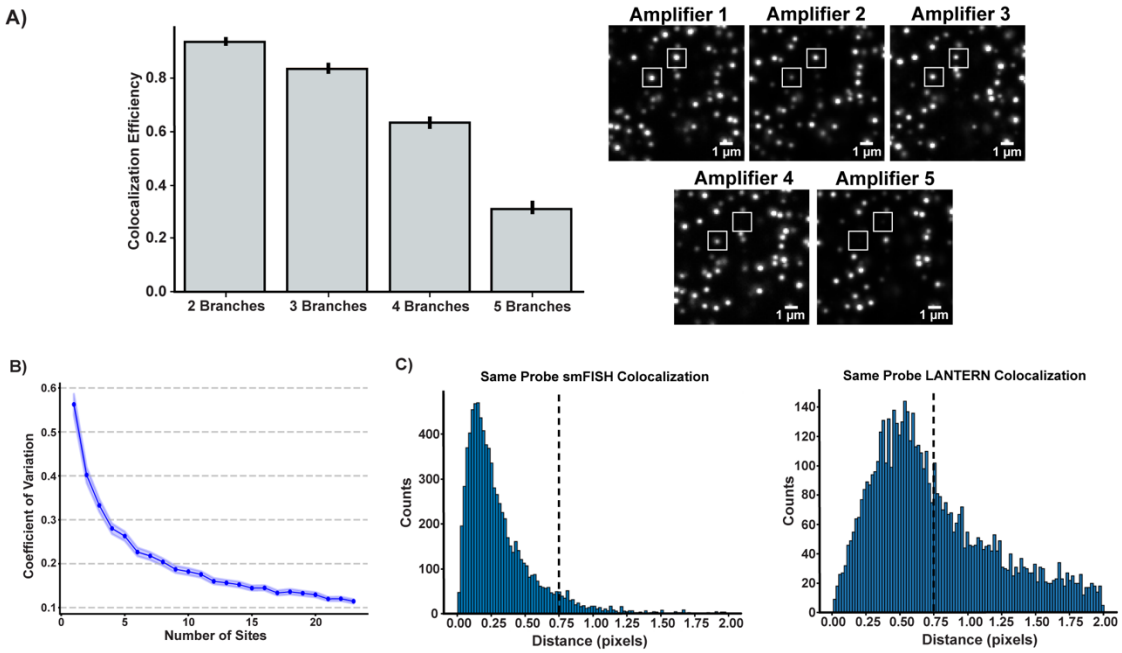


Figure 2.11: A) Changes in colocalization efficiency of each spot looking at any *N* number of branches (left) and representative image depicting heterogeneity in branching from single probe (right). Bar plot represents the mean colocalization with changing amplifier reference. Error bar corresponds to 95% C.I. B) Simulation of bDNA growth, modeling binding event probabilities as a binomial process (p = 0.75) shows reduced CV as the number of initial binding sites increases (1000 trials, 95% C.I. from bootstrapping). C) Distribution of colocalization distance between binding sites in standard smFISH (left) and LANTERN amplifiers (right). Black dashes line indicate cutoff at 0.75 pixels. 1 pixel = 103 nm.

Aside from variable branching, we observed high dispersion in colocalization distance across branches. Two primary factors may explain this separation. First, steric constraints arising from the bDNA structures could limit how closely the amplifiers can assemble. Considering the persistence length of dsDNA is around 150 bp[31] and each amplifier binding site is 30 bp, an amplifier unit that is ligated and occupied on both ends by additional amplifiers will exhibit substantial rigidity. The structural rigidity may result in amplifiers assembling in certain directions creating physical distance from each other. Second, if the initial primary probe padlock is not ligated, the branches may not remain effectively anchored to the primary probe, potentially untethering from the substrate and increasing the observed separation. When compared against smFISH colocalization between two sites on the same probe, we observed a much smaller dispersion in colocalization compared to LANTERN (**Figure 2.11C**). As the density of gene targets increases, the combined effects of high distance between branches, noise, and incomplete amplification due to high variance in branch growth can lead to poor decoding outcomes when attempting to colocalize spots. Furthermore, the additional variable of chromatic aberrations, which arise when colocalizing spots across different channels, can further hinder colocalization accuracy.

2.4 Discussion

Herein, we investigated the potential of both RCA and LANTERN for highly multiplexed signal amplification and detection of numerous RNA species. While RCA proved to be highly specific and compatible with seqFISH, we observed poor RNA detection efficiencies and amplification inefficiencies when targeting highly abundant biomolecules such as proteins and glycans. In addition, we found that RCA-generated amplicons can diffuse away from the RNA target, limiting its utility for precise subcellular localization. This challenge becomes especially significant when using a split barcode or

detection design, where the presence of two or more probes on a target determines its identity. Furthermore, the generated amplicons are not diffraction-limited (> 200 nm) which reduces spatial resolution.

We introduced LANTERN as an alternative amplification strategy to ClampFISH, replacing click chemistry with ligase enzymes and demonstrating significantly higher multiplexing by profiling more than 1,000 genes in a single experiment. Compared to RCA, LANTERN exhibited superior detection efficiency and successfully amplified dense targets such as Rn18s. Moreover, LANTERN-generated amplicons are diffraction-limited and remained colocalized with smFISH signals, making it a viable strategy for subcellular localization and split barcode/detection systems. We demonstrated its ability to profile up to 3,000 genes in NIH 3T3 cells. Moreover, we observed significantly high detection efficiency in the 150-plex experiment and moderate detection efficiency in the 1000-plex experiment. However, it became evident that LANTERN is not yet scalable for whole-transcriptome profiling due to large and variable distances between branches, and incomplete detection of all unique branches corresponding to a given barcode. We believe the bDNA design could be further optimized by shifting away from padlock-based architectures to a system that is less rigid, or a system that can tether to many targets as opposed to just the primary probe padlock. Also, if we can increase the number of initial binding sites, the rate of completion between branches will be significantly improved. Furthermore, we still want a bDNA system that is amenable to stringent wash conditions used in seqFISH and has increased number of initial amplifier binding sites. In the next chapter, we will discuss the development of Signal amPlificAtion by Recursive Crosslinking (SPARC), which addresses some of the design constraints in LANTERN and showcases a novel approach to *in situ* signal amplification.

2.6 Methods

Animals. All animal care and experiments were carried out in accordance with Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. 6-7 week old C57BL/6J male mice were obtained from The Jackson Laboratory (USA).

Readout probe design. Readout probes were designed as previously described.[22,29,32] In brief, 15 nts sequences were randomly generated and a local BLAST query was performed on each probe against the mm10 transcriptome to ensure specificity. Any probe that has 10 contiguous matching sequences were removed. Each probe had a GC content ranging from 40–60%. Readout probes were empirically assessed in-house to determine which sequences had minimal off-target labeling and noise in NIH 3T3 cells.

RNA targeting primary probe design for RCA padlocks. Transcript-specific primary probes were designed as previously described with some modifications. [22,29,32] Using the mm10 masked genome and annotation from UCSC, probe sets targeting exons within the CDS region of 151 unique transcripts for NIH3T3 were generated. For the 150 genes experiment, each probe was 32 nts in length while the Eef2 experiments were 36 nts in length. Each probe was spaced by 2 nts, and had a GC content ranging from 40–65%. A local BLAST query was performed on each probe against the mouse transcriptome to ensure specificity. BLAST hits on any sequences other than the target transcript with a 15 - 17 nts match or more were considered off target and discarded. To minimize cross-hybridization between probe sets, a local BLAST database was constructed from the probe sequences and probes with hits of 15 -17 nts or longer were removed by dropping the one of the matched probes from the larger probe set. From the above selection criteria, 12 primary probes were selected for each transcript target for the 150 genes experiment. For the Eef2 experiments, 8 probes were used. The final probes were split to two equal parts and placed accordingly to our padlock probe for subsequent ligation.

RNA targeting primary probe design for LANTERN padlocks. Transcript-specific primary probes were designed as previously described with some modifications.[22,29,32] Using the mm10 masked genome and annotation from UCSC, probe sets targeting exons within the CDS region of 3001 unique transcripts for NIH3T3 were generated. Each probe was 42 nts in length and had a GC content ranging from 40–65%, with subsequent probes spaced by 2 nts. A local BLAST query was performed on each probe against the mouse transcriptome to ensure specificity. BLAST hits on any sequences other than the target transcript with a 17 nts match or more were considered off target and discarded. To minimize cross-hybridization between probe sets, a local BLAST database was constructed from the probe sequences and probes with hits of 17 nts or longer were removed by dropping the one of the matched probes from the larger probe set. From the above selection criteria, 12-16 primary probes were selected for each transcript target. The final probes were split to two equal parts and placed accordingly to our padlock probe for subsequent ligation.

Antibody and lectin targeting primary probe design for RCA padlocks. Antibody-specific primary probes were designed as previously described with some modifications.[32] A set of 18 or 30 nts sequences were generated and a local BLAST query was performed on each probe against the mm10 transcriptome to ensure specificity. Any probe that has 10 contiguous matching sequences were removed. The sequence for rGRFT was 30 nts

in length, while Lamin-B had 18 nts. Each probe had a GC content ranging from 40–60%. Padlock probes were designed against these sequences and split to enable downstream ligation.

SeqFISH and standard smFISH probe design. RNA seqFISH padlock probes for target genes were designed similarly to our previous studies.[22,29,32] In brief, unique RNA species were encoded across 3 unique channels (647, 561, and 488 nm), similarly to RNA SPOTs,[22] where they are called at specific channels and hybridization rounds corresponding to their assigned barcodes. To implement this, RNA targeting sequences (32, 36, or 42 nts), four unique readout probe binding sites (15 nts) encoded for each RNA target, a T7 RNAP site at the 3' end, and a pair of primer binding sites at the 5′ and 3′ ends of the probe for probe generation were concatenated. If the padlock probe required perfectly blunted ends matching its RNA target, then a BspQI site is incorporated on the 3' end before the T7 site. Standard smFISH probes were designed similarly to seqFISH probes with slight modifications. In the standard design, each unique target contains one or two copies of a single unique readout binding site that is called at a given hybridization round.

Amplifier Screen Design. The amplifier screening probes were designed similarly to seqFISH and standard smFISH probes. In brief, 90 unique RNAs with 24 probes per target were used. Two sets of amplifiers were assigned per RNA target. For each probe, a 42 nts RNA targeting sequences, two unique amplifier binding sites (30 nts) encoded for each RNA target, a T7 RNAP site at the 3' end, and a pair of primer binding sites at the 5′ and 3′ ends of the probe for probe generation were concatenated.

Click primary probe synthesis*.* Primary probes were ordered as oligoarray complex pools from Twist Bioscience and were constructed as previously described with some modifications.[22,29,32] In brief, limited PCR cycles were used to amplify the probe sequences from the oligo complex pool. Then, the amplified PCR products were purified using QIAquick PCR Purification Kit (Qiagen, 28104) according to the manufacturer's instructions. The PCR products underwent *in vitro* transcription (NEB, E2050S) supplemented with RNase Inhibitor (ThermoFisher Scientific, AM2694) and Pyrophosphatase (NEB, M0361S), then treated with DNAse I (NEB, M0303S) to remove dsDNA template. The RNA product was further purified using SPRI beads (Beckman Coulter, B23318). Post purification, the RNA product underwent reverse transcription (ThermoFisher Scientific, EP7051) with a hybrid forward primer containing a 3' alkyne with both ribonucleotides and deoxyribonucleotides (purchased from IDT) and supplemented with RNase inhibitor (ThermoFisher Scientific, AM2694) and Pyrophosphatase (NEB, M0361S). After reverse transcription, the probes were alkaline hydrolyzed with 250 mM NaOH at 65°C for 20 minutes to degrade the RNA templates and cleave the forward primer leaving only the alkyne at the 5' end of the probe. The alkyne containing probes were SPRI bead purified, then underwent 3' azido-dATP (Jena Bioscience, NU-1707S) addition using terminal deoxynucleotidyl transferase (Thermo Fisher, EP0161). Next, the probes underwent final purification with SPRI beads, then resuspended in water. The probes were stored at –20°C until later use.

SplintR primary probe synthesis. Primary probes were ordered as oligoarray complex pools from Twist Bioscience and were constructed as previously described with some modifications.[22,29,32] In brief, limited PCR cycles were used to amplify the probe sequences

from the oligo complex pool. Then, the amplified PCR products were purified using QIAquick PCR Purification Kit (Qiagen, 28104) according to the manufacturer's instructions. The PCR products underwent *in vitro* transcription (NEB, E2050S) supplemented with RNase Inhibitor (ThermoFisher Scientific, AM2694) and Pyrophosphatase (NEB, M0361S), then treated with DNAse I (NEB, M0303S) to remove dsDNA template. The RNA product was further purified using SPRI beads (Beckman Coulter, B23318). Post purification, the RNA product underwent reverse transcription (ThermoFisher Scientific, EP7051) with a forward primer containing a 3' deoxyuridine (purchased from IDT) and supplemented with RNase inhibitor (Thermo Fisher, AM2694) and Pyrophosphatase (NEB, M0361S). After reverse transcription, the probes were alkaline hydrolyzed with 250 mM NaOH at 65°C for 20 minutes to degrade the RNA templates, then SPRI bead purified. The probes underwent 50 U/mL USER (NEB, M5505S) and 500 U/mL BspQI (NEB, R0712S) digestion in 1X rCutSmart buffer containing primers reverse complement to the cut sites for 18 h at 37 °C and 6 h at 50 °C. Post digestion, probes were purified again using SPRI beads. Probes were loaded onto 15% TBE-Urea PAGE gels with 1X TBE-Urea sample loading buffer. The gel ran at 100 V (24 mA) for 5 min, then 135 V (24 mA) for 5 h. The desired band was cut under Safe Imager™ 2.0 Blue-Light Transilluminator, then incubated in 1X TE buffer for 24 h at 42 C. The probes from the gel and the solution were isolated using Spin-X centrifuge tube filters. The gel pieces were rehydrated once using 1X TE and incubated at 42 °C for 1 h for further isolation. Probe solutions were ethanol precipitated, then resuspended in water and stored at -20 °C for later use.

Readout probe synthesis. Readout probes as 5' amine-modified ssDNA (15 nts) was ordered from Integrated DNA Technologies (IDT). The synthesis of dye-conjugated readout probes was performed as previously described with slight modifications.[29,32] Briefly, 5 nmols of amine-modified oligonucleotides were resuspended in 0.5 M sodium bicarbonate buffer (100 μM). Then, 5 molar excess of AlexaFluor 647 NHS ester (ThermoFisher Scientific, A20006), Cy3B NHS ester (Cytiva, PA63101), or single isoform AlexaFluor 488 TFP ester (ThermoFisher Scientific, A30005) in anhydrous DMSO (ThermoFisher Scientific, D12345) was added to the reaction and incubated at 37°C for 1h in the dark. Two consecutive additions of 5 molar excess dye were added, with each addition being at intervals of 1h (DMSO ~15% v/v). Then, acetic acid was added at the same molar amount of sodium bicarbonate to quench the buffer. The dye-conjugated ssDNA probes were subjected to ethanol precipitation with 4 μg/mL linear acrylamide (ThermoFisher Scientific, AM9520) and HPLC purification. The probes were lyophilized, then resuspended in water. The readout probes were quantified using Nanodrop and a 1 mM working stock was made. All the readout probes were kept at –20°C in amber tubes for later use.

In-gel fluorescence. Approximately 10 pmol of alkyne and azide modified oligonucleotides were mixed with 1 nmol of azide-AF488 (ThermoFisher Scientific, A10266) or alkyne-Cy5 (Vector Labs, CCT-TA116-1). Then, the CuAAc reaction was performed at 37 °C for 1h with 75 μM CuSO4, 375 μM BTTAA, 2x SSC, 3 mM ascorbate and 0.25%Triton X-100. The reaction mix was loaded onto a 15% TBE-Urea PAGE gel and ran at 100V (24 mA) for 5 min then 140V (24 mA) for 5 h. The gel was imaged using a Typhoon FLA 9000 at AF488 and Cy5 settings. After imaging, the gel was stained with 1:10,000 SYBR gold for 5 min then imaged using Safe Imager™ 2.0 Blue-Light Transilluminator.

Mass shift assay. Approximately 30 pmol of alkyne and azide modified oligonucleotides were mixed with 300 pmol of DBCO-AmC6-TTG (25 mer) in 1X PBS. The reaction proceeded for 2 h at 65 °C, then 2 h at 80 °C. The reaction mixture was loaded onto a 15% TBE-Urea PAGE gel and ran at 100 V (24 mA) for 5 min, then 150 V (24 mA) for 3 h. The gel was stained with 1:10,000 SYBR Gold in 1X TBE buffer for 5 min, then imaged under Safe Imager™ 2.0 Blue-Light Transilluminator.

Coverslip functionalization for cell culture and tissues. No 1.5H coverslips (Bioscience Tools, CSHP-No1.5-24x60) were plasma cleaned (PDC-001, Harrick Plasma) for 5 minutes. Next, coverslips were incubated with 100 µg/mL of Poly-D-lysine (Sigma, P6407) in 1X PBS at 37 °C for 2 hours. Coverslips were rinsed with deionized water, air-dried, and UV-treated in a biosafety cabinet.

Cell culture preparation. NIH/3T3 cells (ATCC) were cultured as previously described.[29] In brief, cells were placed on PDL-functionalized coverslips in high glucose DMEM media (ThermoFisher Scientific, 10569010) with 10% CCS (Cytiva, SH30087.043) and 0.1% Penicillin-Streptomycin (ThermoFisher Scientific, 15070063) for 16 hours. Cells were washed with 1X DPBS (ThermoFisher Scientific, J67802.K2) and fixed with 4% PFA (ThermoFisher Scientific 28908) in 1X PBS (ThermoFisher Scientific, AM9624) for 10 min at room temperature. Fixed cells were rinsed with 1X DPBS and incubated with 70% Ethanol in water at -20 °C overnight.

Mouse brain slice preparation. Mouse brain slices were prepared as previously described.[29,32] Mice were perfused for 8 min with perfusion buffer (10 U ml^{-1} heparin, 0.5% $NaNO_2$ (w/v) in 0.1 M PBS at 4 °C). Mice were then perfused with fresh 4% PFA in 0.1 M PBS buffer at 4 °C for 8 min. The mouse brain was removed from the skull and immediately placed in a 4% PFA buffer for 2 h at room temperature under gentle mixing. The brain was then immersed in 4 °C 30% RNase-free sucrose (Amresco, 0335) in 1X PBS until the brain sank. After the brain sank, the brain was frozen in a dry ice–isopropanol bath in OCT medium and stored at –80 °C. Ten-micrometer sections were cut using a cryotome and immediately placed on the functionalized coverslips.

Antibody conjugation. The Lamin-B1 antibody (Abcam, ab220797) and rGRFT lectin (TCI, R0229) were conjugated as previously described with slight modifications.[32] First, the antibody stock was diluted in 4 times its initial volume with ice-cold 1X PBS buffer. The solution was transferred to a 50kDa MWCO Amicon filter (Millipore Sigma, UFC905008) and concentrated. Then, the solution underwent buffer exchange 3 times with 4 times the volume of the antibody solution using ice-cold 1X PBS. Next, DBCO-PEG_4-NHS (Sigma-Aldrich 764019) in DMF was diluted in 1X PBS and 10 molar equivalents were added. The reaction was incubated for 6 hours at 4 °C. The DBCO-modified antibody was buffer exchanged with 1X PBS as mentioned previously, then 10 molar equivalents of azide-modified-oligonucleotides were added and incubated for 48 hours at 4 °C. The oligo-conjugated antibody was buffer exchanged as previously described and concentrated to desired volume. The antibody solution is supplemented 0.02 % sodium azide and stored at 4 °C for later use.

General smFISH of RNA. SmFISH experiments were performed as previously described with a few modifications.[29,32] In brief, samples were dried under compressed nitrogen to remove the 70% ethanol solution. A custom, in-house generated flow cell was attached to the coverslip. The sample was rinsed with 1X PBS three times, then post-fixed twice using 7.5 mM $BS(PEG)_5$ (ThermoFisher Scientific, A35396) in 1X PBS for 15 min at room temperature. Next, the sample was treated twice with 100 mM N-(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature. Sample was rinsed three times with 50% wash buffer (50% formamide, 2X SSC, and 0.1% Triton-X100), then blocked with pre-hybridization buffer (50% formamide, 200 nt TTG repeat sequence, 0.1 mg/mL yeast tRNA, 4x SSC, 10% 500kDa Dextran Sulfate) for 1h at 37 °C. Then, the sample was hybridized with ~5 nM/probe in pre-hybridization buffer at 37 °C for 16-36 h. After primary probe hybridization, samples were rinsed 3 times with 50% wash buffer then incubated at 37°C for 30 min. This was repeated once more. Then, the sample was rinsed several times with 4X SSC, then hybridized with readouts at 100 nM in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate, 4X SSC) for 15 min at room temperature. The sample was rinsed several times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100), stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase (Sigma ,G2133), and 1:500 catalase (Sigma, C3155)) prior to imaging.

RCA RNA experiments. Fixed cells were dried with compressed nitrogen to remove 70% ethanol. A custom, in-house generated flow cell was attached to the coverslip. Sample was rinsed three times with 10 or 20% wash buffer (10 or 20% formamide, 2X SSC, and 0.1% Triton-X100), then blocked with pre-hybridization buffer (10 or 20% formamide, 200 nt TTG repeat sequence, 0.1 mg/mL yeast tRNA, 4x SSC, 10% 6.5-10 kDa Dextran Sulfate) for 1h at 37 °C. If the RNA binding probe was 16+16 nts, then 10% formamide was used. If the RNA binding probe was 18+18 nts, then 20% formamide was used. The sample was hybridized with ~5 nM/probe in pre-hybridization buffer at 37 °C for 16-36 h. After primary probe hybridization, the sample was rinsed 3 times with 10-20% wash buffer, then two times with 5 min incubation in between at room temperature. Then, the sample was rinsed 3 times with 4X SSC and 3 times with 1X PBST (1X PBS, 0.1% Triton X-100). Sample was rinsed again with 1X SplintR buffer twice, then ligated using SplintR mix (NEB, M0375S) containing 1X SplintR buffer, 1:100 RNase Inhibitor (ThermoFisher Scientific, AM2694), and 1:10 SplintR ligase for 3h at 37 °C. Post ligation, the sample was rinsed 5 times with 60% wash buffer (60% formamide, 2X SSC, and 0.1% Triton-X100), then incubated once at 37 °C for 30 min. Sample was rinsed 3 times with 4X SSC, then 5 times with 1X PBST. Sample was incubated in RCA (ThermoFisher Scientific, A39390) pre-mix solution (1X EquiPhi29 buffer, 1 mM DTT, 0.2 mg/mL BSA, and 0.2 U/μL EquiPhi29 DNA polymerase) for 5 min at room temperature. Then, RCA complete mix (1X EquiPhi29 buffer, 1 mM DTT, 500 μM each of dNTPs, 100 μM Aminoallyl-dUTP, 1 μM of 3' phosphorothioated primer, 0.2 mg/mL BSA, and 1 U/μL EquiPhi29 DNA polymerase) was added and incubated at 42 °C for 2h. Post RCA, 7.5 mM of $BS(PEG)_5$ (ThermoFisher Scientific, A35396) in 1X PBS was added and incubate at room temp for 30 min twice. The sample was rinse several times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100), stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase (Sigma ,G2133), and 1:500 catalase (Sigma, C3155)) prior to imaging.

LANTERN RNA experiments. Fixed cells were dried with compressed nitrogen to remove 70% ethanol. A custom, in-house generated flow cell was attached to the coverslip. The sample was rinsed with 1X PBS three times, then post-fixed twice using 7.5 mM $BS(PEG)_5$ (ThermoFisher Scientific A35396) in 1X PBS for 15 min at room temperature. Next, the sample was treated twice with 100 mM N-(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature. Sample was rinsed three times with 25% wash buffer (25% formamide, 2X SSC, and 0.1% Triton-X100), then blocked with pre-hybridization buffer (25% formamide, 200 nts TTG repeat sequence, 0.1 mg/mL yeast tRNA, 4x SSC, 10% 500 kDa Dextran Sulfate) for 1h at 37 °C. Then, the sample was hybridized with ~3-5 nM/probe (azide and alkyne modified) in pre-hybridization buffer at 37 °C for 16-36 h. After primary probe hybridization, samples were rinsed 3 times with 30% wash buffer then incubated at 37 °C for 30 min. Afterwards, the sample was rinsed with 4X SSC three times, then three times with 1X PBST (1X PBS, 0.1% Triton X-100). Post washing, the sample was incubated with exonuclease digestion solution (1X rCutSmart Buffer (NEB, B6004S), 150 mM NaCl, 600 U/mL RecJF (NEB, M0264S), 400 U/mL ExoI (NEB, M0293S), 1:50 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific, AM2694), and 5% DMSO) for 1h at 37 °C. Then, the sample was washed 3 times with 2X SSCT (2X SSC, 0.25% Triton X-100) and incubated at room temperature for 5 min. The solution was replaced with CuAAc mix (75 μM $CuSO_4$, 375 μM BTTAA, 0.25% Triton X-100, 5% DMSO, 100 mM sodium ascorbate, and 2X SSC) and incubated for 1h at 37 °C. The sample was washed with 60% wash buffer (60% formamide, 2X SSC, and 0.1% Triton-X100) several times then with 2X SSC. The sample was then placed in an automated fluidics system for 8-10 exponential amplification cycles. Amplification cycles are described briefly as follows: the sample was initially washed with 2X SSC, then washed with hybridization buffer (10% ethylene carbonate, 1M betaine, 40 kDa dextran sulfate, and 4X SSC). The sample was hybridized with secondary amplifier solution (hybridization buffer with 75 nM/amplifier) for 45 min at 37 °C, then rinsed three times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100) and 1X PBST. Then the sample was rinsed once with ligation buffer (1X rCutSmart, 7.5% PEG 6000, 1:200 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific, AM2694), and 300 mM NaCl) and incubated with ligation mix (1X rCutSmart Buffer, 7.5% PEG 6000, 1:100 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific, AM2694), 1 mM ATP, 300 mM NaCl, and 150,000 U/mL of T3 DNA ligase (NEB, M0317S)) for 45 min at 37 °C. Once the amplifier was ligated, an exonuclease solution (1X rCutSmart Buffer, 300 mM NaCl, 1:100 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific, AM2694), 60 U/mL ExoI, and 240 U/mL RecJF) was added and incubated for 30 min at room temperature to digest amplifiers non-specifically bound to the cell. The sample was then washed with 60% wash buffer (60% formamide, 2X SSC, and 0.1% Triton-X100) for 5 min at room temperature, then rinsed with 2X SSC. This cycling procedure is repeated for tertiary amplifiers and subsequent rounds until desired amplification factor is reached.

