UNIVERSITY OF CALIFORNIA SAN DIEGO

Understanding diverse RNA function using quantum dot based fluorescence in situ hybridization

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Norman Huang

Committee in charge:

Professor Sheng Zhong, Chair
Professor Prashant Mali
Professor Peter Ying Xiao Wang
Professor Jin Zhang
Professor Kun Zhang

2020

The dissertation of Norman Huang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

To my family and friends who supported me throughout these years and their unending patience.

EPIGRAPH

If you try and lose then it isn't your fault. But if you don't try and we lose, then it's all your fault.

—Orson Scott Card

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank and acknowledge Professor Sheng Zhong for his support as the chair of my committee and as my advisor for the past five and half years. His consistent positive and persevering attitude towards science and innovation continually motivated me and his entrepreneurial spirit is truly inspiring. I would also like to thank the other members of my committee, Professor Prashant Mali, Professor Peter Ying Xiao Wang, Professor Jin Zhang and Professor Kun Zhang for their contributions and wonderful discussion.

My PhD would not have been possible without the wonderful collaborative help from my fellow lab mates in the Zhong lab. Special thanks for Fernando Biase for his initial mentorship and guidance during the first year of my degree. His patient explanations and work ethic and prowess in experimental procedures helped me a lot in my development and understanding of molecular biology. Thanks to many experimental biologists that worked with me such as Kathia Zaleta-Rivera who patiently worked with me on many of my projects and helped speed up much of the experiment process. Xiaochen Fan for her fun conversation and help when going through arduous experimental procedures and collaboration on the cell surface work. My undergrads Yiqun Jiang and Jingyao Chen for their time and effort when we processed hundreds of tissue samples. Tri Nguyen for his wonderful advice and knowledge as well as the thought provoking discussions on politics and media.

Thanks to the computational biologists in the lab that helped me along in attacking more complicated problems with sequencing and image analysis. Jia Lu who shared our windowless office for 3 years, Xiaoyi Cao who maintained our computational infrastructure, Marcelo Rivas-Astroza who was also a wonderful roommate, Weizhong Chen and Zhangming Yan who helped me a lot with R and all the other computational biologists that helped me and to all the other lab members that helped give me advice Lucie Herbet, Kara Johnson, Zhijie Qi, Xingzhao Wen, Chien-ju Chen (William), and Xuerui Huang.

Last but not least, the incredible support from my friends and family throughout my PhD

without in San Francisco, in San Diego, and around the world. More importantly, I would like to thank my parents were the real trailblazers, who immigrated and made a new life in a foreign country at a time when the world was much larger and information was less connected. I would finally like to thank Jasmine Ou for her unwavering support throughout my ups and downs in my life and I could not end without emphasizing that. This dissertation is dedicated to them.

Chapter 3, in part is currently being prepared for submission for publication of the material. Huang, Norman, Zhong, Sheng. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the material as it appears in Genome-wide colocalization of RNA DNA interactions and fusion RNA pairs in Proceedings of the National Academy of Sciences of the United States of America 2018. Norman Huang, Zhangming Yan, Weixin Wu, Weizhong Chen, Yiqun Jiang, Jingyao Chen, Xuerui Huang, Xingzhao Wen, Jie Xu, Qiushi Jin, Kang Zhang, Zhen Chen, Shu Chien,and Sheng Zhong, *Proceedings of the National Academy of Sciences of the United States of America*, 2016. The dissertation author was contributing investigator and author of this paper.

Chapter 5 in part is currently being prepared for submission for publication of the material. Norman Huang, Xiaochen Fan, Kathia Zaleta Rivera, Tri C. Nguyen, Jiarong Zhou, Yingjun Luo, Jie Gao, Ronnie H. Fang, Zhen Bouman Chen, Liangfang Zhang, Sheng Zhong Naturally occurring cell surface-display of genome encoded RNAs and their impacts on cell-cell interactions. The dissertation author was the primary investigator and author of this paper.

Chapter 6 in part is currently being prepared for submission for publication of the material. Norman Huang, Xiaochen Fan, Kathia Zaleta Rivera, Tri C. Nguyen, Jiarong Zhou, Yingjun Luo, Jie Gao, Ronnie H. Fang, Zhen Bouman Chen, Liangfang Zhang, Sheng Zhong Naturally occurring cell surface-display of genome encoded RNAs and their impacts on cell-cell interactions. The dissertation author was the primary investigator and author of this paper.

VITA

2013	B. S. in Bioengineering *specialization in Biotechnology*, University of California, San Diego
2020	Ph. D. in Bioengineering, University of California, San Diego

PUBLICATIONS

Norman Huang, Zhangming Yan, Weixin Wu, Weizhong Chen, Yiqun Jiang, Jingyao Chen,Xuerui Huang, Xingzhao Wen, Jie Xu, Qiushi Jin, Kang Zhang, Zhen Chen, Shu Chien,and Sheng Zhong. Genome-wide colocalization of RNA DNA interactions and fusion RNA pairs.*Proceedings of the National Academy of Sciences of the United States of America*,116(8):33283337, 2019.

Weizhong Chen, Zhangming Yan, Simin Li, Norman Huang, Xuerui Huang, Jin Zhang,and Sheng Zhong. RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome.*iScience*, 4:204215, 2018.

T. C. Nguyen, X. Cao, P. Yu, S. Xiao, J. Lu, F. H. Biase, B. Sridhar, N. Huang, K. Zhang, and S. Zhong. Mapping rna-rna interactome and rna structure in vivo by mario. *Nature Communications*, 7:12023, 2016.

Rupak Doshi, Beverly R. Chen, Cecile Rose T. Vibat, Norman Huang, Chang Wook Lee,and Geoffrey Chang. In vitro nanobody discovery for integral membrane protein targets.*Scientific Reports*, 4(October), 2014.

ABSTRACT OF THE DISSERTATION

Understanding diverse RNA function using quantum dot based fluorescence in situ hybridization

by

Norman Huang

Doctor of Philosophy in Bioengineering

University of California San Diego, 2020

Professor Sheng Zhong, Chair

Ribonucleic acids (RNAs) serve as one of the main building blocks of organic life and with the advent of new understanding beyond an intermediary for protein coding, understanding it's function becomes imperative. Previous tools study RNA expression from cell populations such as tissue or cell lines, however, an expansive area of research has developed for the individual cells. The recognition of cell heterogeneity in normal and diseased tissues further highlighted the importance of accurate RNA detection in single cells. Single molecule RNA fluorescence in situ hybridization (smRNA-FISH) has revolutionized RNA detection and quantification by visualizing single RNA transcripts.

The work in this dissertation improves upon smRNA-FISH by developing quantum dot based fluorescent in situ technologies that uses hybridization of quantum dot-labeled single stranded DNA oligonucleotides to the RNA targets for visualization and quantification. In chapter 2, we demonstrate our proof-of-concept quantum dot single molecule RNA-FISH (qdot-smRNA-FISH) by comparing the sensitivity and specificity of qdot-smRNA-FISH against commercialized smRNA-FISH. In chapter 3, we describe qdot-smRNA-FISH for detection of isoforms (isoFISH), where different from smFISH that requires dozens of hybridization probes per gene, isoFISH requires only two hybridization probes per splicing variant. In chapter 4, we used the co-localization of multi-color quantum dot FISH (CoQ-FISH) to observe for the first time a fusion transcript formation of Eml4-Alk in non-small cell lung cancer biopsy samples from trans-splicing and not chromosomal rearrangement. In chapter 5, we develop cell surface qDot RNA-FISH (Surface-FISH), a method for visualization of RNAs on the cell surface. We identified RNA targets using RNA-seq on membrane coated nanoparticles from EL4 and applied Surface-FISH to observe regions of Malat1 and Neat1. In chapter 6, we developed Surface-PromFISH that uses randomized antisense probes to promiscuously label as many csRNAs as possible and made it compatible with co-staining of cell membrane proteins. We subjected the hybridization targets to RNA sequencing (Surface-FISHseq), which led to 51 putative PBMC csRNAs and we identified a potential function where antisense blocking of FNDC3B and CTSS lowers monocyte attachment.

Chapter 1

Introduction

Many of the scientific discoveries of a given gene are obtained by studying its transcription. The information obtained from this biological phenomenon known as gene transcription are used to infer important parameters related to gene functions, from the structure of the coding sequence, all the way to the potential interactions at the protein level. While most of the information of gene expression has been collected from cell populations, for example, cell culture, or tissue samples, an expansive area or research has directed the attention for the individual cells. Moreover, the recognition of cell heterogeneity in normal and diseased tissues further highlighted the importance of accurate RNA detection and quantification in single cells. An example would be if a given cell population contains two populations, one with high expression of a gene and one with low expression. In a bulk analysis, this would lead to an erroneous interpretation of average expression for that given transcript. This makes the detection of transcript expression an essential tools to characterize cells. Additionally, With the commercialization of single cell RNA sequencing and further developments in PCR technology, these collective methodologies have produced a wealth of information about relative changes in the abundance of mRNAs in various cells. However, most of the data represents expression of transcripts in a pool of cells that has undergone amplification and some normalization to a constitutive expressed housekeeping RNA or a control population,

which constrains these assays since cannot identify changes in transcript localization nor single cell quantity, which may play important roles in understanding cellular function. Single cell RNA fluorescence in situ hybridization (smRNA-FISH), since its inception in 2008, revolutionized RNA detection and quantification by detecting every single RNA molecule of a gene. This tool is widely used in disease diagnosis, pharmacogenomics, and drug development. However, this technology requires multiple probes or amplification probes to provide a sensitive and specific signal. We developed a technology that overcomes this drawback. This quantum dot RNA-FISH technology uses hybridization of quantum dot-labeled single stranded DNA oligonucleotides to the RNA targets for microscopic visualization and quantification. We found that contrary to a common belief, quantum dots can be used to quantify single molecules, when they are subjected to a specific experimental design. In chapter 2, we describe a quantum dot based single molecule RNA fluorescence in situ hybridization (qFISH) method that is comparable in sensitivity and specificity to commercialized single molecule FISH, but uses 10x less oligonucleotide probes allowing for detection of smaller regions and lower cost.

Most of multi-exonic human genes are alternatively spliced. Abnormal splicing variants contribute to explain a number of genetic disorders and are implicated in tumorigenesis. However, it remains challenging to detect and quantify splicing variants in single cells. In chapter 3, we describe a single-molecule fluorescence in situ hybridization (smFISH) based technique called Iso-FISH to visualize the transcripts of a splicing variant. Different from the classical smFISH that requires dozens of hybridization probes per gene, Iso-FISH requires only two hybridization probes per splicing variant. This reduction of probe number was enabled by utilizing quantum dots as the fluorophore and is essential for differentiating the multiple splicing variants of the same gene. In a test case, Iso-FISH detected both splicing variants of FoxP1 (FoxP1-18, FoxP1-18b) in human embryonic stem (ES) cells and only one variant (FoxP1-18) in HEK cells, confirming the ES cell-specific expression of FoxP1-18b. Iso-FISH offers a simple alternative to single-cell qPCR and single-cell RNA-seq, saving the trouble of single-cell isolation and nucleic acid amplification.

Iso-FISH will be instrumental to expand our understanding of cellular heterogeneity of alternative splicing.

Fusion transcripts are used as biomarkers in companion diagnoses. Although more than 15,000 fusion RNAs have been identified from diverse cancer types, few common features have been reported. In chapter 4, we compared CoQ-FISH with DNA-FISH, RNA-seq and iMARGI. We were able to validate a tumor tissue sample with EML4-ALK fusion transcripts but without apparent translocation on the genomic level suggesting a novel RNA-poise model, where spatial proximity of RNA and DNA could poise for the creation of fusion transcripts. Among the top 10 most significant RNA-DNA interactions found with iMARGI in normal cells, 5 co-localized with the gene pairs that formed fusion RNAs in cancer. Furthermore, throughout the genome, the frequency of a gene pair to exhibit RNA-DNA interactions is positively correlated with the probability of this gene pair to present documented fusion transcripts in cancer. To test whether RNA-DNA interactions in normal cells are predictive of fusion RNAs, we analyzed these in a validation cohort of 96 lung cancer samples using RNA-seq. 37 out of 42 fusion transcripts in the validation cohort were found to exhibit RNA-DNA interactions in normal cells. Finally, by combining RNA-seq, single-molecule RNA FISH, and DNA FISH, we detected a cancer sample with EML4-ALK fusion RNA without forming the EML4-ALK fusion gene. Collectively, these data suggest a novel RNA-poise model, where spatial proximity of RNA and DNA could poise for the creation of fusion transcripts.

Cells use surface molecules for self-representation, intercellular recognition and interactions. The discoveries of cell-surface receptors, glycans, and functional lipids made fundamental impacts to biology. However, less is known about the potential presence of host-genome transcribed RNAs on the cell surface. Here, we developed a series of technologies to test for any RNAs that are stably attached to the outside surface of cell membranes and exposed to the extracellular space, hereafter called cell surface RNAs (csRNAs). In chapter 5, we explore the concept that RNAs may be present on the cell surface. We developed Surface-seq to selectively

sequence the RNAs that are attached on the outside cell surface, which led to the prioritization of 167 csRNAs from cultured lymphoblasts. We developed Surface-FISH for RNA fluorescence in situ hybridization on the outside surface of intact cells, which led to validation of 2 of the Surface-seq prioritized RNAs. In chapter 6, to test for any cell-type specificity of csRNA presenting (csRNA+) cells, we developed Surface-PromFISH that uses randomized antisense probes to promiscuously label as many csRNAs as possible and made it compatible with co-staining of cell membrane proteins. Among human peripheral blood mononuclear cells (PBMCs), csRNA+ cells were enriched with CD14 presenting (CD14+) cells based on both single-cell imaging and flow cytometer analyses, suggesting csRNAs are not uniformly presented in all cell types. We subjected the csRNA+ cells and csRNA absent (csRNA-) PBMCs to single-cell RNA sequencing. Monocytes, as characterized by LYZ and CD14 transcripts, were enriched in csRNA+ cells, consistent with co-staining of the CD14 protein and csRNAs on cell surface. We subjected the hybridization targets of Surface-PromFISH from cell membranes to RNA sequencing (Surface-FISHseq), which led to 51 putative PBMC csRNAs. We prioritized 9 of them based on co-expression with monocyte markers in the single-cell RNA-seq data for functional validation. We tested whether any csRNA affects monocytes capability to attach to endothelial cells (M-EC attachment) by designing antisense probes against each csRNA. Antisense probes of FNDC3B and CTSS csRNAs reduced M-EC attachment levels. Moreover, the probes targeting the specific regions that were covered by Surface-FISHseq reads within FNDC3B and CTSS mRNAs reduced M-EC attachments whereas the probes targeting other regions of the same mRNAs did not. Taken together, these data suggest cell-type specific presentation of csRNAs and the exposed regions of these csRNAs impact cell-cell interactions.

Chapter 2

Development, validation and optimization of single molecule RNA detection using quantum dot-based oligonucleotides

2.1 Introduction

There is a rapidly increasing interest in understanding single cell gene expression as RNA expression can serve as indicators for cell behavior and potential disease markers[19]. Most traditional methods for gene expression analysis are tested on average cell populations which adds to potential biases due to the assumptions of homogeneity and the requirement of amplification to detect less expressed genes[20]. These studies have previously been able to uncover many insights into RNA regulations, but recent studies performed to detect whether there are single cell heterogeneity in cell populations have shown that used average expression could cause a loss in expression information from subsets of potentially different cell populations, such as when there may be two different populations of cells but one expresses a certain transcript lowly while the second expresses the same transcript highly, the average becomes insignificant[102,

153, 39, 137, 5, 15, 81, 19, 32, 46, 78, 103]. This transcript variability is a significant issue in both research and clinical diagnostics since misidentifying different cell populations could cause a drastically different response to perturbation of cell function or disease response. To our knowledge, there is only one microscopy-based technique that can reliably count RNA copy numbers in single cells, RNA-FISH and later single molecule RNA fluorescence in situ hybridization (smRNA-FISH)[102, 32]. Fluorescence in situ hybridization (FISH) uses short single stranded DNA sequences known as oligonucleotides coupled with modifiable fluorophores as probes to visualize the presence or absence of complementary sequences in cells or tissues under a fluorescent microscope. FISH offers the advantages of high sensitivity and specificity in recognizing localized signals of targeted DNA sequences or RNA transcripts at the single-cell level. smRNA-FISH revolutionized RNA detection and quantification by detecting single RNA transcripts in single cells. Many new RNA-FISH based technology have been developed to sensitively detect RNA expression in single cells such as Cas-9 mediated FISH (CASFISH) using fluorophore bound sgRNAs and dCAS9, oligopaint-FISH using fluorophore bound primary oligonucleotides for targeted SNP detection and secondary oligonucleotides to enhance labeling efficiency and sensitivity, single molecule RNA FISH by rolling cycling amplification (RCA) using padlock probes targeting to reverse transcribed cDNA with different alleles followed by ligation, cycling amplification and specific fluorophore-couple probe hybridization, sequential barcoding of multiplex different mRNAs by repeat rounds of hybridization, imaging, and stripping and others for sub-cellular detection of RNA transcripts[54, 25, 28, 6, 67, 72, 83, 82, 50]. One drawback is that the current methods of smRNA-FISH still requires the use of oligonucleotide pools or more commonly 30-40 oligonucleotides to achieve a good localized signal to background that is capable of sensitively and specifically distinguish true RNA molecules. Under conventional instrumentation such as fluorescence microscopes, this technology cannot work with fewer probes, due to the difficulty to separate signal from noise.

One way to lower the number of required probes is to increase the signal from each given

fluorophore, in this case we used fluorescent quantum dots conjugated oligonucleotides to create a better signal to noise. Considerations were made in the delivery and conjugation method for the quantum dot to the ssDNA, which may influence the overall size of the molecule as well as other effects such as aggregation and steric hindrance. For delivery into the cell, the size of QDs is also like that of traditional fluorophores such as GFP of about 2 nm 10 nm. The 40 base long oligonucleotides contain a C30 carbon spacer between the 5 amine functionality and the DNA sequence to minimize the steric hindrance from the QDs during hybridization. Various methods of conjugations are available including thiol-containing ssDNA can be directly attached to QD surface by ligand exchange, positively charged protein, such as streptavidin, can be electrostatically absorbed onto QDs with a negatively charged surface, followed by conjugation with biotin-functionalized ssDNA, water-soluble silanized QDs containing functional groups can be conjugated with DNA via bifunctional linkers, QDs capped with mercapoacetic acid ligands can be conjugated with primary-amine modified ssDNA, or carboxyl coated QDs conjugated with ssDNA modified on the 5 end with a primary amine[93, 43, 88, 41, 98, 3, 145, 152]. Here carboxyl and biotin QDs were selected for their size, conjugation stability, and controllable conjugation based on the salt and pH conditions[127]. In this chapter, we validate the detection of RNA transcripts using quantum dot fluorescence in situ hybridization (qFISH).

2.2 Results

2.2.1 Optimization of QD conjugation to oligos

Here we first optimized the conjugation of oligonucleotides to quantum dots to ensure we will maximize the detected signal of each hybridized pair. Oligonucleotides were designed 30-40 bases long and synthesized to contain one biotin attached to the 5 end (Tab. 2.1). As QDs were coated with streptavidin (Invitrogen), labeling was achieved at room temperature for 30 minutes at ratio of 0.5M of oligonucleotides per 1M of QDs (Fig. 2.1). FISH experiments were conducted

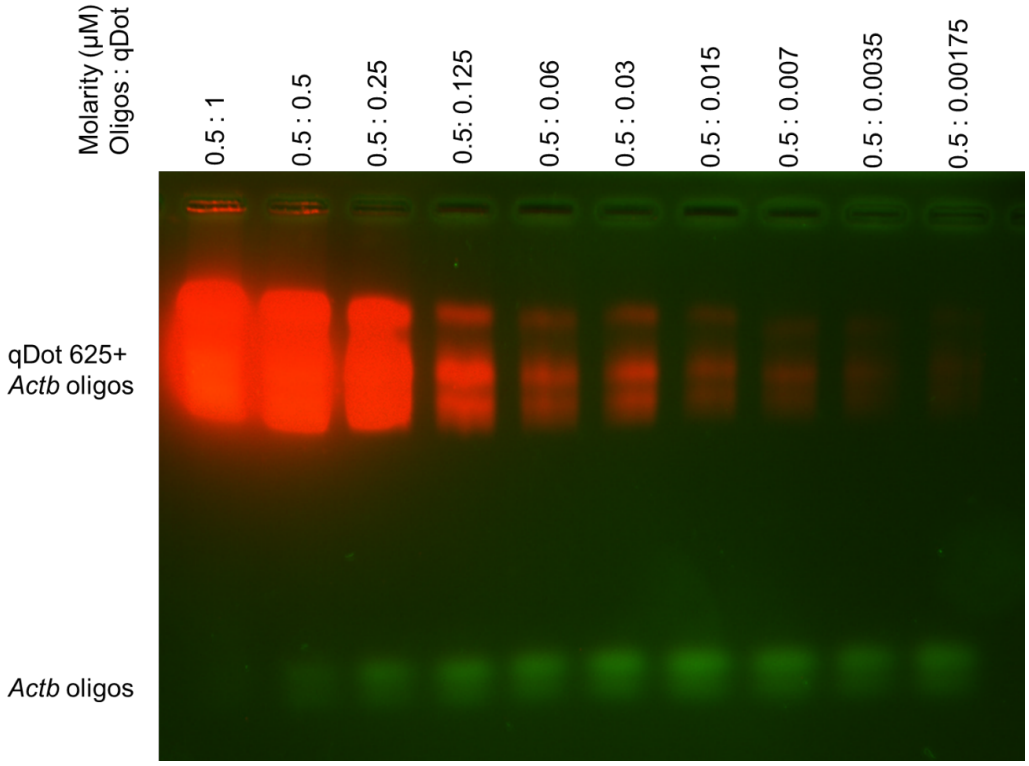

Figure 2.1: Titration of quantum dots and oligonucleotides for testing the optimal mixture for probe labeling.

using a modified version of an established smRNA-FISH protocol[120]. Following imaging, raw image stacks were processed in ImageJ[14] first by applying the Laplacian of Gaussian filter[111], second by counting the 3D rendered signal spots[13] at incremental intervals of 500 fluorescence arbitrary units (FAU). The number of spots in each image stack was obtained at the threshold interval of three consecutive equal counts (Fig. 2.2). We observed that selecting fluorescence intensities beyond this plateau one would have reduced the number of spots identified in the order of one unit, with the potential increase in the chance of not identifying a real spot. This criteria allowed us to quantify spots reliably irrespective of the image background.

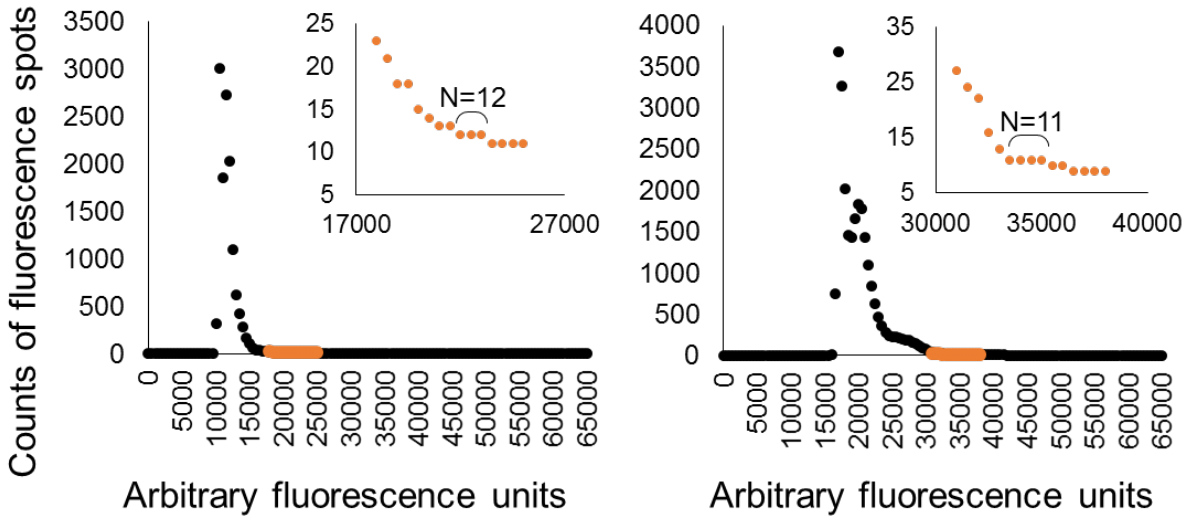

Figure 2.2: Count of spots at progressive fluorescence thresholds. Inset images are the counts around the first plateau of three consecutive counts.

2.2.2 Detection specificity between organic dye smRNA-FISH and qFISH

We explored the specificity of a 40 probe system designed with Alexafluor 555-bound oligonucleotides compared to 5 quantum dot 565 with biotin-streptavidin bound oligonucleotide probes targeting Actb transcripts. To show that the same transcripts are targeted we performed hybridization on the same mouse embryonic stem cells with a mixture of Alexafluor and QD bound oligonucleotides to observe co-localization of the signal through different fluorescent filter sets.

Initially, we tested the ability of qFISH to quantify RNA molecules using five oligonucleotides coupled with QDs. Using a modified version of smRNA-FISH protocol11 we were able to image spots representative of Actb RNA molecules under the microscope. Following our image quantification processing, we identified 119 (mean standard deviation) mRNA molecules per cell imaged (N=82) (Fig. 2.3b). To access the accuracy of probing an RNA molecule with our qdot-smRNA-FISH, we designed 43 probes and labeled them with Alexa555 to compare the co-localization of spots with different set of dyes and probes targeting the same RNA (Table 2.1,

Fig. 2.3c). Co-localization could not be resulted by cross imaging of dyes between filters as the dyes were excited at different wavelengths (Fig. 2.4). Using five probes labeled with QDs, we were able to identify almost all the same spots that 43 probes labeled with Alexa555 identified (Fig. 2.3d). Thus, qdot-smRNA-FISH performed with five probes achieved accuracy equivalent to smRNA-FISH with 43 probes labeled with organic dyes.

Table 2.1: Table denotes the quantum dot conjugated oligonucelotide sequences used to detect single ActB transcripts in mESC cells compared to Stellaris smRNA-FISH.

Gene symbol	Accession	Attachment	Oligonucleotide 5- 3	CG %
Actb	NM_007393.3	5 Alexa555	tgcaaagaagctgtgctcgc	55
Actb	NM_007393.3	5 Alexa555	tgtggaccggcaacgaagga	60
Actb	NM_007393.3	5 Alexa555	atatcgtcatccatggcgaa	45
Actb	NM_007393.3	5 Alexa555	acgatggaggggaatacagc	55
Actb	NM_007393.3	5 Alexa555	cacataggagtccttctgac	50
Actb	NM_007393.3	5 Alexa555	gtacttcagggtcaggatac	50
Actb	NM_007393.3	5 Alexa555	gttggtaacaatgccatgtt	40
Actb	NM_007393.3	5 Alexa555	acacgcagctcattgtagaa	45
Actb	NM_007393.3	5 Alexa555	tgatctgggtcatcttttca	40
Actb	NM_007393.3	5 Alexa555	ggggtgttgaaggtctcaaa	50
Actb	NM_007393.3	5 Alexa555	ctggatggctacgtacatgg	55
Actb	NM_007393.3	5 Alexa555	agaggcatacagggacagca	55
Actb	NM_007393.3	5 Alexa555	catcacaatgcctgtggtac	50

Table 2.1 Quantum dot ActB probe sequences continued from previous page.