Lamin-b1 RCA in NIH 3T3 cells. Seeded coverslips were dried under compressed nitrogen and custom flow cells were attached. Samples were permeabilized with 0.3% Triton-X100 in PBS for 15 min at room temperature and then incubated with blocking buffer containing 5 mg/mL UltraPure BSA (ThermoFisher Scientific, AM2616), 0.3% Triton-X100, 0.1% 6.5-10k dextran sulfate (Sigma, D4911-50G), and 0.5 mg/mL sheared salmon sperm (ThermoFisher Scientific, AM9680) in 1X PBS for an additional 30 min. Samples were incubated with labeling solution consisting of 1:100 Lamin-B1-conjugated antibody in blocking buffer with 1:100 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific,

AM2694) at 4°C overnight. Samples were washed three times at 5 min intervals with 1X PBST (1X PBS, 0.1% Triton X-100, pH 7.4), then post-fixed with 4% formaldehyde (ThermoFisher Scientific, 28908) in 1X PBS at room temperature for 5 min. The samples were washed with 1X PBS six times, then post-fixed again with 1.5 mM $BS(PEG)_9$ (ThermoFisher Scientific, A35396) in 1X PBS at room temperature for 20 min. The fixative was quenched with 100 mM Tris-HCl (Alfa Aesar, J62848) at room temperature for 5 min. Samples were rinsed with 1X PBST three times, then the antibody was optionally visualized by hybridizing 100 nM readout probes in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate, 4X SSC) for 15 min at room temperature. Post hybridization, the sample was washed with 10% wash buffer 3 times with 5 min incubations. Samples were stained with 1 μg/mL DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) prior to imaging. Post visualization, readouts are stripped using 60% wash buffer (60% formamide, 2X SSC, and 0.1% Triton-X100) by three 5 min incubation periods at room temperature, then rinsed several times with 1X PBS. Samples were rinsed with 1X PBST three times and once with 1X Quick Ligase buffer (NEB, M2200S). Quick ligation master mix containing 1X Quick Ligation Buffer, 1:20 Quick Ligase (NEB, M2200S), and 20 nM/oligo padlock probes were added to the sample and incubated at room temperature for 1 h. Post ligation, samples were rinsed three times with 1X PBS. RCA master mix containing 1X EquiPhi29 buffer, 500 μM dNTPs each (ThermoFisher Scientific, R1121), 1 mM DTT, 100 μM aminoallyl-dUTPs (ThermoFisher Scientific, R0091), 2 mg/ml BSA, and 1:10 of EquiPhi29 enzyme (ThermoFisher Scientific A39390) was added and incubated at 42 °C for 2 h. After RCA, 12.5 mM $BS(PEG)_9$ (ThermoFisher Scientific, 21582) in 1X PBS was added for 30 min, followed by the addition of 100 mM Tris-HCl for 5 min. The amplicons were washed three times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100). Readout probes at 100 nM were hybridized against the target in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate, 4X SSC) for 15 min at room temperature. Post hybridization, the sample was washed with 10% wash buffer 3 times with 5 min incubations. Samples were stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) prior to imaging.

rGRFT RCA in mouse brain tissue. Tissues were permeabilized with 70% EtOH for 1h at 4 °C, then dried using compressed nitrogen. A custom, in-house generated flow cell was attached to the coverslip. Samples were rinsed three times with TBST (1X Tris Buffered Saline, 0.1% Triton X-100, pH 7.4), then incubated in blocking solution containing 5 mg/mL UltraPure BSA (ThermoFisher, AM2616), 0.3% Triton-X100, 0.1% 6.5-10k dextran sulfate (Sigma, D4911-50G), and 0.5 mg/mL sheared salmon sperm (ThermoFisher, AM9680) in 1X TBS for 1h at room temperature. Samples were rinsed three additional times with TBST, then incubated with 20 μg/ml oligo-conjugated rGRFT in TBST with 1:100 SUPERaseIn RNase Inhibitor (ThermoFisher Scientific, AM2694) for 1h at room temperature. After incubation, samples were washed three times with TBST at 5 min intervals, then three additional rinses with PBS. Post wash, samples were fixed with 4% formaldehyde in 1X PBS at room temperature for 5 min. The samples were then washed with 1X PBS six times. Subsequently, samples were post-fixed with 1.5 mM $BS(PEG)_9$ (ThermoFisher Scientific, A35396) in 1X PBS at room temperature for 20 min, followed by quenching with 100 mM Tris-HCl pH 7.4 (Alfa Aesar, J62848) at room temperature for 5 min. Samples were rinsed with 1X PBST three times, then the lectin was

optionally visualized by hybridizing 100 nM readout probes in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate, 4X SSC) for 15 min at room temperature. Post hybridization, the sample was washed with 10% wash buffer 3 times with 5 min incubations. Samples were stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) prior to imaging. Post visualization, readouts were stripped using 60% wash buffer (60% formamide, 2X SSC, and 0.1% Triton-X100) by three 5 min incubation periods at room temperature, then rinsed several times with 1X PBS. Samples were rinsed with 1X PBST three times and once with 1X Quick Ligase buffer (NEB, M2200S). Quick ligation master mix containing 1X Quick Ligation Buffer, 1:20 Quick Ligase (NEB, M2200S), and 20 nM/oligo padlock probes were added to the sample and incubated at room temperature for 1 h. Post ligation, samples were rinsed three times with 1X PBS. RCA master mix containing 1X EquiPhi29 buffer, 500 μM dNTPs each (ThermoFisher Scientific, R1121), 1 mM DTT, 100 μM aminoallyl-dUTPs (ThermoFisher Scientific, R0091), 2 mg/ml BSA, and 1:10 of EquiPhi29 enzyme (ThermoFisher Scientific, A39390) was added and incubated at 42 °C for 2 h. After RCA, 12.5 mM $BS(PEG)_9$ (ThermoFisher Scientific, 21582) in 1X PBS was added for 30 min, followed by the addition of 100 mM Tris-HCl for 5 min. The amplicons were washed three times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100). Readout probes at 100 nM were hybridized against the target in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate, 4X SSC) for 15 min at room temperature. Post hybridization, the sample was washed with 10% wash buffer 3 times with 5 min incubations. Samples were stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) prior to imaging.

Screening or seqFISH imaging routine. In summary, the flow cell on the sample was initially connected to an automated fluidics system. The region of interest (ROI) was identified using nuclei signals stained with 3 µg/ml DAPI in 2X SSC (D8417; Sigma). For screening or seqFISH experiments, blank images containing beads were first imaged before the first round of serial hybridization. Each serial hybridization buffer contained 2 (for screen) or 3 (for seqFISH) unique readout sequences conjugated to either Alexa Fluor 647 (100 nM), Cy3B (100 nM), or Alexa Fluor 488 (100 nM) in 10% hybridization buffer (10% formamide, 10% 6-10kDa dextran sulfate, 0.1% Triton X-100, 4X SSC). A total of 150 µl of serial hybridization buffers for 8-20 rounds of seqFISH imaging (depending on number of pseudocolors), with a repeat for round 1 and cytoplasmic stain with polyT-647 (10-22 rounds in total), was pipetted into a 96-well plate. The same was performed for screening experiment except for 45 serial hybridization rounds. During each serial hybridization, the automated sampler moved to the designated well and aspirated the 150 µl hybridization buffer through a multichannel fluidic valve (EZ1213-820-4; IDEX Health & Science) to the flow cell (requires ~30 µl) using a syringe pump (63133-01, Hamilton Company). The solution was incubated for 20 minutes at room temperature. After serial hybridization, the sample was washed with approximately 1000 µl of 15% wash buffer (15% formamide and 0.1% Triton X-100 in 2× SSC) to remove excess readout probes and non-specific binding. The sample was then rinsed with about 200 µl of 4X SSC before being stained with DAPI solution (3 µg/mL of DAPI, 2X SSC) for roughly 30 seconds. An anti-bleaching buffer solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) was flowed through the samples. Imaging was performed using a Leica DMi8 microscope with various components, including a

confocal scanner unit (Yokogawa CSU-W1 or Andor Dragonfly 400), a sCMOS camera (Andor Zyla 4.2 Plus), a 63× oil objective lens (Leica 1.40 NA), and a motorized stage (ASI MS2000). Lasers from CNI and filter sets from Semrock or High Power Laser Engine and filter sets from Andor were utilized, and snapshots were acquired with 0.25-0.50 μm z-steps for five to seven z-slices per field of view across different fluorescent channels (647-nm, 561-nm, 488-nm, and 405-nm). After imaging, a stripping buffer (60% formamide and 0.1% Triton-X 100 in 2X SSC) was flowed through the sample for 1 minute, followed by a 3-minute incubation time then repeated twice more. Post readout removal, the sample was rinsed with 2X SSC solution. The serial hybridization, imaging, and signal extinguishing steps were repeated until the desired rounds were reached. The integration of the automated fluidics delivery system and imaging was controlled by a custom-written script in μManager.

Image processing. Image registration was performed similarly to previous reports with a few modifications.[33] Translation shifts were obtained using phase cross correlation (scikit-image) on DAPI stained images. Images were further aligned using affine transformation (OpenCV) on fiducial beads for each channel. Chromatic aberrations were corrected using the 647 nm channel as reference for affine transformation. Once the images and channels were aligned, a 7 x 7 high pass Gaussian filter was applied followed by a 3 x 3 low pass Gaussian filter. Image intensities across channels and serial hybridizations were used unscaled or were normalized by 80-99.999% percentile clipping and rescaling between 0 and 1.

Spot calling. Spot calling was performed similarly to previous reports.[33] Using DAOStarFinder (Photutils), we measured sub-pixel centroids and extracted various spot characteristics, including flux, peak amplitude, sharpness, bilateral and four-fold symmetry, as well as Gaussian-fit symmetry. The full width at half maximum (FWHM) was numerically optimized to determine the optimal parameter for spot detection. We then calculated the total spot area, an additional feature, by isolating each spot in a 7×7 bounding box and applying a local adaptive threshold with a Gaussian kernel. The resulting binary image was summed to obtain the spot's total area.

***Cell segmentation*.** Post seqFISH imaging, cells were stained with polyT-AF647 (IDT) in 2X SSC for 20 min and DAPI for 30 s at room temperature. Then, the sample was washed with 15% wash buffer (15% formamide, 0.1% Triton X-100, and 2X SSC). Next, anti-bleaching buffer solution (100 mM Tris HCl pH 8, 4X SSC, 2 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:500 catalase) was flowed through the sample prior to imaging. Whole cell masks or nuclear masks were generated using Cellpose 2.0 on polyT-AF647 or DAPI stained images, respectively.[34] Masks near the edges of the image were discarded to remove illumination bias or partial cells from the analysis. Additionally, 2 pixels were deleted between two or more masks that touch. Spots were mapped to each cell mask to assign each spot to a cell id.

Colocalization analysis. Sub-pixel centroids of spots obtained from image processing and spot calling (see Methods) were used for colocalization analysis by performing a nearest neighbor search algorithm with 1-2 pixel search radius (sklearn). Colocalization efficiency was calculated as the number of spots that colocalize referenced to the seed that was used for the search. The pixel distance between each spot to its neighbor was also stored for distance dispersion analysis.

Decoding. Super-resolved, mapped spots undergo SVM-embedded, feature-based symmetrical nearest neighbor decoding as previously described with a few modifications.[33] An SVM model with an RBF or polynomial kernel is used to assign probabilities to each spot, based on its measured characteristics from spot calling, indicating whether it represents a true signal or noise. False spots are filtered based on user defined likelihood cutoff. Using the cleaned-up data, a radial search (2 - 3 pixels) across barcoding rounds scores potential spots on distance, intensity, and size, particularly if multiple candidates are available for codeword assignment. Best spots are selected for each round based on their individual scores. A total codeword score is then computed by summing those individual scores, and an ambiguity score (indicating the number of additional nearby spots) is also assigned. The overall codeword score is normalized by the ambiguity score to assign confidence to the overall scores. Identified codewords (barcodes) undergo parity checks (checksum) and spot set consistency filtering (≥ 3 appearances in any given seed). If only one parity bit is missing, the barcode can still be assigned, otherwise, it remains unidentifiable. If there are any codewords that have overlapping spots that passed all criteria, then their overall codeword score was used to pick the best one. If they had the same score, then the codeword with the smallest total distance between spots was used. In the first decoding round following the above-mentioned criteria, the top 10% of highly expressed genes relative to the dataset with complete parity (codes with missing parity is ignored) is decoded first to reduce crowdedness. Unused spots (not assigned to any barcode) are resubmitted for 2 additional decoding rounds to maximize decoding outcomes and assigning clashing codewords. The second round still utilizes complete codes with parity, and the last round allows codes with loss in parity signal. Decoded spots are sorted by codeword score and subsampled to calculate FDR. Spots with FDR ≤10% are used to generate the final gene-by-cell matrix for downstream analysis.

SVM training. Quick-pass decoding assigns labeled probabilities to each spot on whether they are true or noisy signal. When 500 to 500,000 potentially noisy spots are identified, the classifier is employed. True spots are then downsampled to match the number of noisy spots, and the data is split into 80% for training and 20% for validation. Features are normalized using the MinMax Scaler, and GridSearchCV (with 8-fold cross-validation) is used to tune the C, gamma, and degree parameters for polynomial or radial basis function kernels. Finally, test set performance is evaluated using the scaling parameters derived from the training data.

False discovery rate. False discovery rate is defined as follows:

$$\mathrm{FDR} = \frac{\#\ \text{of Real Barcodes x } \dfrac{\text{Blank Counts}}{\#\ \text{of Blank Codes}}}{\text{Real Barcode Counts}}$$

Number of real barcodes are defined as the number of gene-coding barcodes in the codebook while number of blank codes correspond to the number of unassigned barcodes in the codebook. Blank counts are the number of decoded blank barcodes, and the real barcode counts are the number of decoded gene-assigned barcodes. This essentially uses the frequency of blank code counts to estimate the number of false codes that may be detected as a true code (assuming frequency of error is the same). This value is

normalized by the detection of all real barcode counts which includes both true positives and false positives.

RNA-seq on NIH 3T3. RNA was isolated from P9 NIH 3T3 cells using RNeasy kits (Qiagen, 74104) following manufacturer's instructions. The library was sequenced at 50 million reads with PE50. The reads were aligned against the GRCm38 mouse transcriptome using kallisto[35] to obtain TPM expression levels.

RNA-seq on E9.5 embryos. Sectioned OCT-embedded mouse embryos on PDL-functionalized coverslips were obtained and air-dried until the OCT became opaque. Using RNAse cleaned razors, the section was slowly removed from the coverslip to form a swiss roll. The roll was placed into an Eppendorf tube, then RNA was extracted using RNeasy kits (Qiagen, 74104) following manufacturer's instructions. The library was sequenced at 50 million reads with PE50. The reads were aligned against the GRCm38 mouse transcriptome using kallisto[35] to obtain TPM expression levels.

Simulation. Branched DNA growth was simulated for each initial site (ranging from 1 to 23) by running binomial trials (n=1000) with a set probability ($p = 0.75$). The resulting data was stored and used to compute the coefficient of variation (CV) and bootstrap confidence intervals.

pyFISH Tools: PyFISH tools is the general seqFISH and smFISH processing pipeline used in this study which can be found at https://github.com/klcolon/pyfish_tools.

Sequences. Primary probe sequences are available upon request.

2.7 References

1. Wu, C. *et al.* RollFISH achieves robust quantification of single-molecule RNA biomarkers in paraffin-embedded tumor tissue samples. *Commun. Biol.* **1**, 1–8 (2018).
2. Gyllborg, D. *et al.* Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* **48**, e112 (2020).
3. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
4. Zeng, H. *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446 (2023).
5. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014).

6. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).

7. Chen, X., Sun, Y.-C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018).

8. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

9. Schulte, S. J., Fornace, M. E., Hall, J. K., Shin, G. J. & Pierce, N. A. HCR spectral imaging: 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent samples. *Development* **151**, dev202307 (2024).

10. Wang, F. *et al.* RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).

11. Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533 (2019).

12. Attar, S. *et al.* Efficient and highly amplified imaging of nucleic acid targets in cellular and histopathological samples with pSABER. *Nat. Methods* 1–10 (2024) doi:10.1038/s41592-024-02512-2.

13. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).

14. Rouhanifard, S. H. *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry–based amplification. *Nat. Biotechnol.* **37**, 84–89 (2019).

15. Dardani, I. *et al.* ClampFISH 2.0 enables rapid, scalable amplified RNA detection in situ. *Nat. Methods* **19**, 1403–1410 (2022).

16. He, S. *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).

17. Khafizov, R. *et al.* Sub-cellular Imaging of the Entire Protein-Coding Human Transcriptome (18933-plex) on FFPE Tissue Using Spatial Molecular Imaging. 2024.11.27.625536 Preprint at https://doi.org/10.1101/2024.11.27.625536 (2024).

18. Deng, R., Zhang, K., Sun, Y., Ren, X. & Li, J. Highly specific imaging of mRNA in single cells by target RNA-initiated rolling circle amplification. *Chem. Sci.* **8**, 3668–3675 (2017).

19. Jin, J., Vaud, S., Zhelkovsky, A. M., Posfai, J. & McReynolds, L. A. Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Res.* **44**, e116 (2016).

20. Winz, M.-L., Linder, E. C., André, T., Becker, J. & Jäschke, A. Nucleotidyl transferase assisted DNA labeling with different click chemistries. *Nucleic Acids Res.* **43**, e110 (2015).

21. Sui, X. *et al.* Scalable spatial single-cell transcriptomics and translatomics in 3D thick tissue blocks. 2024.08.05.606553 Preprint at https://doi.org/10.1101/2024.08.05.606553 (2024).

22. Eng, C.-H. L., Shah, S., Thomassie, J. & Cai, L. Profiling the transcriptome by RNA SPOTs. *Nat. Methods* **14**, 1153–1155 (2017).

23. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).

24. Frieda, K. L. *et al.* Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).

25. Janesick, A. *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).

26. Chang, T. *et al.* Color-barcoded Super-multiplex RNA FISH in Three-dimensional Tissues through Single-round Imaging. 2024.06.29.601330 Preprint at https://doi.org/10.1101/2024.06.29.601330 (2024).

27. Alon, S. *et al.* Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, (2021).

28. Xenium Prime 5K Gene Expression Workflow, Analysis & Data Highlights - Official 10x Genomics Support. *10x Genomics* https://www.10xgenomics.com/support/in-situ-gene-expression/documentation/steps/assay/xenium-prime-5k-gene-expression-workflow-analysis-data-highlights.

29. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235 (2019).

30. Torre, E. *et al.* Rare Cell Detection by Single-Cell RNA Sequencing as Guided by Single-Molecule RNA FISH. *Cell Syst.* **6**, 171-179.e5 (2018).

31. Mitchell, J. S., Glowacki, J., Grandchamp, A. E., Manning, R. S. & Maddocks, J. H. Sequence-Dependent Persistence Lengths of DNA. *J. Chem. Theory Comput.* **13**, 1539–1555 (2017).

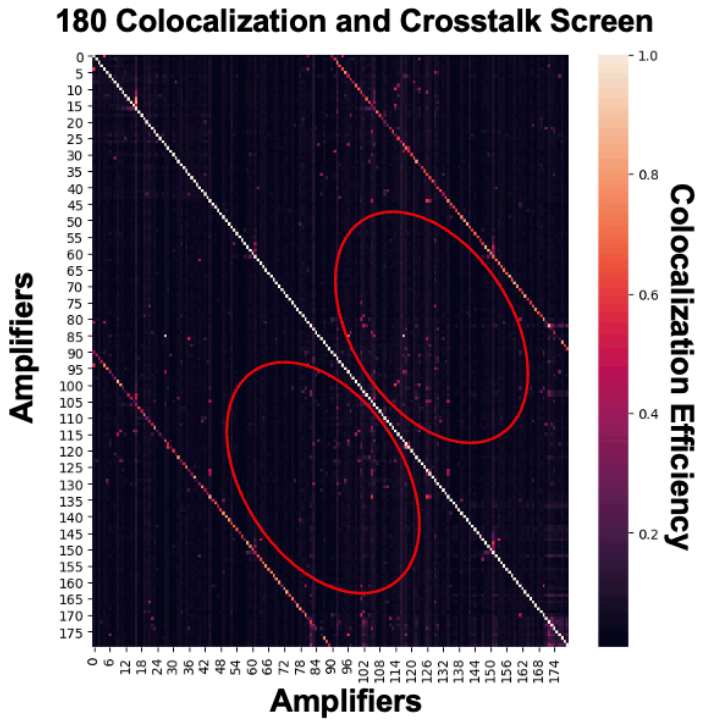
32. Takei, Y. *et al.* Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586–594 (2021).

33. Polonsky, M. *et al.* Spatial transcriptomics defines injury specific microenvironments and cellular interactions in kidney regeneration and disease. *Nat Commun* **15**, 7010 (2024).

34. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634–1641 (2022).

35. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).

2.8 Supplementary Figures and Tables



Supplemental Figure 2.1: Heatmap showing colocalization efficiency between all amplifiers. Red circles indicate regions where amplifier colocalization should not occur, yet we observe similar signals.

Name	Sequence
Forward primer for pool generation	GCCCCATCATGTGCCTTTCCT
Reverse primer for pool generation	GGCCGGTAATACGACTCACTATAG
USER forward primer for reverse transcription	GCCCCATCATGTGCCTTTCC/3deoxyU/
Alkyne forward primer for reverse transcription	GCCCCrATCArUGTGrCCTTTCrC/35OCTdU/
BspQI cleavage primer	GTGATAGGCTCTTCCCGG
USER cleavage primer	AGGAAAGGCACATGATGGGGC
RCA phosphorothioate primer	ATGATGGGCGCGTAGA*G*T

Table 2.1: Primer sequences.

Name	Sequence
Secondary1	/5Phos/ATTTCAGAGGTCTTTtcAAATAGTTGCCCAATGTCCAGGTGTAAATTcta atcAAATAGTTGCCCAATGTCCAGGTGTAAATTctAACACTGGTTGAAGC
Secondary2	/5Phos/TACAACGTACGATTTtcAAACCGAAAGCTTCAGAGGGACTTAGTATTct aatcAAACCGAAAGCTTCAGAGGGACTTAGTATTctAAAGTGTCTTACAGC
Secondary3	/5Phos/ATCACATACGCCTTTtcAAAATTGCGATCGAGGTCTAATTGCTGGTTcta atcAAAATTGCGATCGAGGTCTAATTGCTGGTTctAACAGTGAGGGAATC
Secondary4	/5Phos/GTCTACTGTATGTTTtcAAAGACGTGCTAACTGACACTTGTTACGTTcta atcAAAGACGTGCTAACTGACACTTGTTACGTTctAAAGTGCATACCAGC
Secondary5	/5Phos/TATTGAAGACGGTTTtcAAACAACAAGTGAGAGCTGATCGGTCTATTct aatcAAACAACAAGTGAGAGCTGATCGGTCTATTctAAGCCAAACGCTATC
Secondary6	/5Phos/ATGTGAAGCTACTTTtcAAAGGACACTTTATCGACTTATTGTGCCTTcta atcAAAGGACACTTTATCGACTTATTGTGCCTTctAACTTGTAGCATGTC
Secondary7	/5Phos/TCAAGGAAGCACTTTtcAAAATATGCTCGCCAGTAAAGTGGAAGCTTct aatcAAAATATGCTCGCCAGTAAAGTGGAAGCTTctAACTATGGTGCTAGC
Secondary8	/5Phos/AAATGCCGGAAGTTTtcAAATGGTTGGAGTTAGCACCTATACTGGTTct aatcAAATGGTTGGAGTTAGCACCTATACTGGTTctAACAGTCTCATAGAC
Secondary9	/5Phos/GATACACTGATCTTTtcAAATTGGATCTGGAAGTAATTATGGCGCTTcta atcAAATTGGATCTGGAAGTAATTATGGCGCTTctAAGATAGGAACGCAC
Secondary10	/5Phos/ATACTAGCTGCATTTtcAAAGCTTAATACGGAGCTGTTTTGCCTGTTcta atcAAAGCTTAATACGGAGCTGTTTTGCCTGTTctAAGCGAATTTTGCAC
Secondary11	/5Phos/CCCATATCCAGTTTTtcAAATGCATTACGGATGTTACCGCCTAATTTcta atcAAATGCATTACGGATGTTACCGCCTAATTTctAAATGAGGACAACTC
Secondary12	/5Phos/ATTCCGTCCTGTTTTtcAAACGGCAGTTCCTTGAGTATGGTAGACTTcta atcAAACGGCAGTTCCTTGAGTATGGTAGACTTctAAGCGACTAGAAATC
Secondary13	/5Phos/CTGCGAGAAACGTTTtcAAATAACCGCATTTCGATCGCGACCTAATTct aatcAAATAACCGCATTTCGATCGCGACCTAATTctAAATACGAAGGCTAC
Secondary14	/5Phos/GAACTACTAAGCTTTtcAAAGAAATAGTGGCAGATGCGGCATAATTTct aatcAAAGAAATAGTGGCAGATGCGGCATAATTTctAAGTAAGTGCGTCAC
Secondary15	/5Phos/GGACTAATTCGGTTTtcAAATGTCCCTCAGTTGATTGATCAACGGTTcta atcAAATGTCCCTCAGTTGATTGATCAACGGTTctAATGCGTCAGATAAC
Secondary16	/5Phos/GTCAGGACAATGTTTtcAAAGAGCAGGCATTGGAACCCTTACTAGTTct aatcAAAGAGCAGGCATTGGAACCCTTACTAGTTctAAATAATCAGTGGGC
Secondary17	/5Phos/TGGTGTACGGATTTTtcAAAGCACAGTATGATGTCGCAACTACTCTTcta atcAAAGCACAGTATGATGTCGCAACTACTCTTctAAGAGACTGATCAAC
Secondary18	/5Phos/TAAATGATGGCCTTTtcAAACGAGTAGAATCAGCTTCGCTACTTATTcta atcAAACGAGTAGAATCAGCTTCGCTACTTATTctAAAGAGATCACTGAC
Secondary19	/5Phos/GTTTTGCAGCCCTTTtcAAACATACGCCTGTCGAAACCTGCCTACTTct aatcAAACATACGCCTGTCGAAACCTGCCTACTTctAAGATCATGCACATC
Secondary20	/5Phos/CCCATTTTACTCCTCtcCACCTACTCACTCCAGTACCTTCCCAATAActa atcCACCTACTCACTCCAGTACCTTCCCAATAActTTCCCAATAATCCTC
Secondary21	/5Phos/CTCTATTTTCTCCATtcCTCAATCCCACATTAGATCAACACCATCTCcta atcCTCAATCCCACATTAGATCAACACCATCTCctACACCATCTCTCTAC
Secondary22	/5Phos/AGAGGGCAAAGATTTtcAAAGAGAGAAGTACCGCATGGTAAGGTTTTct aatcAAAGAGAGAAGTACCGCATGGTAAGGTTTTctAACGTACCAATAGTC
Secondary23	/5Phos/GTGCCAGGTAAGTTTtcAAAACCCTAGAAGGCGTCCTTATGCGTGTTct aatcAAAACCCTAGAAGGCGTCCTTATGCGTGTTctAAAGCGGAGGTATTC
Secondary24	/5Phos/CCTCATCAAATACTTtcCATTCATACATCCTCGAACAACATATCCCTcta atcCATTCATACATCCTCGAACAACATATCCCTctACATATCCCTTTATC
Secondary25	/5Phos/GGTATGCTTGAATTTtcAAATCGTGCGAGAGAGTTGTAAGGAGTGTTct aatcAAATCGTGCGAGAGAGTTGTAAGGAGTGTTctAATATGGGCACCATC
Secondary26	/5Phos/GGGACTTGTCACTTTtcAAAACGGCAAGTTCCGCATAAGATGACCTTct aatcAAAACGGCAAGTTCCGCATAAGATGACCTTctAAACACTTGACTTGC
Secondary27	/5Phos/ATTCAAGTGAGCTTTtcAAATGTCATAGATCCGATACTGAAGCGCTTcta atcAAATGTCATAGATCCGATACTGAAGCGCTTctAAAATACAACACCGC
Secondary28	/5Phos/AGCACAGCACCTTTTtcAAAAAAAGGCCGCTAGTACCACAACACGTTct aatcAAAAAAAGGCCGCTAGTACCACAACACGTTctAAGAGCGTAAAACTC

Secondary29	/5Phos/CAGATGCACTAATTTtcAAAACTTGATGTTCCGATCTGCGATGTATTcta atcAAAACTTGATGTTCCGATCTGCGATGTATTctAATGTCATGGTCAAC
Secondary30	/5Phos/GGAGGTCTAACGTTTtcAAAGCGTTCTGTAATGATGGGCAAATCCTTct aatcAAAGCGTTCTGTAATGATGGGCAAATCCTTctAAGTAAAGGCTGATC
Secondary31	/5Phos/CTAAACCCTCCTACCtcATCCCACTTTCATTTGTTCCATCAATTCATctaa tcATCCCACTTTCATTTGTTCCATCAATTCATctATCAATTCATTCCAC
Secondary32	/5Phos/GCTACCCAACTGTTTtcAAAAGCAGACCTTTGGCTTTGCTAAACCTTcta atcAAAAGCAGACCTTTGGCTTTGCTAAACCTTctAAGTCGCTAAAGTTC
Secondary33	/5Phos/CCAATCTACTCGTTTtcAAAACTCCTAACTTGGCGGGCATATGTTTTcta atcAAAACTCCTAACTTGGCGGGCATATGTTTTctAAAGAGCGTCGATTC
Secondary34	/5Phos/GGAACCTGACACTTTtcAAACAGATAACGTCGGCCCAGGATAGATTTct aatcAAACAGATAACGTCGGCCCAGGATAGATTTctAAGTAATACGCCATC
Secondary35	/5Phos/TTTGGTCACATCTTTtcAAAGCGATCTACTCTGCATCTAATCCCATTcta atcAAAGCGATCTACTCTGCATCTAATCCCATTctAAAAGGCATAACAGC
Secondary36	/5Phos/TGTGAGTGAACGTTTtcAAAGTCAGCAGAATCGATCCACGGGATATTct aatcAAAGTCAGCAGAATCGATCCACGGGATATTctAATCAGGTTCTAGGC
Secondary37	/5Phos/ATTGACGTCGACTTTtcAAAGTTGATTGGGTTGCGCCTTTCAAATTTcta atcAAAGTTGATTGGGTTGCGCCTTTCAAATTTctAAACTCGCTAATGAC
Secondary38	/5Phos/GACTCAGTACCATTTtcAAAAACAATCGGCTTGATGCTACTCGTGTTcta atcAAAAACAATCGGCTTGATGCTACTCGTGTTctAATTGATTCCAGTGC
Secondary39	/5Phos/GTTGCAAGTTGATTTtcAAACAGGTATTATCCGCGTCATCGAATGTTcta atcAAACAGGTATTATCCGCGTCATCGAATGTTctAAGGTTGTGAGTATC
Secondary40	/5Phos/CTTTGCCGTGTCTTTtcAAATGGACGGTTACCGAAGTGCGAACTCTTct aatcAAATGGACGGTTACCGAAGTGCGAACTCTTctAACACCAATCTGTGC
Secondary41	/5Phos/CTAACTGCGGGATTTtcAAAAATGCATGGTGGGTATTGGGTCGACTTct aatcAAAAATGCATGGTGGGTATTGGGTCGACTTctAATACTGAATCGCTC
Secondary42	/5Phos/CGGAAAGCTGAATTTtcAAATATCCGGGAACAGTGTATTGGCCTTTTct aatcAAATATCCGGGAACAGTGTATTGGCCTTTTctAAGTACCTCAAAAGC
Secondary43	/5Phos/TTTTGGGAACCTTTTtcAAAGACTACGATACTGACACGGTTCAACTTcta atcAAAGACTACGATACTGACACGGTTCAACTTctAACCGATTGATAAGC
Secondary44	/5Phos/TCTCCTATCCTACCAtcCTCAAACTTATTCACGATCTCTTACCTTAActaa tcCTCAAACTTATTCACGATCTCTTACCTTAActCTTACCTTAACCTAC
Secondary45	/5Phos/TTACTATTCTCCATCtcTCAATCAATTCACCAGATTTCTACAACCTActaa tcTCAATCAATTCACCAGATTTCTACAACCTActCTACAACCTACCCAC
Secondary46	/5Phos/TCCTCTCCCTTATCAtcTTCCTATCCACCCACGTTCTTTTTTTCTCCctaa tcTTCCTATCCACCCACGTTCTTTTTTTCTCCctTTTTTCTCCTTCAC
Secondary47	/5Phos/CTATACTATCCATCCtcCCCACTTTCTCATATGTTACCTAAATCCCTctaa tcCCCACTTTCTCATATGTTACCTAAATCCCTctCTAAATCCCTATCTC
Secondary48	/5Phos/CCCTTCTCATACCAAtcATTCCCTCCTATCCTGACACCCAATACACTcta atcATTCCCTCCTATCCTGACACCCAATACACTctCCAATACACTTCCTC
Secondary49	/5Phos/GTTGTGACATCTTTTtcAAACGGGTTGGGAATGTTACCTCAGATCTTcta atcAAACGGGTTGGGAATGTTACCTCAGATCTTctAAGCATGCATGAAAC
Secondary50	/5Phos/CGCTGGACATAGTTTtcAAAAAAGGTGTGGAGGCACTTTCTGGACTTct aatcAAAAAAGGTGTGGAGGCACTTTCTGGACTTctAATTCATCCTTTGGC
Secondary51	/5Phos/GATCGTGGGAAGTTTtcAAAAAGTTGACGTCTGTCCAAGGCAAAATTct aatcAAAAAGTTGACGTCTGTCCAAGGCAAAATTctAACAATACGGGTAAC
Secondary52	/5Phos/GGAAGTTGGCGATTTtcAAAGTGCCTTTGGGTGTGTGATCCACTCTTct aatcAAAGTGCCTTTGGGTGTGTGATCCACTCTTctAATCTACGGAAGAAC
Secondary53	/5Phos/AAGCCTTCTTAGTTTtcAAACCGAACCTTATCGATATCCCCTAAGTTcta atcAAACCGAACCTTATCGATATCCCCTAAGTTctAAGAACTACAGTTGC
Secondary54	/5Phos/CAATCATTAACTTCCtcCATTAAATCCCTTTCGATCCATAAACTAACctaa tcCATTAAATCCCTTTCGATCCATAAACTAACctATAAACTAACTCCAC
Secondary55	/5Phos/TCCCTATACAAACACtcTTACATCTCCCACTAGTTCCTCCACCCAATcta atcTTACATCTCCCACTAGTTCCTCCACCCAATctTCCACCCAATATCAC
Secondary56	/5Phos/AACCTCTCTTCTACAtcCCCTTTACTTCTCCCGATTCATTCAACACCcta atcCCCTTTACTTCTCCCGATTCATTCAACACCctATTCAACACCTCCTC
Secondary57	/5Phos/CTGGTTATCTCGTTTtcAAACCCTGTAAGAACGTGAGGATTGCTCTTcta atcAAACCCTGTAAGAACGTGAGGATTGCTCTTctAAGTGAAAGAGGTGC

Secondary58	/5Phos/ATCCCAAATAACACCtcTCCACCTCACACTCCGTCATTAATTCCCACcta atcTCCACCTCACACTCCGTCATTAATTCCCACctTAATTCCCACTTTAC
Secondary59	/5Phos/CCTTTCCAAATCACCtcCCTTTCTACTATACAGTTCTCTCCCATACTcta atcCCTTTCTACTATACAGTTCTCTCCCATACTctCTCCCATACTTCTAC
Secondary60	/5Phos/ACCTCTTCTCACCCAtcTTCCCACACCAATACGACATTTACCTCACCcta atcTTCCCACACCAATACGACATTTACCTCACCctTTACCTCACCTTCAC

Table 2.2: Top 60 secondary amplifier sequences.

Name	Sequence
Tertiary1	/5Phos/ATTGGGCAACTATTTtcAAAGACCTCTGAAATGCTTCAACCAGTGTTctaatcAA AGACCTCTGAAATGCTTCAACCAGTGTTctAATTTACACCTGGAC
Tertiary2	/5Phos/TGAAGCTTTCGGTTTtcAAATCGTACGTTGTAGCTGTAAGACACTTTctaatcAA ATCGTACGTTGTAGCTGTAAGACACTTTctAATACTAAGTCCCTC
Tertiary3	/5Phos/CTCGATCGCAATTTTtcAAAGGCGTATGTGATGATTCCCTCACTGTTctaatcAA AGGCGTATGTGATGATTCCCTCACTGTTctAACCAGCAATTAGAC
Tertiary4	/5Phos/AGTTAGCACGTCTTTtcAAACATACAGTAGACGCTGGTATGCACTTTctaatcAA ACATACAGTAGACGCTGGTATGCACTTTctAACGTAACAAGTGTC
Tertiary5	/5Phos/TCTCACTTGTTGTTTtcAAACCGTCTTCAATAGATAGCGTTTGGCTTctaatcAAA CCGTCTTCAATAGATAGCGTTTGGCTTctAATAGACCGATCAGC
Tertiary6	/5Phos/GATAAAGTGTCCTTTtcAAAGTAGCTTCACATGACATGCTACAAGTTctaatcAA AGTAGCTTCACATGACATGCTACAAGTTctAAGGCACAATAAGTC
Tertiary7	/5Phos/TGGCGAGCATATTTTtcAAAGTGCTTCCTTGAGCTAGCACCATAGTTctaatcAA AGTGCTTCCTTGAGCTAGCACCATAGTTctAAGCTTCCACTTTAC
Tertiary8	/5Phos/TAACTCCAACCATTTtcAAACTTCCGGCATTTGTCTATGAGACTGTTctaatcAAA CTTCCGGCATTTGTCTATGAGACTGTTctAACCAGTATAGGTGC
Tertiary9	/5Phos/TTCCAGATCCAATTTtcAAAGATCAGTGTATCGTGCGTTCCTATCTTctaatcAAA GATCAGTGTATCGTGCGTTCCTATCTTctAAGCGCCATAATTAC
Tertiary10	/5Phos/TCCGTATTAAGCTTTtcAAATGCAGCTAGTATGTGCAAAATTCGCTTctaatcAA ATGCAGCTAGTATGTGCAAAATTCGCTTctAACAGGCAAAACAGC
Tertiary11	/5Phos/ATCCGTAATGCATTTtcAAAACTGGATATGGGGAGTTGTCCTCATTTctaatcAA AACTGGATATGGGGAGTTGTCCTCATTTctAAATTAGGCGGTAAC
Tertiary12	/5Phos/AAGGAACTGCCGTTTtcAAAACAGGACGGAATGATTTCTAGTCGCTTctaatcAA AACAGGACGGAATGATTTCTAGTCGCTTctAAGTCTACCATACTC
Tertiary13	/5Phos/GAAATGCGGTTATTTtcAAACGTTTCTCGCAGGTAGCCTTCGTATTTctaatcAA ACGTTTCTCGCAGGTAGCCTTCGTATTTctAATTAGGTCGCGATC
Tertiary14	/5Phos/TGCCACTATTTCTTTtcAAAGCTTAGTAGTTCGTGACGCACTTACTTctaatcAAA GCTTAGTAGTTCGTGACGCACTTACTTctAAATTATGCCGCATC
Tertiary15	/5Phos/AACTGAGGGACATTTtcAAACCGAATTAGTCCGTTATCTGACGCATTctaatcAA ACCGAATTAGTCCGTTATCTGACGCATTctAACCGTTGATCAATC
Tertiary16	/5Phos/CAATGCCTGCTCTTTtcAAACATTGTCCTGACGCCCACTGATTATTTctaatcAA ACATTGTCCTGACGCCCACTGATTATTTctAACTAGTAAGGGTTC
Tertiary17	/5Phos/ATCATACTGTGCTTTtcAAAATCCGTACACCAGTTGATCAGTCTCTTctaatcAAA ATCCGTACACCAGTTGATCAGTCTCTTctAAGAGTAGTTGCGAC
Tertiary18	/5Phos/TGATTCTACTCGTTTtcAAAGGCCATCATTTAGTCAGTGATCTCTTTctaatcAAA GGCCATCATTTAGTCAGTGATCTCTTTctAATAAGTAGCGAAGC
Tertiary19	/5Phos/GACAGGCGTATGTTTtcAAAGGGCTGCAAAACGATGTGCATGATCTTctaatcA AAGGGCTGCAAAACGATGTGCATGATCTTctAAGTAGGCAGGTTTC
Tertiary20	/5Phos/TGGAGTGAGTAGGTGtcGAGGAGTAAAATGGGGAGGATTATTGGGAActaatc GAGGAGTAAAATGGGGAGGATTATTGGGAActTTATTGGGAAGGTAC
Tertiary21	/5Phos/TAATGTGGGATTGAGtcATGGAGAAAATAGAGGTAGAGAGATGGTGTctaatcA TGGAGAAAATAGAGGTAGAGAGATGGTGTctGAGATGGTGTTGATC
Tertiary22	/5Phos/GGTACTTCTCTCTTTtcAAATCTTTGCCCTCTGACTATTGGTACGTTctaatcAAA TCTTTGCCCTCTGACTATTGGTACGTTctAAAACCTTACCATGC
Tertiary23	/5Phos/GCCTTCTAGGGTTTTtcAAACTTACCTGGCACGAATACCTCCGCTTTctaatcAA ACTTACCTGGCACGAATACCTCCGCTTTctAACACGCATAAGGAC

Tertiary24	/5Phos/GAGGATGTATGAATGtcAAGTATTTGATGAGGGATAAAGGGATATGTctaatcA AGTATTTGATGAGGGATAAAGGGATATGTctAGGGATATGTTGTTC
Tertiary25	/5Phos/TCTCTCGCACGATTTtcAAATTCAAGCATACCGATGGTGCCCATATTctaatcAA ATTCAAGCATACCGATGGTGCCCATATTctAACACTCCTTACAAC
Tertiary26	/5Phos/GGAACTTGCCGTTTTtcAAAGTGACAAGTCCCGCAAGTCAAGTGTTTctaatcAA AGTGACAAGTCCCGCAAGTCAAGTGTTTctAAGGTCATCTTATGC
Tertiary27	/5Phos/GGATCTATGACATTTtcAAAGCTCACTTGAATGCGGTGTTGTATTTTctaatcAA AGCTCACTTGAATGCGGTGTTGTATTTTctAAGCGCTTCAGTATC
Tertiary28	/5Phos/TAGCGGCCTTTTTTTtcAAAAGGTGCTGTGCTGAGTTTTACGCTCTTctaatcAA AAGGTGCTGTGCTGAGTTTTACGCTCTTctAACGTGTTGTGGTAC
Tertiary29	/5Phos/GGAACATCAAGTTTTtcAAATTAGTGCATCTGGTTGACCATGACATTctaatcAA ATTAGTGCATCTGGTTGACCATGACATTctAATACATCGCAGATC
Tertiary30	/5Phos/ATTACAGAACGCTTTtcAAACGTTAGACCTCCGATCAGCCTTTACTTctaatcAA ACGTTAGACCTCCGATCAGCCTTTACTTctAAGGATTTGCCCATC
Tertiary31	/5Phos/AAATGAAAGTGGGATtcGGTAGGAGGGTTTAGGTGGAATGAATTGATctaatcG GTAGGAGGGTTTAGGTGGAATGAATTGATctATGAATTGATGGAAC
Tertiary32	/5Phos/CAAAGGTCTGCTTTTtcAAACAGTTGGGTAGCGAACTTTAGCGACTTctaatcAA ACAGTTGGGTAGCGAACTTTAGCGACTTctAAGGTTTAGCAAAGC
Tertiary33	/5Phos/CAAGTTAGGAGTTTTtcAAACGAGTAGATTGGGAATCGACGCTCTTTctaatcAA ACGAGTAGATTGGGAATCGACGCTCTTTctAAAACATATGCCCGC
Tertiary34	/5Phos/CGACGTTATCTGTTTtcAAAGTGTCAGGTTCCGATGGCGTATTACTTctaatcAA AGTGTCAGGTTCCGATGGCGTATTACTTctAAATCTATCCTGGGC
Tertiary35	/5Phos/AGAGTAGATCGCTTTtcAAAGATGTGACCAAAGCTGTTATGCCTTTTctaatcAA AGATGTGACCAAAGCTGTTATGCCTTTTctAATGGGATTAGATGC
Tertiary36	/5Phos/GATTCTGCTGACTTTtcAAACGTTCACTCACAGCCTAGAACCTGATTctaatcAA ACGTTCACTCACAGCCTAGAACCTGATTctAATATCCCGTGGATC
Tertiary37	/5Phos/AACCCAATCAACTTTtcAAAGTCGACGTCAATGTCATTAGCGAGTTTctaatcAA AGTCGACGTCAATGTCATTAGCGAGTTTctAAATTTGAAAGGCGC
Tertiary38	/5Phos/AAGCCGATTGTTTTTtcAAATGGTACTGAGTCGCACTGGAATCAATTctaatcAA ATGGTACTGAGTCGCACTGGAATCAATTctAACACGAGTAGCATC
Tertiary39	/5Phos/GGATAATACCTGTTTtcAAATCAACTTGCAACGATACTCACAACCTTctaatcAA ATCAACTTGCAACGATACTCACAACCTTctAACATTCGATGACGC
Tertiary40	/5Phos/GGTAACCGTCCATTTtcAAAGACACGGCAAAGGCACAGATTGGTGTTctaatcA AAGACACGGCAAAGGCACAGATTGGTGTTctAAGAGTTCGCACTTC
Tertiary41	/5Phos/CCACCATGCATTTTTtcAAATCCCGCAGTTAGGAGCGATTCAGTATTctaatcAA ATCCCGCAGTTAGGAGCGATTCAGTATTctAAGTCGACCCAATAC
Tertiary42	/5Phos/TGTTCCCGGATATTTtcAAATTCAGCTTTCCGGCTTTTGAGGTACTTctaatcAA ATTCAGCTTTCCGGCTTTTGAGGTACTTctAAAAGGCCAATACAC
Tertiary43	/5Phos/AGTATCGTAGTCTTTtcAAAAGGTTCCCAAAAGCTTATCAATCGGTTctaatcAA AAGGTTCCCAAAAGCTTATCAATCGGTTctAAGTTGAACCGTGTC
Tertiary44	/5Phos/GTGAATAAGTTTGAGtcTGGTAGGATAGGAGAGTAGGTTAAGGTAAGctaatcT GGTAGGATAGGAGAGTAGGTTAAGGTAAGctTAAGGTAAGAGATC
Tertiary45	/5Phos/TGGTGAATTGATTGAtcGATGGAGAATAGTAAGTGGGTAGGTTGTAGctaatcG ATGGAGAATAGTAAGTGGGTAGGTTGTAGctTAGGTTGTAGAAATC
Tertiary46	/5Phos/GTGGGTGGATAGGAAtcTGATAAGGGAGAGGAGTGAAGGAGAAAAAActaatc TGATAAGGGAGAGGAGTGAAGGAGAAAAAActGGAGAAAAAAAGAAC
Tertiary47	/5Phos/ATATGAGAAAGTGGGtcGGATGGATAGTATAGGAGATAGGGATTTAGctaatcG GATGGATAGTATAGGAGATAGGGATTTAGctAGGGATTTAGGTAAC
Tertiary48	/5Phos/AGGATAGGAGGGAATtcTTGGTATGAGAAGGGGAGGAAGTGTATTGGctaatc TTGGTATGAGAAGGGGAGGAAGTGTATTGGctAGTGTATTGGGTGTC
Tertiary49	/5Phos/ATTCCCAACCCGTTTtcAAAAGATGTCACAACGTTTCATGCATGCTTctaatcAA AAGATGTCACAACGTTTCATGCATGCTTctAAGATCTGAGGTAAC
Tertiary50	/5Phos/CTCCACACCTTTTTTtcAAACTATGTCCAGCGGCCAAAGGATGAATTctaatcAA ACTATGTCCAGCGGCCAAAGGATGAATTctAAGTCCAGAAAGTGC
Tertiary51	/5Phos/AGACGTCAACTTTTTtcAAACTTCCCACGATCGTTACCCGTATTGTTctaatcAA ACTTCCCACGATCGTTACCCGTATTGTTctAATTTTGCCTTGGAC
Tertiary52	/5Phos/ACCCAAAGGCACTTTtcAAATCGCCAACTTCCGTTCTTCCGTAGATTctaatcAA ATCGCCAACTTCCGTTCTTCCGTAGATTctAAGAGTGGATCACAC

Tertiary53	/5Phos/GATAAGGTTCGGTTTtcAAACTAAGAAGGCTTGCAACTGTAGTTCTTctaatcAA ACTAAGAAGGCTTGCAACTGTAGTTCTTctAACTTAGGGGATATC
Tertiary54	/5Phos/GAAAGGGATTTAATGtcGGAAGTTAATGATTGGTGGAGTTAGTTTATctaatcGG AAGTTAATGATTGGTGGAGTTAGTTTATctGTTAGTTTATGGATC
Tertiary55	/5Phos/TAGTGGGAGATGTAAtcGTGTTTGTATAGGGAGTGATATTGGGTGGActaatcG TGTTTGTATAGGGAGTGATATTGGGTGGActATTGGGTGGAGGAAC
Tertiary56	/5Phos/GGGAGAAGTAAAGGGtcTGTAGAAGAGAGGTTGAGGAGGTGTTGAATctaatc TGTAGAAGAGAGGTTGAGGAGGTGTTGAATctGGTGTTGAATGAATC
Tertiary57	/5Phos/GTTCTTACAGGGTTTtcAAACGAGATAACCAGGCACCTCTTTCACTTctaatcAA ACGAGATAACCAGGCACCTCTTTCACTTctAAGAGCAATCCTCAC
Tertiary58	/5Phos/GGAGTGTGAGGTGGAtcGGTGTTATTTGGGATGTAAAGTGGGAATTActaatcG GTGTTATTTGGGATGTAAAGTGGGAATTActGTGGGAATTAATGAC
Tertiary59	/5Phos/TGTATAGTAGAAAGGtcGGTGATTTGGAAAGGGTAGAAGTATGGGAGctaatcG GTGATTTGGAAAGGGTAGAAGTATGGGAGctAGTATGGGAGAGAAC
Tertiary60	/5Phos/GTATTGGTGTGGGAAtcTGGGTGAGAAGAGGTGTGAAGGTGAGGTAActaatc TGGGTGAGAAGAGGTGTGAAGGTGAGGTAActGGTGAGGTAAATGTC

Table 2.3: Top 60 tertiary amplifier sequences.

Name	Probe Sequence
Amp1_647	/5AmMC6/ATTTCAGAGGTCTTT
Amp2_561	/5AmMC6/TACAACGTACGATTT
Amp3_488	/5AmMC6/ATCACATACGCCTTT
Amp4_647	/5AmMC6/GTCTACTGTATGTTT
Amp5_561	/5AmMC6/TATTGAAGACGGTTT
Amp6_488	/5AmMC6/ATGTGAAGCTACTTT
Amp7_647	/5AmMC6/TCAAGGAAGCACTTT
Amp8_561	/5AmMC6/AAATGCCGGAAGTTT
Amp9_488	/5AmMC6/GATACACTGATCTTT
Amp10_647	/5AmMC6/ATACTAGCTGCATTT
Amp11_561	/5AmMC6/CCCATATCCAGTTTT
Amp12_488	/5AmMC6/ATTCCGTCCTGTTTT
Amp13_647	/5AmMC6/CTGCGAGAAACGTTT
Amp14_561	/5AmMC6/GAACTACTAAGCTTT
Amp15_488	/5AmMC6/GGACTAATTCGGTTT
Amp16_647	/5AmMC6/GTCAGGACAATGTTT
Amp17_561	/5AmMC6/TGGTGTACGGATTTT
Amp18_488	/5AmMC6/TAAATGATGGCCTTT
Amp19_647	/5AmMC6/GTTTTGCAGCCCTTT
Amp20_561	/5AmMC6/CCCATTTTACTCCTC
Amp21_488	/5AmMC6/CTCTATTTTCTCCAT
Amp22_647	/5AmMC6/AGAGGGCAAAGATTT
Amp23_561	/5AmMC6/GTGCCAGGTAAGTTT
Amp24_488	/5AmMC6/CCTCATCAAATACTT
Amp25_647	/5AmMC6/GGTATGCTTGAATTT

Amp26_561	/5AmMC6/GGGACTTGTCACTTT
Amp27_488	/5AmMC6/ATTCAAGTGAGCTTT
Amp28_647	/5AmMC6/AGCACAGCACCTTTT
Amp29_561	/5AmMC6/CAGATGCACTAATTT
Amp30_488	/5AmMC6/GGAGGTCTAACGTTT
Amp31_647	/5AmMC6/CTAAACCCTCCTACC
Amp32_561	/5AmMC6/GCTACCCAACTGTTT
Amp33_488	/5AmMC6/CGATTCCCAATCTAC
Amp34_647	/5AmMC6/GGAACCTGACACTTT
Amp35_561	/5AmMC6/TTTGGTCACATCTTT
Amp36_488	/5AmMC6/TGTGAGTGAACGTTT
Amp37_647	/5AmMC6/ATTGACGTCGACTTT
Amp38_561	/5AmMC6/GACTCAGTACCATTT
Amp39_488	/5AmMC6/GTTGCAAGTTGATTT
Amp40_647	/5AmMC6/CTTTGCCGTGTCTTT
Amp41_561	/5AmMC6/TCCTAACTGCGGGAT
Amp42_488	/5AmMC6/CGGAAAGCTGAATTT
Amp43_647	/5AmMC6/TTTTGGGAACCTTTT
Amp44_561	/5AmMC6/TCTCCTATCCTACCA
Amp45_488	/5AmMC6/TTACTATTCTCCATC
Amp46_647	/5AmMC6/TCCTCTCCCTTATCA
Amp47_561	/5AmMC6/CTATACTATCCATCC
Amp48_488	/5AmMC6/CCCTTCTCATACCAA
Amp49_647	/5AmMC6/GTTGTGACATCTTTT
Amp50_561	/5AmMC6/CGCTGGACATAGTTT
Amp51_488	/5AmMC6/GATCGTGGGAAGTTT
Amp52_647	/5AmMC6/GGAAGTTGGCGATTT
Amp53_561	/5AmMC6/AAGCCTTCTTAGTTT
Amp54_488	/5AmMC6/CAATCATTAACTTCC
Amp55_647	/5AmMC6/TCCCTATACAAACAC
Amp56_561	/5AmMC6/AACCTCTCTTCTACA
Amp57_488	/5AmMC6/CTGGTTATCTCGTTT
Amp58_647	/5AmMC6/ATCCCAAATAACACC
Amp59_561	/5AmMC6/CCTTTCCAAATCACC
Amp60_488	/5AmMC6/ACCTCTTCTCACCCA

Table 2.4: Readout sequences for LANTERN.