Gene symbol	Accession	Attachment	Oligonucleotide 5- 3	CG %
Actb	NM_007393.3	5 Alexa555	tcgtagatgggcacagtgtg	55
Actb	NM_007393.3	5 Alexa555	atggcgtgagggagagcata	55
Actb	NM_007393.3	5 Alexa555	atcttcatgaggtagtctgt	40
Actb	NM_007393.3	5 Alexa555	ctgtggtggtgaagctgtag	55
Actb	NM_007393.3	5 Alexa555	ttgatgtcacgcacgatttc	45
Actb	NM_007393.3	5 Alexa555	atctcctgctcgaagtctag	50
Actb	NM_007393.3	5 Alexa555	tagtttcatggatgccacag	45
Actb	NM_007393.3	5 Alexa555	ggtctttacggatgtcaacg	50
Actb	NM_007393.3	5 Alexa555	gacagcactgtgttggcata	50
Actb	NM_007393.3	5 Alexa555	tgggtacatggtggtaccac	55
Actb	NM_007393.3	5 Alexa555	agagcagtaatctccttctg	45
Actb	NM_007393.3	5 Alexa555	gatcttcatggtgctaggag	50
Actb	NM_007393.3	5 Alexa555	gctcaggaggagcaatgatc	55
Actb	NM_007393.3	5 Alexa555	ccgatccacacagagtactt	50
Actb	NM_007393.3	5 Alexa555	acagtgaggccaggatggag	60
Actb	NM_007393.3	5 Alexa555	gatccacatctgctggaagg	55
Actb	NM_007393.3	5 Alexa555	actcatcgtactcctgcttg	50
Actb	NM_007393.3	5 Alexa555	tagaagcacttgcggtgcac	55
Actb	NM_007393.3	5 Alexa555	aacgcagctcagtaacagtc	50

Table 2.1 Quantum dot ActB probe sequences continued from previous page.

Gene symbol	Accession	Attachment	Oligonucleotide 5- 3	CG %
Actb	NM_007393.3	5 Alexa555	ggttttgtcaaagaaagggt	40
Actb	NM_007393.3	5 Alexa555	ttcaccgttccagtttttaa	35
Actb	NM_007393.3	5 Alexa555	atgtttgctccaaccaactg	45
Actb	NM_007393.3	5 Alexa555	tcagccacatttgtagaact	40
Actb	NM_007393.3	5 Alexa555	cctgtaaccacttatttcat	35
Actb	NM_007393.3	5 Alexa555	cttttgggagggtgagggac	60
Actb	NM_007393.3	5 Alexa555	cacagaagcaatgctgtcac	50
Actb	NM_007393.3	5 Alexa555	aaaaagggaggcctcagacc	55
Actb	NM_007393.3	5 Alexa555	gaccaaagccttcatacatc	45
Actb	NM_007393.3	5 Alexa555	ttggtctcaagtcagtgtac	45
Actb	NM_007393.3	5 Alexa555	gtgtaaggtaaggtgtgcac	50
Actb	NM_007393.3	5 qDot 565	tctcaaacatgatctgggtcatcttttcac	40
Actb	NM_007393.3	5 qDot 565	gttggcatagaggtctttacggatgtcaac	46.7
Actb	NM_007393.3	5 qDot 565	caatgatcttgatcttcatggtgctaggag	43.3
Actb	NM_007393.3	5 qDot 565	ttttgtcaaagaaagggtgtaaaacgcagc	40
Actb	NM_007393.3	5 qDot 565	cgttccagtttttaaatcctgagtcaaaag	36.7

2.2.3 Detection sensitivity between smRNA-FISH and qFISH

Experimentally we observed the high fluorescence yields allowing 3-5 QD-DNA probes to be visualized using conventional microscopy. More specifically, we show in a proof-of-concept experiment that QDDNA probes with different emission spectra used in a one-step hybridization experiment to visualize in-situ single gene fusion transcripts with similar sensitivity compared to smRNA-FISH. The requirement of 40 20mer probes was primarily due to the necessity for specificity towards the desired target by increasing the signal above any nonspecific background. Here we were able to demonstrate the same sensitivity and specificity to target transcripts using down to two quantum dot conjugated oligonucleotides. We first compared the sensitivity for detecting the housekeeping gene Gapdh using the commercialized Stellaris smRNA-FISH which uses 30-40 20mer probes against our qFISH technology using only two 40mer probes. Here we found that both technologies detected similar numbers of transcripts on average per cell, around 275.

Table 2.2: Table denotes the quantum dot conjugated oligonucelotide sequences used to detect single Gapdh transcripts in HEK293T cells compared to Stellaris smRNA-FISH.

Sequence Name	Sequence
GAPDH-1	C ATT GCT GAT GAT CTT GAG GCT GTT GTC ATA CTT C
GAPDH-2	G ATG TCA TCA TAT TTG GCA GGT TTT TCT AGA CGG C

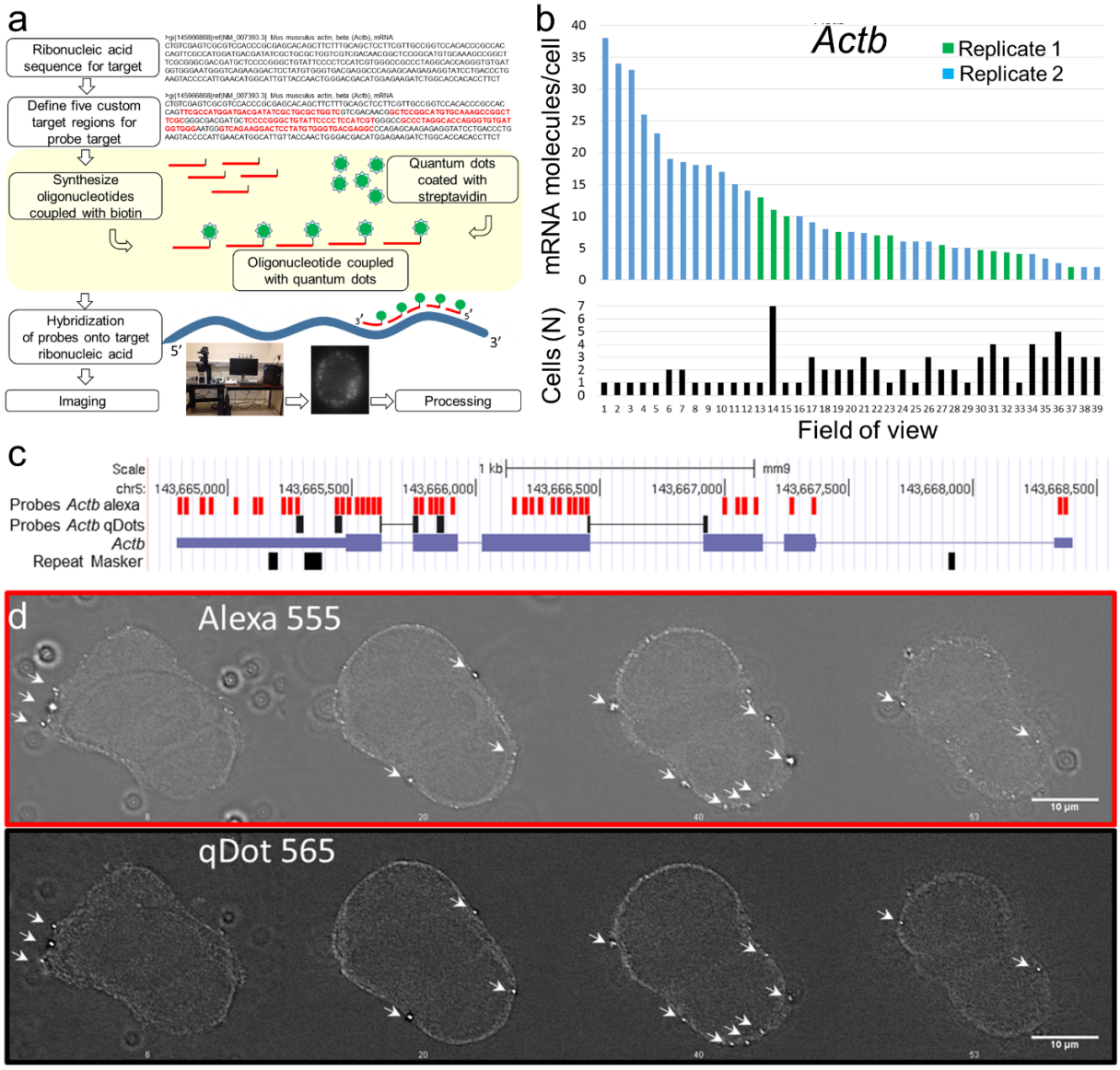

Figure 2.3: Comparing specificity of smRNA-FISH with qFISH in mESCs by sequential hybridization. (A) Genomic positions for smRNA-FISH probes labeled with organic dyes (Alexa555) and qFISH for the Actb. (B) Comparison between smRNA-FISH and qFISH detection of Actb transcripts using mESCs cells. Arrows indicate visualized co-localization of qFISH and smRNA-FISH signals.

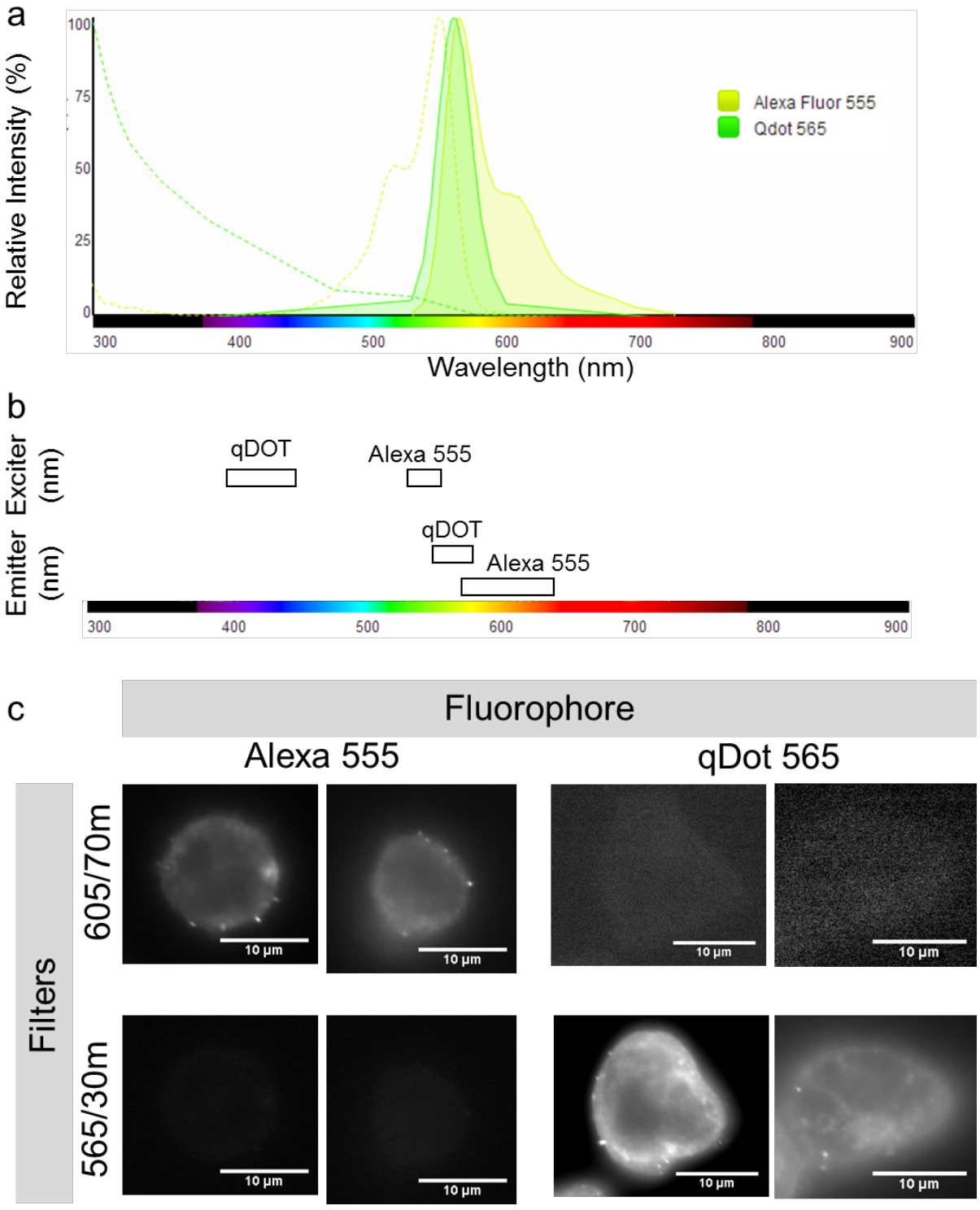

Figure 2.4: Distinctions of Alexa 555 and QD 565 excitation and emission. (a,b) The excitation wave lengths of QDs and Alexa 555 were distinct. (c) RNA-FISH signals of Alexa 555 and QD 565 acquired with corresponding and exchanged emission filters

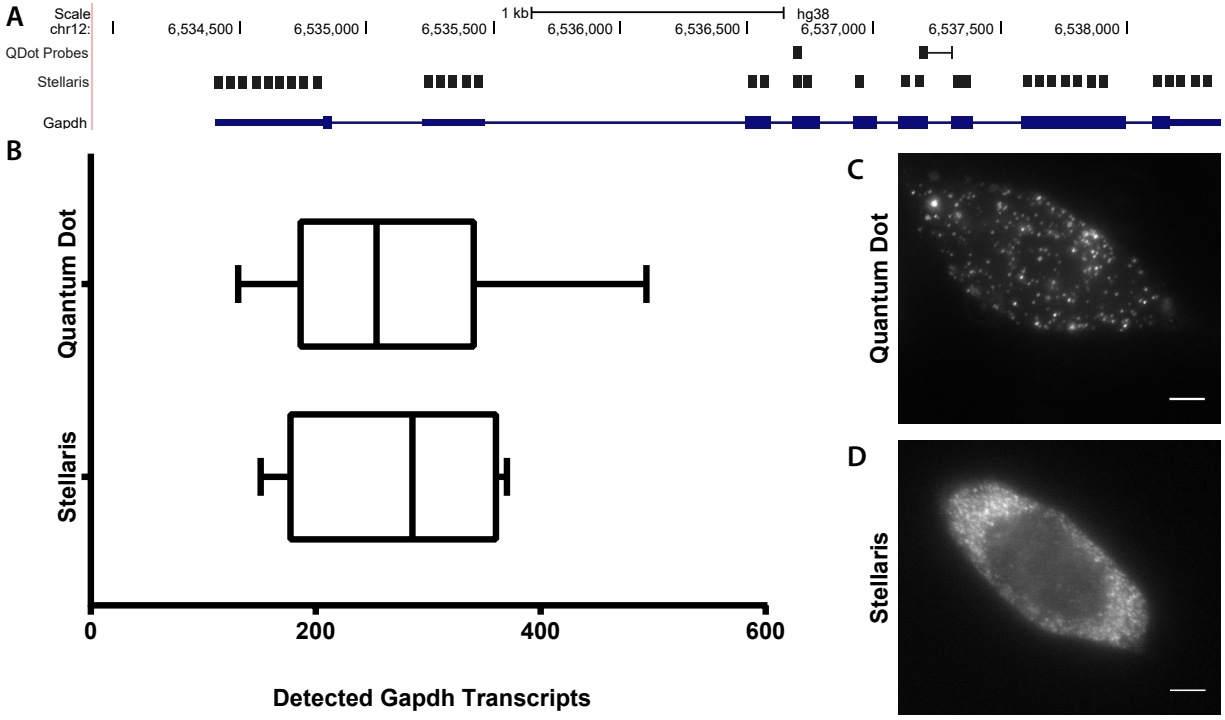

Figure 2.5: Detection of Gapdh transcripts using smRNA-FISH compared with qFISH. (A) Genomic positions for smRNA-FISH probes labeled with organic dyes (Stellaris) and probes labeled with QDs for the same gene (Gapdh). (B) Comparison between Stellaris smRNA-FISH and qFISH detection of Gapdh transcripts using HEK293 cells. (D) Plot of cell counts comparing organic (quasar 570) to inorganic (quantum dot) dyes against Gapdh. Counts were conducted on 10 cells for Stellaris with an average of 273 Gapdh transcripts and 24 cells for qFISH with an average of 275 transcripts detected. Error bars represent min to max range.

2.3 Discussion

Although very important, studying the gene expression of single cells brings along technical challenges. Currently, only one technique can obtain information of gene expression of a single cell without the introduction of technical biases. This technique is the fluorescence in situ hybridization, and a recent technological improvement has allowed researchers to image a single molecule of ribonucleic acid of a gene of interest using a microscope 2,3,4. Even though this advancement was a great step towards the study of gene expression, it suffers from low sensitivity, high cost, and low throughput (often one gene at a time).

We developed a technology to detect single RNA molecules in cells and thus digitally quantify the RNA transcripts of any gene. This technology uses hybridization of quantum dot-labeled single stranded DNA oligonucleotides to the RNA targets for microscopic visualization and quantification. To our knowledge, this is the first time that quantum dots are used in quantifying single molecules. Previously, quantum dots were thought not applicable to quantification of single molecules due a well-known blinking problem. The design of our technology resolved this problem. Our technology reduces the costs of the competing technology that uses organic dyes and many folds more oligonucleotides for single molecule RNA-FISH to less than 30%, while offering clearer, easier to detect, and more reliable signals. Moreover, our technology allows for simultaneous detection of the RNAs of multiple genes in the same cells using conventional instrumentation (without super resolution imaging equipment, $500,000).

2.4 Methods

2.4.1 Cell Culture and Fixation

The human embryonic kidney cell line HEK293T was maintained in Dulbeccos Modified Eagles Medium (DMEM) (ATCC 30-2002) with the following added components 10% Fetal

Bovine Serum (heat inactivated) (ATCC 30-2020), 2mM L-glutamine (ATCC 30-2214), 1% Penicillin/Streptomycin. ES cells were seeded on glass-bottom micro-chamber (1.5, Lab Teck) previously coated with poly-D-lysine (5M, Sigma) and laminin (0.01mgl1, Sigma). Cultures were kept at 37C in 5% CO2 and maintained at 80% confluence on 10cm TC-treated plates. To prepare for fixation and imaging, 10,000 HEK293T cells were seeded on glass (No. 1.5) bottom micro-chamber (Lab Tek II) previously coated with poly-d-lysine (5 uM, Sigma) and laminin (0.01 mg/ul, Sigma). Following incubation for 2 hours, cells were washed in nuclease free PBS. Fixation was perform44ed with incubating in 4% formaldehyde, 1x PBS for 10 minutes. 10% acetic acid may be added to help permeabilization of nuclear membrane. Cells are then permeabilized and stored with 70% EtOH at 4C overnight.

2.4.2 Oligonucleotide sequence design

Oligonucleotides were synthesized The oligonucleotides for hybridization were designed using Primer3 and manually adjusted based on the sequences available from Ensembl Genome Browser. The oligonucleotides contained a C30 carbon spacer between the 5 amine functionality and the DNA sequence to minimize the steric hindrance from the QDs during their hybridization to the target mRNA. For the Alexafluor comparison, oligonucleotides were synthesized with a biotin attached to their 5-end (IDT). Additionally, oligonucleotides are 40 bases long to increase specificity to the target. Sequences were checked using NCBI Blast for possible cross-hybridization targets.

2.4.3 Preparation of QDDNA Complexes

1 uL of carboxyl coated QDs (qDot 525, 605, or 705 nm; .08 uM) were reacted with 15 uL of 5 amine-modified DNA oligonucleotides (100 uM) in the presence of 8 uL EDC (50 mM), 8 uL NHS (50 mM) in 200 uL borate buffer (50 mM; pH 8.0). The reaction mixture was

mixed by rotating in respective lo-bind micro centrifuge tube overnight at room temperature. The QDDNA conjugate was filtered through a .22 um membrane to remove larger aggregated at 2000 rcf for 15 minutes and was then purified using the membrane filter (100 kDa molecular weight cut-off, Milipore, USA) to remove the excess EDC and unbound DNA at 2500 rcf. SILANE filtration step may be used at this point to remove any excess unbound QDs. A second filtration is performed through a .22 um membrane to remove any remaining aggregates. For the biotin labeled probes, labelling was achieved by incubation of oligonucleotides and dyes (Alexa 555 or QDs, Invitrogen) coupled with streptavidin at room temperature for 30min at a ratio of 0.5M of oligonucleotides per 1M of dye.

2.4.4 In Situ Hybridization Procedure

Prior fixed HEK293T cells on Lab Tek II glass chamber are rehydrated with 50% Formimide, 2x SSC. Filtered QD-DNA complexes are added sequentially to hybridization solution containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNAse-free BSA, 40 ug E.coli tRNA, 2x SSC, 50% formamide, 30 ng of probe. Cells are incubated overnight in a covered water bath at 37 C. QD-DNA complexes not hybridized are removed via two washes (SSC 2x, formamide 50%) at 37C for 30 minutes each wash on rocker. Cells are images using 1x PBS.

2.4.5 Fluorescence microscopy

Wide field fluorescence imaging was conducted in an Olympus IX83 inverted microscope equipped with appropriate cubes for the dyes used and 60x oil immersion objective (NA=1.4, Olympus). Images were captured with ORCA-R2 CCD camera (Hamamatsu) at intervals of 0.2um on the z-axis.

2.4.6 Image Processing

Raw images were processed with with ImageJ by first generating a PSF and applying deconvolution using the Richardson-Lucy method34. Furthermore, the Laplacian of Gaussian filter35 is applied. 3D object counter was used for the counting the 3D rendered spots.

Chapter 3

Detection of single molecule RNA Isoform transcripts

3.1 Introduction

Alternative splicing (AS) constitutes a key regulatory layer of gene expression [62, 141]. More than 88% of human protein coding genes can be alternatively spliced. In normal conditions, AS produces cell type-specific RNA isoforms, resulting in protein isoforms suitable for cell-type specific activities [113, 36]. For example, an AS isoform of the FoxP1 transcription factor, FoxP1-18b, regulate pluripotency by promoting the expression of pluripotency related transcription factors and suppressing differentiation related genes [37]. However, abnormal AS events are causes of a number of human genetic disorders [131]. Furthermore, modulating the relative abundances of the splicing variants of a disease associated gene has demonstrated promises for treatment ([7]).

Splicing variant detection assays of the bulk of cells [133, 134, 74, 115, 23], single-cell RNA-seq based methods [122, 4], single-cell in situ methods include SpliceRCA [104] and plasmonic nanoparticle [76] at single cell level poses further challenges. Most methods for single

cell analysis require amplification of RNA before measurement which adds to the technical biases [124]. To our knowledge, there is only one microscopy-based technique that can reliably count RNA copy numbers in single cells [101], smRNA-FISH. Single cell RNA fluorescence in situ hybridization (smRNA-FISH) revolutionized RNA detection and quantification by detecting single RNA transcripts in single cells. However, this technology is incapable of assaying alternatively spliced RNAs. The current standard for detection of single RNA molecules is smRNA-FISH, which requires the use of 40 20mer oligonucleotides to achieve a good localized signal to background that is capable of sensitively and specifically distinguish true RNA molecules. It would be difficult to distinguish positive localized signals from background even with an ideal case of fitting seven 20mer probes that may fit within average exon, which has an average length of 145 nt [71], using smRNA-FISH. Under conventional instrumentation, this technology cannot work with fewer probes, due to the difficulty to separate signal from noise.

To achieve this, it was necessary to have a fluorophore with high fluorescent yields at low quantities. Quantum dots (QDs) are known to have high fluorescence quantum yields and optical stability against photo bleaching [105]. We decided to test the use of QD for single molecule imaging, however, we also recognize that although QDs were used to image protein molecules [121, 154], it is commonly accepted that QDs are not applicable to count single molecules. This is due to the well-known blinking problem, that the signals are intermittent [90, 107]. If QDs were used to label single molecules, at any time of imaging, a random subset of QDs would be invisible, making a random subset of single molecules undetectable. We tested single QD-DNA hybridized probes and found that blinking events did not hinder the sensitivity through introduction of chemical components and multiple acquisitions. Next it was needed to ensure the specificity of the detected isoform to minimize the risk of false positives. In smRNA-FISH specificity was ensured using an aggregation of fluorophores to the target RNA and the brightest signals had the highest chance of being a true positive. Similarly, we explored the use of an aggregation of QD-oligonucleotides, but found that using five or three QD-oligonucleotide conjugates were too

space-consuming for alternatively spliced isoforms with smaller exons. Instead, the isoform of interest is labelled at the two exon junctions spanning the exon of interest where the flanking region anneals to the exon sequences directly adjacent on the spliced transcript.

Considerations were made in the delivery and conjugation method for the QD to the ssDNA, which may influence the overall size of the molecule as well as other effects such as aggregation and steric hindrance. For delivery into the cell, the size of QDs is also similar to that of traditional fluorophores such as GFP of about 2 nm 10 nm. The 40 nt long oligonucleotides contains a C30 carbon spacer between the 5 amine functionality and the DNA sequence to minimize the steric hindrance from the QDs during hybridization. Furthermore, various methods of conjugations are available including thiol-containing ssDNA can be directly attached to QD surface by ligand exchange, positively charged protein, such as streptavidin, can be electrostatically absorbed onto QDs with a negatively charged surface, followed by conjugation with biotin-functionalized ssDNA, water-soluble silanized QDs containing functional groups can be conjugated with DNA via bifunctional linkers, QDs capped with mercapoacetic acid ligands can be conjugated with primary-amine modified ssDNA, or carboxyl coated QDs conjugated with ssDNA modified on the 5 end with a primary amine [2, 40, 42, 89, 92, 97, 144, 151]. Here carboxyl QDs were selected for their smaller size and controllable conjugation based on the salt and pH conditions [126].

Experimentally we observed the high fluorescence yields allowing single QDs to be visualized using conventional microscopy. More specifically, we show in a proof-of-concept experiment that two QDDNA probes with different emission spectra used in a one-step hybridization experiment to visualize in-situ the alternatively spliced isoform of FoxP1 (Iso-FISH). One isoform of FoxP1 (NM_001244810) was found to only be expressed in pluripotent cell lines, such as iPSC and ES cells, while this specific isoform is not expressed in differentiated cell lines, such as HEK293T [37]. There are few validated isoforms discovered that offer a clear distinction among cell types, which makes FoxP1 an ideal standard. Additionally, we independently verified

the expression using qPCR and sequencing data.

3.2 Results

3.2.1 Detection of FoxP1 isoforms using Iso-FISH

Development of Iso-FISH

We designed multiple QD-labeled probes targeting a particular isoform RNA of interest (Figure. 3.1). We used oligonucleotides spanning the spliced exons on the isoform of interest, FoxP1. Each oligonucleotide was designed to be 40 bases and synthesized to contain one amino group attached to the 5 end (Table. 3.1). A spacer is also included before the amino group to account for possible steric hinderance. As QDs were coated with carboxyl groups, labeling was achieved at room temperature for 30 minutes at ratio of 1.5 nmol of oligonucleotides per 50 pmol of QDs (Fig. S1). We optimized a hybridization and imaging protocol for Iso-FISH by testing a number of variations of reagents and parameters from the hybridization protocol of standard RNA-FISH.