Name	Sequence
RCA-SP1-647	/5AmMC6/TAATCAAAGGCCGCA
RCA-SP2-561	/5AmMC6/TGCTCCCGGATTACA
RCA-SP3-488	/5AmMC6/CACTCGTACAGCGTT
RCA-SP4-647	/5AmMC6/CCCTCGTACGCCTAT
RCA-SP5-561	/5AmMC6/AATCCAGACGGGATG
RCA-SP6-488	/5AmMC6/TAGAATTCGCGGCCA
RCA-SP7-647	/5AmMC6/AATCAGGATACGGCG
RCA-SP8-561	/5AmMC6/GGTCTCGGATATGCA
RCA-SP9-488	/5AmMC6/TGCGTACGACAATGT
RCA-SP10-647	/5AmMC6/TATCATACGACGTGG
RCA-SP11-561	/5AmMC6/TCAATTGGGCGCTCA
RCA-SP12-488	/5AmMC6/CTCTAGGATCGTACG
RCA-SP13-647	/5AmMC6/TGACTCGCACTGGGA
RCA-SP14-561	/5AmMC6/TGGTCGAGGTATTCG
RCA-SP15-488	/5AmMC6/ATATTCAGGCTAGCG
RCA-SP16-647	/5AmMC6/ACGCTAGTAACGTCT
RCA-SP17-561	/5AmMC6/CCTAGGCATATTCGA
RCA-SP18-488	/5AmMC6/AGTACTCTTACGCGA
RCA-SP19-647	/5AmMC6/GTCGTAATAGCACGC
RCA-SP20-561	/5AmMC6/AGCACCTTTACATCG
RCA-SP21-488	/5AmMC6/CCACGATCGCTGATT
RCA-SP22-647	/5AmMC6/TGGCCCATACTCGTT
RCA-SP23-561	/5AmMC6/CTTCAGCGATAGGCA
RCA-SP24-488	/5AmMC6/AGTCGTATACGCGTG

Table 2.5: Readout sequences for RCA.

Name	Sequence
Lamin-B1 oligo	AAACAAATCTCCAACCTC
rGRFT oligo	GTTTGGAGGAGATCGGAGCTACTTTCCAGC

Table 2.6: Sequences conjugated to antibodies.

Chapter III

SPARC enables enzyme-free and photochemical signal amplification for multiplexed detection of transcript and proteins

This was a collaborative effort with Carsten Tischbirek. Contribution to figures that was not primarily obtained by me is outlined in the figure caption.

3.1 Abstract

Single-molecule Fluorescent *In Situ* Hybridization (smFISH) is a versatile, widely adopted method for detecting various biomolecules in cells while preserving their native spatial context. Despite its strengths, smFISH often suffers from a low signal-to-noise ratio, which can vary depending on tissue type and gene targets. It also has limited throughput due to the need for long exposure times or high-magnification objectives. These limitations can be addressed by a signal amplification strategy that (1) significantly improves signal-to-noise ratios in suboptimal samples, (2) reduces imaging duration by shortening exposure times or permitting the use of lower magnification objectives, and (3) requires fewer probes to detect genes of interest. In this work, we present Signal amPlificAtion by Recursive Crosslinking (SPARC), a method that relies on iterative amplifier deposition and bDNA assembly, with concurrent photocrosslinking to the extracellular matrix to form highly stable structures. SPARC is enzyme-free and uses short amplifier strands for enhanced diffusion (< 50 nts). It also incorporates multiple benzophenone groups into the amplifiers, enabling rapid UV crosslinking (< 5 minutes) and the amplifiers have significant stability at room temperature. Using this approach, we demonstrate more than 500-fold *in situ* signal amplification. We further show that SPARC-assembled structures are highly stable and remain intact under stringent denaturing conditions. With SPARC, we demonstrate its multiplexing capabilities by profiling ~150 genes in NIH 3T3 cells and mouse brain tissues. Moreover, SPARC not only amplifies single transcript molecules, but also amplifies antibody signals, demonstrating its versatility.

3.2 Introduction

In recent years, imaging-based spatial transcriptomics has emerged as a powerful approach for studying the spatial organization of gene expression, providing critical insights into cellular heterogeneity, organization, and function. Despite the widespread adoption of spatial transcriptomic techniques that rely on single-molecule fluorescent *in situ* hybridization (smFISH), these methods still face several challenges. These limitations include a low signal-to-noise ratio in suboptimal samples, the need for multiple probes to detect specific genes, and low sample throughput due to long exposure times and the requirement for high-magnification objectives. To circumvent these challenges, many researchers have focused on developing robust *in situ* amplification strategies, which can be broadly classified into enzymatic and non-enzymatic approaches. Enzymatic amplification methods commonly involve Rolling Circle Amplification (RCA), such as RollFISH,[1] HybISS,[2] STARmap,[3,4] FISSEQ,[5,6] and BaristaSeq,[7] among others. Although these methods are powerful, they can be limited by suboptimal detection efficiencies and the formation of large (>200 nm) "DNA nanoballs," which can reduce spatial resolution.[3,6,8–10] Alternatively, non-enzymatic methods such as third-generation HCR[11] and HCR spectral imaging[12] can achieve high detection efficiency in deep tissue. However, HCR has limited capacity for multiplexing thousands of genes and often limited by extensive incubation times regarding signal removal.[13] In addition to RCA- and HCR-based methods, branched DNA (bDNA) approaches have gained considerable momentum. Techniques such as RNAScope,[14] SABER,[15,16] ClampFISH,[17,18] π-FISH,[19] and SMI probes[20,21] exemplify these strategies. Notably, most *in situ* bDNA methods, with the exception of ClampFISH, are not compatible with harsh denaturing wash conditions. SMI probes, on the other hand, rely on *ex situ* generation of bDNA strands, which can be problematic for samples thicker than 10 microns. Therefore, we sought create a novel, enzyme-free and hydrogel-free

amplification strategy that utilizes very short amplifier strands to generate unique DNA assemblies that is highly stable and amenable to denaturing wash conditions. By using short amplifier strands, we can reduce non-specific binding, improve diffusion capabilities, and lower costs. Furthermore, we wanted to create an amplification platform that can translate to other biomolecules such as proteins.

Here, we present a novel method called Signal amPlificAtion by Recursive Crosslinking (SPARC), which enables robust, amplified detection of both transcripts and proteins. SPARC begins by hybridizing photoreactive amplifiers to a primary probe bound to some target. Upon binding, the amplifier is crosslinked to the extracellular matrix, then undergoes toehold-mediated strand displacement to remove the amplifier and re-open the amplifier binding site. This process can happen iteratively to deposit many amplifiers near the target. Then, bDNA structures are assembled on top of the deposited amplifier scaffold with each amplifier layer crosslinked to the extracellular matrix, following an exponential growth process (**Figure 3.1A**). With this, SPARC can achieve both linear and exponential growth of DNA nanostructures. SPARC offers several advantages and unique features over current methods. First, SPARC demonstrates its ability to repeatedly deposit amplifier strands near the primary probe which has not been explicitly shown in current literature. Through this repeated targeting of primary probes, more amplifiers accumulate at the true target, thereby increasing amplification specificity and reducing non-specific signals. This iterative amplification process also increases the number of initial binding sites for secondary amplifiers, which should theoretically reduce the overall variance in exponential bDNA assembly across different branches (**Supplementary Figure 3.1**). Second, the SPARC-generated structures are highly stable, withstanding multiple denaturing wash conditions, making it amenable to seqFISH. Theoretically, even if the original target degrades, the crosslinked DNA structure remains intact as long as the extracellular matrix

is preserved. Finally, SPARC uses short amplifier strands (< 50 nts), which enhances their diffusion in samples, further improving detection sensitivity and reducing hybridization time. Using SPARC, we demonstrate its ability to amplify signal and detect various RNA targets and proteins in intact cells and tissues.

3.3 Results

3.3.1 SPARC demonstrates iterative deposition of amplifiers near the primary probe and exponential signal amplification

In our SPARC amplifier design, we utilized benzophenones (BP) as our photo-crosslinking agent due to its ubiquitous use in chemical biology and amenability in aqueous solutions.[22–24] To enhance photo-crosslinking efficiency, we conjugated ~2-3 benzophenones to each amplifier as previous reports indicate that multiple benzophenones improve crosslinking in aqueous environments.[25] Specifically, we took advantage of the 3' tailing ability of terminal deoxynucleotidyl transferase to append multiple primary amine–modified dATPs, then reacted them with succinimidyl ester–conjugated BP (**Figure 3.1B**). To test the efficiency of photocrosslinking, we generated BP-amplifiers with USER cleavage sites to determine if the amplifier is retained after cleaving it from the bound primary probe. Even after USER cleavage and stringent denaturing wash conditions, BP-amplifiers exhibited near quantitative retention (**Supplementary Figure 3.2A**). Beyond photo-crosslinking, we integrated a 10-nt toehold into the amplifier design to enable strand displacement, allowing the amplifier to dissociate from the primary probe target. This allows the re-hybridization of BP-amplifiers to the bound primary probe. As expected, we observed efficient displacement of our amplifier and rebinding of a second amplifier onto the same target. Moreover, we show that unless displacement occurs, another amplifier is unable to bind to its target (**Supplementary**

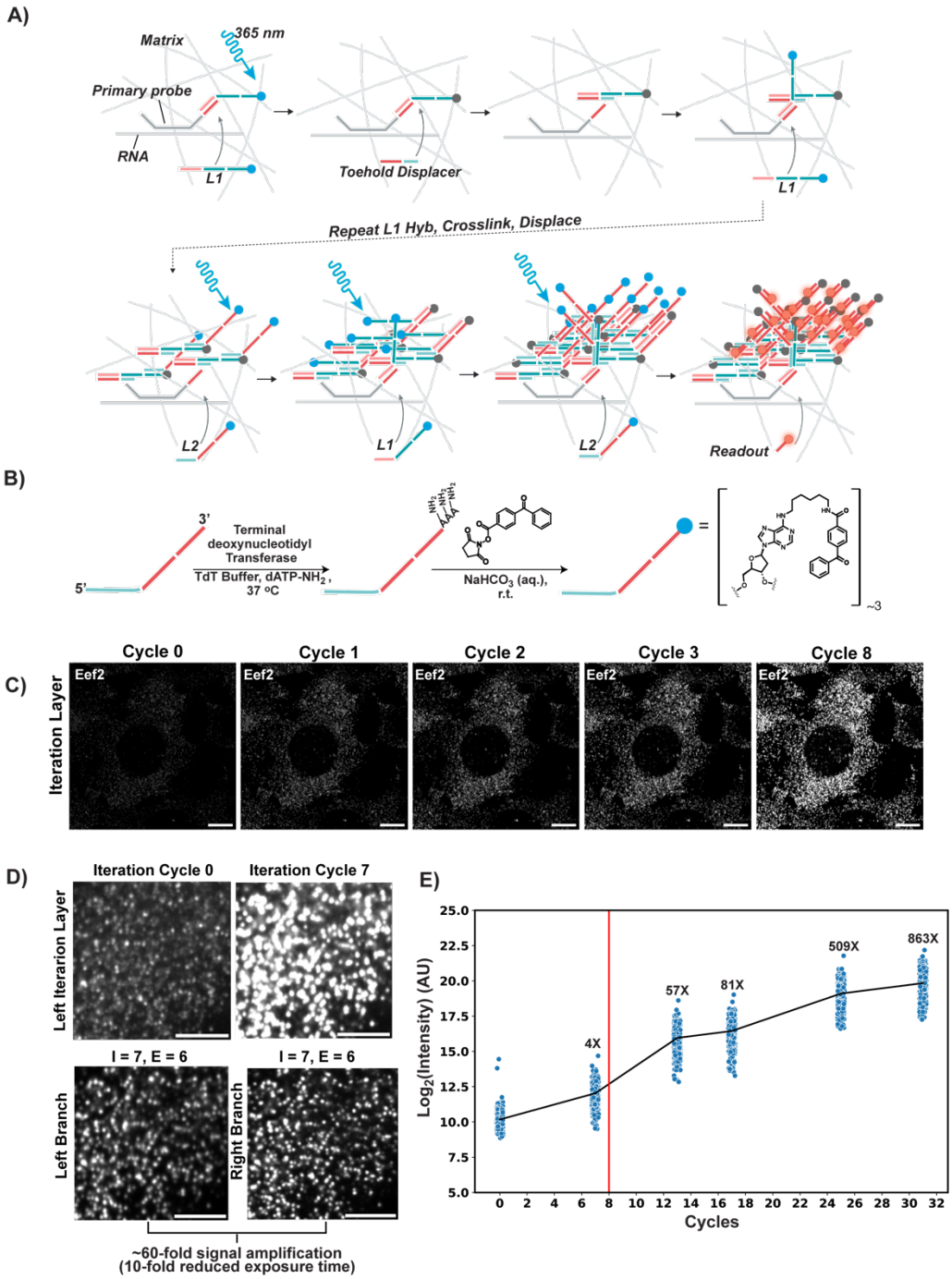


Figure 3.1: A) Schematic of SPARC. B) Amplifier synthesis scheme. C) Representative image of iterative probe deposition showing linear signal amplification. Scale bar = 10 microns. D) Representative images showing colocalization with smFISH spots and signal amplification after linear deposition cycle (top). Representative images for exponential bDNA assembly on the deposited amplifier layer showing high colocalization between two unique branches. I = iteration, E = Exponential, Scale bar = 10 microns. E) SPARC growth curve showing >500-fold signal amplification (red line = start of exponential cycle).

Figure 3.2B-C). Coupling our photo-crosslinking and displacement scheme, we observed a roughly linear increase in signal up to eight repeated cycles of crosslinking and

displacement, with areas having more bound probes showing greater increase in signal (**Figure 3.1C**). Moreover, we observe that the amplified signal colocalizes significantly with the initial spots observed. Next, we proceeded to assemble bDNA structures atop the deposited amplifier layer, iteratively crosslinking to the extracellular matrix after each hybridization cycle. This achieved nearly 60-fold signal amplification while maintaining its appearance as diffraction-limited spots and colocalizing with smFISH signals. Moreover, we observed high colocalization between two simultaneously generated branches from the same primary probe (**Figure 3.1D**). As an additional note, we also found that cultured cells require fixation different from 4% paraformaldehyde to prevent sample degradation. We observed that 3% glyoxal fixation followed by sodium borohydride treatment leads to high sample stability for cell culture making it amenable to SPARC. Considering 3% glyoxal fixation is less commonly used in FISH-based experiments, we validated its effectiveness through serial smFISH experiments, demonstrating high concordance with RNA-seq measurements (**Supplementary Figure 3.3**).

With our findings, we wanted to test the limits of signal amplification using SPARC. To our surprise, we were able to achieve > 500-fold signal amplification (**Figure 3.1E**) with subsequent cycles providing reduced amplification gain. With this significant growth in bDNA structures, we were curious to know the average distance between branches after super-resolving the spots. We observed the branches generally colocalize below 0.5 pixel (< 50 nm), which is comparable to smFISH, with variance in distance between branches becoming larger after ~25 cycles (**Supplementary Figure 3.4A**). The variance in branch distance is most likely due to compromised sample integrity as we observe changes in cellular morphology and degradation after many cycles. We measured branch distance because of its importance in combinatorial smFISH-based methods, where barcodes must colocalize to enable target decoding.[26] Our results indicate high spatial resolution for

efficient decoding of many targets. We also observed increased variance in branch growth beyond 10 exponential cycles as the correlation in intensity across branches decreased (**Supplementary Figure 3.4B**). This is expected considering the binding events are independent, and differing binding probabilities would lead to increasing variability in intensity (growth) as additional exponential cycles are performed. Hence, we limited our experiments to eight iterations and six exponential cycles. Under these conditions, we have also successfully amplified four probes bound to Eef2 via SPARC and observed good colocalization with smFISH (approx. 70%), demonstrating future applications of SPARC to detect small RNA targets (**Supplementary Figure 3.5**).

3.3.2 SPARC enables multiplexed signal amplification of RNA targets

Next, we assessed the multiplexing capability of SPARC by amplifying 16 unique RNA species in NIH 3T3 cells. We initially found that mixing many amplifiers together resulted in poor amplification and substantial noise. Upon *in vitro* testing, we observed aggregates that can potentially be micelles when the overall concentration of the amplifiers reached beyond 1 μM (**Supplemental Figure 3.6**). The formation could be a result of the lipophilic benzophenone group, as aggregates were not observed with purely amine modified amplifiers. With this knowledge, we performed all multiplexing experiments under 1 μM total amplifiers during hybridization. Furthermore, we examined several crowding agents to enhance hybridization kinetics and noted that neutral agents, such as polymeric polyethylene glycols (PEG), significantly strengthen BP-amplifier interactions. At high concentrations of BP-amplifiers (3 μM), large liquid-like condensates appeared. When we used higher molecular weight PEG (20 kDa), net-like structures emerged and underwent a phase transition after one hour, creating a multi-phasic system. Although we currently lack mechanistic insight into these phase transitions, we found that anionic crowding agents like dextran sulfate provided the most effective results (**Supplemental Figure 3.7**).

When conducting our 16-plex experiments, we observed variability in specificity arising from our amplifiers, with some showing very minimal non-specific signals while others showing substantial noise by measuring the ratio of amplifiers that colocalize with smFISH (**Figure 3.2A, Supplementary Figure 3.8A**). We believe that this variability can be further improved by screening more amplifiers with minimal off-target binding and

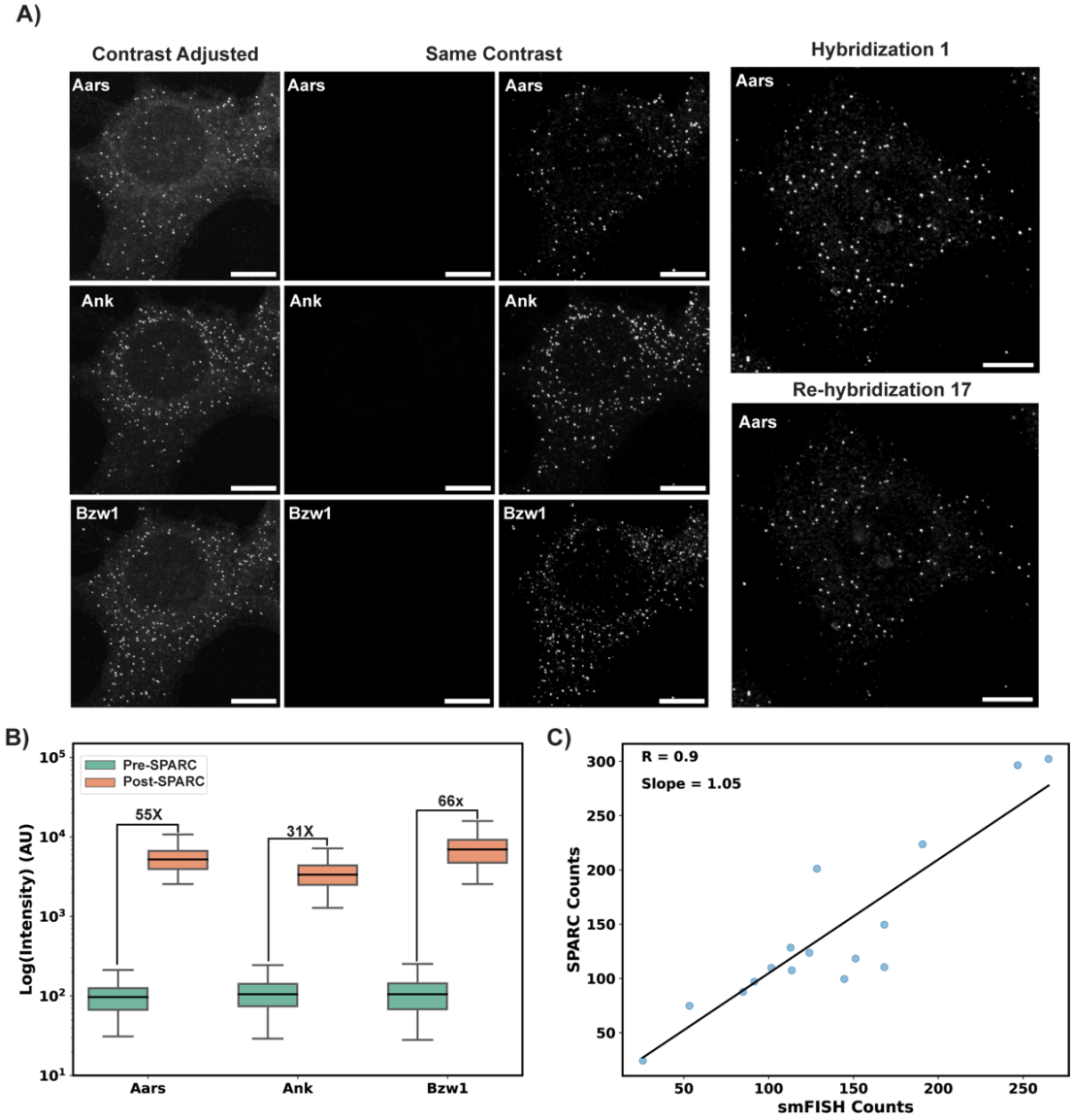


Figure 3.2: A) Representative images depicting pre- and post- SPARC along with significant colocalization with smFISH (left panel, scale bar = 10 microns). Representative images also show retention of signal after 16 denaturing wash cycles highlighting the stability of SPARC generated structures (right panel, scale bar = 10 microns). B) Quantification of pre- and post-SPARC showing high amplification factors from a representative experiment. Box plots were generated from individual colocalizing points between conditions across differing field of views. Horizontal line indicates median. C) Regression between SPARC generated structures and smFISH of the same RNA targets from the same cell shows high correlation.

improving the amplifier synthesis to consistently generate 2-3 BP at the 3' end. Based on our results regarding amplifier interactions, we find that if too many BPs exist in our amplifiers, then that leads to increased non-specific binding to the sample. Regardless, we achieved significant signal amplification for most genes, with bDNA structures remaining intact through multiple denaturing wash cycles to remove fluorescent signal (**Figure 3.2A)**. Amplification factors ranged from ~30- to ~70-fold across different amplifiers, reflecting possible variations in amplifier synthesis or binding efficiency (**Figure 3.2B**). Still, SPARC showed high concordance with smFISH counts (R = 0.9, Slope = 1.08) and bulk RNA-seq (**Figure 3.2B, Supplementary Figure 3.7B**).

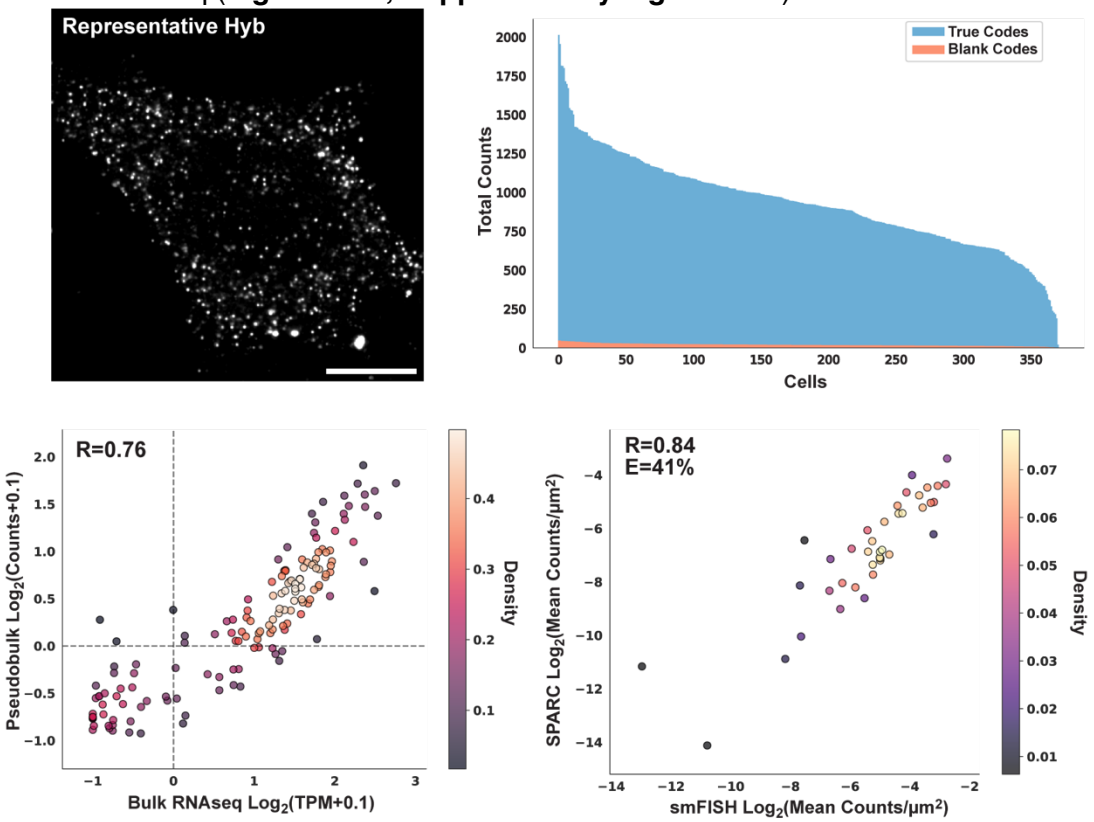


Figure 3.3: Representative hybridization cycle of 150-plex SPARC in NIH 3T3 cells (top left, scale bar = 10 microns) and counts distribution (top right, FDR ~ 5%). Multiplexing experiment shows high correlation with RNA-seq (bottom left) and standard smFISH (bottom right). We measured the detection efficiency of SPARC to be ~ 40 %, which represents the slope of the regression line derived from non-logged data.

Given the high stability of SPARC-generated DNA structures under stringent denaturing wash conditions, we performed 150-plex seqFISH SPARC in NIH 3T3 cells using 6 pseudocolors and 3 barcoding rounds with 1 additional parity round. We obtained

a moderate detection efficiency of 41% with high correlations with bulk RNA-seq and standard smFISH measurements (**Figure 3.3**). These results highlight the amenability of SPARC with seqFISH applications to obtain biologically relevant information. In addition to cell culture, we performed SPARC in intact mouse brain slices and achieved more than 30-fold signal amplification (**Figure 3.4A**). Additionally, unlike cell culture, we observed that tissue samples do not require 3% glyoxal fixation and standard 4% paraformaldehyde is sufficient for high sample stability. Leveraging the capability of SPARC to amplify signal in tissues, we performed 161-plex seqFISH SPARC in mouse brain slices using the same encoding scheme as the 150-plex experiment conducted in NIH 3T3 cells. These genes were selected from 10X Genomics cell-typing panel to identify various cell-types in the mouse brain. Regardless of the variability in amplifier quality as previously mentioned, we identified known cell types and preserved cytoarchitecture of the mouse hippocampus by proper identification of the CA1, CA2, CA3, and Dentate Gyrus, along with a few cortical

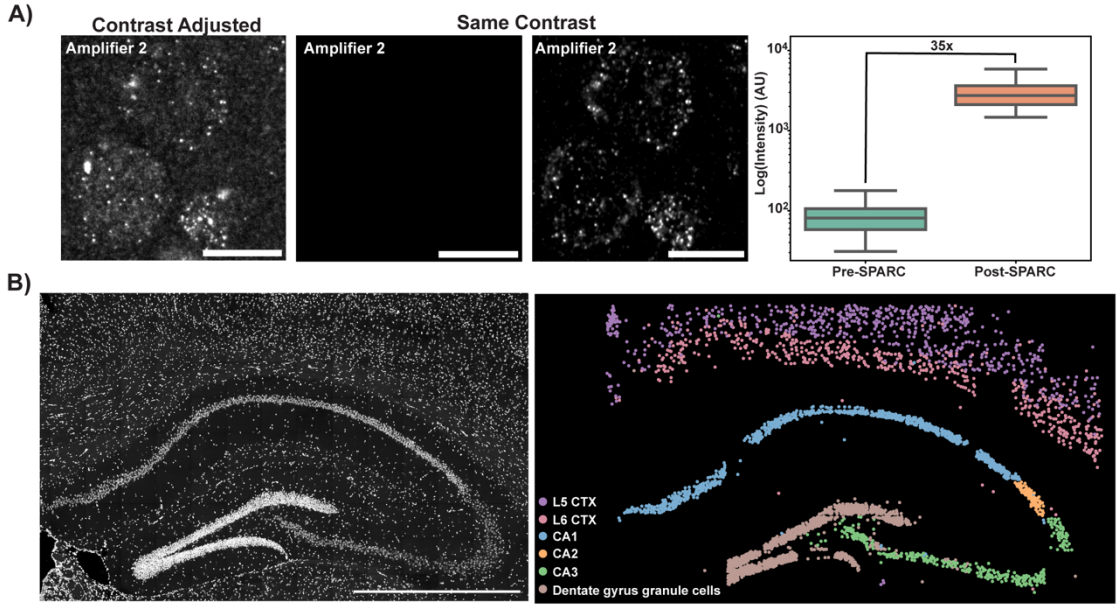


Figure 3.4: A) Representative images depicting pre- and post- SPARC (left) and quantification showing 35-fold signal amplification obtained from various colocalizing spots across conditions. Quantification was performed from a single representative dataset (right). Horizontal line indicates media. Scale bar = 10 microns. B) Representative image showing DAPI stained mouse hippocampus (left) and cell-type identification (right). Scale bar = 1 mm.

layers such as layer 5 and 6 (**Figure 3.4B**). These results indicate the utility of SPARC for tissue applications.

3.3.3 SPARC enables amplification of antibody signal

Inspired by methods like Immuno-SABER,[16] we next evaluated the ability of SPARC to amplify antibody signals and form stable DNA structures. In HeLa cells, we performed SPARC on an oligonucleotide-conjugated Lamin-B1 antibody. Considering Lamin-B1 has distinct spatial patterning around the nuclear lamina, we reasoned that this will be a good target to assess non-specific amplification from SPARC. We observed >50-fold signal amplification while preserving the nuclear lamina's spatial pattern, closely

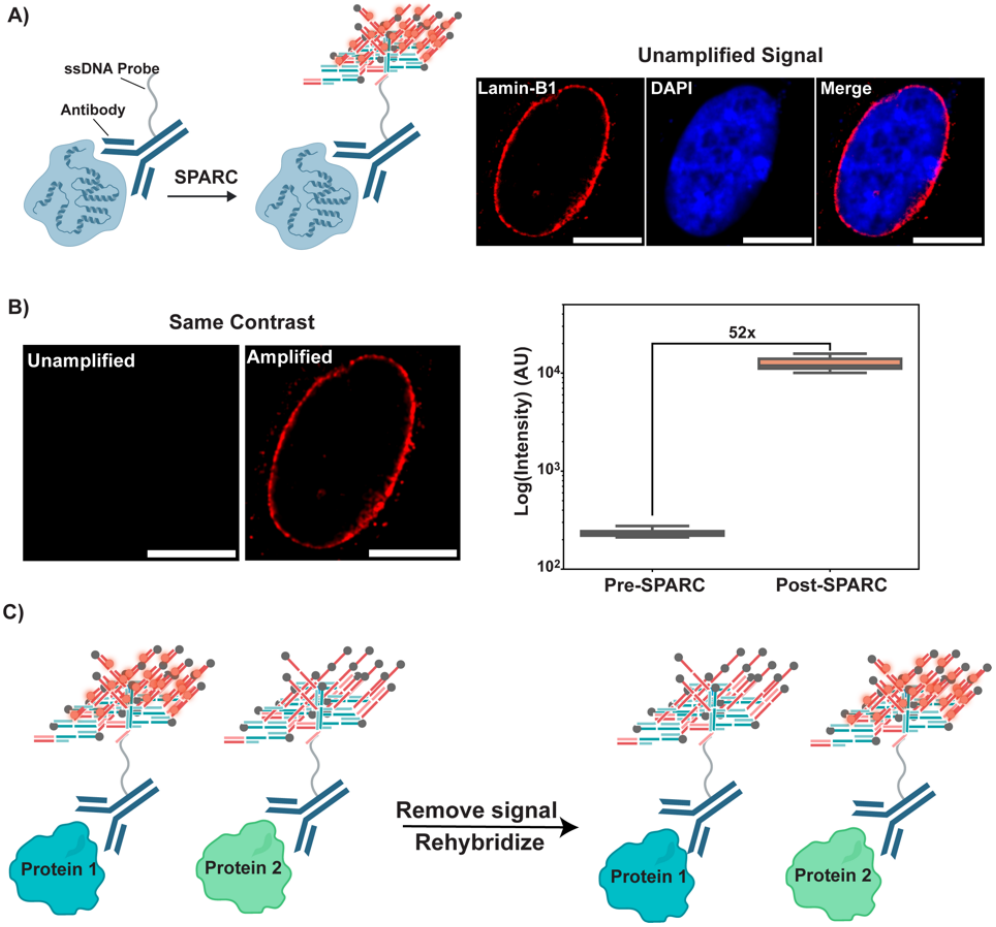


Figure 3.5: A) Schematic of antibody amplification (left panel) and standard immunolabeling of the nuclear lamina in HeLa cells (right panel). Scale bar = 10 microns. B) Representative image depicting SPARC on Lamin-B1antibody and preservation of spatial patterning (left panel). Scale bar = 10 microns. Quantification was performed on segmented masks that overlap across conditions in a single representative dataset (right panel). Horizonal line represents the median.

matching unamplified controls (**Figure 3.5A-B**). This highlights the utility of SPARC to not only amplify RNA targets, but also proteins. This can be especially useful when targeting lowly expressed proteins or to amplify signal from weak binding antibodies. Moreover, unlike Immuno-SABER, we are generating highly stable bDNA structures which is amenable to repeated denaturing wash conditions. This can enable multiplexed amplification of antibodies and serial hybridizations of fluorescent readouts which can be removed under denaturing wash conditions (**Figure 3.5C**).

3.4 Discussion

In this work, we introduce SPARC, an *in situ* signal amplification method that utilizes photochemical iterative deposition of amplifiers and bDNA assembly. We highlight the technical ability to iteratively deposit short and unique amplifier strands near the primary probe for linear signal amplification. Moreover, by employing exponential bDNA assembly atop the deposited amplifiers and photo-crosslinking the strands to the extracellular matrix, SPARC substantially increases signal intensity and form highly stable structures while preserving high spatial resolution. We also highlight the ability of SPARC to amplify signal with very few probes, demonstrating that just 4 primary probes on an RNA target can provide sufficient detection. Unlike some bDNA based methods that utilizes large amplifier strands (>100 nts),[14-16,19-21] our approach utilizes very short (< 50 nts) strands, which enhances diffusion in thicker tissues. This ensures broader applicability in challenging sample types. We demonstrated the utility of SPARC for multiplexing applications using seqFISH, successfully profiling more than 150 genes in intact cells and tissues. Furthermore, we observed strong correlations with both RNA-seq and standard smFISH measurements, with seqFISH-SPARC achieving a detection efficiency of around 40%. We were also able to recapitulate the cytoarchitecture of the mouse hippocampus using SPARC, highlighting its potential for tissue-based applications.

Beyond RNA targets, SPARC can also amplify protein signals, demonstrating broad applicability for other biomolecules. Its performance in multiplexed experiments, while retaining single-molecule spatial accuracy, and its compatibility with various biomolecules underscore the value of SPARC for spatial biology research. Future work will involve screening additional amplifiers to further improve detection efficiency and specificity, as well as extending the method to more complex biological specimens.

3.5 Methods

Animals. All animal care and experiments were carried out in accordance with Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. 6-7 week old C57BL/6J male mice were obtained from The Jackson Laboratory (USA).

Coverslip functionalization for cell culture. No 1.5H coverslips (Bioscience Tools, CSHP-No1.5-24x60) were plasma cleaned (PDC-001, Harrick Plasma) for 5 minutes. Next, coverslips were incubated with 100 µg/mL of Poly-D-lysine (Sigma, P6407) in 1X PBS at 37 C for 2 hours. Coverslips were rinsed with deionized water, air-dried, and UV-treated in a biosafety cabinet.

Coverslip functionalization for tissues. No 1.5H coverslips (Bioscience Tools, CSHP-No1.5-24x60) were plasma cleaned (PDC-001, Harrick Plasma) for 5 minutes. Next, coverslips were incubated with 2% (3-Aminopropyl)triethoxysilane (ThermoFisher, 80370) in anhydrous acetone for 1h at room temperature. Coverslips were rinsed with 100% ethanol then dried at 90 ^{o}C for 5 min.

Cell culture preparation. NIH/3T3 cells (ATCC) were cultured as previously described.[26] In brief, cells were placed on PDL-functionalized coverslips in high glucose DMEM media (ThermoFisher Scientific, 10569010) with 10% CCS (Cytiva, SH30087.043) and 0.1% Penicillin-Streptomycin (ThermoFisher Scientific, 15070063) for 16 hours. HeLa cells (ATCC) were placed on PDL-functionalized coverslips in high glucose DMEM media (ThermoFisher Scientific, 10569010) with 10% FBS (ThermoFisher Scientific, A5670701) and 0.1% Penicillin-Streptomycin (ThermoFisher Scientific, 15070063) for 16 hours. Cells were washed with 1x DPBS (ThermoFisher Scientific, J67802.K2) and fixed with fresh 3% Glyoxal (Sigma, 50649) containing 150 mM NaCl, 45 mM NaOH, and 0.8% Acetic Acid for 10 minutes at room temperature as described previously with slight modifications.[26] Alternatively, cells were fixed with fresh 4% PFA (ThermoFisher Scientific, 28908) in 1X PBS (ThermoFisher Scientific, AM9624) for 10 min at room temperature. Fixed cells were rinsed with 1X DPBS and incubated with 70% ethanol in water at -20 ^{o}C overnight.

Mouse brain slice preparation. Mouse brain slices were prepared as previously described with slight modifications.[26,28] The mouse brain was dissected out of the skull and immediately placed in Optimal Cutting Temperature (OCT) compound (Fisher, 23-730-571). The brain was flash-frozen in 2-methylbutane slurry (approx. -160 °C), which was cooled using liquid nitrogen, and stored at −80 °C. Sections at 8-12 µm were cut with a cryostat (Leica, CM3050) and immediately placed on aminosilane-functionalized

coverslips. Tissue slices were fixed with 4% PFA in 1X PBS for 10 minutes, then rinsed with 70% ethanol. Slices were stored at -80 °C for later use.

Readout probe design. Readout probes were designed as previously described.[26,28] In brief, 15 nts sequences were randomly generated and a local BLAST query was performed on each probe against the mm10 transcriptome to ensure specificity. Any probe that has 10 contiguous matching sequences were removed. Each probe had a GC content ranging from 40–60%. Readout probes were empirically assessed in-house to determine which sequences had minimal off-target labeling and noise in NIH 3T3 cells.

RNA Targeting Probe Design for Mouse. Transcript-specific primary probes were designed as previously described with some modifications.[26] Using the mm10 masked genome and annotation from UCSC, probe sets targeting exons within the CDS region of 41 unique transcripts for NIH3T3 and 161 unique transcripts for mouse brain were generated. Each probe was 35-nts in length, with subsequent probes being spaced by 2 nts and had a GC content ranging from 40–65%. A local BLAST query was performed on each probe against the mouse transcriptome to ensure specificity. BLAST hits on any sequences other than the target transcript with a 17-nt match or more were considered off target and discarded. To minimize cross-hybridization between probe sets, a local BLAST database was constructed from the probe sequences and probes with hits of 17-nt or longer were removed by dropping the one of the matched probes from the larger probe set. From the above selection criteria, 24 primary probes were selected for each transcript target.

SeqFISH and standard smFISH probe design. RNA seqFISH probes for marker genes were designed similarly to our previous studies.[26,28] In brief, unique RNA species were encoded in a single channel (488 nm) where they are called at specific hybridization rounds of their assigned barcodes. To implement this, 35-nt RNA targeting sequences, four 15-nt unique amplifier/readout probe binding sites encoded for each RNA target, a T7 site at the 3’ end, and a pair of primer binding sites at the 5′ and 3′ ends of the probe for probe generation were concatenated. Marker genes were selected from 10X Genomics mouse brain panel. Serial smFISH probes were designed similarly to seqFISH probes with slight modifications. In the serial design, each unique target contains one or two copies of a single unique readout binding site that is called at a given hybridization round. In the case of amplification, a single unique amplifier binding site was added.

Primary Probe Synthesis. Primary probes were ordered as oligoarray complex pools from Twist Bioscience and were constructed as previously described with some modifications.[26,28] Briefly, limited PCR cycles were used to amplify the probe sequences from the oligo complex pool. Then, the amplified PCR products were purified using QIAquick PCR Purification Kit (28104; Qiagen) according to the manufacturer’s instructions. The PCR products were used as the template for *in vitro* transcription (E2050S; NEB) supplemented with SUPERase·In™ RNase Inhibitor (ThermoFisher Scientific, AM2694) and Pyrophosphatase (NEB, M0361S). Then, the DNA template was removed by DNAse I digestion (NEB, M0303S). The RNA product was further purified using SPRI beads (Beckman Coulter, B23318) following manufacturer’s instructions with a few modifications. In brief, RNA strands were captured using 2x SPRI (twice the starting volume) and washed with 80% EtOH on a magnetic rack, then eluted using nuclease-free water. Post purification, the RNA product underwent reverse transcription (ThermoFisher Scientific, EP7051) with a hybrid forward primer containing both ribonucleotides and

deoxyribonucleotides (**Supplementary Table 1**) and supplemented with SUPERase·In™ RNase Inhibitor and Pyrophosphatase. After reverse transcription, the probes were alkaline hydrolyzed by 250 mM NaOH at 65°C for 20 minutes to degrade the RNA templates and remove the forward primer. The solution was quenched by adding acetic acid at a final concentration of 250 mM. Next, the probes underwent final purification with SPRI beads (as mentioned previously), then resuspended in water. The probes were stored at −20°C until later use.

Amplifier Synthesis. Standard desalted, single-stranded DNA oligonucleotides (ssDNA) were ordered from IDT and purified using SPRI beads (Beckman Coulter, B23318) according to manufacturer's protocol with few modifications. The ssDNA was purified by adding 2x SPRI beads (twice the initial volume) and mixing the solution. Then, isopropanol (50% v/v) was added and incubated for 5 min at room temperature. Captured ssDNA strands were washed with 80% EtOH. The ssDNA was eluted by adding nuclease-free water and incubating at 37 °C for 5 min. For 3' tailing of amplifiers, 10 µM purified ssDNA, 35 µM N6-(6-Aminohexyl)-dATP (Jena Bioscience, NU-835L), 1 U/µl Terminal Deoxynucleotidyl Transferase (ThermoFisher Scientific, EP0162), and enzyme reaction buffer (ThermoFisher Scientific, EP0162) were incubated for 18 hours at 37 °C, then heat inactivated at 75 °C for 20 minutes. Following SPRI bead purification (same procedure as described previously), modified ssDNA was resuspended in 500 mM $NaHCO_3$. A 20 mM solution of N-Succinimidyl 4-Benzoylbenzoate (TCI America, S0863) in anhydrous DMSO was added to the solution mix 3 times at 1-hour intervals (3.7 mM final) at 37 °C. Then, the sodium bicarbonate solution was quenched by adding 17 M acetic acid. The modified strands were precipitated using ethanol with co-precipitate, 4 µg/mL linear acrylamide (ThermoFisher Scientific, AM9520). Next, the sample was further purified with SPRI beads as mentioned previously. Modified ssDNA strands underwent final purification with HPLC (Agilent, 1200 series) using a C18 reverse-phase column (XBridge Oligonucleotide BEH C18, 130 Å, 2.5 µm, 4.6 x 50 mm) following an A-B step gradient at 1 mL/min (90% A for 2 min, 85% A for 3 min, 55% A for 7 min; A = 100 mM Triethylammonium acetate in H_2O, B = Acetonitrile). The retention time of our modified ssDNA strands were 6.5-7.5 min. Purified probes were lyophilized (Labconco Freezone 2.5 L) and resuspended in RNAse-free H_2O. The concentration of our final amplifier strands was determined using Quant-iT™ OliGreen™ ssDNA Assay Kit (ThermoFisher Scientific, O7582). All amplifier probes were stored in amber tubes at -20 °C.

Readout probe synthesis. Readout probes as 5' amine-modified ssDNA (15 nts) was ordered from Integrated DNA Technologies (IDT). The synthesis of dye-conjugated readout probes was performed as previously described with slight modifications.[26,28] Briefly, 5 nmols of amine-modified oligonucleotides were resuspended in 0.5 M sodium bicarbonate buffer (100 µM). Then, 5 molar excess of AlexaFluor 647 NHS ester (Thermo Fisher, A20006), Cy3B NHS ester (Cytiva, PA63101), or single isoform AlexaFluor 488 TFP ester (Thermo Fisher, A30005) in anhydrous DMSO (Thermo Fisher, D12345) was added to the reaction and incubated at 37°C for 1h in the dark. Two consecutive additions of 5 molar excess of dye were added, with each addition being at intervals of 1h (DMSO ~15% v/v). Then, acetic acid was added at the same molar amount of sodium bicarbonate to quench the buffer. The dye-conjugated ssDNA probes were subjected to ethanol precipitation with 4 µg/mL linear acrylamide (Thermo Fisher, AM9520) and purified using HPLC. The probes were lyophilized, then resuspended in water. The readout probes were quantified using Nanodrop and a 1 mM working stock was made. All the readout probes were kept at –20°C in amber tubes for later use.

Antibody conjugation. The Lamin-B1 antibody (Abcam, ab220797) as previously described with slight modifications.[28,29] The antibody was buffer exchanged with ice-cold 1X PBS using a 50kDa MWCO Amicon filter (Millipore Sigma, UFC905008). Next, DBCO-PEG_4-NHS (Sigma-Aldrich 764019) in DMF was diluted in 1X PBS, then 10 molar equivalents were added to the antibody solution. The reaction was incubated for 6 hours at 4°C. The DBCO-modified antibody was buffer exchanged with ice-cold 1X PBS, then 10 molar equivalents of azide-modified-oligonucleotides (**Supplementary Table 1**) were added and incubated for 48 hours at 4°C. The oligo-conjugated antibody was buffer exchanged as previously and concentrated to desired volume. The antibody solution is supplemented 0.02% sodium azide and stored at 4 °C for later use.

NIH 3T3 RNA hybridization. SmFISH experiments were performed as previously described with a few modifications.[26,28,29] In brief, glyoxal-fixed NIH 3T3 samples were dried under compressed nitrogen to remove the 70% ethanol solution. A custom, in-house generated flow cell was attached to the coverslip. The sample underwent 0.1% sodium borohydride treatment for 15 min at room temperature while continuously mixing. Post-reduction the sample was rinsed with 1X PBS, then post-fixed twice using 7.5 mM $BS(PEG)_5$ (ThermoFisher Scientific, A35396) in 1X PBS for 15 min at room temperature. Next, the sample was treated twice with 100 mM N-(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature. Sample was rinsed with 50% wash buffer (50% formamide, 2X SSC, and 0.1% Triton-X100), then blocked with hybridization buffer (50% formamide, 200 nts TTG repeat sequence (**Supplementary Table 1**), 0.1 mg/mL yeast tRNA, 4X SSC, 10% 500kDa Dextran Sulfate) for 1h at 37 °C. Then, the sample was hybridized with ~5 nM/probe in hybridization buffer in a humid chamber at 37 °C for 16 h. After primary probe hybridization, samples were rinsed 3 times with 50% wash buffer then incubated at 37 °C for 30 min and repeated once more. Then, the sample was rinsed several times with 4X SSC. The sample was then amplified using Cell Culture SPARC protocol or the signal was read directly. For direct readout of primary probes, the sample was hybridized with readouts at 100 nM in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate (Sigma, D4911), 4X SSC) for 15 min at room temperature. The sample was rinsed several times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100), stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution containing 100 mM Tris HCl pH 8, 4X SSC, 5 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase (Sigma ,G2133), and 1:1000 catalase (Sigma, C3155) prior to imaging.

Lamin-b1 SPARC in HeLa cells. PFA fixed HeLa cells were dried under compressed nitrogen to remove the 70% EtOH solution. A custom, in-house generated flow cell was attached to the coverslip, then the sample was rinsed with 1X PBS. The sample was incubated with blocking buffer containing 1% UltraPure BSA (Ambion, AM2616), 0.1% 6.5-10 kDa dextran sulfate (Sigma Aldrich, D4911), 0.5 mg/mL sheared salmon sperm DNA (Invitrogen, AM9680), 1U/μL SUPERase•In RNase Inhibitor (Invitrogen,AM2694), 0.01% sodium azide (G Biosciences, 786-299), and 0.3% Triton X-100 (Sigma-Aldrich, 93443) in 1X PBS for 90 min at room temperature. Then, the sample was incubated with antibody-oligonucleotide conjugated in blocking buffer overnight at 4 °C. Post immunolabeling, the sample was rinsed with 1X PBS several times. The antibody was fixed to the target by incubating 7.5 mM $BS(PEG)_5$ (ThermoFisher Scientific, A35396) in 1X PBS for 15 min at room temperature twice. Post-fixation the sample was treated with 100 mM N-

(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature twice. Prior to SPARC, the antibody signal was read by hybridizing readout probes in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate (Sigma, D4911), 4X SSC) for 15 min at room temperature. The sample was rinsed several times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100), stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution containing 100 mM Tris HCl pH 8, 4X SSC, 5 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase (Sigma ,G2133), and 1:200 catalase (Sigma, C3155) prior to imaging. Post-imaging, the readout was removed using 55% wash buffer (55% formamide, 2X SSC, and 0.1% Triton X-100). The sample was placed on an automated fluidics system for SPARC.

Mouse tissue RNA hybridization. SmFISH experiments were performed as previously described with a few modifications.[26,28,29] In brief, mouse brain slices on aminosilane treated coverslips were incubated in 70% EtOH overnight at 4 °C. The sample was dried using compressed nitrogen, then treated with 8% Triton X-100 in 1X PBS for 30 min at room temperature. The tissue slice was rinsed with 70% EtOH, then a custom, in-house generated flow cell was attached to the coverslip. The sample was rinsed with 1X PBS, then post-fixed using 7.5 mM $BS(PEG)_5$ (ThermoFisher Scientific, A35396) in 1X PBS for 15 min at room temperature. Next, the sample was treated twice with 100 mM N-(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature. Sample was rinsed with 50% wash buffer (50% formamide, 2X SSC, and 0.1% Triton-X100), then blocked with hybridization buffer (50% formamide, 200 nts TTG repeat sequence (**Supplementary Table 1**), 0.1 mg/mL yeast tRNA, 4X SSC, 10% 500kDa Dextran Sulfate) for 1h at 37 °C. Then, the sample was hybridized with ~5 nM/probe in hybridization buffer in a humid chamber at 37 °C for 36 h. After primary probe hybridization, samples were rinsed with 50% wash buffer then incubated at 37 °C for 30 min and repeated once more. Then, the sample was rinsed several times with 4X SSC, followed by 1X PBS. The sample was treated once more with 100 mM N-(Propionyloxy)succinimide (Sigma, 93535-1G) in 1X PBS for 15 min at room temperature, then rinsed with 4X SSC. The sample was then amplified using Tissue SPARC or readout directly. For direct readout of primary probes, the sample was hybridized with readouts at 100 nM in 10% hybridization buffer (10% formamide, 10% 6.5-10 kDa dextran sulfate (Sigma, D4911), 4X SSC) for 25 min at room temperature. The sample was rinse several times with 10% wash buffer (10% formamide, 2X SSC, and 0.1% Triton-X100), stained with 1 μg/ml DAPI in 2X SSC and replaced with anti-bleaching solution containing 100 mM Tris HCl pH 8, 4X SSC, 5 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase (Sigma ,G2133), and 1:200 catalase (Sigma, C3155) prior to imaging.

In situ USER cleavage. Samples were incubated in 1X rCutSmart buffer (NEB, B6004S) with 50 U/mL USER enzyme (NEB, M5505S) for 30 min at 37 °C. Post-digest, the sample was washed 3 times at 5 min intervals using 55% wash buffer (55% formamide, 0.1% Triton X-100, 2X SSC) at room temperature. Then, the sample was rinsed 3 times 4x SSC.