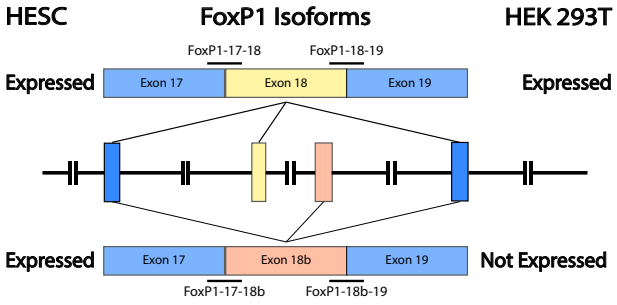

Figure 3.1: Primer design targeting exon junctions on FoxP1. Two oligonucleotide probes were designed targeted each isoform FoxP1-17-18 targets the exon junction of FoxP1 transcript exon 17 and 18, FoxP1-18-19 targets the exon junction of exon 18 and 19. FoxP1-17-18b targets the exon junction of FoxP1 transcript 17 and 18b, FoxP1-18b-19 targets the exon junction of 18b and 19. HESCs were shown to express both isoforms of FoxP1 while HEK293T were shown to only express FoxP1 isoform with exon 18.

Table 3.1: Primer sets for Q-RTPCR and oligonucleotide sequences for Iso-FISH for FoxP1 Exon 18 and 18b Detection in HEK and HESC.

Primer Name	Application	Sequence
FoxP1-18F	qPCR	GAAGGCCACAAAAGATCAGTG
FoxP1-18R	qPCR	TGGAAGCGGTAGTGTATAGAGG
FoxP1-18bF	qPCR	GAAGAGTTTAAACGTGGCCG
FoxP1-18bR	qPCR	CATTGAGAGGTGTGCAGTAGG
FoxP1-17-18	Iso-FISH	CTAAGATTATGACGCACTGCATTCTTCCACGTGGCCGCGT
FoxP1-18-19	Iso-FISH	AATAAGGGAAGGGTTACCACTGATCTTTTGTGGCCTTCGT
FoxP1-17-18b	Iso-FISH	TATGCAGTGAGAGGTTGGTGCGAATGGCACCCTTCCACGT
FoxP1-18b-19	Iso-FISH	TGTTTTTAATAAGGGAAGGGTTATGTGCGACAAGCTCGTC

Detection of FoxP1 isoforms with Iso-FISH

To ensure the validity of using FoxP1 as a standard for isoform expression, we validated the expression of NM_001244810 along with a widely expressed, canonical form of FOXP1 (NM_032682) in HEK293T using qPCR and in both HEK293T and human embryonic stem cell (HESC) lines through sequencing. In conducting qPCR between HEK293T and HESC, there are two considerations (1) biological differences between samples and (2) technical variations between each sample. For this experiment biological differences could be considered negligible since all the cDNA came from the same plate of HEK cells. Technical variation is also minimal each reaction was processed as one sample until the addition of primers. The Ct value is defined as a threshold line that intersects with the amplification curve and is influenced by the components of the master mix, the passive reference dye, and the efficiency of the PCR. All replicates were made in the same master mix and used the same reference dye, thus assuming similar efficiencies, the Ct values are directly comparable.

For qPCR, FoxP1-18F and FoxP1-18R were used to detect exon 18 in NM_032682 and FoxP1-18bF and 18bR to detect exon 18b in NM_001244810. We were able to observe the expected level of expression in HEK293T, where NM_032682 is highly expressed and HESC, where both NM_032682 and NM_001244810 were expressed. We further expand this using Iso-FISH to detect different isoforms of Foxp1. In Iso-FISH, the oligonucleotides were designed

to span the two exon junctions to ensure detection of the specific isoform transcript. Each pair of quantum dot labeled oligonucleotides, FoxP1-17-18 FoxP1-18-19 and FoxP1-17-18b FoxP1-18b-19 were separately labeled for each cell line. Positive transcripts were detected with the union of the detected fluorescent signals from each oligonucleotide probe. We were able to observe the similar expression profiles between Iso-FISH and qPCR in each cell line.

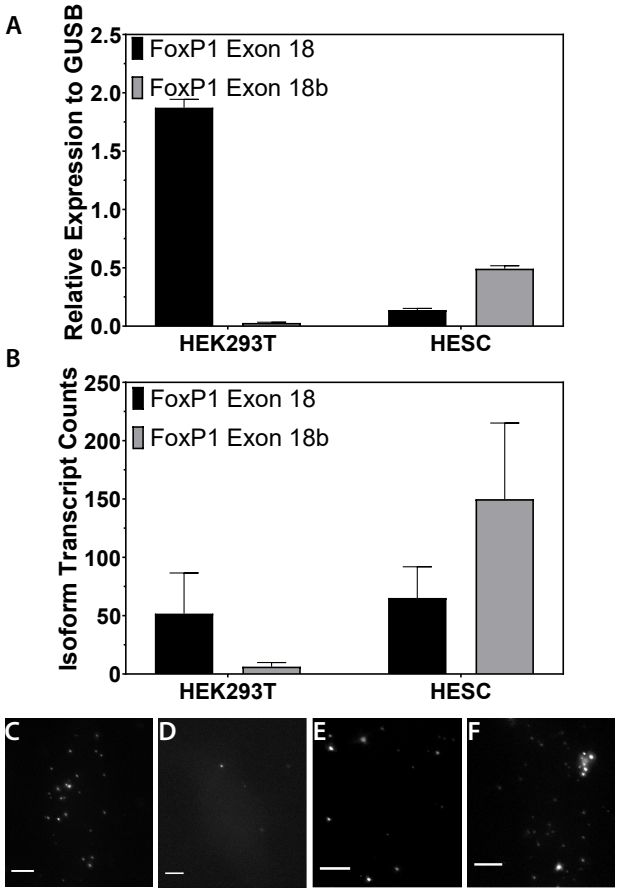

Figure 3.2: Detection of alternatively spliced isoform FoxP1 in human embryonic cell line (HESC) and differentiated cell line HEK293T. (A) Quantitative reverse transcription PCR on probes detecting isoform with exon 18 or 18b in each respective cell line compared to house-keeping gene GUSB confirming expression of 18b isoform in HESC. (B) qDot FISH detection of spliced isoforms in both HEK293T and HESC, showing quantitative single cell expression of isoform 18b in HESC and much lower expression in HEK293T. (C) Detection of FoxP1 exon 18 In HEK293T (D) Detection of FoxP1 exon 18b In HEK293T (E) Detection of FoxP1 exon 18 in HESC (F) Detection of FoxP1 exon 18b In HESC. Scale bar denotes 5 um.

3.3 Discussion

We aimed to exploit the brightness of a single QD to detect single oligonucleotides targeting RNA isoform transcripts. The time interval of for a QD to stay at the on or the off state can range from milliseconds to seconds [115, 133]. A QD typically spends more than half the time at the on state [31, 34, 148]. If necessary, Iso-FISH samples can be imaged multiple times. Combined with experimental data we verified that QDs can be applied to counting single molecules. One major advantage of Iso-FISH lies in its use of less oligonucleotide probes. There is a financial incentive of cost reduction from approximately $370 to ¿$1 per gene. More importantly, it enables targeting the RNA isoform molecules with 200 or less bases. Such RNA molecules are too short to be assayed by traditional smRNA-FISH. This enables single cell and single molecule analysis of RNA splicing, which has never been done before. Another advantage of Iso-FISH is the applicability to assaying the alternatively spliced RNA products using multiple colors in parallel. This has been a difficult task for two reasons. First, while dyes and fluorescent proteins have been the mainstay of fluorescence imaging for decades, their fluorescence is unstable under high photon fluxes necessary to observe individual molecules, yielding only a few seconds of observation before complete loss of signal [121]. The other dyes necessary for targeting different genes could have photo bleached during scanning the first dye (gene). Nanocrystals are more photo stable than organic dyes [77, 105], allowing for repeated scans of the same experimental sample. Second, organic dyes can contaminate each other, especially when 3 or more different dyes are used together. In contrast, QDs have a profile of fluorescence spectra, all of them have a narrow Gaussian distribution of the fluorescence intensity [105]. This feature forms the basis of multiplexing several QDs in one assay. Moreover, the spectral position of emission can be tuned according to the particle size [105]. As the sizes of the particles used for one specific wave length are very similar, the width of the emission curve is also narrow. For commercial QDs, much of the intensity is emitted within a window of

100 nm around the target wave length. This narrow emission of fluorescence of specific QDs allied to the appropriate choice of filters favors the use of QDs in experiments that probe multiple targets with different dyes in the same cell. Experiments have shown dual probing of cells with unambiguous distinction of signal between QDs targeting two different wavelengths [55, 144]. Further multiplexing is possible with QDs targeting three [48] or four [44] wavelengths.

3.4 Methods

3.4.1 Cell Culture and Fixation

The human embryonic kidney cell line HEK293T was maintained in Dulbeccos Modified Eagles Medium (DMEM) (ATCC 30-2002) with the following added components 10% Fetal Bovine Serum (heat inactivated) (ATCC 30-2020), 2mM L-glutamine (ATCC 30-2214), 1% Penicillin/Streptomycin. Cultures were kept at 37C in 5% CO2 and maintained at 80% confluence on 10cm TC-treated plates. To prepare for fixation and imaging, 10,000 HEK293T cells were seeded on glass (No. 1.5) bottom micro-chamber (Lab Tek II) previously coated with poly-d-lysine (5 uM, Sigma) and laminin (0.01 mg/ul, Sigma). Following incubation for 2 hours, cells were washed in nuclease free PBS. Fixation was performed with incubating in 4% formaldehyde, 1x PBS for 10 minutes. 10% acetic acid may be added to help permeabilization of nuclear membrane. Cells are then permeabilized and stored with 70% EtOH at 4C overnight.

3.4.2 Reagents

All reagents were of analytical grade. Qdot Streptavidin Conjugates 525, 605, and 705 nm were purchased from Life Tech.

3.4.3 Oligonucleotide sequence design

The oligonucleotides for hybridization were designed using Primer3 and manually adjusted based on the sequences available from Ensembl Genome Browser. The oligonucleotides contained a C30 carbon spacer between the 5 amine functionality and the DNA sequence to minimize the steric hindrance from the QDs during their hybridization to the target mRNA. Additionally, oligonucleotides are 40 bases long to increase specificity to the target FoxP1 isoform. Sequences were checked using NCBI Blast for possible cross-hybridization targets.

3.4.4 Preparation of QDDNA Complexes

1 uL of carboxyl coated QDs (qDot 525, 605, or 705 nm; .08 uM) were reacted with 15 uL of 5 amine-modified DNA oligonucleotides (100 uM) in the presence of 8 uL EDC (50 mM), 8 uL NHS (50 mM) in 200 uL borate buffer (50 mM; pH 8.0). The reaction mixture was mixed by rotating in respective lo-bind micro centrifuge tube overnight at room temperature. The QDDNA conjugate was filtered through a .22 um membrane to remove larger aggregated at 2000 rcf for 15 minutes and was then purified using the membrane filter (100 kDa molecular weight cut-off, Milipore, USA) to remove the excess EDC and unbound DNA at 2500 rcf. SILANE filtration step may be used at this point to remove any excess unbound QDs. A second filtration is performed through a .22 um membrane to remove any remaining aggregates.

3.4.5 In Situ Hybridization Procedure

Prior fixed HEK293T cells on Lab Tek II glass chamber are rehydrated with 50% Formimide, 2x SSC. Filtered QD-DNA complexes are added sequentially to hybridization solution containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNAse-free BSA, 40 ug E.coli tRNA, 2x SSC, 50% formamide, 30 ng of probe. Cells are incubated overnight in a covered water bath at 37 C. QD-DNA complexes not hybridized are removed via two washes

(SSC 2x, formamide 50%) at 37C for 30 minutes each wash on rocker. Cells are images using 1x PBS.

3.4.6 Fluorescence microscopy

Wide field fluorescence imaging was conducted in an Olympus IX83 inverted microscope equipped with appropriate cubes for the dyes used and 60x oil immersion objective (NA=1.4, Olympus). Images were captured with ORCA-R2 CCD camera (Hamamatsu) at intervals of 0.2um on the z-axis.

3.4.7 Image Processing

Raw images were processed with with ImageJ [117] by first generating a PSF [65] and applying deconvolution using the Richardson-Lucy method [138]. Furthermore, the Laplacian of Gaussian filter [110] is applied. 3D object counter was used for the counting the 3D rendered spots [12].

3.5 Acknowledgements

Chapter 3, in part is currently being prepared for submission for publication of the material. Huang, Norman, Zhong, Sheng. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Detection of gene fusion transcripts

4.1 Introduction

Fusion transcripts are associated with diverse cancer types and have been proposed as diagnostic biomarkers [38, 91, 66]. Companion tests and targeted therapies have been developed to identify and treat fusion-gene defined cancer subtypes [91, 26]. Efforts towards identification of fusion transcripts in cancer have primarily relied on analyses of RNA sequencing (RNA-seq) data [38, 66, 125, 149, 132, 27]. A recent study analyzed 9,966 RNA-seq datasets across 33 cancer types from The Cancer Genome Atlas (TCGA) and identified more than 15,000 fusion transcripts [26]. Lung cancer remains one of the most common forms of cancer worldwide with more than 80% being non-small cell lung cancer (NSCLC)[114]. Two primary clinical tests are used to identify NSCLC subtypes through the detection of EGFR mutation and oncogenic gene fusions such as EML4-ALK[109]. Since the discovery of EML4-ALK, there has been a multitude of targeted therapies such as ALK inhibitors alectinib, ceritinib, and crizotinib, which has shown great success in treatment. However, these methods, such as real-time qPCR[106, 142], gene expression microarrays[116, 75], and RNA sequencing[24], at single cell level poses further challenges. Most methods for single cell analysis require amplification of RNA before

measurement which adds to the technical biases[20]. The difficulty of resolving single transcripts of gene fusion transcripts stems from the necessity to sensitively and specifically detect the proximity of fused exons of fusion transcripts. In this chapter we aim to demonstrate the ability to detect gene fusion transcripts in both cell line and cancer tissue biopsies using co-localizaton quantum dot fluorescence in situ hybridization (CoQ-FISH). Here we compared CoQ-FISH with DNA-FISH, RNA-seq and iMARGI[143]. Additionally, despite the large number of gene pairs in identified fusion transcripts, it remains formidable to predict what unreported pair of genes may form a new fusion transcript. Recent analyses could not identify any distinct feature of fusion RNA forming gene pairs [69]. Here, we report a characteristic pattern of the two-dimensional distribution of the genomic locations of the gene pairs involving RNA-DNA interactions that provides novel insights to understand the creation of fusion transcripts. We were able to validate a tumor tissue sample with EML4-ALK fusion transcripts but without apparent translocation or apparent fusion on the genomic level suggesting a novel RNA-poise model, where spatial proximity of RNA and DNA could poise for the creation of fusion transcript, wherein the RNA of gene 1 by interacting with the genomic sequence of gene 2 is poised for being spliced into gene 2's nascent transcript and thus creating a fusion transcript.

4.2 Results

It remains formidable to predict what unreported RNA pairs can form new fusion transcripts. By systematic mapping of chromatin-associated RNAs and their respective genomic interaction loci, we obtained genome-wide RNA-DNA interaction maps from two non-cancerous cell types. The gene pairs involved in RNA-DNA interactions in these normal cells exhibited strong overlap with those with cancer-derived fusion transcripts. These data suggest an RNA-poise model, where the spatial proximity of one genes transcripts and the other genes genomic sequence poises for the creation of fusion transcripts. We validated this model with 96 additional

lung cancer samples. One of these additional samples exhibited fusion transcripts without a corresponding fusion gene, suggesting that genome-recombination is not a required step of the RNA-poise model.

4.2.1 Validation of quantum dot based CoQ-FISH by detection Eml4-Alk in H2228 NSCLC cell line

Here we demonstrate the ability of quantum dot based CoQ-FISH to detect Eml4-Alk fusion in a known positive cell line H2228. Lung cancer cell line H2228 has been previously characterized to contain the Eml4-Alk fusion and is used as a positive control for CoQFISH detection. This is compared to Eml4-Alk fusion detection in HEK293T cell line which is not known to contain the Eml4-Alk fusion serving as a negative control sample. We first validated the expression of EML4-ALK fusion using northern blot analysis and qPCR (Fig. 4.1). We then compared this to CoQ-FISH (Fig. 4.2 to find positive expression in all three detection methods.

4.2.2 Cancer-derived Futra pairs that co-localize with RNA-DNA interaction in normal cells do not form fusion transcripts in normal cells

A model that may explain the co-localization of RNA-DNA interactions and Futra pairs is that RNA-DNA interactions in the normal cells poise for creation of fusion transcripts. Recognizing that this model cannot be tested by perturbation due to the very small likelihood for a fusion transcript to occur in a cancer sample, we carried out two other tests. First, we tested whether the cancer-derived Futra pairs were detectable in normal cells. We re-analyzed the merged RNA-seq datasets of more than 75 million 2×100 bp paired-end read pairs from HEK293T cells [146] and ran STAR-Fusion [47] on these datasets, which reported a total of 8 Futra pairs. None of the previously derived 15,144 Futra pairs from TCGA RNA-seq data were detected in HEK293T cells. In addition, we specifically tested for EML4-ALK fusion transcripts, which were reported

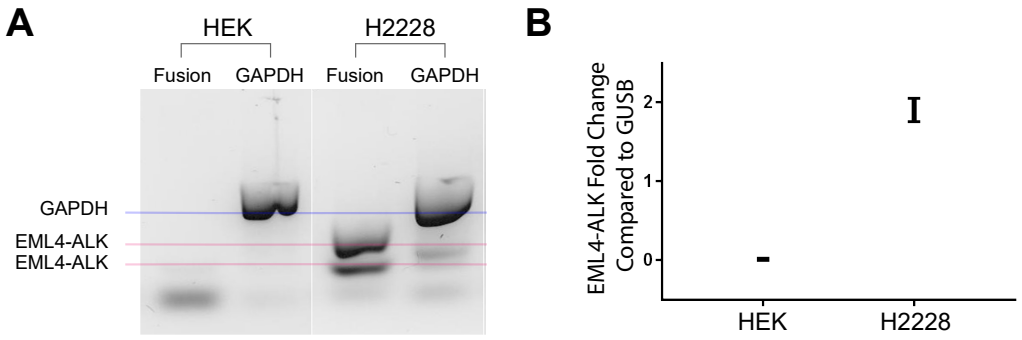

Figure 4.1: Detection of fusion transcripts in H2228 cells. (A) Northern blot analysis of GAPDH (control) and EML4-ALK fusion transcripts in HEK and a NSCLC cell line (H2228). A forward primer GCATAAAGATGTCATCATCAACCAAG against EML4 and a reverse primer TCTTGCCAGCAAAGCAGTAGTTGG against ALK were used in the Fusion lanes. A forward primer ACCACAGTCCATGCCATCAC and a reverse primer TCCACCACCCTGTTGCTGTA against GAPDH were used in the GAPDH Lanes. The two bands in the H2228 Fusion lane correspond to two known variants of the EML4-ALK fusion transcripts. (B) qPCR analysis of EML4-ALK fusion transcripts using house-keeping GUSB mRNA as the internal control. Fold change of EML4-ALK vs. GUSB (y axis) is plotted in HEK and H2228. Error bars: standard deviation derived from 4 replicates.

]

in non-small cell lung carcinoma (NSCLC) [136], and there were RNA-DNA interactions between EML4 RNA and ALK genomic locus in HEK and HFF cells (Fig. 4.4A). Neither PCR nor quantitative PCR analysis detected EML4-ALK fusion transcripts in HEK293T cells (*SI Appendix*, Fig. S8), whereas both assays detected fusion transcripts in a NSCLC cell line (H2228) (*SI Appendix*, Fig. S8). Taken together, these data suggest that, although cancer-derived Futra pairs co-localized with RNA-DNA interactions in normal cells, the fusion transcripts found in cancer are not present in the normal cells.

4.2.3 RNA-DNA interactions in normal cells are predictive of fusion transcripts in new cancer samples

Next, we tested whether the RNA-DNA interactions in normal cells are predictive of fusion transcript formation in cancer. To this end, we analyzed a validation cohort comprising 96

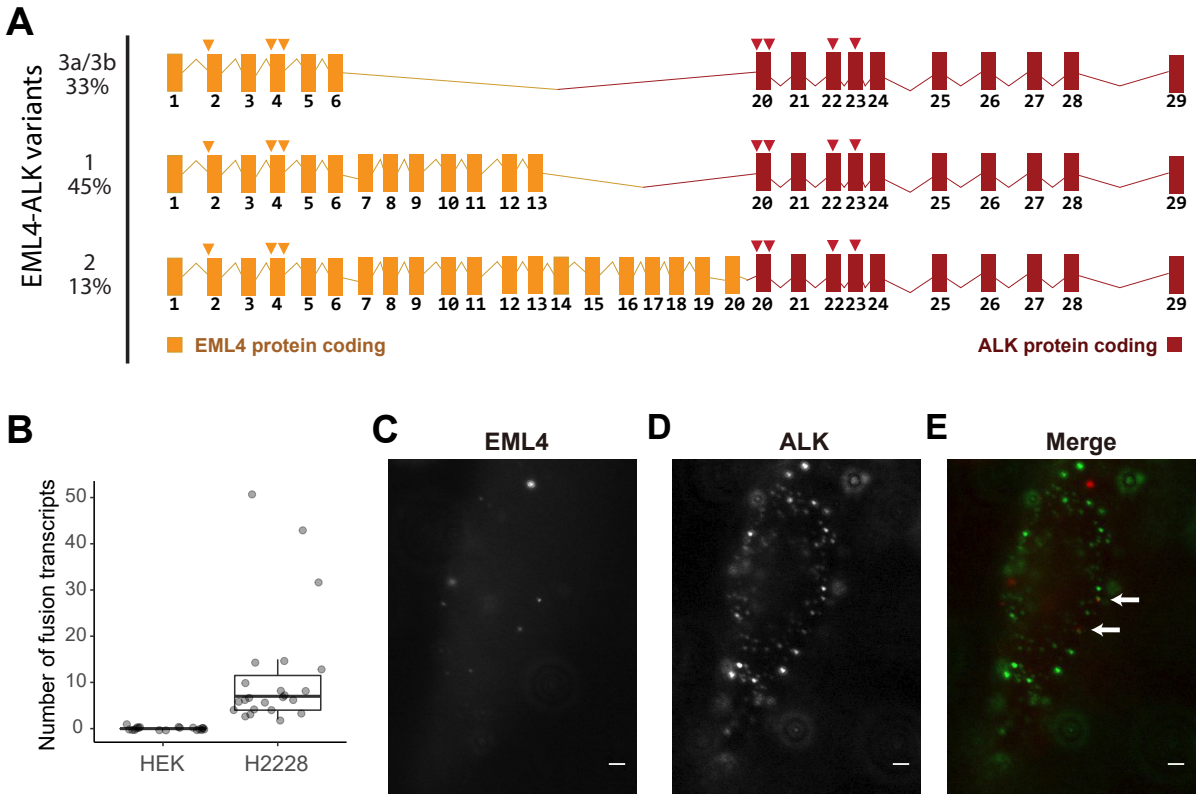

Figure 4.2: CoQ-FISH analysis in HEK and H2228 cells. (A) Schematic view of the top three most abundant variants of EML4-ALK fusion transcripts. Each row shows a variant with its included EML4 exons (yellow boxes) and ALK exons (red boxes). Arrows: targeting locations of FISH probes. Probes labeled with Qdots of 605 nm (yellow arrows) and 705 nm (red arrows) were complementary to EML4 and ALK RNA sequences, respectively. (B) Numbers of co-localized signals (y axis) in 19 single cells (dots) of the HEK and 22 single cells (dots) of H2228. (C-E) Representative images of the EML4 (C) and ALK (D) channels and their merged image (E). Arrows: co-localized signals. Scale bar: 2 μm.

new lung cancer samples from patients who were not part of the TCGA cohorts. We also analyzed a NSCLC cell line (H2228). RNA was extracted and targeted RNA sequencing (RNA-seq) was carried out with Illumina's TruSight RNA Pan-Cancer Panel. Of these 96 samples, 27 did not yield sufficient RNA for sequencing, whereas the other 69 samples produced a sequencing library and yielded on average 3.9 million uniquely aligned read pairs per sample (Fig. 4.3A). STAR-Fusion [47] was applied to this dataset and it reported a total of 42 fusion transcripts from these 69 samples (Fig. 4.3 B and C). These 42 fusion transcripts included EML4-ALK and FRS2-NUP107 fusion, which were also reported from the 9,966 TCGA cancer samples, as well as 40 new fusion transcripts that were not previously documented. The small amount of recurring Futra pairs between these additional cancer samples and TCGA samples is expected from the small fraction of recurring Futra pairs across the TCGA samples.

Among these 42 Futra pairs detected from the validation cohort, 37 (88.1%) co-localized with RNA-DNA interactions in the assayed normal cells, supporting the idea that RNA-DNA interactions in the already assayed normal cells are predictive of Futra pairs in cancer (odds ratio = 106.51, p-value ¡ 2.2×10^{-16}, Chi-squared test). We asked if only intra-chromosomal Futra pairs co-localized with RNA-DNA interactions. Nineteen out of the 42 (45%) detected Futra pairs were inter-chromosomal (Fig. 4.3C), comparable to the proportion (41%) of inter-chromosomal Futra pairs detected from TCGA samples. Eighty-three percent (19 out of 23) intra- and 95% (18 out of 19) inter-chromosomal Futra pairs overlapped with RNA-DNA interactions (Fig. 4.3 D and E), suggesting that the co-localization of RNA-DNA interactions and Futra pairs were not restricted to intra-chromosomal interactions. Taken together, the co-localization of Futra pairs and RNA-DNA interactions, the lack of cancer-derived fusion transcripts in normal cells, and the predictability to additional Futra pairs in new cancer samples support the model where RNA-DNA interactions in normal cells poise for creation of fusion transcripts in cancers. Hereafter, we refer to this model as the *RNA-poise model*. We call the gene pairs with RNA-DNA interactions in normal cells as *fusion susceptible pairs*.