Cell culture SPARC. SPARC was performed using an automated fluidics system with a constant flow rate of 300 µL/min. The process began with a rinse in 4X SSC, followed by hybridization buffer (10% Formamide, 10% 40 kDa Dextran Sulfate (Sigma, 42867), 0.1% Triton X-100, and 4X SSC). The sample was hybridized with 30 nM/amplifier in hybridization buffer at 37°C for 45 minutes. Post-hybridization, the sample was washed for 8 min with 10% wash buffer (10% Formamide, 0.1% Triton X-100, 2X SSC), then with

1X PBS for 2 min. Next, the sample underwent 365 nm irradiation (JSHSG, FIUBIUUBF-A-365NM) at 150 mW/cm² with a 1 min pulse followed by a 1 min cooldown, repeated for five cycles. A subsequent 4 min wash was performed with 30% wash buffer (30% formamide, 0.1% Triton X-100, and 2X SSC). The displacement cycles then commenced, starting with a 2 min wash in 4X SSC. The sample was incubated with 500 nM/displacer in hybridization buffer at 37°C for 15 min, followed by 15 min at room temperature. After amplifier displacement, the sample was washed for 8 min with 30% wash buffer and for 1 min with 4X SSC. This process was repeated for the desired number of deposition cycles. After the deposition phase, the sample was rinsed with hybridization buffer before adding 30 nM amplifier solution, which bound to the deposited amplifier layer. The sample was incubated at 37°C for 45 minutes, followed by the same post-hybridization wash sequence: 8 minutes with 10% wash buffer and 2 minutes with 1X PBS. The irradiation step (365 nm, 150 mW/cm², 1-min pulse, 1-min cooldown, five cycles) was repeated. The sample was then washed with 30% wash buffer for 8 min, followed by a 1 min rinse in 4X SSC. The exponential amplification phase was repeated for the required number of cycles for bDNA assembly. Upon completion, the sample was washed with 1X PBS for 1 minute, subjected to an additional five cycles of UV treatment, and stored in 4X SSC.

Tissue SPARC. SPARC was performed using an automated fluidics system with a constant flow rate of 300 µL/min. The process began with a rinse in 4X SSC, followed by hybridization buffer (10% formamide, 10% 40 kDa Dextran Sulfate (Sigma, 42867), 0.1% Triton X-100, and 4X SSC). The sample was hybridized with 30 nM/amplifier in hybridization buffer at 37°C for 75 minutes. Post-hybridization, the sample was washed for 16 min with 10% wash buffer (10% formamide, 0.1% Triton X-100, 2X SSC), then with 1X PBS for 5 min. Next, the sample underwent 365 nm irradiation (JSHSG, FIUBIUUBF-A-365NM) at 150 mW/cm² with a 1 min pulse followed by a 1 min cooldown, repeated for five cycles. A subsequent 4 min wash was performed with 30% wash buffer (30% formamide, 0.1% Triton X-100, and 2X SSC). The displacement cycles then commenced, starting with a 5 min wash in 4X SSC. The sample was incubated with 500 nM/displacer in hybridization buffer at 37°C for 20 min, followed by 20 min at room temperature. After amplifier displacement, the sample was washed for 16 min with 30% wash buffer and for 2 min with 4X SSC. This process was repeated for the desired number of deposition cycles. After the deposition phase, the sample was rinsed with hybridization buffer before adding 30 nM amplifier solution, which bound to the deposited amplifier layer. The sample was incubated at 37°C for 75 minutes, followed by the same post-hybridization wash sequence: 16 minutes with 10% wash buffer and 5 minutes with 1X PBS. The irradiation step (365 nm, 150 mW/cm², 1-min pulse, 1-min cooldown, five cycles) was repeated. The sample was then washed with 30% wash buffer for 12 min, followed by a 2 min rinse in 4X SSC. The exponential amplification phase was repeated for the required number of cycles for bDNA assembly. Upon completion, the sample was washed with 1X PBS for 1 minute, subjected to an additional five cycles of UV treatment, and stored in 4X SSC.

SmFISH or seqFISH imaging routine. Imaging routine is performed as previously described with a few modifications.[26,28,29] In summary, the flow cell on the sample was initially connected to an automated fluidics system with a flow rate of 250 µL/min. The region of interest (ROI) was identified using nuclei signals stained with 3 µg/ml DAPI in 2X SSC. For serial smFISH or seqFISH experiments, each serial hybridization buffer contained a single unique readout sequence conjugated to Alexa Fluor 488 (100 nM) in 10% hybridization buffer (10% formamide, 10% 6-10kDa dextran sulfate (Sigma, D4911),

0.1% Triton X-100, 4X SSC). A total of 200 µl of serial hybridization buffers for 16 (serial smFISH) or 24 (seqFISH) rounds of imaging, with a repeat for round 1 and cytoplasmic stain with polyT-647 (18 or 26 rounds in total), was pipetted into a 96-well plate. During each serial hybridization, the automated sampler moved to the designated well and aspirated 100 µl hybridization buffer through a multichannel fluidic valve (EZ1213-820-4; IDEX Health & Science) to the flow cell (requires ~30 µl) using a syringe pump (63133-01, Hamilton Company). The solution was incubated for 25 minutes at room temperature. After serial hybridization, the sample was washed with 800 µl of 10% wash buffer (10% Formamide and 0.1% Triton X-100 in 2× SSC) over 6 min period to remove excess readout probes and non-specific binding. The sample was then rinsed with about 200 µl of 4X SSC before being stained with DAPI solution (3 µg/mL of DAPI, 2X SSC) for roughly 30 seconds. An anti-bleaching buffer solution (100 mM Tris HCl pH 8, 4X SSC, 5 mM Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:200 catalase) was flowed through the sample. Imaging was performed using a Leica DMi8 microscope with various components, including a confocal scanner unit (Andor Dragonfly 200), a sCMOS camera (Andor Zyla 4.2 Plus), a 63× oil objective lens (Leica 1.40 NA), and a motorized stage (ASI MS2000). High Power Laser Engine and filter sets from Andor were utilized, and snapshots were acquired with 0.25 µm z-steps for nine z-slices per field. After imaging, a stripping buffer (60% formamide and 0.1% Triton-X 100 in 2X SSC) was flowed through the sample for 1 minute, followed by a 3-minute incubation time then repeated twice more. Post readout removal, the sample was rinsed with 4X SSC solution. The serial hybridization, imaging, and signal extinguishing steps were repeated until the desired rounds were reached. The integration of the automated fluidics delivery system and imaging was controlled by a custom-written script in µManager.

RNA-seq on NIH 3T3. RNA was isolated from P9 NIH 3T3 cells using RNeasy kits (Qiagen, 74104) following manufacturer's instructions. The library was sequenced at 50 million reads with PE50. The reads were aligned against the GRCm38 mouse transcriptome using kallisto[30] to obtain TPM expression levels.

Image processing. Image registration was performed similarly to previous reports with a few modifications.[31] Translation shifts were obtained using phase cross correlation (scikit-image) on DAPI stained images. Once the images were aligned, a 7 x 7 high pass Gaussian filter was applied followed by a 3 x 3 low pass Gaussian filter. Image intensities were normalized by 80-99.999% percentile clipping and rescaling between 0 and 1.

Spot calling. Spot calling was performed similarly to previous reports.[31] Using DAOStarFinder (Photutils), we measured sub-pixel centroids and extracted various spot characteristics, including flux, peak amplitude, sharpness, bilateral and four-fold symmetry, as well as Gaussian-fit symmetry. The full width at half maximum (FWHM) was numerically optimized to determine the optimal parameter for spot detection. We then calculated the total spot area, an additional feature, by isolating each spot in a 7×7 bounding box and applying a local adaptive threshold with a Gaussian kernel. The resulting binary image was summed to obtain the spot's total area.

Cell segmentation. Post seqFISH imaging, cells were stained with polyT-AF647 (**Supplementary Table 1)** in 4X SSC for 20 min and DAPI for 30 s at room temperature. Then, the sample was washed with 10% wash buffer (10% formamide, 0.1% Triton X-100, and 2X SSC). Next, anti-bleaching buffer solution (100 mM Tris HCl pH 8, 4X SSC, 5 mM

Trolox, 10% glucose, 1 mg/mL glucose oxidase, and 1:200 catalase) was flowed through the sample prior to imaging. Whole cell masks were generated using Cellpose 3.0 with a human-in-loop trained model on polyT-AF647.[32,33] Masks near the edges of the image were discarded to remove illumination bias or partial cells from the analysis. Additionally, 2 pixels were deleted between two or more masks that touch. Spots were mapped to each cell mask to assign each spot to a cell id.

Colocalization analysis. Sub-pixel centroids of spots obtained from image processing and spot calling (see Methods) were used for colocalization analysis by performing a nearest neighbor search algorithm with 2-pixel search radius (sklearn). Colocalization efficiency was calculated as the number of spots that colocalize referenced to the seed that was used for the search. The pixel distance between each spot to its neighbor was also stored for distance dispersion analysis.

Decoding. Super-resolved, mapped spots undergo SVM-embedded, feature-based symmetrical nearest neighbor decoding as previously described with a few modifications.[31] An SVM model with a radial basis function or polynomial kernel was used to assign probabilities to each spot, based on its measured characteristics from spot calling, indicating whether it represents a true signal or noise. False spots are filtered based on user defined likelihood cutoff. Using the cleaned-up data, a radial search (2 pixels) across barcoding rounds scores potential spots on distance, intensity, and size, particularly if multiple candidates are available for codeword assignment. Best spots are selected from each round based on their individual scores. A total codeword score is then computed by summing those individual scores. Additionally, an ambiguity score is assigned, which corresponds to the sum of additional nearby spots that was identified when performing the radial search across barcoding rounds. The overall codeword score is normalized by the ambiguity score to assign confidence to the overall scores. Identified codewords (barcodes) undergo parity checks (checksum) and spot set consistency filtering (≥ 3 appearances in any given seed). If only one parity bit is missing, the barcode can still be assigned, otherwise, it remains unidentifiable. If there are any codewords that have overlapping spots that passed all criteria, then their overall codeword score was used to pick the best one. If they had the same score, then the codeword with the smallest total distance between spots was used. In the first decoding round following the above-mentioned criteria, the top 10% of highly expressed genes relative to the dataset with complete parity (codes with missing parity is ignored) is decoded first to reduce crowdedness. Unused spots (not assigned to any barcode) are resubmitted for 2 additional decoding rounds to maximize decoding outcomes and assigning clashing codewords. The second round still utilizes complete codes with parity, and the last round allows codes with loss in parity signal. Decoded spots are sorted by codeword score and subsampled to calculate FDR. Spots with FDR ≤10% are used to generate the final gene-by-cell matrix for downstream analysis.

SVM training. Quick-pass decoding assigns labeled probabilities to each spot on whether they are true or noisy signal. When 500 to 500,000 potentially noisy spots are identified, the classifier is employed. True spots are then downsampled to match the number of noisy spots, and the data is split into 80% for training and 20% for validation. Features are normalized using the MinMax Scaler, and GridSearchCV (with 8-fold cross-validation) is used to tune the C, gamma, and degree parameters for polynomial or radial basis function

kernels. Finally, test set performance is evaluated using the scaling parameters derived from the training data.

False discovery rate. False discovery rate is defined as follows:

$$\mathrm{FDR} = \frac{\#\text{ of Real Barcodes x } \frac{\text{Blank Counts}}{\#\text{ of Blank Codes}}}{\text{Real Barcode Counts}}$$

Number of real barcodes are defined as the number of gene-coding barcodes in the codebook while number of blank codes correspond to the number of unassigned barcodes in the codebook. Blank counts are the number of decoded blank barcodes, and the real barcode counts are the number of decoded gene-assigned barcodes. This essentially uses the frequency of blank code counts to estimate the number of false codes that may be detected as a true code (assuming frequency of error is the same). This value is normalized by the detection of all real barcode counts which includes both true positives and false positives.

Simulation. Branched DNA growth was simulated for each initial site (ranging from 1 to 23) by running binomial trials (n=1000) with a set probability ($p = 0.75$). The resulting data was stored and used to compute the coefficient of variation (CV) and bootstrap confidence intervals.

pyFISH Tools: PyFISH tools is the general seqFISH and smFISH processing pipeline used in this study which can be found at https://github.com/klcolon/pyfish_tools.

Sequences. Primers, antibody sequences, amplifier sequences, and readout sequences can be found in Supplementary Table and Figures. Primary probe sequences are available upon request.

3.6 References

1. Wu, C. *et al.* RollFISH achieves robust quantification of single-molecule RNA biomarkers in paraffin-embedded tumor tissue samples. *Commun. Biol.* **1**, 1–8 (2018).
2. Gyllborg, D. *et al.* Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* **48**, e112 (2020).
3. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
4. Zeng, H. *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446 (2023).

5. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).

6. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014).

7. Chen, X., Sun, Y.-C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018).

8. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

9. Deng, R., Zhang, K., Sun, Y., Ren, X. & Li, J. Highly specific imaging of mRNA in single cells by target RNA-initiated rolling circle amplification. *Chem. Sci.* **8**, 3668–3675 (2017).

10. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).

11. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

12. Schulte, S. J., Fornace, M. E., Hall, J. K., Shin, G. J. & Pierce, N. A. HCR spectral imaging: 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent samples. *Development* **151**, dev202307 (2024).

13. 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).

14. Wang, F. *et al.* RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).

15. Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533 (2019).

16. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).

17. Dardani, I. *et al.* ClampFISH 2.0 enables rapid, scalable amplified RNA detection in situ. *Nat. Methods* **19**, 1403–1410 (2022).

18. Rouhanifard, S. H. *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry–based amplification. *Nat. Biotechnol.* **37**, 84–89 (2019).

19. Tao, Y. *et al.* Highly efficient and robust π-FISH rainbow for multiplexed in situ detection of diverse biomolecules. *Nat. Commun.* **14**, 443 (2023).

20. He, S. *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).

21. Khafizov, R. *et al.* Sub-cellular Imaging of the Entire Protein-Coding Human Transcriptome (18933-plex) on FFPE Tissue Using Spatial Molecular Imaging. 2024.11.27.625536 Preprint at https://doi.org/10.1101/2024.11.27.625536 (2024).

22. Dormán, G., Nakamura, H., Pulsipher, A. & Prestwich, G. D. The Life of Pi Star: Exploring the Exciting and Forbidden Worlds of the Benzophenone Photophore. *Chem. Rev.* **116**, 15284–15398 (2016).

23. Hassan, M. M. & Olaoye, O. O. Recent Advances in Chemical Biology Using Benzophenones and Diazirines as Radical Precursors. *Molecules* **25**, 2285 (2020).

24. Burton, N. R., Kim, P. & Backus, K. M. Photoaffinity labelling strategies for mapping the small molecule–protein interactome. *Org. Biomol. Chem.* **19**, 7792–7809 (2021).

25. Jakubovska, J., Tauraitė, D. & Meškys, R. A versatile method for the UVA-induced cross-linking of acetophenone- or benzophenone-functionalized DNA. *Sci. Rep.* **8**, 16484 (2018).

26. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235 (2019).

27. Richter, K. N. *et al.* Glyoxal as an alternative fixative to formaldehyde in immunostaining and super-resolution microscopy. *EMBO J.* **37**, 139–159 (2018).

28. Takei, Y. *et al.* Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586–594 (2021).

29. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).

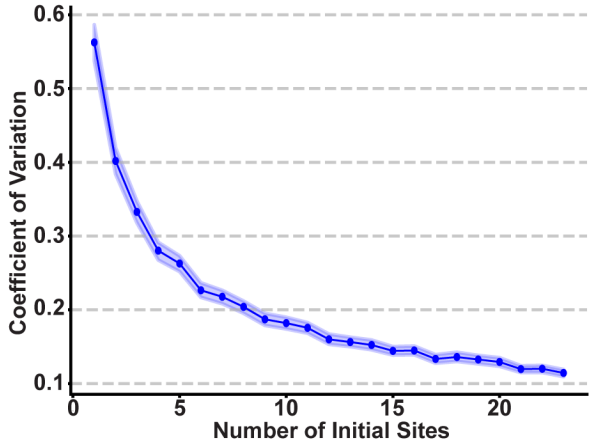
30. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

31. Polonsky, M. *et al.* Spatial transcriptomics defines injury-specific microenvironments in the adult mouse kidney and novel cellular interactions in regeneration and disease. 2023.11.22.568217 Preprint at https://doi.org/10.1101/2023.11.22.568217 (2023).

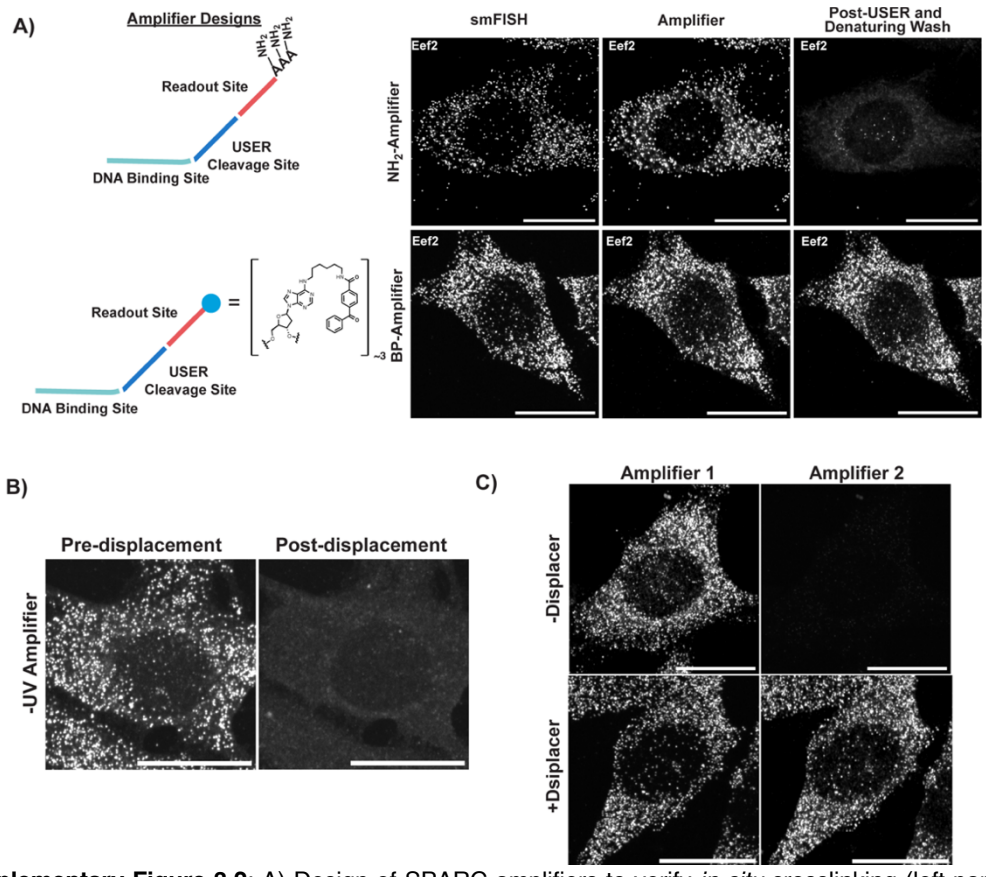
32. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).

33. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. 2024.02.10.579780 Preprint at https://doi.org/10.1101/2024.02.10.579780 (2024).

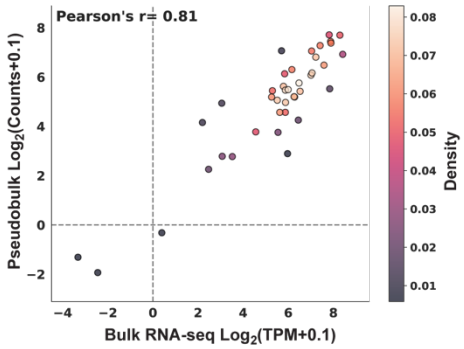
3.7 Supplementary Figures and Tables



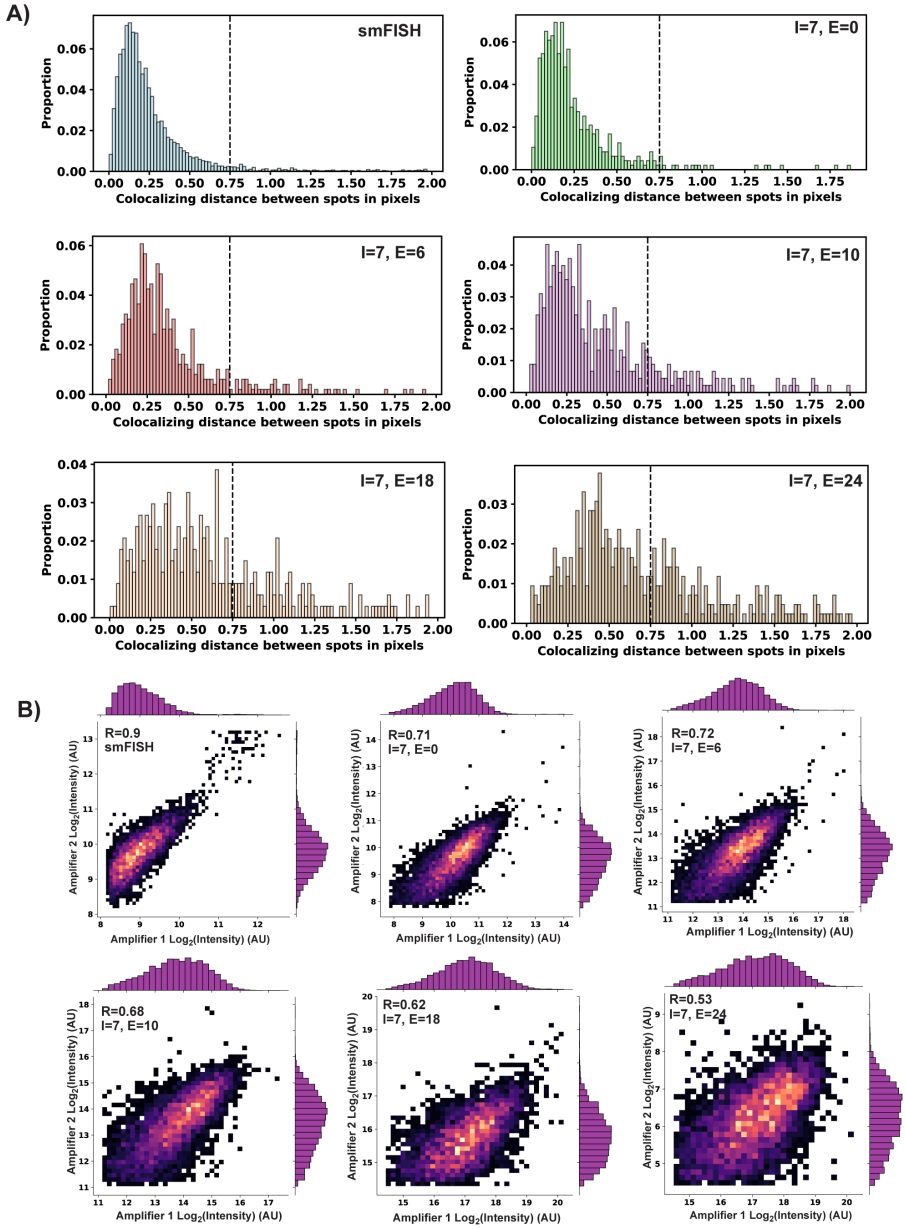
Supplementary Figure 3.1: Simulation demonstrating variance in exponential bDNA growth as a function of amplifier deposition cycles. Binding events are modeled as a binomial process with 75% probability of success (95% confidence interval from bootstrap sampling).



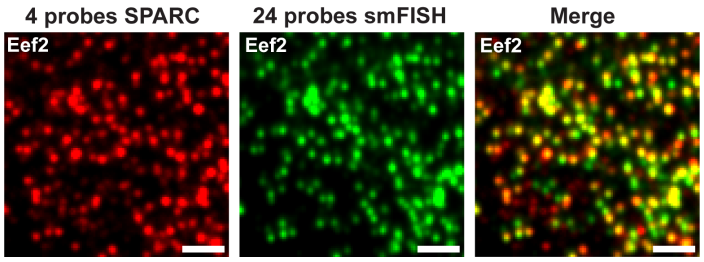
Supplementary Figure 3.2: A) Design of SPARC amplifiers to verify *in situ* crosslinking (left panel). SPARC amplifiers demonstrate high crosslinking efficiency and retention of signal compared to unmodified control amplifier. Scale bar = 20 microns. (Data collected by Katsuya Colón and Carsten Tischbirek) B) Representative images depicting efficient toehold-mediated strand displacement (scale bar = 20 microns). C) Representative images depicting the requirement of toehold-mediated strand displacement to enable hybridization of second amplifier. Scale bar = 20 microns (Data collected by Carsten Tischbirek).



Supplementary Figure 3.3: Bulk RNA-seq correlation against smFISH counts measured from glyoxal-fixed NIH 3T3 cells.



Supplementary Figure 3.4: A) Distribution of measured distance between two different SPARC generated branches on the same primary probe. B) Variance in intensity between branches becomes higher with more exponential cycles. I = Iteration, E = Exponential.



Supplementary Figure 3.5: Representative image depicting visible signal using only 4 probes targeting Eef2 and its colocalization with 24 probes smFISH (Scale bar = 2 microns). Data collected by Carsten Tischbirek.

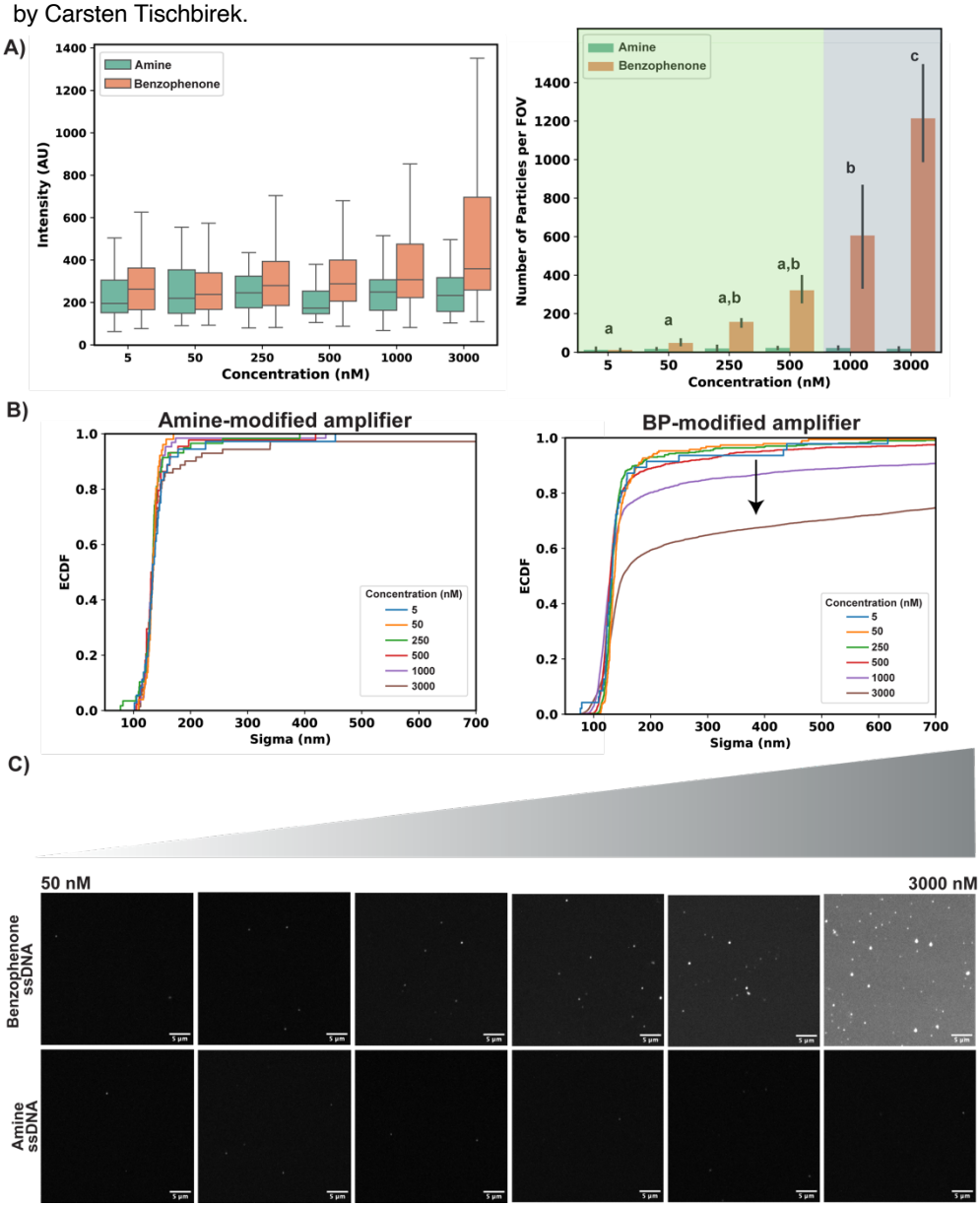
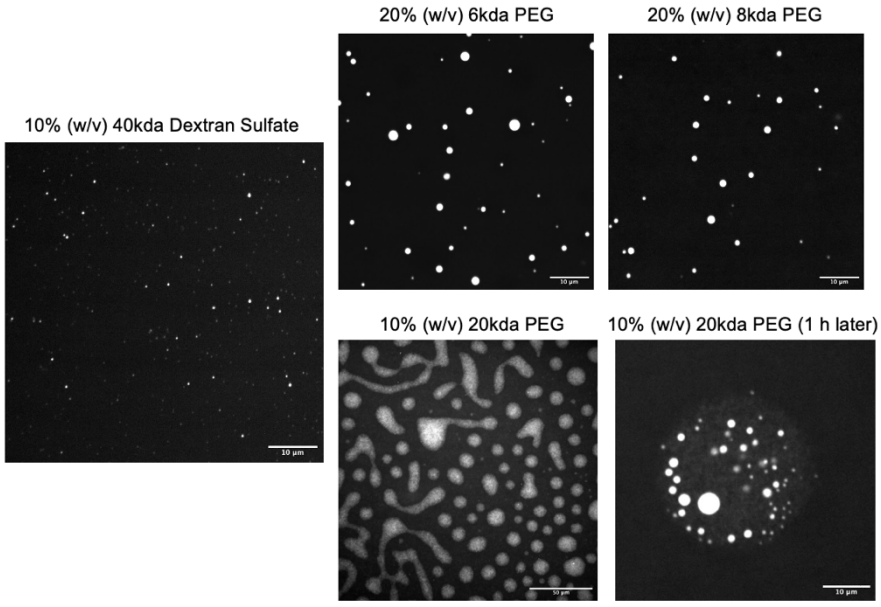
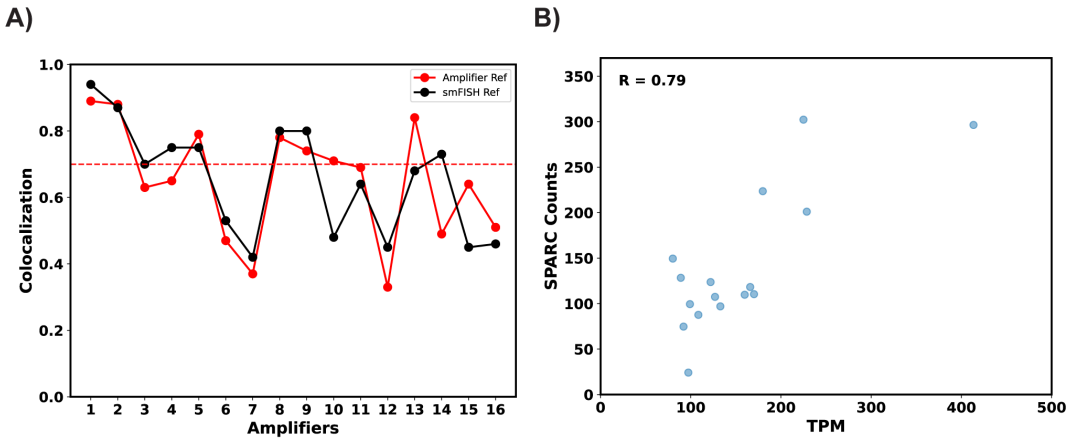


Figure 3.6: A) Intensity of particles at varying concentrations of amine- and benzophenone-modified amplifiers (left). Number of particles at varying concentrations of amine- and benzophenone-modified amplifiers (right). Conditions with non-overlapping unique letters corresponds to $p < 0.05$ using one-way ANOVA and Tukey HSD. Gray shade indicates critical concentration point for aggregate formation. B) ECDF of particle size at varying concentrations of amine- (left) and benzophenone-modified (right) amplifiers. Downward arrow indicating increase in particle size as the concentration of benzophenone amplifier increases. C) Representative images of particle formation as the concentration of benzophenone-modified amplifiers increase. Scale bar = 5 microns. (Data collected by Carsten Tischbirek and Katsuya Colón. Analyzed by Katsuya Colón.)



Supplemental Figure 3.7: Representative images of various crowding agents that affect BP-amplifier interactions (Data collected by Carsten Tischbirek and Katsuya Colón).



Supplemental Figure 3.8: A) Noise is measured as the percentage of amplifiers that colocalize with smFISH. Percentage of smFISH spots that colocalize with amplifier spots correspond to the amount of potentially missed detection of true signal. Dashed red line corresponds to 70% colocalization. B) Bulk RNA-seq correlation against SPARC counts.

Name	Sequence
Forward primer for pool generation	GCCCCATCATGTGCCTTTCC
Reverse primer for pool generation	GGCCGGTAATACGACTCACTATAG
Hybrid primer for RT	GCCCCrATCArUGTGrCCTTTCrC
200 mer TTG	$(TTG)_{66}TG$
PolyT-647	/5Alex647N/TTTTTTTTTTTTTTTTTTTT
Lamin-B1 Sequence	/5AzideN/TTTTAATCAAAGGCCGCA
R2A2 USER cleavage amplifier	CCGTTACAACCGTCA/ideoxyU/G/ideoxyU/G/ideoxyU/A/ideoxyU/G/ideoxyU/g AAACCGTCTTCAATAtt

Supplementary Table 3.1: Primers and sequences used in this study.

Amplifier	L1 Sequence	L2 Sequence
1-L1L2	TGCGGCCTTTGATTACTAGTGATTCGCAC CCTAGTGATTCGCACC	GGTGCGAATCACTAGTAATCAAAGGCC GCATAATCAAAGGCCGCA
2-L1L2	TGTAATCCGGGAGCACGCTTGCGATTTCG ACGCTTGCGATTTCGA	TCGAAATCGCAAGCGTGCTCCCGGATT ACATGCTCCCGGATTACA
3-L1L2	AACGCTGTACGAGTGCGCTTTTACGTACC ACGCTTTTACGTACCA	TGGTACGTAAAAGCGCACTCGTACAGC GTTCACTCGTACAGCGTT
4-L1L2	ATAGGCGTACGAGGGATCTTGCGCGTTG GGATCTTGCGCGTTGGG	CCCAACGCGCAAGATCCCTCGTACGCC TATCCCTCGTACGCCTAT
5-L1L2	CATCCCGTCTGGATTTTCAGAGCCGGTAG ATTCAGAGCCGGTAGA	TCTACCGGCTCTGAAAATCCAGACGGG ATGAATCCAGACGGGATG
6-L1L2	TGGCCGCGAATTCTAATGCGTCGGTGCTA TATGCGTCGGTGCTAT	ATAGCACCGACGCATTAGAATTCGCGG CCATAGAATTCGCGGCCA
7-L1L2	CGCCGTATCCTGATTGTTACGGAACAGCG AGTTACGGAACAGCGA	TCGCTGTTCCGTAACAATCAGGATACGG CGAATCAGGATACGGCG
8-L1L2	TGCATATCCGAGACCGCGTGATCGGTGA TTGCGTGATCGGTGATT	AATCACCGATCACGCGGTCTCGGATAT GCAGGTCTCGGATATGCA
9-L1L2	ACATTGTCGTACGCACGCCGAGCGATATA ACGCCGAGCGATATAA	TTATATCGCTCGGCGTGCGTACGACAAT GTTGCGTACGACAATGT
10-L1L2	CCACGTCGTATGATAACATTCGCAAGGAA CACATTCGCAAGGAAC	GTTCCTTGCGAATGTTATCATACGACGT GGTATCATACGACGTGG
11-L1L2	TGAGCGCCCAATTGAAACTATACCGTCCA AAACTATACCGTCCAA	TTGGACGGTATAGTTTCAATTGGGCGCT CATCAATTGGGCGCTCA
12-L1L2	CGTACGATCCTAGAGTATTGGCTGTTACG ATATTGGCTGTTACGA	TCGTAACAGCCAATACTCTAGGATCGTA CGCTCTAGGATCGTACG
13-L1L2	TCCCAGTGCGAGTCATCTTTACCGACTCA TTCTTTACCGACTCAT	ATGAGTCGGTAAAGATGACTCGCACTG GGATGACTCGCACTGGGA
14-L1L2	CGAATACCTCGACCATTCAATCCGTCGAC ATTCAATCCGTCGACA	TGTCGACGGATTGAATGGTCGAGGTATT CGTGGTCGAGGTATTCG
15-L1L2	CGCTAGCCTGAATATCAAGGTGATTCGTG TCAAGGTGATTCGTGT	ACACGAATCACCTTGATATTCAGGCTAG CGATATTCAGGCTAGCG
16-L1L2	AGACGTTACTAGCGTTGAGTAAAGACCGC CTGAGTAAAGACCGCC	GGCGGTCTTTACTCAACGCTAGTAACGT CTACGCTAGTAACGTCT
17-L1L2	TCGAATATGCCTAGGAAACCGCTATATGC CAAACCGCTATATGCC	GGCATATAGCGGTTTCCTAGGCATATTC GACCTAGGCATATTCGA
18-L1L2	TCGCGTAAGAGTACTTCTGCGTCGAGAGT TTCTGCGTCGAGAGTT	AACTCTCGACGCAGAAGTACTCTTACGC GAAGTACTCTTACGCGA
19-L1L2	GCGTGCTATTACGACACATAACATCCCGG TACATAACATCCCGGT	ACCGGGATGTTATGTGTCGTAATAGCAC GCGTCGTAATAGCACGC

20-L1L2	CGATGTAAAGGTGCTGCGAATACCATACG CGCGAATACCATACGC	GCGTATGGTATTCGCAGCACCTTTACAT CGAGCACCTTTACATCG
21-L1L2	AATCAGCGATCGTGGCAAATCCGCGTGTT ACAAATCCGCGTGTTA	TAACACGCGGATTTGCCACGATCGCTG ATTCCACGATCGCTGATT
22-L1L2	AACGAGTATGGGCCAGCTCTAAGTGTCC GAGCTCTAAGTGTCCGA	TCGGACACTTAGAGCTGGCCCATACTC GTTTGGCCCATACTCGTT
23-L1L2	TGCCTATCGCTGAAGCGCCGTGACGTATT TCGCCGTGACGTATTT	AAATACGTCACGGCGCTTCAGCGATAG GCACTTCAGCGATAGGCA
24-L1L2	CACGCGTATACGACTGTCGATTGATAAGG CGTCGATTGATAAGGC	GCCTTATCAATCGACAGTCGTATACGCG TGAGTCGTATACGCGTG

Supplementary Table 3.2: Amplifier sequences for SPARC.

Displacer Name	L1 Displacers
D1	GAATCACTAGTAATCAAAGGCCGCA
D2	ATCGCAAGCGTGCTCCCGGATTACA
D3	CGTAAAAGCGCACTCGTACAGCGTT
D4	CGCGCAAGATCCCTCGTACGCCTAT
D5	CGGCTCTGAAAATCCAGACGGGATG
D6	ACCGACGCATTAGAATTCGCGGCCA
D7	GTTCCGTAACAATCAGGATACGGCG
D8	CCGATCACGCGGTCTCGGATATGCA
D9	TCGCTCGGCGTGCGTACGACAATGT
D10	TTGCGAATGTTATCATACGACGTGG
D11	CGGTATAGTTTCAATTGGGCGCTCA
D12	ACAGCCAATACTCTAGGATCGTACG
D13	TCGGTAAAGATGACTCGCACTGGGA
D14	ACGGATTGAATGGTCGAGGTATTCG
D15	AATCACCTTGATATTCAGGCTAGCG
D16	TCTTTACTCAACGCTAGTAACGTCT
D17	ATAGCGGTTTCCTAGGCATATTCGA
D18	TCGACGCAGAAGTACTCTTACGCGA
D19	GATGTTATGTGTCGTAATAGCACGC
D20	TGGTATTCGCAGCACCTTTACATCG
D21	CGCGGATTTGCCACGATCGCTGATT
D22	CACTTAGAGCTGGCCCATACTCGTT
D23	CGTCACGGCGCTTCAGCGATAGGCA
D24	ATCAATCGACAGTCGTATACGCGTG

Supplementary Table 3.3: Displacer sequences for SPARC.

Readout name	L2 Readout sequence
SP1	TGCGGCCTTTGATTA
SP2	TGTAATCCGGGAGCA
SP3	AACGCTGTACGAGTG
SP4	ATAGGCGTACGAGGG
SP5	CATCCCGTCTGGATT
SP6	TGGCCGCGAATTCTA
SP7	CGCCGTATCCTGATT
SP8	TGCATATCCGAGACC
SP9	ACATTGTCGTACGCA
SP10	CCACGTCGTATGATA
SP11	TGAGCGCCCAATTGA
SP12	CGTACGATCCTAGAG
SP13	TCCCAGTGCGAGTCA
SP14	CGAATACCTCGACCA
SP15	CGCTAGCCTGAATAT
SP16	AGACGTTACTAGCGT
SP17	TCGAATATGCCTAGG
SP18	TCGCGTAAGAGTACT
SP19	GCGTGCTATTACGAC
SP20	CGATGTAAAGGTGCT
SP21	AATCAGCGATCGTGG
SP22	AACGAGTATGGGCCA
SP23	TGCCTATCGCTGAAG
SP24	CACGCGTATACGACT

Supplementary Table 3.4: Readout sequences for SPARC.

Appendix

Development of pyFISH for seqFISH and smFISH data processing

A.1 Abstract

In this work, we introduce pyFISH, an open-source image preprocessing workflow tailored for standard smFISH and sequential FISH (seqFISH) datasets. PyFISH offers numerous algorithms which include image registration, signal processing, spot detection, cell segmentation, decoding, and more. We also highlight two major enhancements over our prior seqFISH analysis workflow: the integration of DAOStarFinder, a spot detection algorithm from stellar photometry and a novel support vector machine (SVM) embedded, score-based radial search decoding algorithm. PyFISH is formatted for deployment on a high-performance cluster (HPC) and leverages parallel processing to accelerate the analysis of large-scale datasets. Using pyFISH, we demonstrate reduced false discovery rates for the previously published seqFISH+ dataset to below 5% while retaining moderate detection efficiencies.

A.2 Introduction

The field of spatial biology has gained significant momentum in the recent past, with new spatial platforms and methodologies arising from numerous academic labs and industrial partners. Our lab has pioneered the field of spatial biology and developed a combinatorial barcoding method termed sequential FISH (seqFISH) that allowed us to probe hundreds to thousands of genes in intact cells.[1–3] Moreover, we developed RNA[4] and DNA[5] seqFISH+ which allowed us to profile 10,000 unique transcripts and >3,000 DNA loci, respectively. Based on these technologies, we sought to improve and package our previous seqFISH preprocessing workflow by developing pyFISH.

Image preprocessing for seqFISH involves many steps, which includes image registration, potential chromatic aberration corrections, signal processing, spot detection, cell segmentation, mask corrections, decoding, and more. PyFISH essentially provides algorithms for all these steps, streamlining seqFISH data preprocessing (**Figure A.1A**).

We also incorporated two major advancements that substantially enhanced seqFISH decoding outcomes which includes a spot detection algorithm previously used in stellar

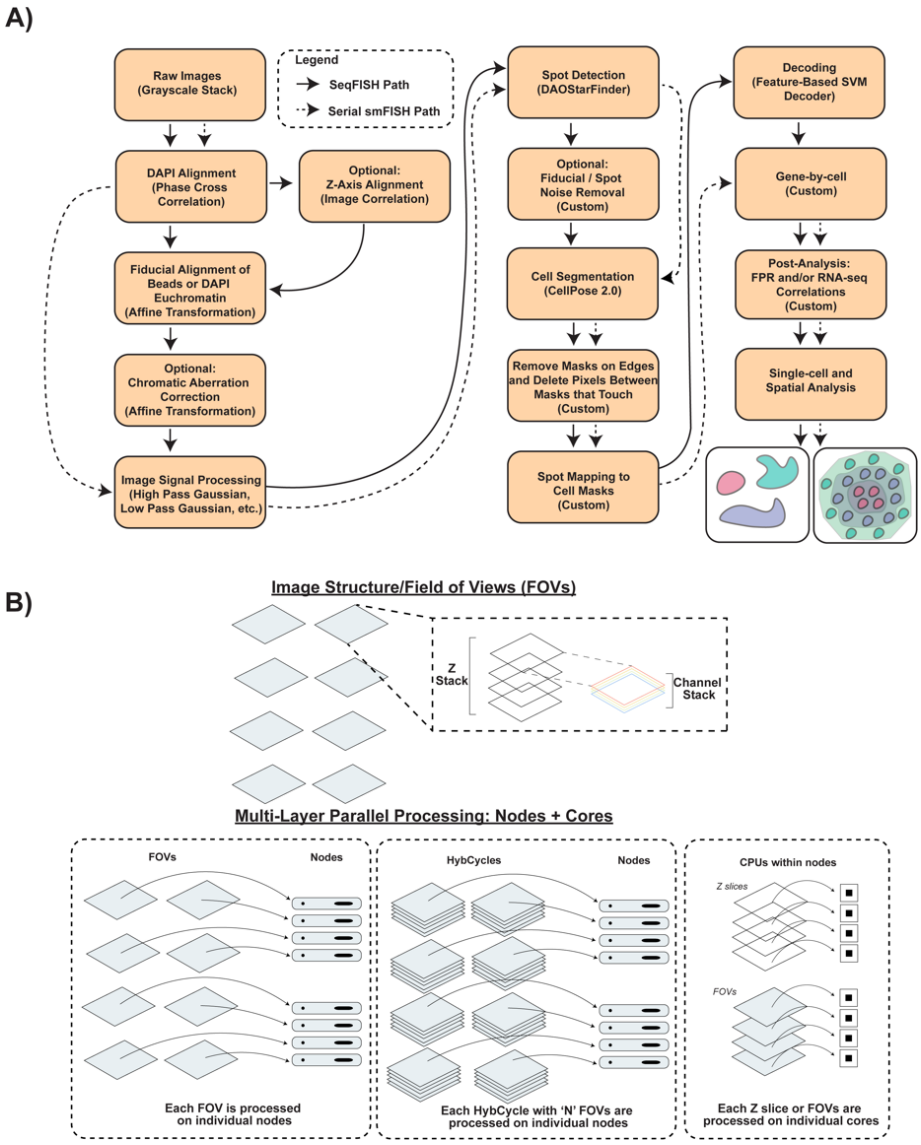


Figure A.1: A) General pyFISH workflow for seqFISH or standard smFISH analysis. B) Graphic of multi-layer parallel processing for enhanced data preprocessing speeds.

photometry called DAOStarFinder[6] and a novel support vector machine (SVM) embedded, score-based radial search decoding algorithm that reduces false discovery rates. Designed for versatility, pyFISH seamlessly integrates with the Slurm workload manager, making it apt for deployment on a high-performance cluster (HPC). One of the pipeline's hallmarks is its adept use of asynchronous parallel processing, which translates to

markedly accelerated processing speeds. Additionally, the parallel job submission capabilities of the HPC enables an additional layer of parallel processing (**Figure A.1B**), allowing for the preprocessing of up to 1 TB or more of seqFISH data in just a few days or less. With this, we not only improved data preprocessing throughput significantly, but further improved the data quality of previously published seqFISH+ results.

A.3 Results

A.3.1 Image processing suite

In pyFISH, we provide three options for image registration: (1) phase cross-correlation of DAPI-stained images, (2) scale-invariant feature transform (SIFT) for DAPI-stained images, and (3) affine transformation using fiducial markers such as fluorescent beads or punctates in DAPI-stained images. Generally, phase cross correlation or SIFT can obtain alignment accuracies of less than 100 nm. However, if distortions occur in DAPI stained images, then phase cross correlation or SIFT (although more robust) may not work as efficiently. With affine transformation using fiducial markers, we can obtain less than 30 nm alignment accuracy using seqFISH+ data (**Figure A.2A**). This added alignment accuracy can improve decoding outcomes if spots are super-resolved. Our fiducial alignment algorithm begins by first obtaining sub-pixel centroids of spots in images containing only fluorescent fiducial markers. If DAPI stained images are used, then the algorithm will attempt to pick bright heterochromatin spots. Once these coordinates are obtained, the algorithm searches for these fiducial markers in the moving images using a radial search with a user defined pixel distance cutoff across barcoding rounds. Once a set of candidate reference and moving fiducials are identified, an affine transformation matrix is calculated while utilizing random sample consensus (RANSAC) algorithm to obtain the best set of fiducial markers that yields the lowest alignment error. This transformation matrix is then applied to the image for image registration. Additionally, if

barcodes are encoded across channels, such as in RNA SPOTs,[7] then chromatic aberrations must be corrected for efficient colocalization of spots. Using a similar approach to fiducial marker alignment, we can utilize fiducial markers that are fluorescent across the desired channels and use a single channel as reference for affine transformation. Utilizing this approach, we can obtain chromatic aberration alignment accuracies of approximately 30 nm (**Figure A.2B**).

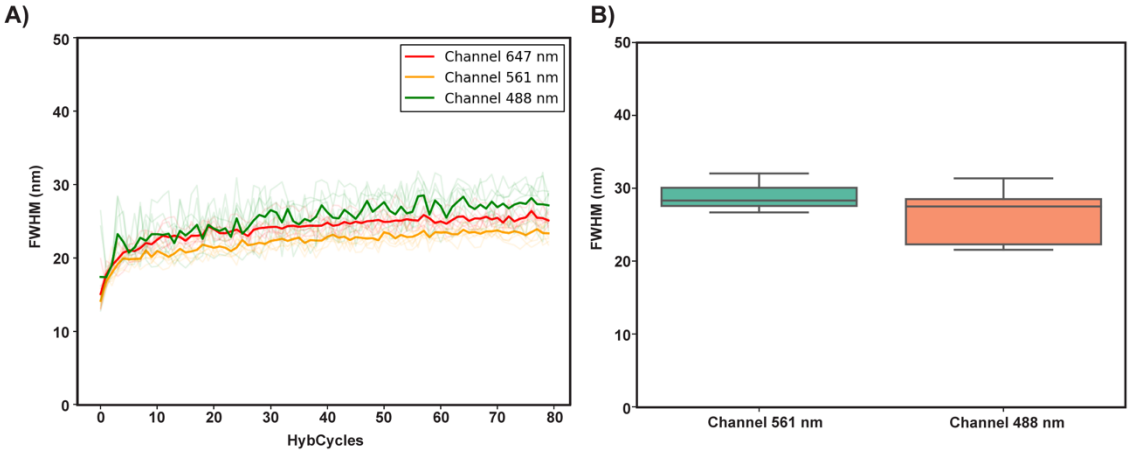


Figure A.2: A) Alignment error per channel across hybridization rounds. Dark-colored lines represent the mean alignment error, whereas the more transparent lines represent different fields of view. B) Box plot representing alignment error of spots across channels after chromatic aberration correction. Quantification was performed on varying fields of view. Horizontal line corresponds to median.

Although we should expect sub-nanometer alignment accuracy with fiducial markers, we find that is generally not the case when performing seqFISH experiments. This could be attributed to slight bleaching of beads which can cause distortions in gaussian fits or z-plane drift. In some cases, z-plane drifts of 250 nm can occur even with autofocus enabled. To solve this problem, pyFISH enables z-axis alignment with normalized cross-correlational analysis of DAPI stained reference images. During z-axis alignment the algorithm searches for the best matching z plane across all hybridization rounds or time series images and clip z-slices so that each z-slice corresponds better with the subsequent rounds. However, if there are distortions in DAPI stained images, then more sophisticated algorithms are required.

Post alignment, images typically undergo various signal processing techniques to boost desired signal. We tested various algorithms, such as Richardson-Lucy (RL) deconvolution, high-pass median or Gaussian filters, low-pass median or Gaussian filters, rolling ball subtraction, blank/dark image subtraction, and more. We found that a standard signal processing sequence starting with a 7x7 2D high-pass Gaussian filter, followed by a 3x3 2D low-pass Gaussian filter, and percentile intensity scaling, typically produces decent results for cell culture images (**Figure A.3**). It is important to acknowledge that various samples necessitate distinct pre-processing approaches. To cater to individual user requirements, the pipeline includes methods such as TopHat transformation to remove large fluorescent blobs (ex. lipofuscin), rolling ball subtraction, blank/dark image subtraction, boosted high-pass Gaussian filter, RL deconvolution, and autofluorescent/noisy spot location removal by user defined search radius. This flexibility ensures that the pipeline can be tailored to address the unique needs of each sample effectively.