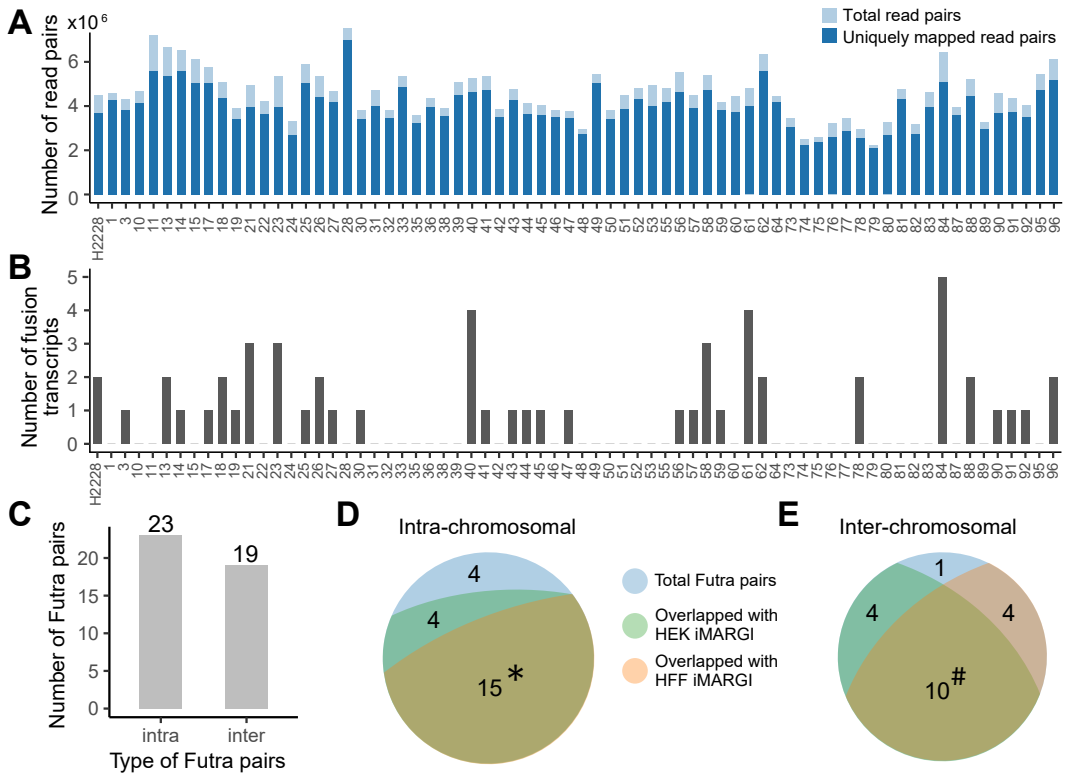

Figure 4.3: Fusion transcripts detected from the new lung cancer samples. (A) The number of RNA-seq read pairs (light blue bar) and the uniquely mapped read pairs (dark blue bar) of each sample (column). (B) Number of detected fusion transcripts (y axis) in each sample (columns). (C) Numbers of intra- and inter-chromosomal Futra pairs detected from the 69 cancer samples. (D-E) Intersections of these intra- (D) and inter-chromosomal (E) Futra pairs to RNA-DNA interactions in HEK (green) and HFF (pink). $*$: The 15 intra-chromosomal Futra pairs that overlap with RNA-DNA interactions in both HEK and HFF (yellow-green) are ALK:EML4, RP11-557H15.4:SGK1, FRS2:NUP107, ACTN4:ERCC2, LRCH1:RP11-29G8.3, CUX1:TRRAP, CEP70:GSK3B, NIPBL:WDR70, KCTD1:SS18, CHST11:NTN4, RP1-148H17.1:TOP1, LMO7:LRCH1, NAV3:RP1-34H18.1, LPP:PPM1L, and MTOR-AS1:RERE. #: The 10 inter-chromosomal Futra pairs that overlap with RNA-DNA interactions in both HEK and HFF (yellow-green) are LIN52:PI4KA, LPP:OSBPL6, FCGBP:MT-RNR2, KMT2B:MALAT1, ETV6:TTC3, KTN1:MALAT1, FLNA:MALAT1, COL1A2:MALAT1, COL1A1:MALAT1, and MALAT1:TNFRSF10B.

4.2.4 RNA-DNA interaction between EML4 and ALK correlates with an RNA fusion without fusion gene in tumor

We tested whether genome re-arrangement is a prerequisite step for the creation of fusion transcripts from fusion susceptible pairs by choosing EML4-ALK fusion transcripts for this test

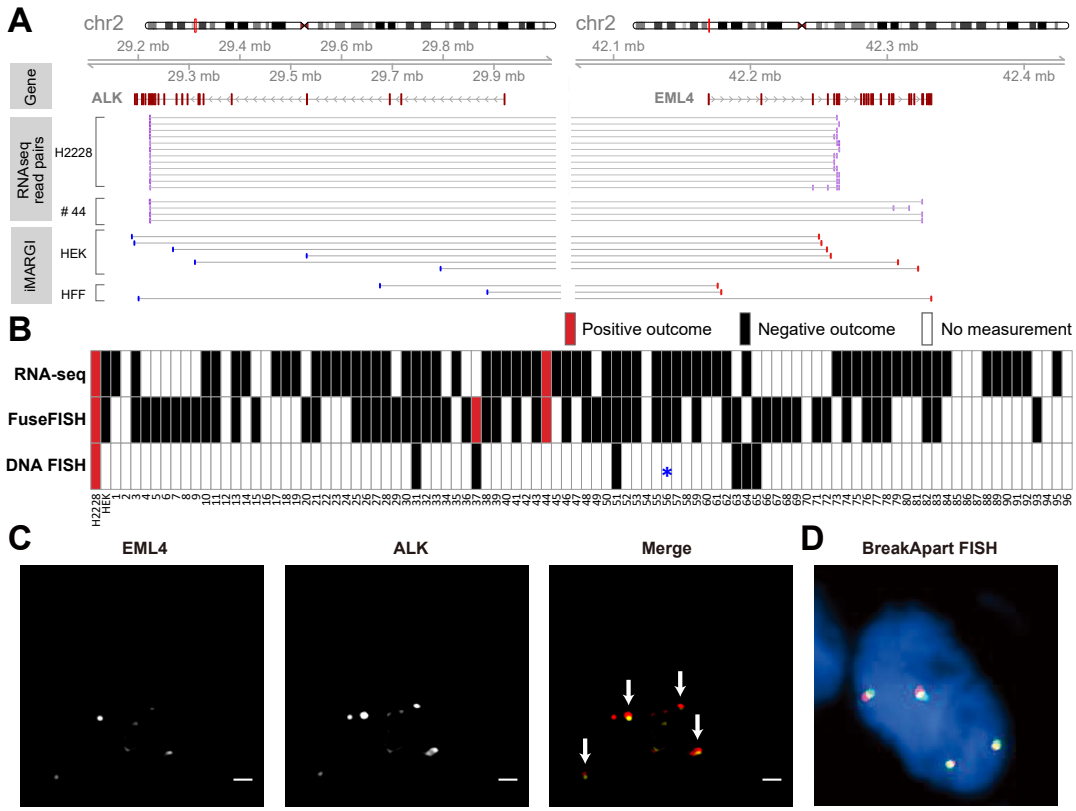

Figure 4.4: RNA fusion and DNA break point comparisons. (A) Middle tracks: RNA-seq read pairs (purple bars) aligned to ALK (left) and EML4 (right). Each pair of paired-end reads is linked by a horizontal line. Lower tracks: iMARGI read pairs aligned to the two genes. Red bars: RNA-end. Blue bars: DNA-end. Thin grey lines: pairing information of iMARGI read pairs. (B) Positive (red) and negative (black) outcomes of detection of EML4-ALK fusion transcripts based on RNA-seq and CoQ-FISH, and DNA FISH (ALK Break Apart FISH, a DNA re-arrangement test) in each cancer sample (column). White box: no measurement. $*$: Detected a partial deletion of ALK gene without re-arrangement. (C) CoQ-FISH images of the #37 cancer sample from the EML4 and ALK channels. Arrows: co-localized FISH signals indicating fusion RNA. Scale Bar: 2 μm. (D) A representative image of ALK Break Apart FISH, with co-localized red and yellow signals that indicate integral ALK gene without re-arrangement.

because EML4-ALK is a fusion susceptible pair (Fig. 4.4A), EML4-ALK fusion transcripts are detected in one of our new tumor samples (Sample #44) (Fig. 4.4B) and there is an FDA approved diagnosis kit (Vysis ALK Break Apart FISH) based on DNA FISH detection of the EML4-ALK fusion gene. Break Apart assays were performed by Knight Diagnostic Laboratories at the Oregon Health & Science University according to standardized protocols. We subjected

the remaining tissue from Sample #44 for DNA FISH analysis. None of our 8 attempts yielded any DNA FISH signal in the remaining tissue from either control or ALK probes. We therefore could not ascertain whether there was genome re-arrangement in the only sample with detectable EML4-ALK fusion transcripts.

In order to identify other cancer samples that express EML4-ALK fusion transcripts, we re-analyzed our collection of 96 lung cancer samples with CoQ-FISH, a single-molecule fluorescence in situ hybridization (sm-FISH) based method for the detection of fusion transcripts [118, 86]. We carried out quantum dot-labelled sm-FISH [95] by labelling EML4 and ALK transcripts with quantum dots at 705 nm and 605 nm, respectively (Fig. 4.2). The FISH probes were designed to hybridize to the consensus exons shared among all 28 variants of EML4-ALK fusion transcripts that have been identified to date [89]. Following prior literature [118, 86], fusion transcripts were detected by the co-localized sm-FISH signals targeting EML4 and ALK transcripts. In a positive control test, an average of 12 co-localized sm-FISH signals per cell were detected in a total of 22 H2228 cells (Fig. 4.2), that were known to express EML4-ALK fusion transcripts (Fig. 4.1) [87]. In contrast, HEK293T cells exhibited on average zero co-localized signals per cell from 19 cells (Fig. 4.2), consistent with the lack of such a fusion transcript in HEK293T cells (Fig. 4.1).

In our collection of 96 tumor samples, only 57 had remaining tissues for CoQ-FISH analysis. These 57 samples included 39 that yielded RNA-seq data and 18 that did not yield RNA-seq data (Fig. 4.4B). The CoQ-FISH analysis detected EML4-ALK fusion transcripts in two samples, including Sample #44 which was also analyzed by RNA-seq and Sample #37 which did not yield RNA-seq data (Fig. 4.4 B and C). To test whether Sample #37 had ALK- related fusion genes, we subjected it together with other 6 randomly selected samples (#18, #51, #56 #57, #63, #65) for DNA re-combination analysis using Vysis ALK Break Apart FISH. None of these seven samples exhibited ALK- related fusion genes. More specifically, one sample (#57) failed to generate DNA FISH signals from four attempts; one sample (#56, negative for EML4-ALK

fusion transcripts by RNA-seq and CoQ-FISH analyses) exhibited a partial deletion of the ALK gene, but no sign of ALK- related fusion genes (star, Fig. 4.4B). The other five samples, including Sample #37, exhibited integral ALK genes (Fig. 4.4 B and D). Taken together, the lung cancer Sample #37 expressed EML4-ALK fusion transcripts without having an EML4-ALK fusion gene. These data suggest that genome re-arrangement is not a prerequisite step for the creation of fusion transcripts from fusion susceptible pairs. In other words, the RNA-poise model does not require alterations of the DNA.

4.3 Discussion

4.3.1 Many genome rearrangement independent fusion transcripts

Our RNA FISH and DNA FISH analyses revealed a cancer sample that contained a fusion transcript without the corresponding fusion gene. Such an example, although not often seen in literature may not be a rare case [135]. The lack of reports are likely attributable to the research attentions paid to the other side of the coin, i.e., the fusion transcripts created by fusion genes [91]. Indeed, approximately 36 65% of fusion transcripts derived from cancer RNA-seq data were attributable to genome rearrangement[38]. However, this is likely an overestimate because when low quality whole genome sequencing (WGS) data were removed, only approximately 30 45% fusion transcripts had corresponding WGS reads (High Pass bars, Figure S1A of [38]). These published results are consistent with the notion that fusion genes do not account for all observed fusion transcripts and suggest the occurrence of fusion transcripts independent of genome re-arrangement. However, we recognize that to date, the sheer amount of validated fusion RNAs independent of genome re-arrangement remains limited, which warrants future investigation.

4.3.2 The RNA-poise model allows for splicing errors

Fusion transcripts can be created by two processes. The better recognized process is through transcription of a fusion gene, that was created by genome-rearrangement. The less recognized process is by RNA splicing errors, where two separate transcripts were spliced together (trans-splicing) [79]. Trans-splicing does not involve genome-rearrangement. A theoretical gap in the splicing error model is that trans-splicing can only happen to two RNA molecules that are close to each other in the 3-dimensional (3D) space; however, except for neighboring genes [135] the chances for two RNA molecules transcribed from distant chromosomal locations to meet in space are small. Therefore, it remains difficult to perceive a biophysical process in which fusion transcripts are created by splicing errors.

The RNA-poise model fills this theoretical gap. The pre-installation of Gene 1's transcripts on Gene 2's genomic sequence positions Gene 2's nascent transcripts spatially close to Gene 1's transcripts, allowing for the possibility of trans-splicing. Furthermore, the majority of splicing events are co-transcriptional. The availability of transcripts of Gene 1 during Gene 2's transcription allows for the opportunity of making co-transcriptional trans-splicing.

4.3.3 Breaking down the RNA-poise model by RNA-DNA interactions

Remote RNA-DNA interactions could be created by at least two means. First, the caRNA can target specific genomic sequences, which could be mediated by tethering molecules (RNA targeting, Fig. 4.5). Second, the spatial proximity of the genomic sequences in 3D space could bring the nascent transcripts of one gene to the genomic sequence of another gene (RNA confinement, Fig. 4.5). Both means of RNA-DNA interactions provide spatial proximity between two RNA molecules and thus allows for splicing errors. In addition, the spatial proximity of two genes in the RNA confinement model could enhance the chances of genome re-arrangement of the spatially close genomic sequences and thus creating fusion genes [97]. Thus, the RNA-poise

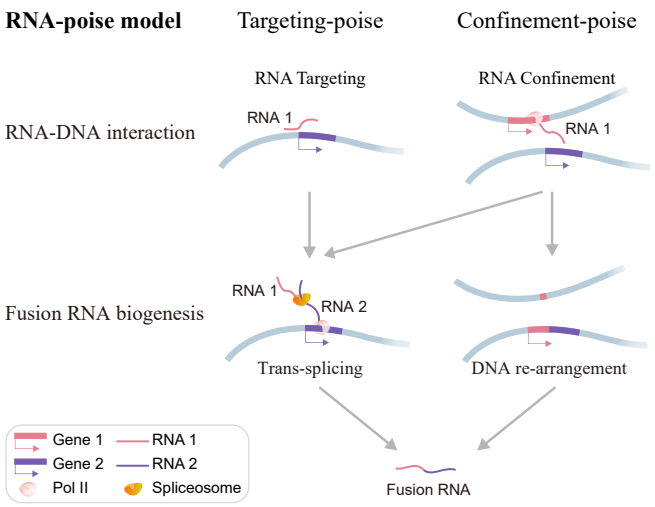

Figure 4.5: RNA-poise model. In this model, the transcripts of one gene (RNA 1, red bar) can exhibit spatial proximity to another gene (RNA 2, purple bar) due to tethering (RNA targeting) or spatial proximity of the two genes (RNA confinement). Both cases could enhance splicing errors (grey arrows), whereas the proximity of genomic sequences may also facilitate gene fusion (grey arrow on the right), which subsequently produces fusion RNA.

model can be regarded as a union of two sub-models depending on the process of RNA-DNA interaction. One sub-model (Targeting-poise, Fig. 4.5) could only create fusion transcripts by trans-splicing. The other sub-model (Confinement-poise, Fig. 4.5) could create fusion transcripts by either trans-splicing or creation of fusion genes.

4.4 Methods

4.4.1 Reference genome and gene annotations

Human genome assembly hg38/GRCh38 and Ensembl gene annotation release 84 were used throughout all data analyses.

4.4.2 Public RNA-seq data

HEK293T RNA-seq datasets were downloaded from NCBI BioProject database (Accession numbers: SRR2992206, SRR2992207, SRR2992208 under the project number PRJNA305831) [147]. The three datasets were merged.

4.4.3 TCGA derived fusion transcripts

The TCGA RNA-seq derived fusion transcripts were downloaded from Tumor Fusion Gene Data Portal (http://www.tumorfusions.org) [51]. Tier 1 and tier 2 fusion transcripts were used in our analyses. Genomic coordinates were converted to hg38 by liftOver. Following Davidson et al. [27], Futra pairs within 200 kb on hg38 were removed. The data of Futra pairs were mainly processed using R [99] with Bioconductor packages GenomicRanges [73] and InteractionSet [84].

4.4.4 Visualization of Futra pairs

Heatmaps of the count matrix were plotted using Bioconductor package ComplexHeatmap [45]. Genomic plots of Futra pairs were created with GIVE [16].

4.4.5 CoQ-FISH analysis

Probe design

Oligonucleotide probes were designed to hybridize to exons 2 - 6 of EML4 RNA and exons 20 - 23 of ALK RNA. These exons were chosen because they were present in all the observed variations of EML4-ALK fusion genes. These probes were 35 - 40 nt in lenghs, with similar GC contents and melting temperatures.

Conjugation of quantum dots to oligonucleotide probes

Oligos were modified on the 5' end with a primary amino group and a spacer of 30 carbons to minimize steric hindrance of probe-RNA hybridization. These probes were conjugated with quantum dots through the amino group using EDC reaction [22]. The probes were subsequently purified with 0.2 μm membrane filtration and 100,000 molecular weight cut-off (MWCO). The retentate of the 100,000 MWCO was subjected with dynabeads MyOne SILANE purification to remove any remaining unconjugated probes. A subsequent 0.2 μm membrane filtration was used to remove any final aggregates.

Hybridization of adherent cell lines

Probe hybridization in H2228 cells was carried out as previously described [118, 95]. Briefly, probes were added to the hybridization solution and incubated with the cells at 37 °C overnight. Cells were washed and resuspended in 1$\times$ PBS for imaging.

Hybridization of tissue samples

Probe hybridization in tissue were carried out as previously described [85]. Briefly, hybridization solution with probes were added to the surface of parafilm to form droplets. A tissue slice (5 - 10 μm in thickness) fixed on a glass cover slip was gently placed over the hybridization solution. The mixture was incubated at 37 °C overnight. The tissue was subsequently washed with wash buffer and resuspended in 1x PBS for imaging.

Imaging and analysis

Cells or tissues were imaged in 1$\times$ PBS through wide field fluorescence imaging using an Olympus IX83 inverted microscope at 60$\times$ oil immersion objective (NA=1.4). Image processing was carried out as previously described [102]. Briefly, single transcripts were detected using an automated thresholding algorithm that searches for robust thresholds. where counts do not change

within a range. Fusion transcripts were determined by searching for co-localization of detection transcripts by overlap between predicted centers within a radius.

4.4.6 RNA sequencing and analysis

RNA was extracted with Trizol from H2228 cell line and lung cancer tissue samples of the approximate size 3 mm$\times$3 mm$\times$30 μm per sample. RNA-sequencing was carried out using the TruSight RNA Pan-Cancer Panel (Illumina) following the manufacturer's protocol. All the RNA-seq data, including HEK293T public data, H2228 cell line and lung cancer sample sequencing data, were mapped to the human genome (hg38) using STAR (v2.5.4b) with default parameters [29]. Fusion transcripts were called using STAR-Fusion (v0.8.0) [47] requiring both numbers of supporting discordant read pair and junction spanning read larger than zero and the sum of them larger than 2.

4.5 Acknowledgements

Chapter 4, in part, is a reprint of the material as it appears in Genome-wide colocalization of RNA DNA interactions and fusion RNA pairs in Proceedings of the National Academy of Sciences of the United States of America 2018. Norman Huang, Zhangming Yan, Weixin Wu, Weizhong Chen, Yiqun Jiang, Jingyao Chen, Xuerui Huang, Xingzhao Wen, Jie Xu, Qiushi Jin, Kang Zhang, Zhen Chen, Shu Chien,and Sheng Zhong, *Proceedings of the National Academy of Sciences of the United States of America*, 2016. The dissertation author was contributing investigator and author of this paper.

Chapter 5

Detection and Validation of Cell Surface RNA with Surface-FISH

5.1 Introduction

Human transcriptome analysis estimates that 60-70% of the genetic information is transcribed by non-coding RNAs (ncRNAs)[10, 35, 17] fulfilling complex functions during human development and physiological activity. The localization of these ncRNAs can reveal disease and differentiation mechanisms and will help to understand their functions. Less explored is the possibility that RNA molecules may appear on the cell surface membrane, which may play crucial role as well. Thus, we proposed that some specialized cells, such as cancer and immune cells, utilize a sophisticated identity control system via ncRNA that regulates cell communication networks and signal transduction through the cell surface membrane assisting proteins and other molecules. The concept that intercellular communication is mediated via RNA through the cell surface has not been described yet, however is well understood that they utilize vesicular systems to communicate. To demonstrate the existence of cell surface membrane RNAs, isolation of highly purified membrane is critical. Thus, to identify potential cell membrane RNA targets we used

current technology, such as cell membrane cloaking nanoparticles method[49]. This technology is crucial because not only isolate the cell membrane, but maintain the membrane orientation intact and minimize the contamination of intracellular and extracellular RNAs producing high yield of cell surface membrane RNAs. The resulting membrane fragments were subjected to RNA purification for library preparation and sequencing. The sequencing data determined more than 3500 possible RNA candidates associated with the cell membrane that had positive FPKM through the sequencing. However the cell surface specificity was still lacking. To increase this specificity, we proposed that cell surface RNAs are exposing their 3 and 5 ends through the protein membrane and could be captured by their specific adaptors. This novel approach not only captured cell surface specific RNAs, but it narrowed down the number of RNA candidates. Thus, intersection data sets of all sequenced samples determined seven potential lncRNA candidates and among them Malat1.

In order to validate the sequencing findings we developed a novel single molecule qDot RNA FISH technology for visual confirmation of RNAs on the cell surface membrane (Surface-FISH). We chose two candidates with varying FPKM to validate their existence in the cell surface membrane. According to RNA-Seq, Malat1 was enriched with a FPKM score of 4.3, while Abl1 was enriched with a FPKM score of 0.75. Using these as controls, we targeted the two genes with qDot bound oligos where the expectation is to see more membrane bound RNA for Malat1 compared to Abl1 which had a weaker signal from the sequencing results.

To ensure only detection of RNA transcripts on the cell surface using Surface-FISH, we addressed the cell membrane integrity to avoid permeabilization of our hybridization probes. The first consideration for this condition was in the design of oligonucleotide probes. Similar to single molecule fluorescence in situ hybridization (smFISH), Surface-FISH uses a set of 3-5 oligonucleotides to detect an RNA transcript of interest. Oligonucleotide probes were designed to be 35-40 bases long, the probes are modified on the 5 end with a primary amino group and a spacer or 30 carbons to minimize steric hindrance during hybridization. The probes are bound to

carboxyl modified quantum dots about 10 20 nm in size, which do not readily diffuse through the membrane. Probes were designed towards a region on Malat1 where our sequencing result provided the most read coverage since the region may be more accessible to probe hybridization. The second consideration was the absence of a permeation reagent after the fixation of the cells. Permeation of the membrane is often required for allowing fluorophores such as quantum dots to enter the cell membrane, while a lack of permeation should prevent entrance into the cytoplasm of the cell.

Two approaches were taken to ensure detection of Malat1 on the surface of EL4. In the first approach, we asked if Malat1 did appear on the cell surface. It was assumed that fixed cells without permeation would not allow hybridization probes conjugated to quantum dots to enter the cell. This means that any detection of Malat1 would be on the surface of the cell since the oligonucleotides can only access outside of the cell membrane. Comparative experiment was performed between, unfixed, fixed, and fixed with permeating factor to demonstrate that detection of Malat1 in the first two conditions were on the surface. Despite evidence showing membrane surface Malat1 transcript by forgoing permeation, there is evidence showing fixation may partially damage the cell membrane.

The second approach alleviates the assumption of an intact membrane by direct determination membrane integrity using the unique combination of CellTracker fluorophores with absorbance dyes. Ideally, without permeation of the cell membrane after fixation, cell membranes should remain intact. However, it was observed that a proportion of cells may have partially damaged membranes after buffer exchange from the fixation solution to saline. This led to increased difficulty in determining whether oligonucleotides were hybridizing to RNA inside or outside the cell. To bypass the need to assume membrane integrity, this technique was conceived since the absorbance dye Acid Blue 9 is impermeable to intact membrane, while the CellTracker Orange is membrane permeable. When damage occurs to the cell membrane, the absorbance dye can leak through and quench the fluorescence from CellTracker Orange, allowing for clear probe

visualization of membrane integrity. We can show that in cells with fully intact membranes that Malat1 was detected on the surface of the cell.

5.2 Results

5.2.1 Designing Surface RNA Oligo Probes

Our goal is to determine the existence of these RNA on the cell surface and identify potential candidates. We first attempted to identify potential membrane RNA targets using a membrane isolation protocol that removes the internal components of a cell and causes the membrane to wrap around nanoparticles with the orientation remaining intact. Afterwards, the resulting membrane fragments purified for RNA and sequenced using RNA-seq. The resulting sequencing are candidates for RNA associated with the cell membrane. We have determined more than 1500 possible candidates that had a positive FPKM through the sequencing. To validate the sequencing findings a new technology needed to be developed allowing for visual confirmation of RNA on the membrane surface. We have developed a method to validate existence of membrane bound RNAs using single molecule qDot RNA FISH. we chose two candidates with varying FPKM to validate if they exist on the membrane surface. According to RNA-Seq, Malat1 was enriched with a FPKM score of 4.3, while Abl1 was enriched with a FPKM score of 0.75 shown in Table 5.1. Using these as controls, we targeted the two genes with qDot bound oligos where the expectation is to see more membrane bound RNA for Malat1 compared to Abl1 which had a weaker signal from the sequencing results.

Table 5.1: Gene candidates for validation using Surface-FISH.

Gene Name	locus	length	FPKM	Biotype
Malat1	chr19:5795689-5802671	6982	4.35836	lincRNA
Abl1	chr2:31688536-31804362	7367	0.751804	protein_coding

5.2.2 Characterizing cell surface RNA by sequencing using Surface-seq and validation with surface-FISH

Surface-Seq

In order to analyze what RNAs may be present on cell surface, we developed a technology called Surface-seq to selectively construct RNA sequencing libraries based on any RNAs that are stably associated with the extracellular layer of the cell membrane. Surface-seq is based on a revolutionary technology that that peels off cell membranes and tightly assemble the membrane onto polymeric cores to form membrane-coated nanoparticles (MCNP)[50]. This technology is particularly suited for our purpose because it keeps the inside-outside orientation of the membrane with the surface molecules on the membrane facing outside, and the process of membrane purification and coating the cell membrane onto the polymeric core ensured rigorous removal of intracellular contents[50].

Table 5.2: Probe sequences for Surface-FISH. C30 represents a 30 carbon spacer and N bases refer to equal mixture of A, C, T, or G bases.

Sequence Name	Sequence
Neat1-1	Amine-C30-GAATGTCTGAGTTCTTGGCCAGCCTGGTTTACAAA
Neat1-2	Amine-C30-CAATCCAAGGCTTCTGGTTGAGGAATGGTGATAG
Neat1-3	Amine-C30-CACATTTGGGAGGCAGTCATTAGTAGATCTTTGAA
Neat1-4	Amine-C30-GTTTTCAGTTAAGAATCCCTCTGACCAATGCAG
Neat1-5	Amine-C30-CAAGGTTTTAAGTGACCCCTTAACCTCAGAGTGAG
Malat1-1	Amine-C30-GTTACAAATAAACACAAGTATACAATGCACAAGAAG
Malat1-2	Amine-C30-GCCACTTCCTTTGTTCCTATAGTAGTTATTAAGAT
Malat1-3	Amine-C30-CTTGATAATATAAAAGCTATCACCAGAAGAAATTCC
Malat1-4	Amine-C30-CTACAATCTATATTCATCCAACAGCTTCAGAAGAG
Malat1-5	Amine-C30-CTGAAATCATAAACTAAACAATTACCTAACACCCC
Malat1_Mut	Amine-C30-GTTACAAATAAACANNNNNNTACAATGCACAAGAA
Neat1_Mut	Amine-C30-GAATGTCTGAGTTCTTNNNNNNCCTGGTTTACAAA

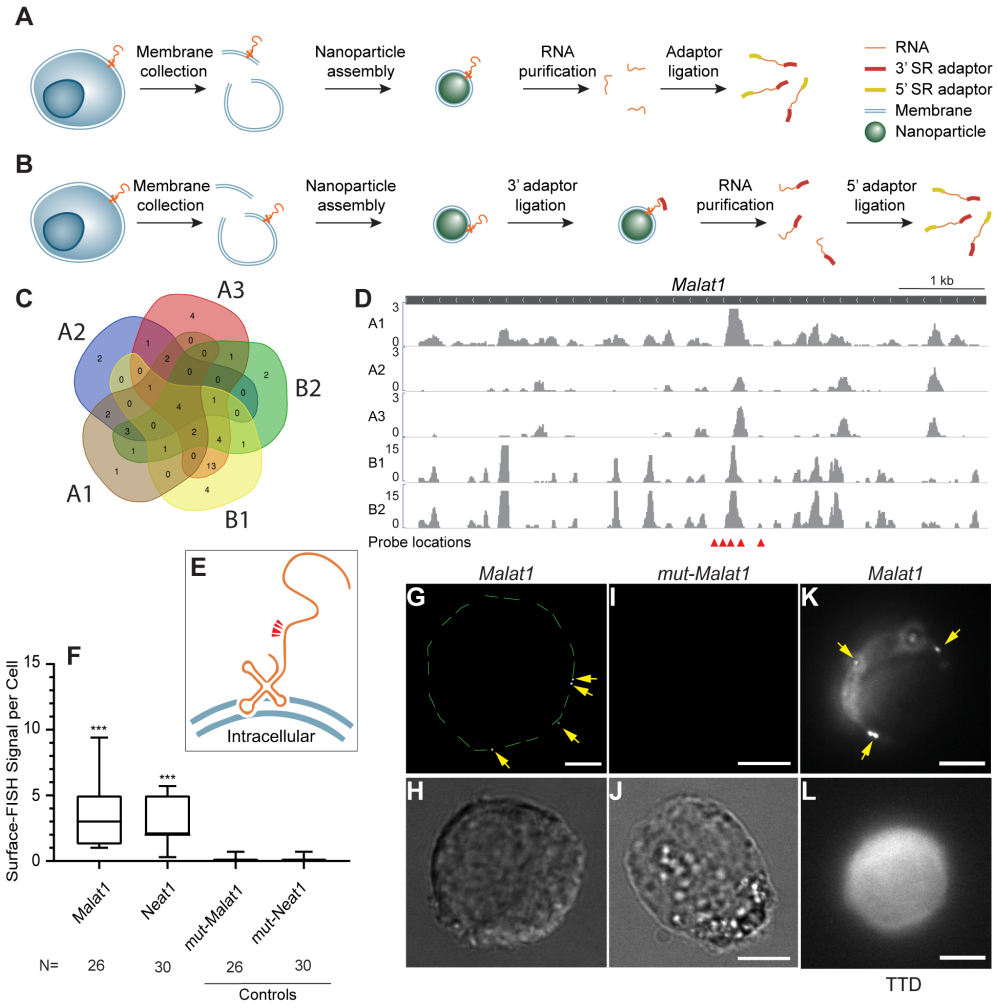

Figure 5.1: Surface-seq and Surface-FISH. (A-B) Schematic pipelines of the two variations of Surface-seq techniques. (C) Venn diagram of the non-coding RNAs identified by the 5 Surface-seq experiments, indexed by A1, A2, A3 (3 libraries based on Surface-seq v1) and B1, B2 (2 libraries based on Surface-seq v2). (D) Read coverages from the 5 Surface-seq libraries on the Malat1 gene, indexed by A1, A2, A3, B1, B2. Red arrowheads: locations of Surface-FISH probes. (E) A hypothetical model of the relative positions of Surface-FISH probes (red arrowheads) on a membrane-bound Malat1 transcript. (F) Box plots of the numbers of Surface-FISH signals per cell (y axis) for Malat1, Neat1, and 3 controls (dArt4, Malat1-r, Neat1-r) (columns). N: number of cells. (G) Malat1 Surface-FISH (Arrows) and DIC (J) images of the same cell. Green dashed line outlines the rim of the cell. (H) dArt4 (control) Surface-FISH (upper panel) and DIC (K) images of the same cell. Surface-FISH (I) and TTD (L) images of the same cell. Arrows: Malat1 Surface-FISH signals. The TTD image was produced by a membrane permeable dye used in conjunction with a membrane impermeable quencher, indicting a cell with intact cell membrane. Scale bar = 5 m. ***: p-value ¡ .0001, Kruskal-Wallis test.

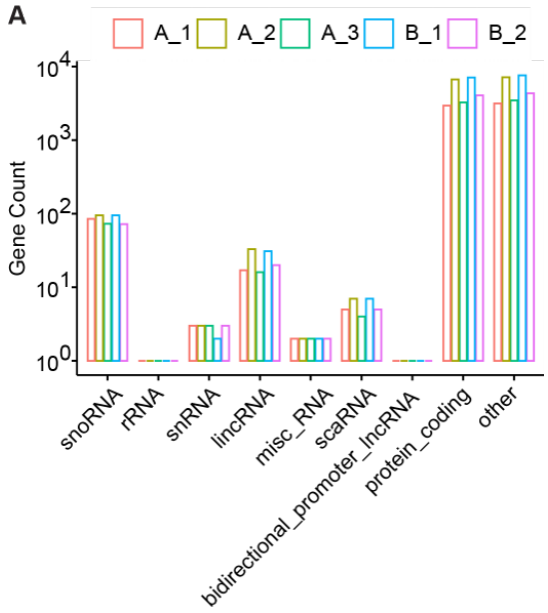

Figure 5.2: (A) Summary of counts for major RNA types in transcripts detected (FPKM threshold =1) in the 5 Surface-seq datasets, indexed by A1, A2, A3, B1, B2. Replicate experiments are annotated by different colored bar.

Leveraging the MCNPs, we developed two variations of the Surface-seq technology. In the first variation, after MCNPs were assembled and washed, we used phenol-chloroform to extract RNAs, which were subsequently quantified and constructed into a sequencing library (Figure 5.1A). In the second variation, we ligated the 3 sequencing adaptor to any RNAs that are exposed on the MCNPs surface, before extracting the RNAs for sequencing library construction (Figure 5.1B).

We generated 3 experimental repeats based on Surface-seq variation 1 and 2 experimental repeats based on Surface-seq variation 2. Each repeat was started with approximately 400 million EL4 cells (Table S-SurfaceSeq-1). Each sequencing library revealed 20 to 33 lincRNAs, and the intersection of the 5 Surface-seq libraries included 4 lincRNAs (Figure 5.1C, Figure 5.2), which are Malat1, Neat1, Snhg20, and mir181a-1 (Figure S-Neat1). Taking Malat1 for example, the Surface-seq reads were not uniformly spread across the entire lincRNA (Figure 5.1D). Instead, the Surface-seq reads were enriched on specific regions, especially near the center of the Malat1

transcript. These data provided candidate mexRNAs for further validation.

Surface-FISH

To test candidate mexRNAs, we developed a technique to carry out single molecule RNA FISH on the cell surface (Surface-FISH). Surface-FISH hybridizes five 40 nt quantum-dot-labeled oligonucleotide probes against a target transcript (arrows in Figure 5.1D-F) following a previously established protocol[95] except that the cell membrane is not permeabilized. Cells are subsequently washed and subjected to microscopic analysis. We tested two Surface-seq prioritized lincRNAs, Malat1 (Figure 5.1D,G), Neat1, with the drosophila Art4 transcript as a control (dArt4) (Figure 5.1H). We included two additional controls, by mutating the 6 bases at the center of the 40 nt probes to Malat1 (mut-Malat1) and mutating the 6 central bases of the Neat1 probes (mut-Neat1) (Figure 5.1F, Table 5.2). We examined 20 to 30 single cells for each target transcript or control experiment (number of cells (N) in Figure 5.1F). Malat1 and Neat1 exhibited Surface-FISH signals in nearly every cell, ranging from 1 to 10 signals per cell, whereas the two controls did not exhibit any signal in the majority of cells (median = 0) (p-values ¡ 0.0001, Wilcoxon rank tests) (Figure 5.1I).

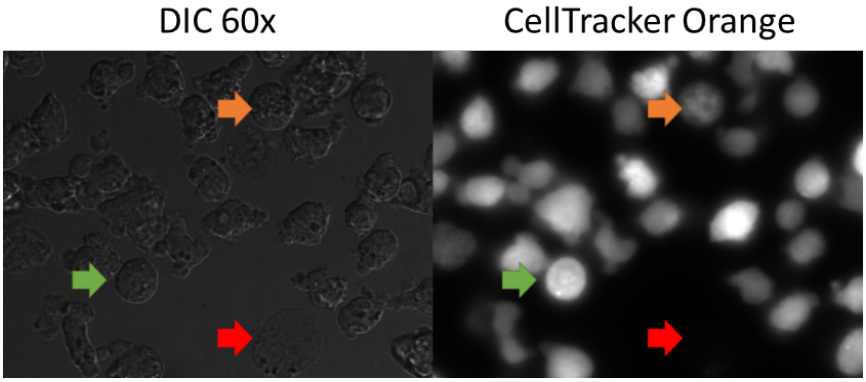

Figure 5.3: TTD microscopy of EL4 cells using Acid Blue 9 (AB9) as quencher dye and CellTracker Orange as cell labeling dye. Green arrow indicates a cell that has intact membrane where membrane impermeant AB9 does not affect CellTracker Signal. Orange arrow indicates cell with partially damaged membrane that has lowered fluorescent signal. Red arrow indicates a cell with damaged membrane where the CellTracker signal is fully suppressed.

To ensure that the Surface-FISH signals are not a result of RNA release due to membrane leakage, we combined Malat1 Surface-FISH with a transmission-through-dye (TTD) microscopic analysis, where only live cells with intact membranes are fluorescently labeled[63, 68] (Figure 5.3). Malat1 FISH signals appeared on cells with perfectly intact membranes as indicated by TTD staining (Figure 5.1J-K). These data suggest that specific genome-coded transcripts are presented on the surface of live cells.

5.2.3 Additional validation of Malat1 on the cell surface and effect of fixation on detection of cell surface RNA

A critical step in our RNA FISH studies is to maintain the cell membrane integrity to target cell membrane specific RNAs and prevent the entry of oligonucleotide probes inside the cells. Fixation of the cells with 4% paraformaldehyde (PFA) without permeating will prevent probe intake. Additionally, quantum dots are very large inorganic fluorophores that cant enter the cells without previous permeation.

Thus, we designed Surface-FISH experiments with EL4 lymphoma suspension cells and oligonucleotide probes: mMalat1 and hAbl1 (negative control) (Table 5.2) conjugated with qDot 705. Experiments were carried in PFA treated (fixed) and non-treated (unfixed) EL4 lymphoma cells as follows:

1. Fixed and non-permeated, this condition prevents the cells to intake the conjugated QD-oligonucleotides and target only Malat1 on the cell surface membrane. Thus, we expect to target extracellular Malat1 (Figure 5.4). For Abl1 we expect low to no detection of transcripts on the surface (Figure 5.5).

2. Unfixed and non-permeated, this condition prevents the cells to intake the conjugated QD-oligonucleotides and target only Malat1 on the cell surface membrane, however we expect that formamide, which is an organic solvent and a component of the hybridization buffer

might also cause some permeability (Figure 5.6). Abl1 again should have less transcripts since it had a much lower expression from cell surface sequencing analysis (Figure 5.7).

3. Fixed and permeated, this condition serves as control and allow the cells to intake the conjugated QD-oligonucleotides. This condition will target intracellular and extracellular Malat1 (Figure 5.8), while for Abl1 we expect to target intracellular transcripts Abl1 (Figure 5.9.

Detection of Malat1 and Abl1 in fixed and non-permeated EL4 cells

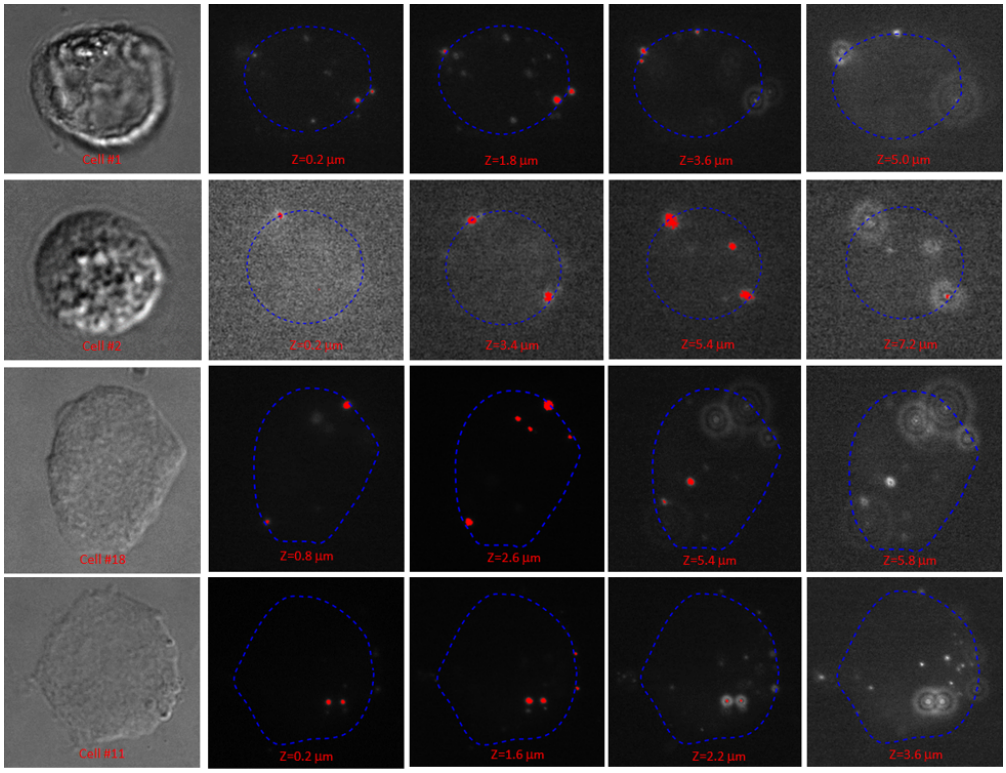

Figure 5.4: Surface-FISH experiments on fixed and non-permeated EL4 lymphoma cells using Malat1 oligonucleotide probes conjugated to 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

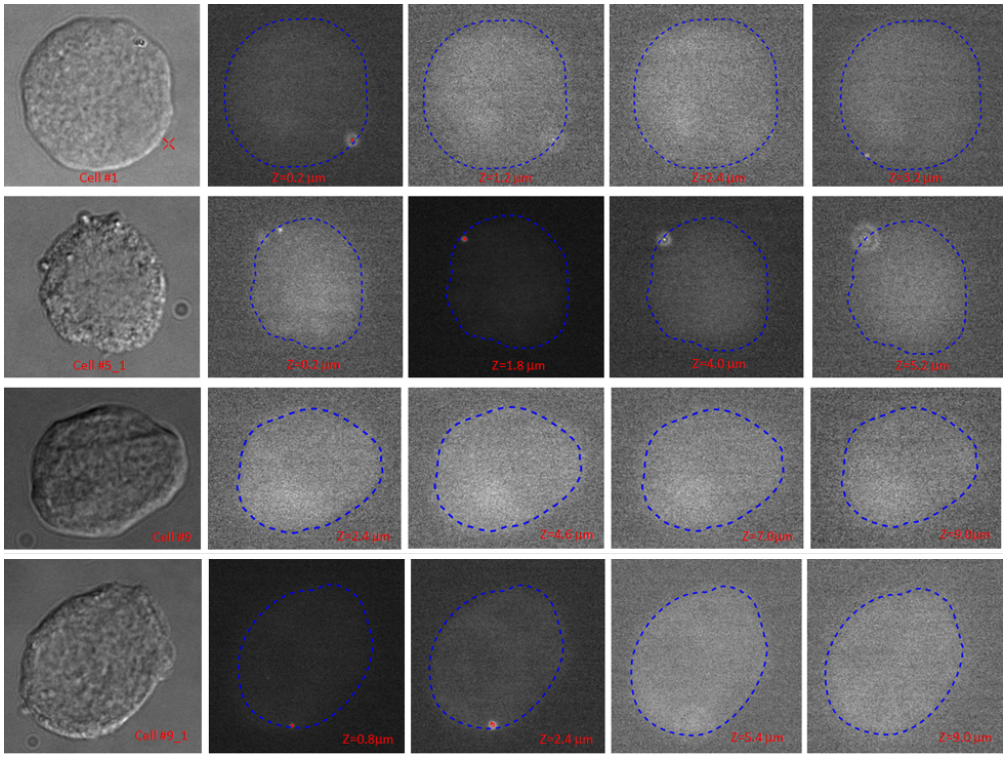

Figure 5.5: Surface-FISH experiments on fixed and non-permeabilized EL4 lymphoma cells using AbL1 oligonucleotide probes conjugated to 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

Detection of Malat1 and Abl1 in unfixed and non-permeated EL4 cells

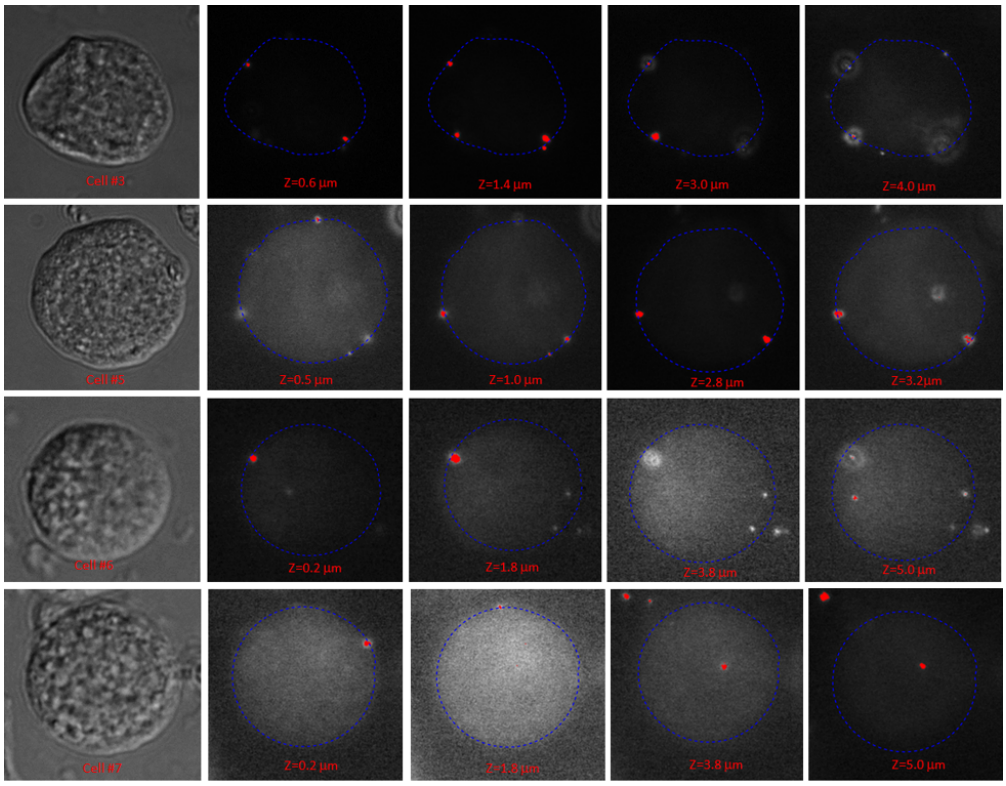

Figure 5.6: Surface-FISH experiments on unfixed and non-permeated EL4 lymphoma cells using Malat1 oligonucleotide probes conjugated to 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

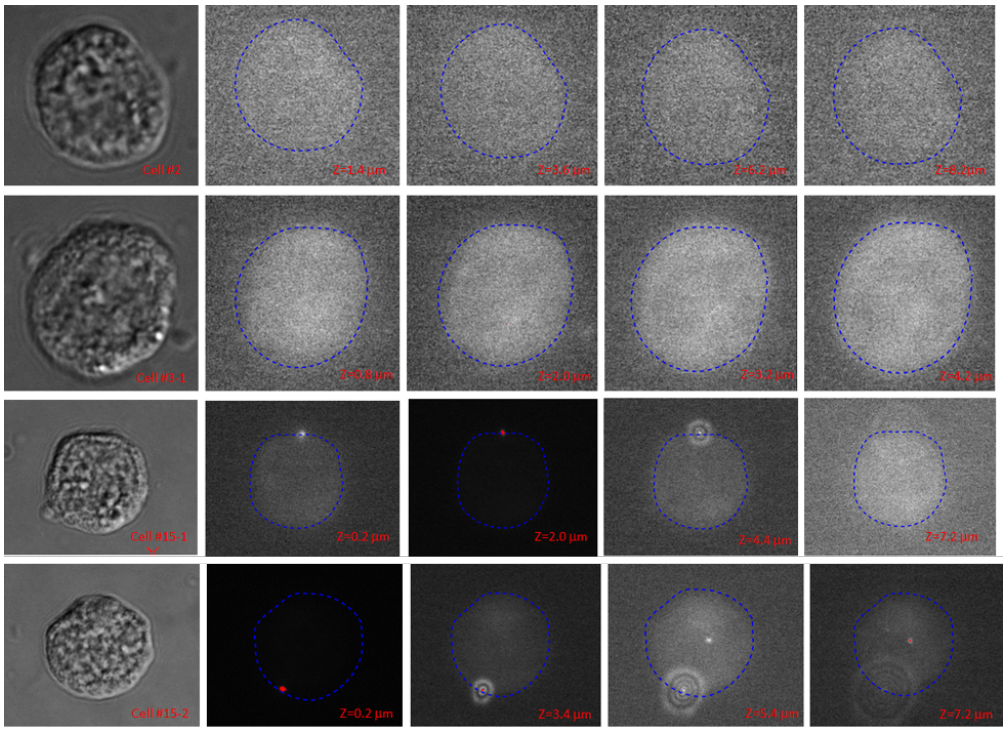

Figure 5.7: Surface-FISH experiments on unfixed and non-permeabilized EL4 lymphoma cells using AbL1 oligonucleotide probes conjugated to 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

Detection of Malat1 and Abl1 in fixed and permeated EL4 cells

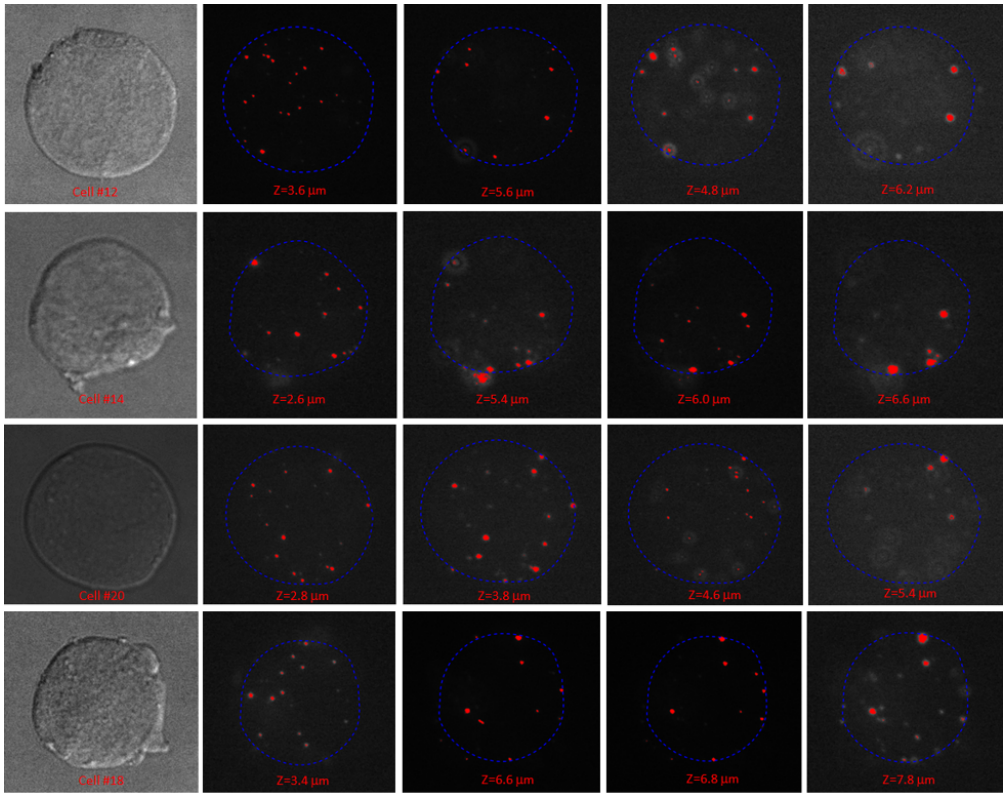

Figure 5.8: Surface-FISH experiments on fixed and permeated EL4 lymphoma cells with Malat1 oligonucleotide probes conjugated 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

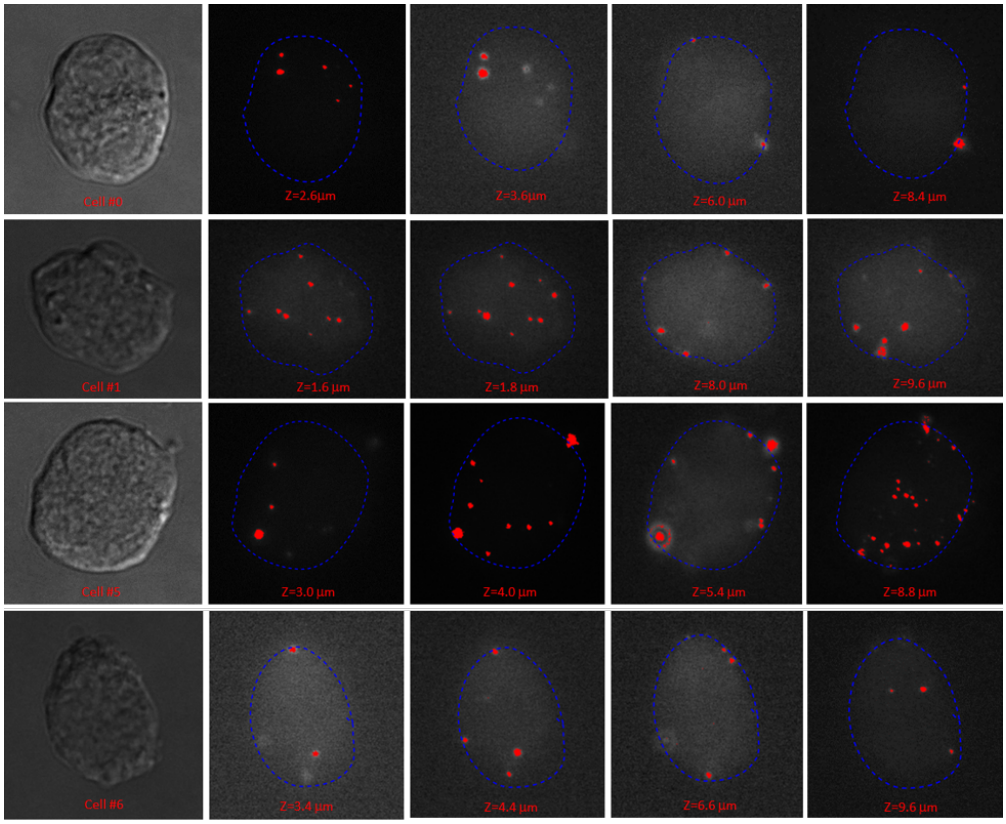

Figure 5.9: Surface-FISH experiments on fixed and permeated EL4 lymphoma cells with Abl1 oligonucleotide probes conjugated 705 Quantum dots. The figure represents 20 imaged single cells and shows the optical sections (z) from the bottom, middle and top of each cell. Optical sections (Z stacks) of the 20 single cells were acquired using ImageJ software and the blue dashed lines were drawn using Inkscape software. Threshold of the fluorescent signal was selected by visualization of the image using ImageJ software, so white dots in some of the imaged cells as well as some brighter images are background noise due to the lack of fluorescent signal.