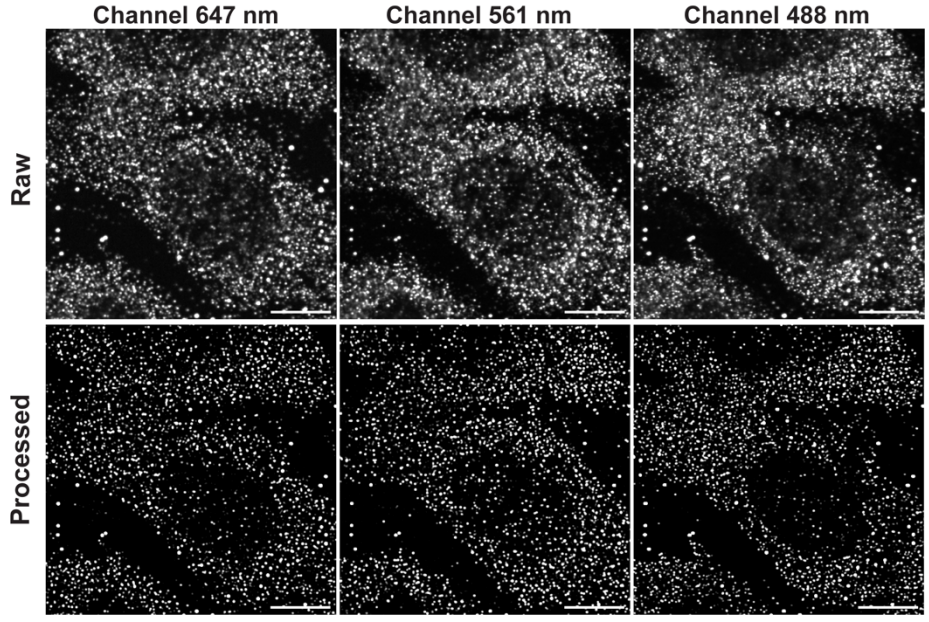


Figure A.3: Representative figure depicting signal processed images from seqFISH+ dataset after 7x7 high pass Gaussian filter, 3x3 low pass Gaussian filter, and intensity scaling. Scale bar = 10 microns.

Finally, pyFISH utilizes Cellpose, a deep learning algorithm developed by Pachitariu lab, to generate cell masks.[8,9] Cellpose is a robust algorithm that already comes packaged with a pre-trained model that is readily deployable across various tissue types. It utilizes intensity gradients of stained cells, similar to the watershed algorithm, coupled with a U-Net type deep learning architecture. Cellpose also offers human-in-loop model training which makes segmenting various tissue types more robust. After mask generation, cell masks that are touching the borders of the FOV are deleted since they will be partial masks and will result in differences regarding depth (number of barcodes decoded). In some cases, if the FOV has uneven illumination near the edges, then removing edge masks can reduce unwanted artifacts where cells on the edges have reduced decoded barcodes. Additionally, users can also choose to delete pixels between the interface of two masks that touch to reduced barcode mixing and improve cell-type clustering outcomes (**Figure A.4**).

Figure A.4: Representative images showing mask generated from polyT stained NIH 3T3 cells using Cellpose. Scale bar = 20 microns.

A.3.2 Spot detection with DAOStarFinder

PyFISH utilizes an algorithm called DAOStarFinder, a tool commonly used in crowded stellar photometry,[6] for detecting and obtaining sub-pixel centroids of fluorescent spots. DAOStarFinder is a highly optimized algorithm capable of obtaining sub-pixel centroids of spots with remarkable speed. It does this by taking a 2D cutout of each spot,

marginalizing along x and y, and calculating a single-step linear least-squares solutions for the amplitude and centroid shift in each 1D direction (using user-defined sigma). Conventionally, sub-pixel centroids can be determined by performing a full 2D Gaussian fit of the cutout via a multi-step nonlinear optimizer, which is a computationally expensive task. To assess the accuracy of DAOStarFinder, the algorithm was benchmarked against the conventional 2D Gaussian fit and radial centering[10] on simulated spots. Simulated spots were either a 6x6 or 5x5 gaussian kernel with or without camera noise. Both overlapping and single spots were generated to assess centroid accuracy of the three algorithms (**Figure A.5A**). The three algorithms performed similarly when assessed on a single spot with or without background noise, however, a substantial difference becomes apparent when assessing centroid accuracy on overlapping spots with DAOStarFinder and radial centering outperforming the standard 2D gaussian fits (**Figure A.5B**). Such results are expected considering the centroid from 2D gaussian fits can get skewed by neighboring spots, hence the initial development of radial centering to overcome such issues. In most cases, radial centering has slightly higher centroid accuracy than DAOStarFinder (**Figure A.5B**). However, DAOStarFinder provides additional features from its algorithm such as symmetry, Gaussian amplitude, flux, and sharpness of each spot. These features are later used to train a support-vector machine (SVM) to distinguish spots arising from non-specific signal from those representing true signal prior to seqFISH decoding. Finally, this fast and accurate algorithm was tested against seqFISH+ images which showed that DAOStarFinder can detect smFISH signal and resolve overlapping spots (**Figure A.5C**). Since DAOStarFinder performs Gaussian fitting, it is critical that the spots have Gaussian features post signal processing (hence the re-convolution), otherwise the algorithm will have difficulty detecting spots. Additionally, DAOStarFinder

also requires numerical optimization of the full-width half maximum parameter to efficiently identify smFISH spots.

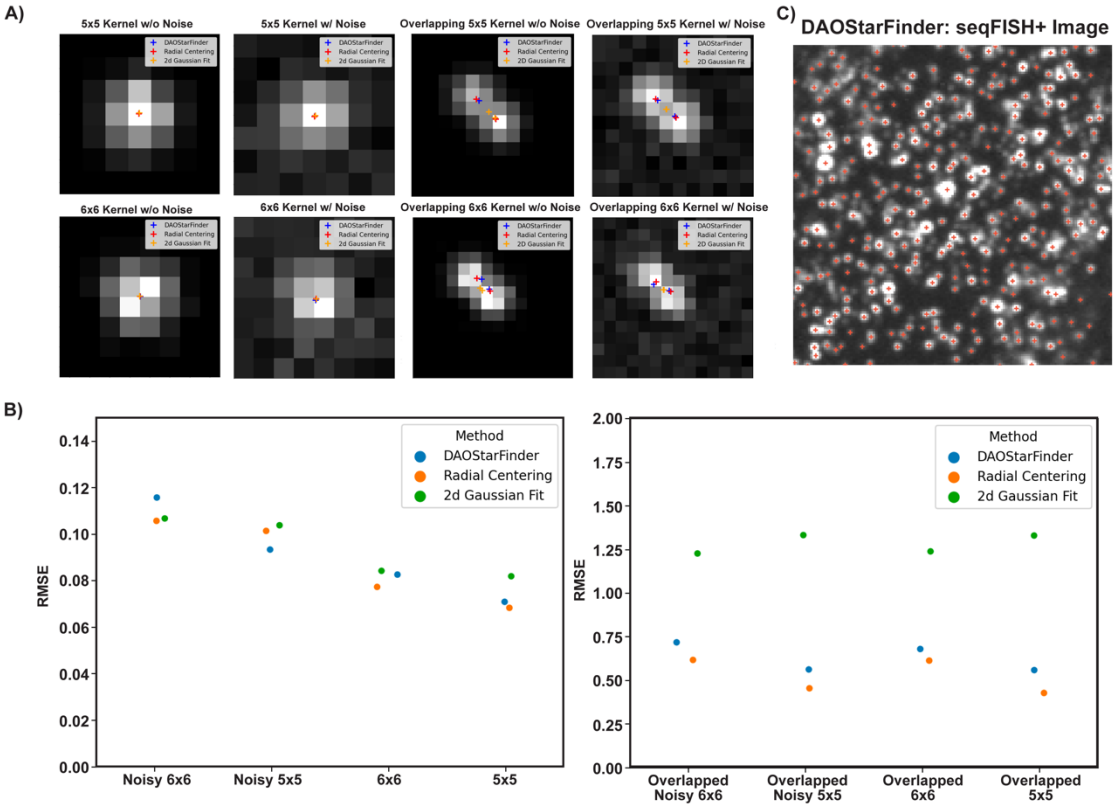


Figure A.5: A) Representative simulated images showing location of predicted centroids. B) RMSE of each spot detection algorithm and centroid predictions. C) Representative image depicting centroid locations obtained from seqFISH+ images.

A.3.3 Feature and Score Based Radial Decoding

In addition to the previously mentioned improvements, we developed a highly generalizable decoding algorithm for seqFISH capable of handling various barcoding strategies, such as within-channel encoding where all barcodes will be encoded in a single channel or across channel encoding where unique barcodes can appear in multiple channels. This algorithm aims to reduce false discovery rates (FDR) and increase decoding accuracy which is achieved by incorporating a feature-based scoring scheme, an SVM model, multi-round decoding, comparative analysis between clashing codewords, priority decoding by gene expression, and automated FDR tuning. Moreover, this decoder utilizes the complete codebook which includes both assigned and unassigned barcodes

as opposed to decoding the assigned barcodes first followed by unassigned barcodes, addressing a potential issue of biased false-positive detection in our previous approaches.[4]

Our decoding algorithm begins by performing a radial search for each spot, using each barcoding round as a reference (termed seeds), similarly to previous methods.[4] These searches run in parallel, comparing each spot in the reference round to other rounds. For every spot within the defined search radius, a score is assigned based on distance, flux, and size (**see Additional Information**). Scoring is particularly important as multiple spots can be found within a single search radius depending on density. Furthermore, true spots can shift slightly due to imaging artifacts, and non-specific signals can introduce noisy spots, complicating the identification of the correct barcode. The highest-scoring spot is selected for each barcoding round, and once these selections are made, an overall codeword score is calculated. Each codeword is then normalized by an ambiguity score, which accounts for additional neighboring spots beyond the expected count for each seed. The weight of the ambiguity score is dependent on the decoding round where it increases at each subsequent round. The ambiguity score is included to account for potential error when decoding dense regions and act as a metric for confidence in picking the correct spots. This can be thought synonymously to cluster density and how it affects q-scores in sequencing technologies. Each codeword is also filtered by a minimum seed criterion, which ensures that all reference barcodes observe the same neighboring spots (**Figure A.6A**). Similar to our previous decoder, this algorithm can recover codes that experience spot call dropouts (since codes are hamming distance 2).[4] Additionally, this algorithm will perform parity checks to ensure that valid codes are picked. Parity checks are essential when distinguishing true signal from noise/blank codes. Here, blank codes are defined as the unassigned barcodes in the codebook.

Detection of blank codes can arise if there are insertion of noisy spots which can arise from non-specific binding of fluorescent readout probes or if the density of spots is high which can lead to improper assignment. Since this algorithm must account for potential noise or ambiguity in the data, parity checks aide in reducing false positives and increase decoding accuracy. Generally, most decoded blank codes arise from codewords that lack parity or have dropouts in spot calls. Finally, if multiple codewords have similar spots picked, their overall codeword score is used to select the best one. However, if two clashing codewords have the same score, then the codeword with lowest distance cost is picked, which corresponds to overall spread of spots (**Figure A.6B**).

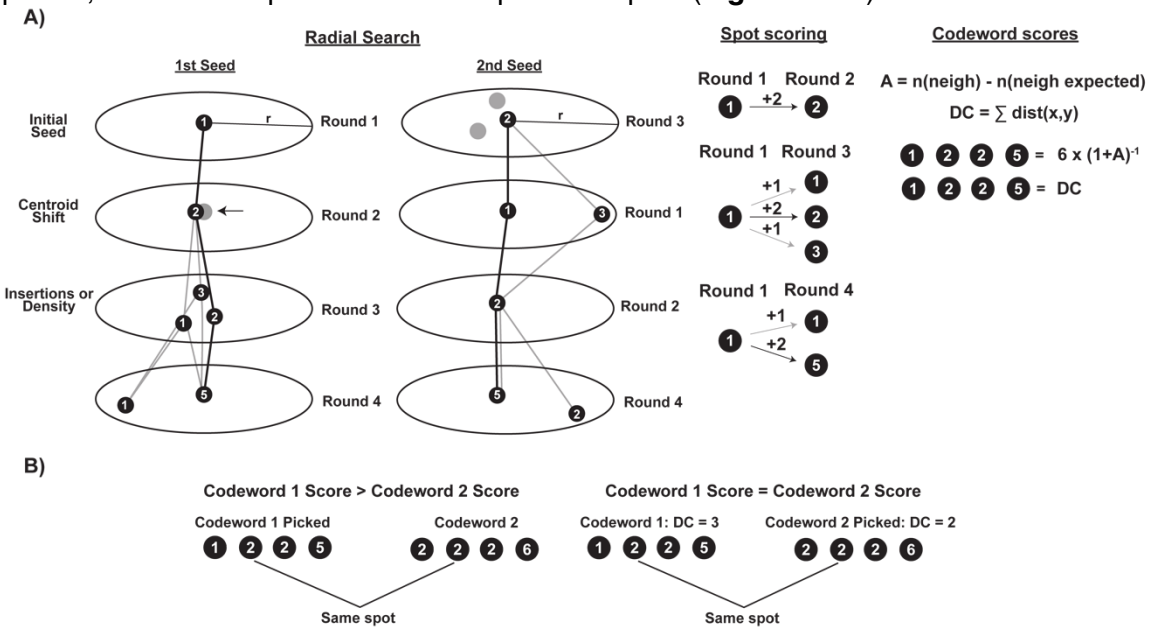


Figure A.6: A) Schematic depicting feature-based radial search algorithm. Initially, the algorithm begins by using spots from a certain barcoding round as reference (seeds) and searches for spots with a user-defined search radius across barcoding rounds. Within a given search radius, there can exist many potential spots, some of which may correspond to another barcode or an undefined barcode (does not correspond to true or blank codes). Spots are chosen based on weighted scores using intensity, distance from reference spot, and size of spot. Scoring is performed to address potential imaging artifacts, such as centroid shifts and noisy signal insertions, and to manage high-density regions where multiple spot choices are available. The sum of spot scores corresponds to the codeword score and it is normalized by the ambiguity score which corresponds to the number of additional spots observed than expected. The distance cost is also calculated as the sum of distances of each chosen spot using round 1 as reference. Probable codewords are determined by min seed criteria where changing the barcoding round reference obtains the same spot combination. Black line represents chosen spots, and gray line shows other potential spot choices. B). Schematic depicting clashing codewords where two potential barcode choices (but not limited to two) can arise from similar spots. If there are multiple choices for a codeword, then the codeword with the high score will be chosen (left). However, if the clashing codewords have the same score, then the codeword that passed min seed criteria with the lowest distance cost (sum of all distances from seed) is used.

This algorithm also performs multi-round decoding. In the initial round of decoding, the algorithm will identify true and noisy spots. Each spot has 6 unique features which includes, roundness by symmetry, roundness based on marginal Gaussian fits, flux, peak amplitude, sharpness, and spot area. These features are used to train an SVM model using the labels obtained from first pass decoding to assign probabilities to each spot on the likelihood that the spot is true signal. The user can define the probability cutoff to filter potentially noisy spots prior to decoding. Post filtering, the algorithm will begin actual decoding on the cleaned-up dataset up to three rounds. Unused spots will be resubmitted each round of decoding to assign leftover spots. The search radii can be adjusted per round to have an expanding search radius. Generally, search radii should be kept <= 2 pixels. In some cases, like hydrogel embedded samples, a search radius of 3 pixels may be required. Finally, the final scores assigned to each spot is sorted from most to least confident. The scores generally follow a monotonic relationship where high scoring codewords are true codes and low scoring codewords are false codes. Using the scores, the FDR can be tuned by the user's choice where low scoring codewords (low quality), which includes both true and false codes, can be excluded from the final gene by cell matrix. Moreover, this algorithm includes an option to decode the top 10% of highly expressed genes in each FOV in the first round of decoding (post SVM filtering) to reduce spot density.

This innovative decoding algorithm for seqFISH offers increased decoding accuracy and reduced FDR, making it an effective tool for various barcoding strategies and applications. We re-analyzed the previously published seqFISH+ data using this new decoder, which had a barcode assignment error rate of 22%, and we obtained an FDR value of less than 5% (**Figure A.7A-B**). This demonstrates the effectiveness of our

algorithm in obtaining low FDR, comparable detection efficiencies, and results compared to previously published results.

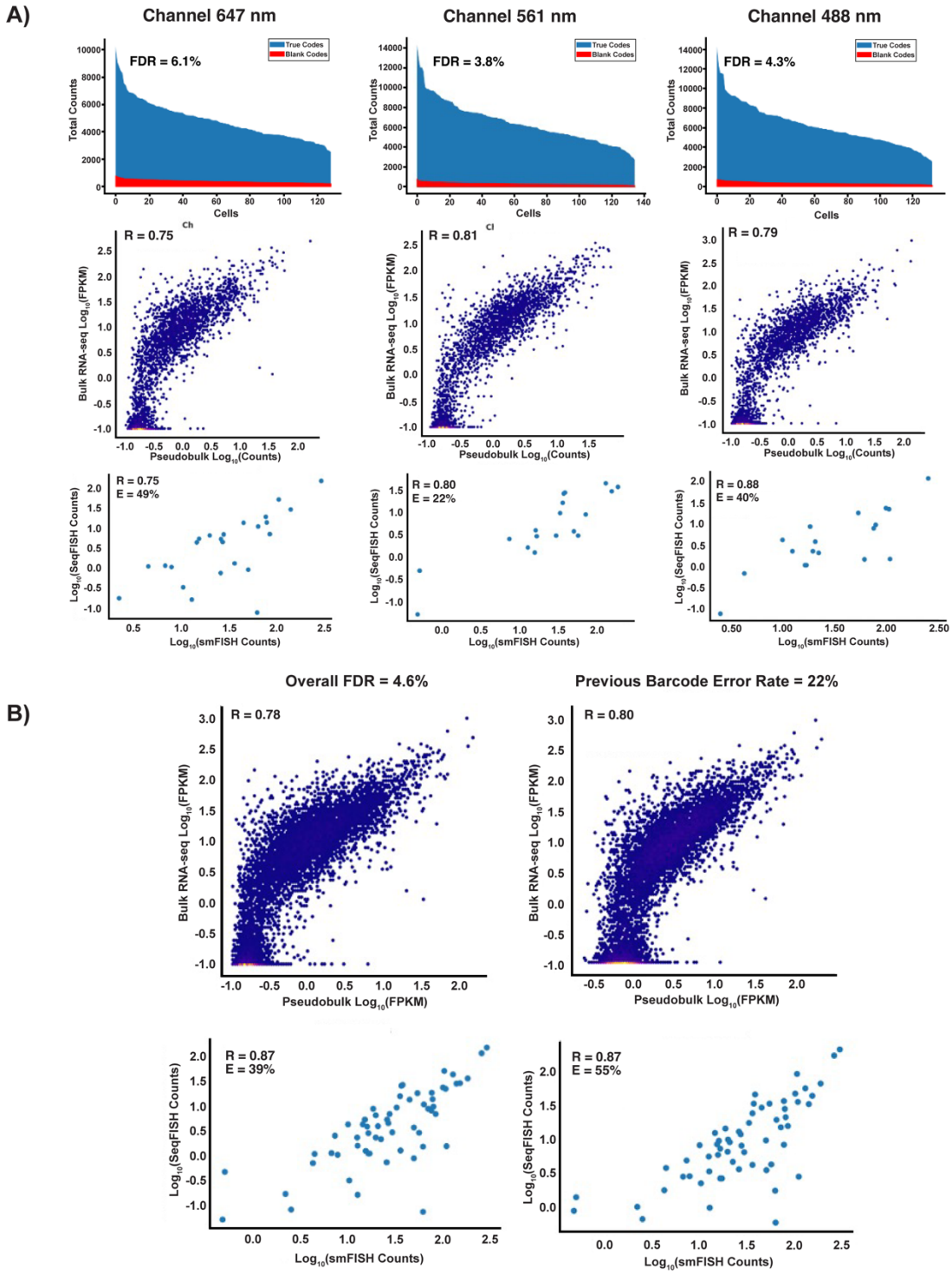


Figure A.7: A) Correlations, FDR, and detection efficiencies obtained from pyFISH workflow for each channel from seqFISH+ dataset. B) Combined FDR was measured to be 4.6% using pyFISH compared to the previously published barcode error rate of 22%. However, the detection efficiency dropped from 55% to 39%.

A.4 Discussion

We have developed pyFISH as a standard workflow for preprocessing various seqFISH and serial smFISH datasets. PyFISH offers a suite of various pre-processing algorithms such as image registration, chromatic aberration corrections, image signal processing, spot detection, cell segmentation, mask corrections, decoding, and more. Furthermore, it offers quality check outputs such as fiducial alignment accuracy, chromatic aberration correction accuracy, matched z axis information, percent spots utilized in decoding, locations of noisy codes (did not pass parity), labels for spots (true, blank, or undefined), and SVM outputs such as median probabilities for true and false spots. These quality check outputs can prove to be useful when troubleshooting experimental datasets. Additional algorithms offered in pyFISH includes colocalization measurements between hybridization rounds and across channels. We highlight two new improvements to pyFISH compared to previously published seqFISH preprocessing workflows, which includes a spot detection algorithm called DAOStarFinder and a novel SVM-embedded, feature-based radial decoder. We utilized pyFISH to reanalyze previously published seqFISH+ dataset and obtained low FDR values with comparable detection efficiencies. This pipeline has been shown to provide adequate results across numerous datasets which include cell culture,[4] mouse testis,[11] and mouse kidneys.[12] With parallel processing capabilities and parallel job submission, pyFISH can process TBs of data within a week. We hope that this pipeline can act as a template for future development in pre-processing for spatial transcriptomic/genomic datasets. Additionally, this open-source pipeline can act as a useful resource for new researchers interested in performing seqFISH.

A.5 Additional Information

Codeword scoring. Scoring begins by first grouping spots together that fall within a given search radius for each barcoding round at a given seed reference. Individual spots are assigned scores based on proximity to a reference spot, along with flux and size of the dot compared to others within the search radius. A score table will be generated which has distance (50%) as the highest weight, followed by brightness (37.5%) then size (12.5%). The range of the score table will always be from 0-2, however, the length of the score table will be equally spaced values from 0-2, where the number of elements is dependent on the number of dots being compared plus one. As such, the max score a dot can have will be 2.0, if the dot was determined to be the closest in distance to reference, brightest, and largest in size.

Once the highest scoring dot is picked for each barcoding round at a given seed, the sum of scores for each dot will be the codeword score (**Eqn. 1**). A codeword that had no other neighbors will have a max score of 2 x number of barcoding rounds.

$$codeword\ score = \sum_{i=1}^{i=\#\ of\ dots} (distance_{score\ i} + brightness_{score\ i} + size_{score\ i})$$

Eqn. 1: Summation of best spot scores.

Additionally, the ambiguity score will also be calculated as the total number of spots being compared minus the expected number of spots (**Eqn. 2**).

$$\textit{ambiguity score} = \text{n}(\textit{dots compared}) - \text{n}(\textit{expected dots})$$

Eqn. 2: Ambiguity assignment based on local spot density.

The total codeword score will be adjusted by the ambiguity score, given a certain weighting factor, as the confidence in codeword assignment will be reduced when there are many potential choices of spots (**Eqn. 3**). The weighting factor is added to increase the strength of ambiguity in subsequent decoding rounds. This is performed to account for the difference in density across multiple decoding rounds. For instance, the spot density will be highest in the first decoding round which would lead to a higher ambiguity score by default. In the subsequent rounds, the density will decrease since decoded spots from prior rounds are removed and so the ambiguity score will be substantially less. In order to have the effect of the ambiguity score to be similar across multiple decoding rounds, a weighting factor is included. It should be noted that spots left over in subsequent decoding rounds typically lead to increased decoding of blank codes. Therefore, applying a higher weight reduces the overall codeword score which ensures proper filtering to obtain high-fidelity codewords for post analysis. This weighting factor was determined empirically to generate a scoring scheme that follows a monotonic relationship where low scores typically correspond to blank/false codes.

$$ambiguity\ adjusted\ codeword\ score = \frac{codeword\ score}{\omega(ambiguity\ score + 1)}$$

Eqn. 3: Adjusting the codeword score based on the ambiguity score.

The overall adjusted codeword score will be scaled from 0-1 by dividing the value by the max achievable codeword score (**Eqn. 4**). The final score will essentially combine dot features, crowdedness, and distances.

$$final\ score = \frac{ambiguity\ adjusted\ codeword\ score}{(max\ dot\ score\ *\ \#\ of\ barcoding\ rounds)/\omega}$$

Eqn. 4: Normalizing final score by putting it on a scale from 0-1.

Simulation of spots. Spots were simulated to compare different sub-pixel fitting algorithms as follows. First, a Gaussian kernel was used as the point spread function to generate a spot on a larger 2D array. Next, camera noise was added using a Poisson model, with each pixel's value serving as the λ parameter. Finally, the fitting algorithms were applied to predict the sub-pixel centroids, and the root mean squared error (RMSE) was computed by comparing these predictions to the ground truth.

False discovery rate. False discovery rate is defined as follows[11,12]:

$$\text{FDR} = \frac{\#\text{ of Real Barcodes x } \frac{\text{Blank Counts}}{\#\text{ of Blank Codes}}}{\text{Real Barcode Counts}}$$

Number of real barcodes are defined as the number of gene-coding barcodes in the codebook while number of blank codes correspond to the number of unassigned barcodes in the codebook. Blank counts are the number of decoded blank barcodes, and the real barcode counts are the number of decoded gene-assigned barcodes. This essentially uses the frequency of blank code counts to estimate the number of false codes that may be detected as a true code (assuming frequency of error is the same). This value is normalized by the detection of all real barcode counts which includes both true positives and false positives.

Barcode assignment error rate. Previously used error rate is defined as follows:

$$\text{Error rate} = \frac{\frac{\text{Blank Counts}}{\#\text{ of Blank Codes}}}{\left(\frac{\text{Real Barcode Counts}}{\#\text{ of Real Barcodes}} + \frac{\text{Blank Counts}}{\#\text{ of Blank Codes}}\right)}$$

Number of real barcodes are defined as the number of gene-coding barcodes in the codebook while number of blank codes correspond to the number of unassigned barcodes in the codebook. Blank counts are the number of decoded blank barcodes, and the real barcode counts are the number of decoded gene-assigned barcodes. This essentially uses the rate at which blank codes and real barcodes appear to estimate the rate at which blank or real codes are misassigned.

pyFISH Github. PyFISH can be found at https://github.com/klcolon/pyfish_tools.

A.6 References

1. Shah, S. *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* **174**, 363-376.e16 (2018).
2. 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).
3. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
4. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235 (2019).
5. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 1–7 (2021) doi:10.1038/s41586-020-03126-2.
6. Stetson, P. B. DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry. *Publ. Astron. Soc. Pac.* **99**, 191 (1987).
7. Eng, C.-H. L., Shah, S., Thomassie, J. & Cai, L. Profiling the transcriptome by RNA SPOTs. *Nat. Methods* **14**, 1153–1155 (2017).
8. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
9. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. 2024.02.10.579780 Preprint at https://doi.org/10.1101/2024.02.10.579780 (2024).
10. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).

11. Chakravorty, A., Simons, B. D., Yoshida, S. & Cai, L. Spatial Transcriptomics Reveals the Temporal Architecture of the Seminiferous Epithelial Cycle and Precise Sertoli-Germ Synchronization. 2024.10.28.620681 Preprint at https://doi.org/10.1101/2024.10.28.620681 (2024).

12. Polonsky, M. *et al.* Spatial transcriptomics defines injury specific microenvironments and cellular interactions in kidney regeneration and disease. *Nat Commun* **15**, 7010 (2024).