5.3 Discussion

The surface of the plasma membrane presents a unique perspective towards how a cell may interact with its external environment. Many studies have been performed centered around the molecules that make up the plasma membrane from lipids to surface markers to membrane channels that affect a host of external cellular functions. While these molecules are important for the characterization of the cell surface, there may be another class that have been largely overlooked in terms of presence, RNA. RNA has largely been perceived as a precursor to protein

synthesis and only recently has new evidence been shown that there are noncoding RNAs that may serve functions independent from proteins. In this chapter, we were able to demonstrate that there is an existence of csRNAs using a nanoparticle isolation technique, which was validated with Surface-FISH where we observed Malat1 and Neat1. Although more work will be needed to uncover additional insights into their roles and functions, we give a small glimpse into the potential for a new bio-molecule on the cell membrane.

5.4 Methods

5.4.1 Suspension Cell fixation and permeation

Cells were grown to confluence in T-75 non-TC treated flasks. Cells were washed twice with 1X PBS to remove excess of media and fixed for 10 minutes with 4% formaldehyde in 1X PBS. Cells are permeated with 70% Ethanol at -20C overnight.

5.4.2 Surface-FISH analysis

Oligonucleotide design

Surface-FISH uses a set of 3-5 oligonucleotides to detect an RNA transcript. Probes were designed towards a region on Malat1 where our sequencing result provided the most read coverage since the region may be more accessible to probe hybridization. Oligonucleotide probes were designed to be 35-40 bases long, the probes are modified on the 5 end with a primary amino group and a spacer of 30 carbons to minimize steric hindrance during hybridization. Consideration is placed on ensuring similar GC content and melting temperature to ensure efficient hybridization of the oligo probes.

Conjugation of oligonucleotides to quantum dots

In lo-bind 2.0 ml Eppendorf tube quantum dots (QD) were diluted 1:10 (.8 uM). From the diluted QD solution, 1 ul was used and mixed with 200 ul of 50 mM Borate buffer pH 8.0, 8 uL of 50mM EDC and 8 uL of 50mM NHS buffer and mixed well with gentle flicking. 15 uL of 100 uM oligonucleotides were immediately added and mixed. Conjugation reaction was incubated at RT with gentle rotating overnight.

Purification of oligonucleotide quantum dots conjugates

QD aggregates are initially removed using a .2 um filtration column centrifuged at 2500 rpm for 15 min. Excess of unconjugated DNA oligonucleotides were removed by ultrafiltration using 100K molecular weight columns (MWCO) and centrifuged at 2500 rpm for 20 min followed by magnetic bead purification. An aliquot of 50 uL of silane magnetic beads was added in a lo-bind 2.0 mL Eppendorf tube and placed on a magnetic rack until complete separation to remove storage buffer. Beads are then resuspended in 200 uL of RLT buffer and placed on a magnetic rack to separate and remove supernatant. Once washed, the beads are resuspended in 175 ul of RLT buffer followed by the QD-DNA conjugate, then 225 ul of isopropanol. The mixture was gently flicked and incubated at room temperature for 10 minutes on a rotator. Tubes were place on a magnetic rack to remove supernatant and washed twice with 70% ethanol. Beads were eluted with 100 ul of borate buffer, magnetically separated and the eluate transferred to a new tube. To remove any newly formed aggregates, the eluate obtained from silane beads was then transferred to a 0.22 m filter columns and centrifuged at 2500 rpm for 15 min.

Hybridization of oligonucleotide quantum dots conjugates to suspension cells

Oligonucleotide-qDot conjugates are added to 500 uL hybridization buffer (containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNAse-free BSA, 40 g E.coli tRNA, 2x SSC, 10% formamide). The solution is gently mixed by flicking and spun down.

Hybridization solution is added to fixed suspension cells and mixed gently by pipetting up and down with a large pipette tip. Mixture is incubated at 37 C overnight. After overnight incubation, cells are spun down at 500 g for 3 minutes and 250 uL of hybridization is removed. An initial wash with 500 uL 2x SSC, 10% formamide is performed by gentle pipetting. Cells are spun down and supernatant is removed and washed again with 2x SSC, 50% formamide. Cells are spun down and supernatant is removed and washed with 2x SSC. Once again remove the supernatant and resuspend in 1x PBS for imaging.

Imaging and Image Processing

Cells are imaged in 1x PBS through wide field fluorescence imaging using the Olympus IX83 inverted microscope at 60x oil immersion objective (NA=1.4). Image processing is performed using imageJ and Matlab with custom programs.

5.5 Acknowledgments

Chapter 5 in part is currently being prepared for submission for publication of the material. Norman Huang, Xiaochen Fan, Kathia Zaleta Rivera, Tri C. Nguyen, Jiarong Zhou, Yingjun Luo, Jie Gao, Ronnie H. Fang, Zhen Bouman Chen, Liangfang Zhang, Sheng Zhong Naturally occurring cell surface-display of genome encoded RNAs and their impacts on cell-cell interactions. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Membrane associated extracellular RNAs and their impacts to cell-cell interactions

6.1 Introduction

Cell surface is the critical interface for a cell to interact with the extracellular environment and other cells. Several important functions including sensing extracellular environment, receiving the extracellular signals, and presenting the cellular identity to other cells are carried out by surface molecules. Every type of known surface molecules including proteins, peptides, lipids, and their chemically modified variations are involved in some of these critical functions. It is less clear whether nucleic acids contribute to mediate functions of the cell membrane.

The human genome encoded RNAs (hgRNA) were not expected to be exposed on the surface of cells with intact membrane[108]. As an exception, the extracellular hgRNAs (exRNAs) that are weakly attached to and be easily washed off from the cell membranes[70, 94]. It is unclear whether such promiscuous surface attachment may mediate any biological functions. Another debatable exception is that ribonuleoproteins can be released to the surface side of leaked membranes of dying cells and serve as antoantigens[108, 129].

However, several lines of prior data alluded to relatively stable RNA-and-cell-surface interaction and the possibility of a fragment of a hgRNA can be exposed to the surface of intact human cells in physiologically normal conditions. First, membrane bound RNAs were discovered in bacteria[11, 112]. These membrane non-coding RNAs are located within the same transcription units of transmembrane protein coding genes (TPCG). Following co-transcription, the non-coding RNA and the transmembrane protein form a ribonucleoprotein, that is localized to the cell membrane[11, 112]. In humans, a few hgRNAs can bind membrane lipids in physiological ionic conditions [80, 60]. Atomic force microscopy (AFM) data suggest that the RNAs attached to the double-layered lipids can be exposed to the milieu, without being encapsulated in any vehicles[58]. Additionally, mRNAs can be tethered to intracellular organelles by membrane-associated RNA-binding proteins to promote organelle-coupled translation[9]. Diverse nuclear coded transcripts can be translocated into mitochondria through mitochondrial membranes[61], making us wonder whether a similar process could happen at cell membrane resulting a display of RNAs on cell surface.

Here, we have tested the presence, cell-type specificity, and functions of hgRNAs that exposed at the surface of mouse and human cells. We call these RNAs membrane associated extracellular RNAs (mexRNA). Throughout this chapter, we will provide clear contexts on whether we are referring to the mexRNAs that are promiscuously attached to cell surface (weak mexRNAs), or those with stable RNA-and-cell-surface interactions (strong mexRNAs).

6.2 Results

6.2.1 Screening and visualization of csRNA in cell lines

Previous literature has described the ability for specific RNAs to bind to cell membrane surfaces but have yet to show the presence of native RNA on the cell surface. Here we present a novel detection of cell surface nucleic acids on the surface of a cell lines and PBMCs using a

randomized set of 5 modified 20 base oligonucleotides (Figure 6.1A). In order to study primary cells, we developed and tested a surface-hybridization method to investigate surface-RNA without requiring metabolic labeling, molecular, genome perturbation, or genome engineering. We demonstrate the presence using imaging flow cytometry, FACS-based single cell RNA-seq and biotin pulldown scRNA-seq towards detection of csRNAs. Here we compared our results with three different selection methods to validate the presence of csRNAs.

We first answered two question, whether we are detecting fluorescent signal that is on the cell membrane and whether this signal is specific to different cell types. We screened several cell lines related to PBMCs and imaged on average 17,000 cell per condition to determine the presence of csRNA using imaging flow cytometry (Figure 6.1B, 6.1C). We tested the following cell lines lymphoblast K562, B cell Daudi, B cell GA-10, B cell Raji, T cell Jurkat, and monocyte THP-1. As a negative control we compared the oligo library to an oligo targeting a cross species gene drosophila Art4. Positive signal was detected at varying degrees for GA10, Raji, THP-1, and K562, while no difference in signal was detected between Daudi and Jurkat cells. The observed csRNA signal only is shown near the periphery of the cell membrane suggesting detection of RNA outside the cell. Additionally, we tested when K562 cells are permeated signal appears throughout the cell cytoplasm as expected. Here we are able to show that using oligo hybridization we were able to screen live cell lines for csRNA.

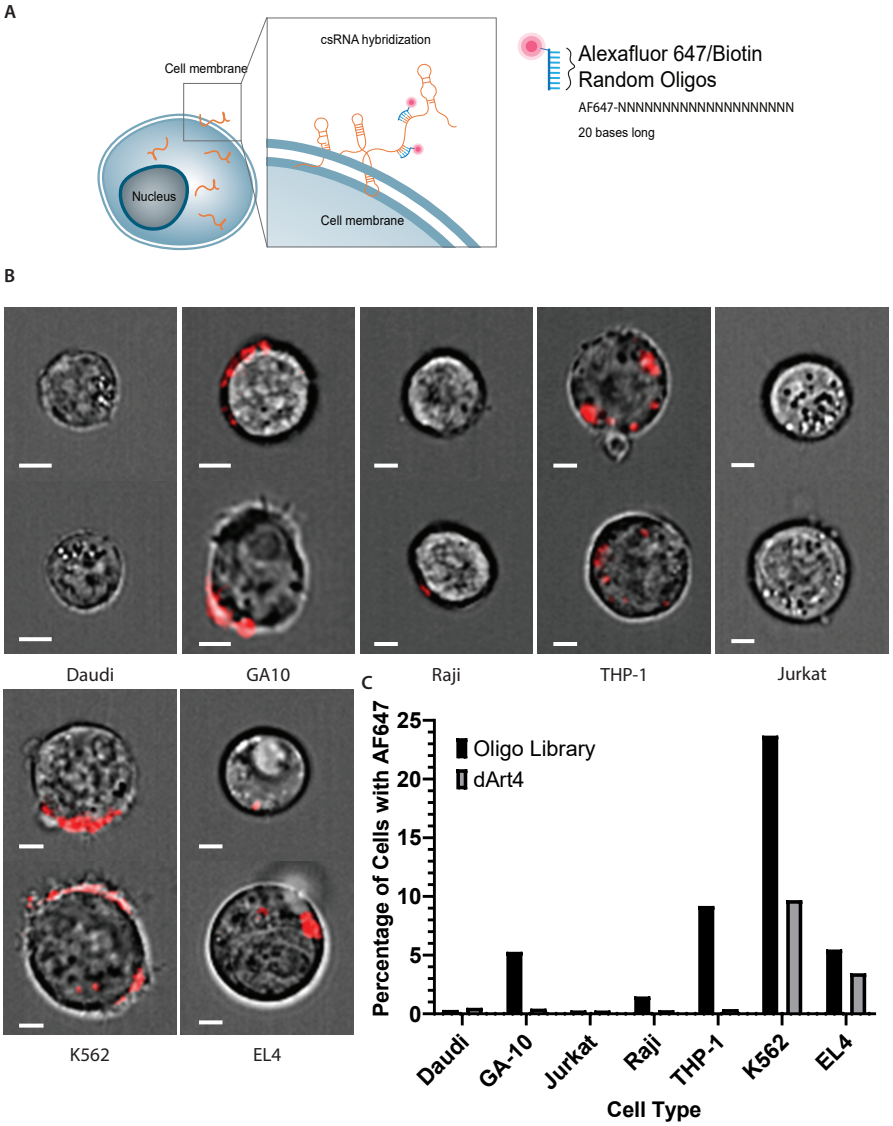

Figure 6.1: Visualization of csRNA on peripheral blood cell lines. (A) Representation of random probe library with AF647 attached to random N oligo library for hybridization on csRNA. (B) On average 17,000 cells were imaged per cell type and the percentage of cells exhibiting csRNA signal where positive signal was gated compared to no fluorescence controls. GA10, Raji, THP-1, K562 and EL4 cell lines showed signal above the negative control dArt4. (C) Representative images of cell lines Daudi, GA10, Jurkat, Raji, THP-1, K562, and EL4. Red signal denotes AF647 oligos bound to the cell surface. Scale bar is 3.5um

6.2.2 Testing cell type variation of mexRNA presentation with Surface-PromFISH

The cell type variability of mexRNA presentation observed so far might be an indication of mexRNAs differential function of the csRNA in the cells. We reasoned that if mexRNAs do not participate any function of their presenting cells, we should not expect to see cell type variability of mexRNA presentation. Thus, the so far observed cell type variability of mexRNA presentation may indicate the possibility of mexRNAs having functional roles. To address the physiological relevance of mexRNA, we examined whether mexRNA is in primary healthy human cells, and whether its presentation also exhibit cell type variability. For this purpose, we used peripheral blood mononuclear cells (PBMCs) due to their heterogeneity and suitability for surface marker characterization. We analyzed PMBCs from 4 research subjects, with approximately 120,000 total PBMCs analyzed for each donor.

Cell type variability of mexRNA presentation could be an indication of mexRNA having functional relevance to the cells. Therefore, we tested whether mexRNA presentation is uniform across different cell types in primary human cells. We chose peripheral blood mononuclear cells (PBMCs) for this test due to their accessibility. To investigate the presence of any possible mexRNAs on primary cells, we analyzed PMBCs from 4 research subjects. We obtained approximately 120,000 PBMCs per donor and split them into 4 aliquots. Each aliquot was used for either the Surface-PromFISH experiment or the 3 control experiments.

In the Surface-PromFISH experiment, we applied a randomized library of fluorescence labeled 20 nt oligonucleotide probes (20nt probe library) to hybridize to putative mexRNAs (Figure 6.2A). We subjected approximately 30,000 PMBCs to a flow cytometry analysis of 6 channels, which were labeled with DAPI, LIVE/DEAD Fixable Blue (Live/dead), antibodies against surface markers for CD3 T cell co-receptor, monocyte marker CD14, and B cell marker CD19, and Surface-PromFISH (Figure 6.2B-C). In each control experiment, replaced Surface-

PromFISH with one of the following 3 controls. First, we used a randomized probe library of 6 nt probes (6-mer library control), which controls for non-specific membrane attachment of fluorophore-probe conjugates. Second, we used a 20 nt probe against the drosophila Art4 (dArt4) RNA, which controls for non-sequence specific binding of a probe to the cell surface (dArt4 control). Third, we used the fluorophore without conjugating any oligonucleotides (fluorophore control). On average, 4.8% PBMCs exhibited Surface-PromFISH signals, that is 27-fold more than any of the control groups (Figure 6.2D) (p-value ¡ 0.005, Kruskal-Wallis test). Less than 3% of CD19+ and CD3-CD14-CD19- cells exhibited Surface-PromFISH signals, whereas on average approximately 5% of CD3+ and more than 12% CD14+ cells exhibited Surface-PromFISH signals (p-value ¡ 0.005, t test) (Figure 6.2D), suggesting that csRNA-presenting cells are enriched in the monocyte subpopulation of PBMCs.

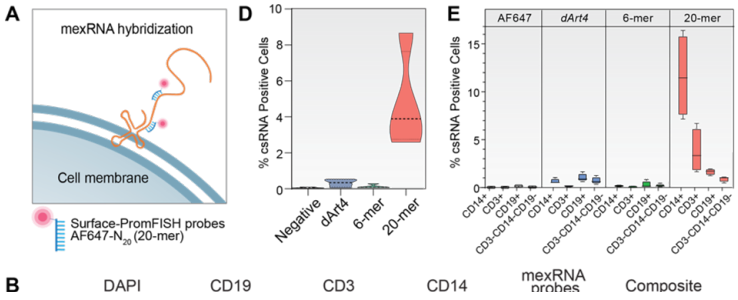

Figure 6.2: Imaging flow cytometry analysis of mexRNAs co-expression with immune cell surface markers. (A) A schematic view of the extracellular hybridization of Alexafluor 647 (AF647, red dot) labeled probes (Surface-PromFISH 20-mer probes) to the exposed single strand regions of a hypothetical mexRNA. (B-C) Fluorescence images of two representative PBMC cells (rows) in 5 channels (columns) including DAPI, antibodies for CD19, CD3, and CD14, mexRNA probes, as well as merged images of all channels (Composite). (D-E) Distributions of the percentages (y axis) of PBMCs (D) exhibiting Surface-PromFISH (red), 6-mer library control (green), dArt4 control (blue), and AF647 fluorophore control (black), and the percentages of CD14+, CD3+, CD19+, and CD3-CD14-CD19- cells (E).

6.2.3 Single-cell transcriptome analysis of candidate mexRNA presenting cells

To characterize the mexRNA presenting cells at single-cell level, we subjected approximately 40 million Surface-PromFISH labeled PBMCs to fluorescence activated cell sorting (FACS) and obtained two cell groups (Surface-PromFISH+ and Surface-PromFISH-). We subjected each group to single-cell RNA sequencing (scRNA-seq) using the 10X genomics platform. A total of 9,058 and 2,488 single cells were sequenced from the Surface-PromFISH+ and the Surface-PromFISH- group, with on average 21,059 sequence reads per cell. Surface-PromFISH+ and Surface-PromFISH- cells were in general separated on a tSNE plot (Figure 6.3A). The majority (87%) of the Surface-PromFISH+ cells were classified as monocytes, whereas the Surface-PromFISH- cells included all cell types found in PBMCs (Figure 6.3D). Specifically, monocyte markers CD14 (Figure 6.3B) and LYZ (Figure 6.4) were detected in large fractions of Surface-PromFISH+ cells but were absent from the vast majority of the Surface-PromFISH- cells. In comparison, T cell markers CD3 (Figure 6.3C), CD8, natural killer (NK) cell marker NKG7, B cell marker MS4A1 were enriched in the Surface-PromFISH- population (Figure 6.4).

In a complementary approach, we classified our sequenced single cells into 11 cell types (Figure 6.3D) using SingleCellNet[130] with a published training dataset of approximately 68,000 PBMC single cells[150]. We used log odds ratio to quantify the association of Surface-PromFISH+ cells with every cell type resulted from this classification (Figure 6.3E). Monocytes and dendritic cells exhibited enrichments to the Surface-PromFISH+ population (p-value ¡ 0.001, Chi-squared test). Taken together, both scRNA-seq and Surface-PromFISH flow cytometry data suggest that a larger proportion of monocytes may present mexRNAs than other cell types in PBMCs.

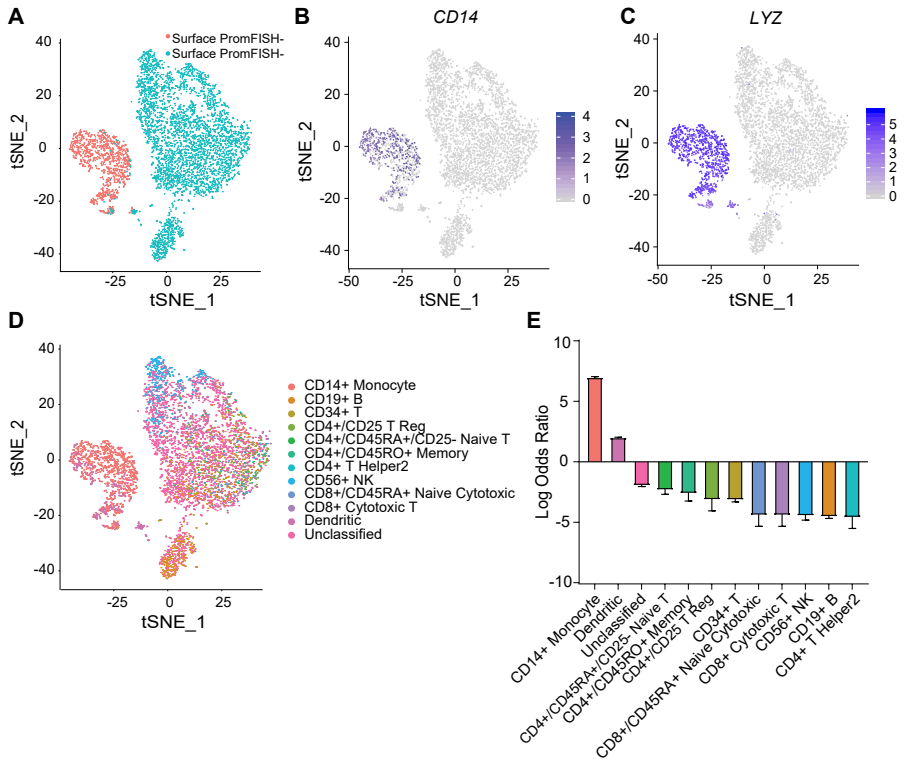

Figure 6.3: Single cell transcriptomes of mexRNA presenting PBMCs. (A-D) tSNE plots of 11,233 single cells resulted from FACS sorting, including 9,043 Surface-PromFISH+ (blue dots) and 2,190 Surface-PromFISH- cells (pink dots) (A), CD14 expressing (dark blue, B) and CD3 expressing cells (dark blue, C), which are classified into 11 cell types (11 colors, D). (E) Association (y axis, log odds ratio ¿ 0: enrichment; log odds ratio ¡ 0, depletion) of Surface-PromFISH+ cells with each cell type (x axis). Error bar: standard error of the log odds ratio.

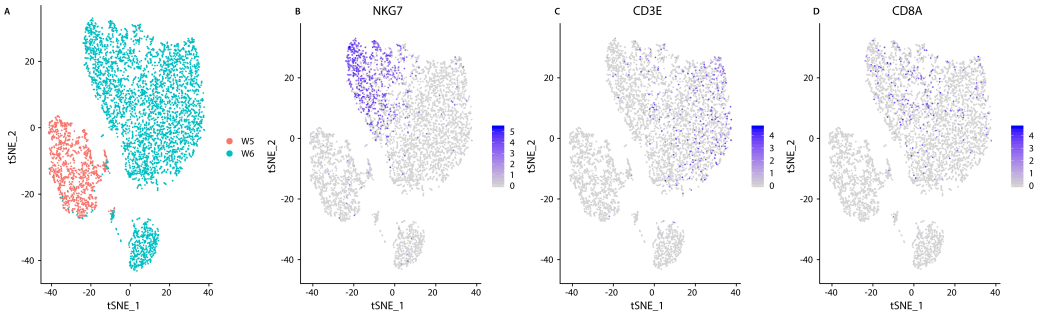

Figure 6.4: tSNE plots of Surface-PromFISH sorted single cells. (A) Surface-PromFISH+ (pink) and Surface-PromFISH- (blue) cells. (B-E) Single cells that express (blue dots) monocyte CD14 (B), CD8 (C), natural killer (NK) cell marker NKG7 (D), and B cell marker MS4A1 (E).

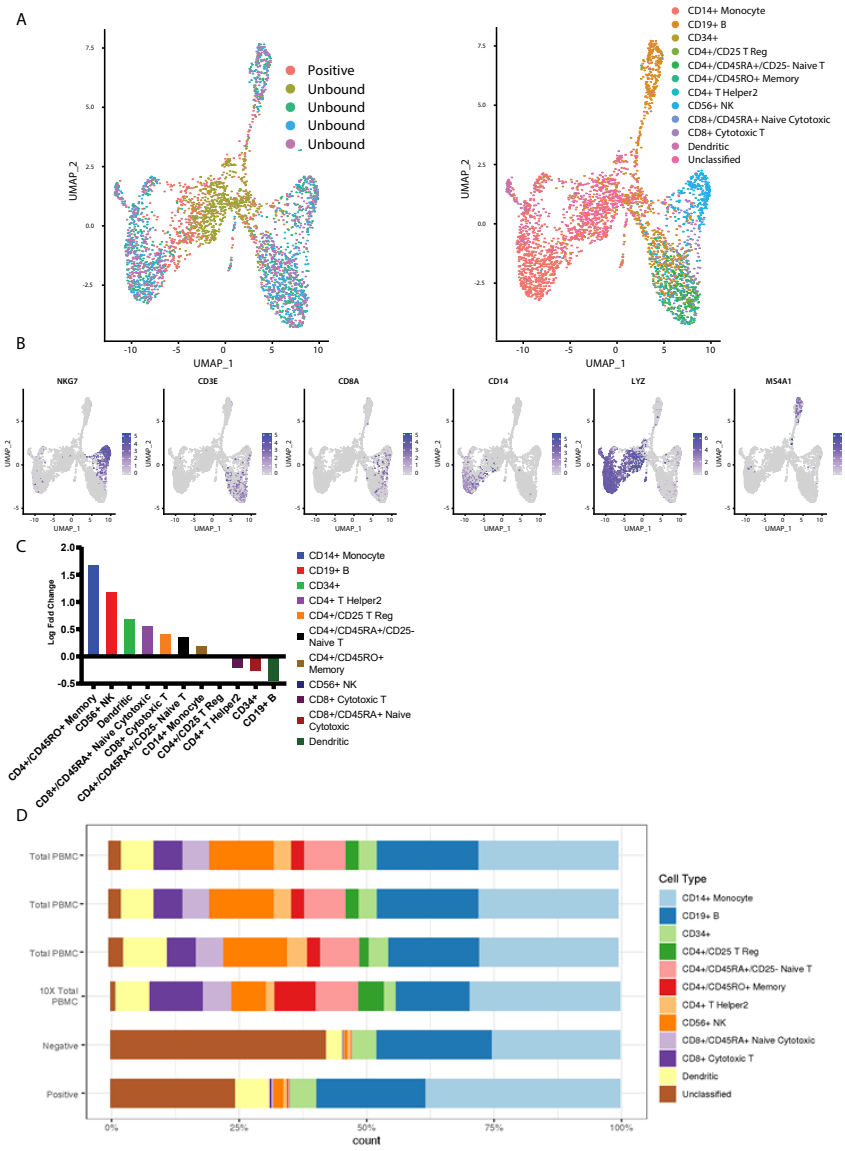

Figure 6.5: Single cell RNA-seq analyis on csRNA presenting PBMCs from biotin pulldown. (A) Biotin pulldown of csRNA positive cells in pink shown more monocytes compared to unbound and total PBMC population. (B) Single cell RNA seq data allows for the classification of cell types based on characteristic marker genes shown in the table for each cell type. Each cell type was clustered together highlighting similar gene expression. (C) Log fold change of biotin pulldown bound vs unbound cells, which show enrichment for monocytes and CD34+ cells. (D) Distribution of cell types in the positive and negative selection using biotin showing enrichment of CD14+ monocytes. Total PBMC from 10x genomics was used as a control to show normal PBMC cell type distributions.

6.2.4 Antisense purification and sequencing of mexRNAs with Surface-FISHseq

To identify mexRNAs presented on the PMBCs, we developed a method called Surface-FISHseq to sequence the Surface-PromFISH identified candidate mexRNAs. The central idea of Surface-FISHseq to purify the probe-RNA hybrids produced by Surface-PromFISH, based on the biotin on the probes, and then subject the purified RNA for sequencing (Figure 6.6A). The advantage of Surface-FISHseq is allowing Surface-PromFISH as a checkpoint to microscopically examine whether the hybridization happened on cell surface as expected.

Figure 6.6: Surface-FISHseq sequencing result from PBMC and functional tests of mexRNAs in human primary monocytes. (A) The Surface-FISHseq experimental workflow. (B) Venn diagram of 3 groups of candidate mexRNAs identified by Surface-FISHseq-membrane (blue), Surface-FISHseq-FACS (green), and Surface-FISHseq-Psoralen (orange). All three groups were analysed by DEseq2, with statistic threshold of FDR ¡ 0.15 and FC ¿ 1 over the control groups. Candidate genes involved in monocyte functions and prioritized for validation were highlighted. (C-F) Read distributions of Surface-FISHseq (red tracks) and control (blue tracks) libraries on FNDC3B (C) and CTSS (D). (D) and (F) demonstrates expanded view of the 3UTR regions of the two genes. Probe track: Locations and IDs of antisense oligonucleotides for functional tests. (G) Effects of Surface-FISH on average monocyte attachment levels (normalized to the no probe control (black)) (y axis) by dArt4 control (red) or antisense probe-sets against 11 Surface-FISHseq identified target transcripts (grey columns). Each probe-set is comprised of 25 x 20 nt antisense oligonucleotide probes. Error bar: standard error of the mean. 8 replicate experiments were performed using monocytes extracted from two independent donors. ***: multiple hypothesis adjusted p ¡ 0.0001, ****: multiple hypothesis adjusted p ¡ 0.00001, Kruskal-Wallis test. (H-I) Effects of individual antisense probes to the average monocyte attachment levels (y axis) in no probe control (black), the control probe against dArt4 (red), and probes indexed by E1-E4, U1-U5 (grey columns, corresponding to the Probe track in panel C) against specific regions of the FNDC3B transcript (F) and by E1, E2, U1-U9 (corresponding to the Probe track in panel D) against specific regions of the CTSS transcript (G). **: multiple hypothesis adjusted p ¡ 0.001, ***: multiple hypothesis adjusted p ¡ 0.0001, Kruskal-Wallis test.

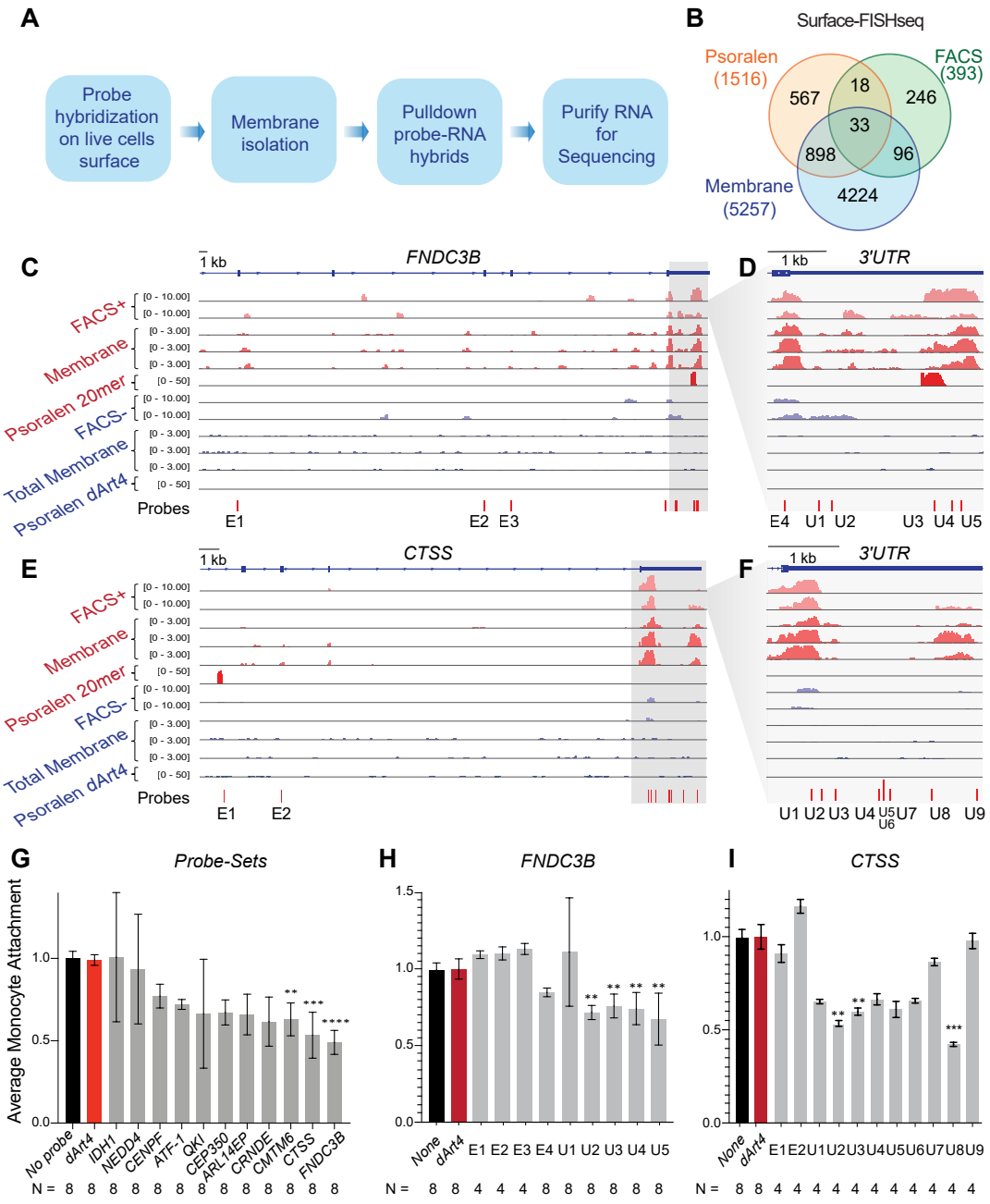
A
Probe hybridization on live cells surface
Membrane isolation
Pulldown probe-RNA hybrids
Purify RNA for Sequencing
B
Surface-FISHseq
Psoralen (1516)
FACS (393)
Membrane (5257)
567
18
246
33
898
96
4224
C
FNDC3B
1 kb
FACS+
Membrane
Psoralen 20mer
FACS-
Total Membrane
Psoralen dArt4
Probes
E1
E2
E3
D
3'UTR
E4
U1
U2
U3
U4
U5
E
CTSS
F
U6
U7
U8
U9
G
Probe-Sets
Average Monocyte Attachment
No probe
dArt4
IDH1
NEDD4
CENPF
ATF-1
QKI
CEP350
ARL14EP
CRNDE
CMTM6
CTSS
FNDC3B
H
FNDC3B
None
I
CTSS

Table 6.1: Data summary for FISHseq. # UMR: number of uniquely mapped reads. UMP: percentage of uniquely mapped reads.

Approaches	Treatment	Group	ID	Donor ID	Total # Reads	# UMR	UMP
Psoralen cross-linking & membrane purification	Probe library pulldown	csRNA	PP_1	9799	5457193	4697546	86.08%
	non-crosslinked library pulldown	Control	PN_C1	9799	234808	95467	40.66%
	non-crosslinked dArt4 pulldown	Control	PN_C2	9799	5180356	4606128	88.92%
	Probe dArt4 pulldown	Control	PN_1	9799	2377320	1909222	80.31%
FACS Pulldown	FACS+ pulldown	csRNA	FP_1	575	2759220	2301510	83.41%
	FACS+ pulldown	csRNA	FP_2	575	2723909	2327848	85.46%
	FACS- pulldown	Control	FN_1	561	2797693	2325819	83.13%
	FACS- pulldown	Control	FN_2	561	2253317	1865421	82.79%
Membrane Purification Pulldown	Mem pulldown	csRNA	MP_1	561	2811824	2099391	74.66%
	Mem pulldown	csRNA	MP_2	575	2850086	2325989	81.61%
	Mem pulldown	csRNA	MP_3	730	3135792	2443210	77.91%
	Total membrane	Control	MT_1	561	5381193	4420311	82.14%
	Total membrane	Control	MT_2	575	4233565	3497718	82.62%
	Total membrane	Control	MT_3	730	3369136	2663567	79.06%

Recognizing that even if mexRNAs exist, their relative amounts are significantly fewer than intracellular RNAs, we reasoned that a successful mexRNA purification procedure will have to highly selective. Thus, we carried out 3 Surface-FISHseq experiments, each with an additional selection step after probe hybridization and wash to remove non-mexRNA presenting cells or intracellular continents.

In the first experiment, we purified cell membrane before pulldown of the probe-RNA hybrids (Surface-FISHseq-membrane). We generated 3 Surface-FISHseq-membrane libraries from 3 PBMC samples of 3 donors (Table 6.1) (Membrane purification tracks in red, Figure 6.6C,D). In addition, we generate 3 control libraries from cell membrane purification followed by RNA sequencing, based on the same PBMC samples (Membrane purification control tracks in blue, Figure 6.6C,D). A comparison of Surface-FISHseq-membrane libraries vs. control libraries resulted in 2,071 RNAs (DEseq2).

In the second experiment, we used FACS to collect Surface-PromFISH+ cells, followed by mexRNA purification and sequencing (Surface-FISHseq-FACS). We generated 2 Surface-FISHseq-FACS libraries from 2 PMBC samples of 2 donors (FACS tracks in red, Figure 6.6C,D). In addition, we generated 2 control libraries using a dART probe from the same 2 PBMC samples (FACS control tracks in blue, Figure 6.6C,D). A comparison of the experimental and the control libraries resulted in 374 RNAs (p-value ¡ 0.05, log2FC ¿ 1, DEseq2).

In the third experiment, we used psoralen to cross-link the hybridized probes and RNAs, purified the cross-linked probe-RNA hybrids and subjected the co-purified RNA to RNA sequencing (Surface-FISHseq-psoralen) (Figure S-poralenSeq). Psoralen only cross-link hybridized nucleic acids and do not cross-link nucleotides with proteins, which allows subsequent stringent washes to remove indirect interactions or promiscuously attached molecules. We generated 4 libraries, including 1 experiment library and 3 control libraries. The first control library is identical to the experiment library except that the psoralen cross-linking step was not performed (Psoralen track in red, Figure 6.6C,D). The second the third control libraries replaced the probes

with a 20 nt probe against dART4 and were carried out with and without psoralen cross-linking, respectively (Psoralen control tracks in blue, Figure 6.6C,D). A comparison of the experimental and the control libraries resulted in 1,516 RNAs (p-value ¡ 0.05, log2FC ¿ 1, DEseq2).

Pairwise comparisons of the three experiments revealed significant overlaps of the genes detected from each experiment. A total of 509 genes were detected by at least 2 out of the 3 Surface-FISHseq experiments, in which 51 genes were identified by both Surface-FISHseq-psoralen and Surface-FISHseq-FACS (Figure 6.6B). These putative mexRNAs became our candidates for testing whether their impact any cellular functions.

6.2.5 Monocyte-EC attachment levels are impaired by blocking specific mexRNAs

Considering that a larger fraction of monocytes likely present mexRNAs as compared to other cell types in PBMCs, we chose monocytes to evaluate if mexRNAs may impact cellular functions. We chose monocyte-endothelial cell (EC) attachment levels as a functional readout to monocytes cellular functions. We collected CD14+ cells from PBMCs by FACS sorting for this test.

We selected 9 genes from the 51 genes in the overlap of Surface-FISHseq-psoralen and Surface-FISHseq-FACS identified mexRNAs to this functional test. These genes are IDH1, ATF1, NEDD4, CENPF, CMTM4, CEP350, ARL14EP, FNDC3B, CTSS (Figure 6.6B). For each gene, we designed a probe set comprised of twenty-five 20nt antisense oligonucleotide probes to target the parts of the transcripts with Surface-FISHseq read coverage (experiment probset). A dArt4 probe set comprised of twenty-five 20nt antisense oligonucleotide probes to dArt4 was used as control (control probe set). We incubated the each probe set with monocytes for 45 minutes before putting the monocytes on confluent HUVEC cells on a plate reader. The degree of monocyte-EC attachment was measured by a normalized fluorescent intensity reflecting the number of attached monocytes.

We carried out 8 replicate experiments. The control probe set did not affect monocyte attachment, neither did the probe sets for IDH1 and NEDD4 (Figure 6.6E). However, the probesets targeting CTSS and FNDC3B reduced monocyte attachment levels (Figure 6.6E), suggesting monocytes capability to attach to ECs are modulated by some mexRNAs.

6.2.6 Specific parts of mexRNAs modulate Monocyte-EC attachment levels

To test if a specific region of a transcript is responsible for the reduced monocyte attachment levels, we repeated the above test with individual probes. We tested 9 probes from the probeset for FNDC3B based on Surface-FISHseq read coverage, including 4 probes targeting 4 exons (Exon 22, 24 -26) and 5 probes targeting the 3UTR (probe track, Figure 6.6C: redraw probe track at the bottom). As controls, we used no probe, a 20nt probe against dArt4. Every probe was tested in 4 to 8 replicate wells. As expected, the dArt4 probe did not change the monocyte attachment levels (red and black bars, Figure 6.6F). Five out of the 9 probes did not affect the attachment levels as well, suggesting that not all parts of the FNDC3B transcripts were responsible for attachment (E1, E2, E3, E4, U1, Figure 6.6F). However, 4 probes reduced the attachment levels, and all the 4 probes targeted the 3 portion of the 3 UTR (U2, U3, U4, U5, Figure 6.6F), consistent with a Surface-FISHseq peak to the 3 end of the 3 UTR. Moreover, U5 that target the most 3 end of the UTR exhibited the largest reduction.

We tested 11 probes from the probe set for CTSS based on Surface-FISHseq read coverage, including 1 probe targeting Exon 5, 1 probe targeting the 4th intron where one Surface-FISHseq exhibited coverage, and 9 probes spanning the 3 UTR where Surface-FISHseq formed two peaks (probe track, Figure 6.6D: redraw probe track at the bottom). As controls, we used no probe, a 20nt probe against dArt4. As expected, the dArt4 probe did not change the monocyte attachment levels as compared to the no probe (red and black bars, Figure 6.6G). Neither the intronic probe (E1) nor exon probe (E2) changed the attachment levels. Three (U2, U3, and U8) out of the 9

probes targeting 3 UTR resulted in reduced attachment levels (Figure 6.6G). U2 and U3 coincided with the first Surface-FISHseq peak. U8 was near the second Surface-FISHseq peak (probe track, Figure 6.6D). Ideally, we would like to design another probe near the center of the second peak. Unfortunately, the relatively degenerative sequence in this peak does not allow for 20 mer probes (our 75nt Surface-FISHseq reads could be uniquely aligned to this region). Taken together, specific regions of the FNDC3B and the CTSS transcripts, when blocked by extracellular oligonucleotides, affected monocyte-EC attachment levels.

6.3 Discussion

Employing de novo RNA metabolic labeling, membrane-coated nanoparticle and in vivo surface hybridization techniques on cell lines and primary PBMCs, we revealed a group of genome encoded transcripts on cell surface. These cell surface RNAs may constitute a novel class of functional substance of the plasma membrane. There has been precedence of cell membrane incorporated RNA in other forms of life. The highly conserved bacterial OLE non-coding RNA family anchors to the membrane by interacting with OAP transmembrane proteins through conserved loop structures[11]. Recently, RNAs modified with sialylated glycans (glycoRNAs) were found in or outside of ER/membrane systems in human lymphoblastoid, leukemia, embryonic stem cells, as well as in mouse liver and spleen[33]. Glycan are an important class of surface regulators for cell-cell and host-pathogen interactions. Therefore, the presence of glycan on RNA provides a strong indication that RNA may be involved in the regulation of cell-cell and host-pathogen interactions. Interestingly, the cell line K562 was demonstrated to present glycoRNA, which was also demonstrated here to present mexRNA[33], however, enrichment and sequencing of glycoRNAs manly identified small RNAs but not mRNAs based on the reported data. These experiments cannot exclude the possibility glycosylation on larger RNA such as mexRNA species. Glycosylated RNA has distinct biophysical potential and lower electro-mobility, therefore, the

size selection by gel electrophoresis used in the experiment may preclude observation of mRNAs. Despite this pre-selection, we still find that there might be a potential glycosylation site in the transcript of FNDC3B. Collectively, these observations imply an active role of hgRNA in the interaction with the plasma membrane, as opposed to the interaction seen between viral RNAs with the lipid bilayers[139, 119].

The mexRNAs exhibited cell type and subpopulation specificity, indicating potential roles of tissue- or cell type-specific functions. We detected mexRNAs on mouse but not human T-lymphoblasts, with enrichment in the CD44hiCD4- cell population. We also observed an enrichment of mexRNAs in CD14+ monocytes. Through visualization with imaging flow cytometry, the mexRNAs colocalized with the CD14 surface protein. A small population of dendritic cells also displayed mexRNA signals, yet for the rest of PBMCs such as B cells ,T cells and natural killer cells, the surface RNA presentation was not statistically significantly above backgrounds. These data suggest that mexRNA is not an artifact of tissue culture and occurs specifically in multiple cultured and primary cell types. Moreover, the surface RNA detection is not due to their higher RNA expression level, since most of the mexRNA candidates examined were not expressed at higher quantity in mexRNA+ versus mexRNA- population in single cell transcriptome, nor are they monocyte-specific transcripts. We speculate that the specificity of mexRNA is regulated at the stage of RNA localization or transportation rather than transcript expression.

Gene Ontology analysis of the mexRNA populations in human monocyte highlights enriched functionality in protein targeting to membranes RNA processing and virus gene expression. A major caveat of applying Gene Ontology analysis on potential mexRNAs is that the gene annotation is based on the functional characterized of proteins. Recent years have seen a surge in functional discoveries of non-coding RNAs or RNAs with both coding and non-coding functions[53]. The non-coding transcripts are responsible for a plethora of cellular roles such as scaffolds, molecular decoys, enzymes, and network regulators[140], that were not well annotated and therefore cannot be retrieved from Gene Ontology analysis. Based on the unique plasma

membrane localization and monocytic enrichment of mexRNAs, it is tempting to speculate that these mRNAs may perform novel monocyte-specific functions in addition to protein-coding.

We tested the possibility of mexRNAs in facilitating monocyte-endothelial attachment, and found that three out of eleven candidates may positively regulate monocyte adhesion. In particular, CTSS and FNDC3B are specifically expressed in monocytes compared to other immune cell types at bulk Human Protein Atlas and single cell RNA-seq levels. Furthermore, these two candidate mexRNAs have reported roles in cell attachment. Cell surface fibronectin mediates the adhesion of many cell types. Specifically FNDC3B, encoding for a fibronectin type III domaincontaining protein 3B, is induced by integrinmediated signaling during monocyte-endothelial adhesion[123]. Similarly, CTSS Cathepins S cyctein protease shape the ECM by collagen degradation. Cathepins regulate communications of cell with the extracellular environment and secreted CTSS is a biomarker for acute inflammation in macrophages[56]. On the other hand, hybridization of primary monocytes with antisense probes against the transcripts of FNDC3B and CTSS caused reduction in attachment, demonstrating gene functionality also at the transcript level. Our result and previous studies on CTSS and FNDC3B suggest that candidate mexRNA may have consensus function with their proteins. Interestingly, the 3UTR of CTSS and FNDC3B transcripts, that had higher signal in the surface FISH-seq datasets, appeared to be more critical for monocyte attachment. FNDC3B and CTSS proteins were mainly found in ER or vesicles that are part of the secretory pathway in cell lines examined Human Protein Atlas. It is plausible that the mexRNAs guide these proteins to the cell surface by attracting signal recognition particles, as has been reported for CD47 3UTR , whose length determines CD47 protein localization to membrane or ER[8]. Under a comparable mechanism, the mexRNA 3UTR may regulate monocyte attachment by tethering the cell adhesion protein to the cell membrane.

Another possible role of the mexRNA is to rapidly adjust local membrane permeability, analogous to lipid pore forming proteins such as FGF2 and gasdermin on surface of macrophages during inflammation[100]. In vitro selection and in vivo examination demonstrated that selected

RNA oligos, typically enriched in poly-G stretches, could bind specifically to liposome or mammalian cell membranes, and penetrate the lipid bilayers with a preference for irregular structures on the membrane[64, 58, 57]. The selected RNAs increase membrane permeability to ions and even larger polar molecules, and may stabilize temporary hydrate pore formation11. In fact, RNAs with binding domains for lipids and tryptophan alone could form passive transport channels that enable tryptophan to pass across diacyl phospholipid membranes[59]. Motif analysis of mexRNAs discovered high frequencies of GC-rich motifs that contained triple-G strings. Further examination of regions of CTSS and FNDC3B potentially responsible for cell attachments were also more conserved than the surrounding regions and exhibited high loop-forming potential. These sequence and structural features supported their interaction with the cell membrane through lipids and potential alteration of the membrane permeability. The 3UTR of CTSS and FNDC3B may therefore facilitate the secretion of the proteins, as part of the full length mRNA or on its own, through the unconventional secretion pathway. Notably, the FNDC3B protein may possess capability to bind RNAs, further potentiating its possible interaction with the FNDC3B transcript[18].

We do not know the mechanism by which mexRNA traffics to the cell surface. Transportation of RNA molecules through the conventional membrane transportation pathway, by firstly entering the ER lumen was previously considered unlikely because of the high polarity of the phosphodiester backbone of RNA molecules. However, the recent discovery of small glycoRNA in membranous organelle[33] would challenge the previous view. This entrance could be facilitated by RNA transportation machineries such as human homologues of the C. elegans SID1&2 transporter that imports dsRNA for RNA interference[128, 96, 1].

The current framework in which membrane structure is understood excludes RNA as an integral component of the membrane. Our discovery of mexRNA suggest that this is an incomplete view and highlights mexRNA at the interaction interphase between the cells and their external environment, where it influence on the permeability, structure and signaling of

biomembrane systems, and thereby participate more broadly in cellular life than previously envisioned.

6.4 Methods

6.4.1 Cell line cuture

Lymphoblast K562, B cell Daudi, B cell GA-10, B cell Raji, T cell Jurkat, and monocyte THP-1 Cells were obtained from ATCC and cultured based on their recommended media and passage guidelines.

6.4.2 Isolation of PBMC from healthy donors

10 ml of the blood samples obtained from the San Diego Blood Bank were transferred to CPT vacutainer tubes. If you don't have a balance tube, use an empty CPT tube by adding/removing PBS. Make sure that they are balance to avoid the tubes to broke. Mix the blood samples immediately prior to centrifugation by inverting the tube 8-10 times. Centrifuge blood samples at room temperature for 25 minutes at 1800g. In the tissue culture hood, carefully open spun CPT tube and carefully discard the serum. Gently pipet up and down around the top and sides of the gel plug to dislodge and PBMCs that are stuck to it without touching the plug. Transfer to new labeled 15 ml falcon tube and centrifuge at 1400 rpm for 5 minutes. Wash PBMCs 2x with PBS or cell media and culture cells for 30 min in suspension flasks prior to labelling.

6.4.3 Surface-Prom-FISH analysis

Antisense library design

Random 20mer libraries were generated with 20 random N bases with Alexafluor 647 conjugated to the 3 end from IDT. Negative control probe design. Random 6mer libraries were

generated with 6 random N bases with Alexafluor 647 conjugated to the 3 end from IDT. dArt4 probe /5Alex647N/TTAATCATAATCGTATTGGG was generated from IDT with Alexafluor 647 conjugated to the 3 end from IDT.

Antisense Probe Hybridization

Denature oligo libraries at 70 C, 5 minutes and keep at 37 C before hybridization. Incubate 1 uL of 100 uM DNA oligonucleotide library per 10 million cells per 10ml for 1 hour at 37C in cell media. Wash twice with 1x PBS, spin down for suspension cells. Stain with recommended amount of live/dead yellow for 30 minutes. Wash twice with 1x PBS.

Imaging and analysis

For imaging with microscopy, cells were imaged in 1x PBS through wide field fluorescence imaging using an Olympus IX83 inverted microscope at 60x oil immersion objective (NA=1.4). For imaging with imaging flow cytometry, cells were imaged using Amnis ImageStreamX Mk II imaging flow cytometer with 60X magnification. Compensation was done using ultracomp beads with the appropriate fluorophores. Image analysis was done using the Amnis IDEAS Software where cells were gated for live/dead using a dye and single focused cells using the areas and gradient RMS separately. PBMCs were labeled and gated for CD14 (monocytes), CD3 (T cells), and CD19 (B cells).

Single cell RNA sequencing and analysis

PBMCs were isolated either through FACS or biotin pulldown for csRNA positive and negative cells. We used 10x genomics to generate single cell RNA-seq libraries, which were sequenced using the Illumina HiSeq4000. Single cell clustering and visualization was done using Seurat 3 and classification of cells were done using singleCellNet. The training set was generated using Zheng 68K PBMC from the 10x database.

6.4.4 Surface-FISHseq Analysis

Antisense Probe Hybridization

Denature oligo libraries at 70 C, 5 minutes and keep at 37 C before hybridization. Incubate cells or PBMCs with 1 uL of 100 uM DNA oligonucleotide library per 10 million cells/ 10ml for 1 hour @ 37C in cell media. Collect the cells and wash in PBS three times by centrifuging at 500g for 5min.

Psoralen Cross-linking

Prepare 0.5 mg/mL AMT solution in PBS: Resuspend AMT in H2O at 1 mg/mL and then add an equal volume of 2 PBS. Chill solution in the dark on ice. Resuspend cell pellet (250 million cells) in 5 mL of ice-cold AMT solution. Incubate the cells on ice for 15 minutes. Transfer samples to a pre-chilled 10-cm tissue culture dish. Place cells on ice under a long-wave UV bulb (350 nm) in a UV Stratalinker 2400 (Stratagene). Cells should be approximately 3-4 cm away from the light source. (6500 uW/cm2). Expose cells to UV light at maximum power for 10 minutes. Mix every 2.5 minutes. Transfer irradiated cells to cold tubes and spin at 330 g for 4 minutes. Discard supernatant.

Membrane Isolation

Resuspend the cells in 1ml hypotonic lysis buffer (10mM Tris-HCl pH=7.5, 10Mm KCl, 2mM MgCl2, 33L protease inhibitor cocktail (Sigma) and 10L RNAsin per 10mL of solution), incubate on ice for 15 min. Disrupt the cells using a 2 ml dounce homogenizer with a tight-fitting pestle (Size B) for 20 passes. Transfer the entire solution into an ultracentrifuge tube, centrifuge at 20,000g for 20min. Transfer the supernatant to another ultracentrifuge tube, centrifuge at 100,000g for 35min. Discard the supernatant, resuspend the pellet with storage buffer (10mM Tris-HCl pH-7.5, 1mM EDTA, and 10L RNAsin per 10mL of solution). Centrifuge at 100,000g

for 35min. Discard the supernatant, resuspend the pellet in storage buffer. Quantify membrane protein concentration by RNA qubit.

csRNA Streptavidin Bead Pulldown

Streptavidin C1 beads were subjected to RNase removal treatment, blocked in BSA/PBST @ 4 degrees until ready. The beads were washed in B&W buffer and incubated with 500 ul membrane extracts per 100 ul beads. Incubation @ 4 degrees for 1 hour and RT for 30 mins. Wash 3x each with urea low salt wash buffer (8M Urea, 10mM Tris-HCl pH 7.5, 1mM EDTA, 0.10% NP-40, 150 mM NaCl, 0.50% SDS) and high salt wash buffer (10 mM Tris-HCl pH 7.5, 4 M NaCl, 1 mM EDTA, 0.2% Tween 20). Wash 1x with PNK wash buffer (20 mM Tris-HCl pH 7.5, 10 mM MgCl2). Reverse Cross-linking by UV in plate at 265 nm in a UV Stratalinker 2400 (Stratagene).

csRNA Library generation and sequencing

Libraries were generated using the NEB Low Input/Single Cell library prep kit. Libraries were sequenced using Illumina Miniseq with High Output cartridge.

Sequencing analysis

All the RNA-seq data were mapped to the human genome (hg38) using STAR (v2.5.4b) with default parameters. DeSeq2 was used for differential gene expression analysis to identify significantly expressed csRNA.

6.4.5 Monocyte Attachment Assay

Monocyte Isolation

CD14+ monocytes were isolated directly from whole blood using Miltenyi Biotec Straight-From LRSC CD14 MicroBead Kit.

Preparation of HUVEC monolayer

Add 100 L of the Gelatin to each well of a 96-well tissue culture treated plate. Incubate for 60 min at 37C in a cell culture incubator. Wash twice with sterile 1X PBS. Aspirate the final wash before use. Add 50,000-100,000 HUVEC cells/well to the Gelatin-coated 96-well plate. Culture cells for 48-72 until the endothelial cells form a monolayer.

Antisense probe blocking and cell staining

Denature oligo target libraries at 70 C, 5 minutes and keep at 37 C before hybridization. Incubate CD14+ monocytes with 1 uL of 100 uM DNA oligonucleotide library per 10 million cells/ 10ml for 1 hour @ 37C in cell media. Collect the cells and wash in PBS three times by centrifuging at 500g for 5min. Incubate with 1 uM Calcein AM for 30 minutes @ 37C in cell media. Wash twice with 1x PBS.

Monocyte attachment

Resuspend the cell pellet at 0.25 - 1.0 x 106 cells/ml in serum free media. Aspirate endothelial culture media and wash once with serum free media. Add 200 L of the cell suspension to each well already containing the endothelial monolayer. Incubate for 30-90 min in a cell culture incubator. Carefully discard or aspirate the media from each well. Gently wash each well 3 times with 250 L 1X PBS.

Monocyte attachment quantification

Cell attachment was quantified using the SpectraMax i3x Multi-Mode Microplate Reader with 132 points per well to measure fluorescence output of Calcein AM.

6.5 Acknowledgments

Chapter 6 in part is currently being prepared for submission for publication of the material. Norman Huang, Xiaochen Fan, Kathia Zaleta Rivera, Tri C. Nguyen, Jiarong Zhou, Yingjun Luo, Jie Gao, Ronnie H. Fang, Zhen Bouman Chen, Liangfang Zhang, Sheng Zhong Naturally occurring cell surface-display of genome encoded RNAs and their impacts on cell-cell interactions. The dissertation author was the primary investigator and author of this paper.

Chapter 7

Final notes

Among the vast amount of nanoscale molecules, semiconductor nanocrystals known as quantum dots are unique for their potentials in many different applications from optical to biosensors. Quantum dots have specific optical properties, high fluorescence yields, and excellent optical stability against photo bleaching making them excellent candidates for detection in biological imaging. In this thesis we utilized these qualities to improve the sensitivity, robustness and cost for the detection of RNA transcripts. Here we were able to present that quantum dot RNA FISH is able to expand the capabilities of RNA detection to offer a versatile tool set to visualize more subtle genetic differences presented in isoform and fusion transcripts. An important finding was the novel detection of Eml4-Alk fusion transcript in NSCLC tissue without the previously known associated gene translocation. This unique fusion transcript event is likely due to a less known process by RNA splicing errors, where two separate transcripts were spliced together. A theoretical gap in the splicing error model is that transsplicing can happen only to two RNA molecules that are close to each other in 3D space; however, except for neighboring genes, the chances for two RNA molecules transcribed from distant chromosomal locations to meet in space are small. Therefore, it remains difficult to perceive a biophysical process in which fusion transcripts are created by splicing errors. Due to our finding we proposed a RNA-poise model

which fills this theoretical gap. Remote RNADNA interactions could be created by at least two means. First, the caRNA can target specific genomic sequences, which could be mediated by tethering molecules (RNA targeting, Fig. 7.1). Second, the spatial proximity of the genomic sequences in 3D space could bring the nascent transcripts of one gene to the genomic sequence of another gene (RNA confinement, Fig. 7.1). Both means of RNADNA interactions provide spatial proximity between two RNA molecules and thus allow for splicing errors. In addition, the spatial proximity of two genes in the RNA confinement model could enhance the chances of genome rearrangement of the spatially close genomic sequences and thus create fusion genes.

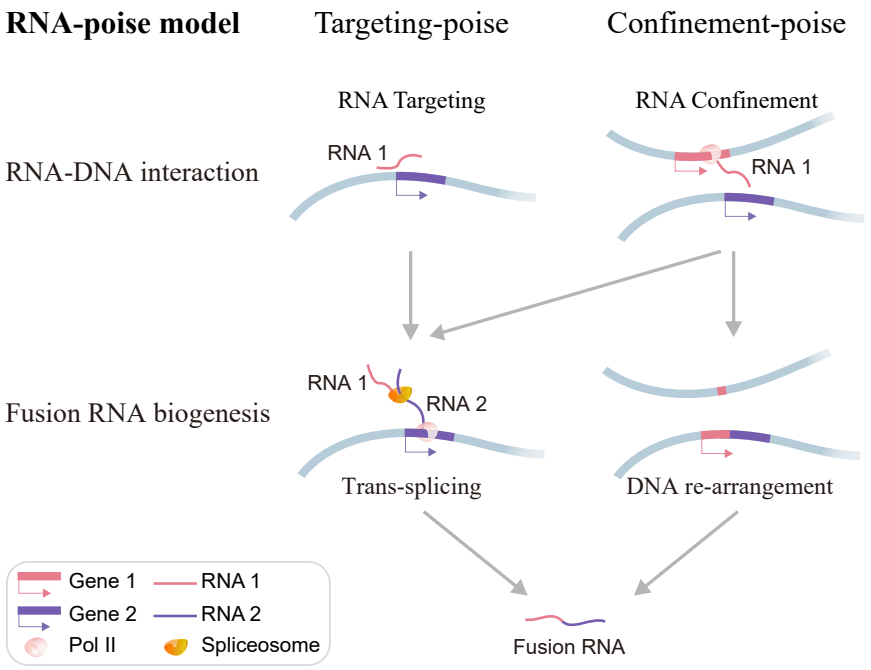

Figure 7.1: RNA-poise model. In this model, the transcripts of one gene (RNA 1, red bar) can exhibit spatial proximity to another gene (RNA 2, purple bar) due to tethering (RNA targeting) or spatial proximity of the two genes (RNA confinement). Both cases could enhance splicing errors (grey arrows), whereas the proximity of genomic sequences may also facilitate gene fusion (grey arrow on the right), which subsequently produces fusion RNA.

Another discovery was the presence of RNAs on the cell surface. Detection of cell surface specific markers are a cornerstone of detection technologies that is used in research and clinical diagnostics. Previous literature has shown detection of CD markers using antibodies

as indicators of cell type and cancer. More recent work demonstrated the ability of specific binding of oligonucleotides to a variety of ligands, such as proteins and other small molecules. Here we were able to demonstrate both detection cell lines and PBMCs. Here we propose that csRNAs are present and may have functional implications in monocyte attachment. We discovered that antisense oligos targeting the 3' UTR of FNDC3B and CTSS diminish the attachment of monocytes from healthy donors.

Overall, we were able to demonstrate quantum dots as a useful tool in for the detection of RNA transcripts in many situations, which led to novel discoveries. This can be expanded towards different applications for detection and monitoring even beyond the detection of RNA.

References

[1] Shu Aizawa, Yuuki Fujiwara, Viorica Raluca Contu, Katsunori Hase, Masayuki Takahashi, Hisae Kikuchi, Chihana Kabuta, Keiji Wada, and Tomohiro Kabuta. Lysosomal putative RNA transporter SIDT2 mediates direct uptake of RNA by lysosomes. *Autophagy*, 12(3):565–578, 2016.

[2] W. R. Algar and U. J. Krull. Adsorption and hybridization of oligonucleotides on mercaptoacetic acid-capped cdse/zns quantum dots and quantum dot-oligonucleotide conjugates. *Langmuir*, 22(26):11346–11352, 2006.

[3] W. Russ Algar and Ulrich J. Krull. Adsorption and hybridization of oligonucleotides on mercaptoacetic acid-capped CdSe/ZnS quantum dots and quantum dot-oligonucleotide conjugates. *Langmuir*, 22(26):11346–11352, 2006.

[4] A. Arzalluz-Luque and A. Conesa. Single-cell rnaseq for the study of isoforms-how is that possible? *Genome Biol*, 19(1):110, 2018.

[5] Gábor Balázsi, Alexander Van Oudenaarden, and James J. Collins. Cellular decision making and biological noise: From microbes to mammals. *Cell*, 144(6):910–925, 2011.

[6] Brian J. Beliveau, Alistair N. Boettiger, Maier S. Avendaño, Ralf Jungmann, Ruth B. McCole, Eric F. Joyce, Caroline Kim-Kiselak, Frédéric Bantignies, Chamith Y. Fonseka, Jelena Erceg, Mohammed A. Hannan, Hien G. Hoang, David Colognori, Jeannie T. Lee, William M. Shih, Peng Yin, Xiaowei Zhuang, and Chao Ting Wu. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature Communications*, 6(May), 2015.

[7] A. J. Bergsma, E. van der Wal, M. Broeders, A. T. van der Ploeg, and W. W. M. Pim Pijnappel. Alternative splicing in genetic diseases: Improved diagnosis and novel treatment options. *Int Rev Cell Mol Biol*, 335:85–141, 2018.

[8] Binyamin D Berkovits and Christine Mayr. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, 522(7556):363–367, jun 2015.

[9] Julien Béthune, Ralf-Peter Jansen, Michael Feldbrügge, and Kathi Zarnack. Membrane-Associated RNA-Binding Proteins Orchestrate Organelle-Coupled Translation. *Trends in Cell Biology*, 29(2):178–188, 2019.

[10] Ewan Birney, John A. Stamatoyannopoulos, Anindya Dutta, Roderic Guigó, Thomas R. Gingeras, Elliott H. Margulies, Zhiping Weng, Michael Snyder, Emmanouil T. Dermitzakis, Robert E. Thurman, Michael S. Kuehn, Christopher M. Taylor, Shane Neph, Christoph M. Koch, Saurabh Asthana, Ankit Malhotra, Ivan Adzhubei, Jason A. Greenbaum, Robert M. Andrews, Paul Flicek, Patrick J. Boyle, Hua Cao, Nigel P. Carter, Gayle K. Clelland, Sean Davis, Nathan Day, Pawandeep Dhami, Shane C. Dillon, Michael O. Dorschner, Heike Fiegler, Paul G. Giresi, Jeff Goldy, Michael Hawrylycz, Andrew Haydock, Richard Humbert, Keith D. James, Brett E. Johnson, Ericka M. Johnson, Tristan T. Frum, Elizabeth R. Rosenzweig, Neerja Karnani, Kirsten Lee, Gregory C. Lefebvre, Patrick A. Navas, Fidencio Neri, Stephen C.J. Parker, Peter J. Sabo, Richard Sandstrom, Anthony Shafer, David Vetrie, Molly Weaver, Sarah Wilcox, Man Yu, Francis S. Collins, Job Dekker, Jason D. Lieb, Thomas D. Tullius, Gregory E. Crawford, Shamil Sunyaev, William S. Noble, Ian Dunham, France Denoeud, Alexandre Reymond, Philipp Kapranov, Joel Rozowsky, Deyou Zheng, Robert Castelo, Adam Frankish, Jennifer Harrow, Srinka Ghosh, Albin Sandelin, Ivo L. Hofacker, Robert Baertsch, Damian Keefe, Sujit Dike, Jill Cheng, Heather A. Hirsch, Edward A. Sekinger, Julien Lagarde, Josep F. Abril, Atif Shahab, Christoph Flamm, Claudia Fried, Jörg Hackermüller, Jana Hertel, Manja Lindemeyer, Kristin Missal, Andrea Tanzer, Stefan Washietl, Jan Korbel, Olof Emanuelsson, Jakob S. Pedersen, Nancy Holroyd, Ruth Taylor, David Swarbreck, Nicholas Matthews, Mark C. Dickson, Daryl J. Thomas, Matthew T. Weirauch, James Gilbert, Jorg Drenkow, Ian Bell, Xiaodong Zhao, K. G. Srinivasan, Wing Kin Sung, Hong Sain Ooi, Kuo Ping Chiu, Sylvain Foissac, Tyler Alioto, Michael Brent, Lior Pachter, Michael L. Tress, Alfonso Valencia, Siew Woh Choo, Chiou Yu Choo, Catherine Ucla, Caroline Manzano, Carine Wyss, Evelyn Cheung, Taane G. Clark, James B. Brown, Madhavan Ganesh, Sandeep Patel, Hari Tammana, Jacqueline Chrast, Charlotte N. Henrichsen, Chikatoshi Kai, Jun Kawai, Ugrappa Nagalakshmi, Jiaqian Wu, Zheng Lian, Jin Lian, Peter Newburger, Xueqing Zhang, Peter Bickel, John S. Mattick, Piero Carninci, Yoshihide Hayashizaki, Sherman Weissman, Tim Hubbard, Richard M. Myers, Jane Rogers, Peter F. Stadler, Todd M. Lowe, Chia Lin Wei, Yijun Ruan, Kevin Struhl, Mark Gerstein, Stylianos E. Antonarakis, Yutao Fu, Eric D. Green, Ula Karaöz, Adam Siepel, James Taylor, Laura A. Liefer, Kris A. Wetterstrand, Peter J. Good, Elise A. Feingold, Mark S. Guyer, Gregory M. Cooper, George Asimenos, Colin N. Dewey, Minmei Hou, Sergey Nikolaev, Juan I. Montoya-Burgos, Ari Löytynoja, Simon Whelan, Fabio Pardi, Tim Massingham, Haiyan Huang, Nancy R. Zhang, Ian Holmes, James C. Mullikin, Abel Ureta-Vidal, Benedict Paten, Michael Seringhaus, Deanna Church, Kate Rosenbloom, W. James Kent, Eric A. Stone, Serafim Batzoglou, Nick Goldman, Ross C. Hardison, David Haussler, Webb Miller, Arend Sidow, Nathan D. Trinklein, Zhengdong D. Zhang, Leah Barrera, Rhona Stuart, David C. King, Adam Ameur, Stefan Enroth, Mark C. Bieda, Jonghwan Kim, Akshay A. Bhinge, Nan Jiang, Jun Liu, Fei Yao, Vinsensius B. Vega, Charlie W.H. Lee, Patrick Ng, Annie Yang, Zarmik Moqtaderi, Zhou Zhu, Xiaoqin Xu, Sharon Squazzo, Matthew J. Oberley, David Inman, Michael A. Singer, Todd A. Richmond, Kyle J. Munn, Alvaro Rada-Iglesias, Ola Wallerman, Jan Komorowski, Joanna C. Fowler, Phillippe Couttet, Alexander W. Bruce, Oliver M. Dovey, Peter D. Ellis, Cordelia F. Langford, David A. Nix, Ghia Euskirchen, Stephen Hartman,

Alexander E. Urban, Peter Kraus, Sara Van Calcar, Nate Heintzman, Tae Hoon Kim, Kun Wang, Chunxu Qu, Gary Hon, Rosa Luna, Christopher K. Glass, M. Geoff Rosenfeld, Shelley Force Aldred, Sara J. Cooper, Anason Halees, Jane M. Lin, Hennady P. Shulha, Xiaoling Zhang, Mousheng Xu, Jaafar N.S. Haidar, Yong Yu, Vishwanath R. Iyer, Roland D. Green, Claes Wadelius, Peggy J. Farnham, Bing Ren, Rachel A. Harte, Angie S. Hinrichs, Heather Trumbower, Hiram Clawson, Jennifer Hillman-Jackson, Ann S. Zweig, Kayla Smith, Archana Thakkapallayil, Galt Barber, Robert M. Kuhn, Donna Karolchik, Lluis Armengol, Christine P. Bird, Paul I.W. De Bakker, Andrew D. Kern, Nuria Lopez-Bigas, Joel D. Martin, Barbara E. Stranger, Abigail Woodroffe, Eugene Davydov, Antigone Dimas, Eduardo Eyras, Ingileif B. Hallgrímsdóttir, Julian Huppert, Michael C. Zody, Gonçalo R. Abecasis, Xavier Estivill, Gerard G. Bouffard, Xiaobin Guan, Nancy F. Hansen, Jacquelyn R. Idol, Valerie V.B. Maduro, Baishali Maskeri, Jennifer C. McDowell, Morgan Park, Pamela J. Thomas, Alice C. Young, Robert W. Blakesley, Donna M. Muzny, Erica Sodergren, David A. Wheeler, Kim C. Worley, Huaiyang Jiang, George M. Weinstock, Richard A. Gibbs, Tina Graves, Robert Fulton, Elaine R. Mardis, Richard K. Wilson, Michele Clamp, James Cuff, Sante Gnerre, David B. Jaffe, Jean L. Chang, Kerstin Lindblad-Toh, Eric S. Lander, Maxim Koriabine, Mikhail Nefedov, Kazutoyo Osoegawa, Yuko Yoshinaga, Baoli Zhu, and Pieter J. De Jong. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, 2007.

[11] Kirsten F. Block, Elena Puerta-Fernandez, Jason G. Wallace, and Ronald R. Breaker. Association of OLE RNA with bacterial membranes via an RNA-protein interaction. *Molecular Microbiology*, 79(1):21–34, 2011.

[12] S. Bolte and F. P. Cordelieres. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy-Oxford*, 224:213–232, 2006.

[13] S. Bolte and FP. Cordelières. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*, 224(April):213–232, 2006.

[14] Roger Bourne and Roger Bourne. ImageJ. *Fundamentals of Digital Imaging in Medicine*, 9(7):185–188, 2010.

[15] Clive G. Bowsher and Peter S. Swain. Environmental sensing, information transfer, and cellular decision-making. *Current Opinion in Biotechnology*, 28:149–155, 2014.

[16] X. Cao, Z. Yan, Q. Wu, A. Zheng, and S. Zhong. Give: portable genome browsers for personal websites. *Genome Biol*, 19(1):92, 2018.

[17] P Carninci, T Kasukawa, S Katayama, J Gough, M C Frith, N Maeda, R Oyama, T Ravasi, B Lenhard, C Wells, R Kodzius, K Shimokawa, V B Bajic, S E Brenner, S Batalov, A R R Forrest, M Zavolan, M J Davis, L G Wilming, V Aidinis, J E Allen, A Ambesi-Impiombato, R Apweiler, R N Aturaliya, T L Bailey, M Bansal, L Baxter, K W Beisel, T Bersano, H Bono, A M Chalk, K P Chiu, V Choudhary, A Christoffels, D R Clutterbuck, M L Crowe, E Dalla, B P Dalrymple, B de Bono, G Della Gatta, D di Bernardo, T Down,

P Engstrom, M Fagiolini, G Faulkner, C F Fletcher, T Fukushima, M Furuno, S Futaki, M Gariboldi, P Georgii-Hemming, T R Gingeras, T Gojobori, R E Green, S Gustincich, M Harbers, Y Hayashi, T K Hensch, N Hirokawa, D Hill, L Huminiecki, M Iacono, K Ikeo, A Iwama, T Ishikawa, M Jakt, A Kanapin, M Katoh, Y Kawasawa, J Kelso, H Kitamura, H Kitano, G Kollias, S P T Krishnan, A Kruger, S K Kummerfeld, I V Kurochkin, L F Lareau, D Lazarevic, L Lipovich, J Liu, S Liuni, S McWilliam, M Madan Babu, M Madera, L Marchionni, H Matsuda, S Matsuzawa, H Miki, F Mignone, S Miyake, K Morris, S Mottagui-Tabar, N Mulder, N Nakano, H Nakauchi, P Ng, R Nilsson, S Nishiguchi, S Nishikawa, F Nori, O Ohara, Y Okazaki, V Orlando, K C Pang, W J Pavan, G Pavesi, G Pesole, N Petrovsky, S Piazza, J Reed, J F Reid, B Z Ring, M Ringwald, B Rost, Y Ruan, S L Salzberg, A Sandelin, C Schneider, C Schönbach, K Sekiguchi, C A M Semple, S Seno, L Sessa, Y Sheng, Y Shibata, H Shimada, K Shimada, D Silva, B Sinclair, S Sperling, E Stupka, K Sugiura, R Sultana, Y Takenaka, K Taki, K Tammoja, S L Tan, S Tang, M S Taylor, J Tegner, S A Teichmann, H R Ueda, E van Nimwegen, R Verardo, C L Wei, K Yagi, H Yamanishi, E Zabarovsky, S Zhu, A Zimmer, W Hide, C Bult, S M Grimmond, R D Teasdale, E T Liu, V Brusic, J Quackenbush, C Wahlestedt, J S Mattick, D A Hume, C Kai, D Sasaki, Y Tomaru, S Fukuda, M Kanamori-Katayama, M Suzuki, J Aoki, T Arakawa, J Iida, K Imamura, M Itoh, T Kato, H Kawaji, N Kawagashira, T Kawashima, M Kojima, S Kondo, H Konno, K Nakano, N Ninomiya, T Nishio, M Okada, C Plessy, K Shibata, T Shiraki, S Suzuki, M Tagami, K Waki, A Watahiki, Y Okamura-Oho, H Suzuki, J Kawai, and Y Hayashizaki. The Transcriptional Landscape of the Mammalian Genome. *Science*, 309(5740):1559 LP – 1563, sep 2005.

[18] Alfredo Castello, Bernd Fischer, Katrin Eichelbaum, Rastislav Horos, Benedikt M Beckmann, Claudia Strein, Norman E Davey, David T Humphreys, Thomas Preiss, Lars M Steinmetz, Jeroen Krijgsveld, and Matthias W Hentze. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149(6):1393–1406, jun 2012.

[19] Sumantra Chatterjee and Nadav Ahituv. Gene Regulatory Elements, Major Drivers of Human Disease. *Annual Review of Genomics and Human Genetics*, 18(1):45–63, 2017.

[20] Geng Chen, Baitang Ning, and Tieliu Shi. Single-cell RNA-seq technologies and related computational data analysis. *Frontiers in Genetics*, 10(APR):1–13, 2019.

[21] Weizhong Chen, Zhangming Yan, Simin Li, Norman Huang, Xuerui Huang, Jin Zhang, and Sheng Zhong. RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome. *iScience*, 4:204–215, 2018.

[22] Y. Choi, H. P. Kim, S. M. Hong, J. Y. Ryu, S. J. Han, and R. Song. In situ visualization of gene expression using polymer-coated quantum-dot-dna conjugates. *Small*, 5(18):2085–91, 2009.

[23] Y. J. Chu and D. R. Corey. Rna sequencing: Platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*, 22(4):271–274, 2012.

[24] Yongjun Chu and David R. Corey. RNA sequencing: Platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*, 22(4):271–274, 2012.

[25] Chenghua Cui, Wei Shu, and Peining Li. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology*, 4(SEP):1–11, 2016.

[26] X. Dai, R. Theobard, H. Cheng, M. Xing, and J. Zhang. Fusion genes: A promising tool combating against cancer. *Biochim Biophys Acta Rev Cancer*, 1869(2):149–160, 2018.

[27] N. M. Davidson, I. J. Majewski, and A. Oshlack. Jaffa: High sensitivity transcriptome-focused fusion gene detection. *Genome Med*, 7(1):43, 2015.

[28] Wulan Deng, Xinghua Shi, Robert Tjian, Timothée Lionnet, and Robert H. Singer. CAS-FISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38):11870–11875, 2015.

[29] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

[30] Rupak Doshi, Beverly R. Chen, Cecile Rose T. Vibat, Norman Huang, Chang Wook Lee, and Geoffrey Chang. In vitro nanobody discovery for integral membrane protein targets. *Scientific Reports*, 4(October), 2014.

[31] N. Durisic, P. W. Wiseman, P. Grutter, and C. D. Heyes. A common mechanism underlies the dark fraction formation and fluorescence blinking of quantum dots. *Acs Nano*, 3(5):1167–1175, 2009.

[32] Andrea M. Femino, Fredric S. Fay, Kevin Fogarty, and Robert H. Singer. Visualization of single RNA transcripts in situ. *Science*, 280(5363):585–590, 1998.

[33] Ryan A. Flynn, Benjamin A. H. Smith, Alex G. Johnson, Kayvon Pedram, Benson M. George, Stacy A. Malaker, Karim Majzoub, Jan E. Carette, and Carolyn R. Bertozzi. Mammalian y rnas are modified at discrete guanosine residues with n-glycans. *bioRxiv*, 2019.

[34] P. A. Frantsuzov, S. Volkan-Kacso, and B. Janko. Universality of the fluorescence intermittency in nanoscale systems: Experiment and theory. *Nano Letters*, 13(2):402–408, 2013.

[35] Martin C. Frith, Michael Pheasant, and John S. Mattick. The amazing complexity of the human transcriptome. *European Journal of Human Genetics*, 13(8):894–897, 2005.

[36] M. V. Fuccillo, C. Foldy, O. Gokce, P. E. Rothwell, G. L. Sun, R. C. Malenka, and T. C. Sudhof. Single-cell mrna profiling reveals cell-type-specific expression of neurexin isoforms. *Neuron*, 87(2):326–40, 2015.

[37] M. Gabut, P. Samavarchi-Tehrani, X. Wang, V. Slobodeniuc, D. O'Hanlon, H. K. Sung, M. Alvarez, S. Talukder, Q. Pan, E. O. Mazzoni, S. Nedelec, H. Wichterle, K. Woltjen, T. R. Hughes, P. W. Zandstra, A. Nagy, J. L. Wrana, and B. J. Blencowe. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, 147(1):132–46, 2011.

[38] Q. Gao, W. W. Liang, S. M. Foltz, G. Mutharasu, R. G. Jayasinghe, S. Cao, W. W. Liao, S. M. Reynolds, M. A. Wyczalkowski, L. Yao, L. Yu, S. Q. Sun, Group Fusion Analysis Working, Network Cancer Genome Atlas Research, K. Chen, A. J. Lazar, R. C. Fields, M. C. Wendl, B. A. Van Tine, R. Vij, F. Chen, M. Nykter, I. Shmulevich, and L. Ding. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep*, 23(1):227–238 e3, 2018.

[39] K. A. Geiler-Samerotte, C. R. Bauer, S. Li, N. Ziv, D. Gresham, and M. L. Siegal. The details in the distributions: Why and how to study phenotypic variability. *Current Opinion in Biotechnology*, 24(4):752–759, 2013.

[40] D. Gerion, F. Pinaud, S. C. Williams, W. J. Parak, D. Zanchet, S. Weiss, and A. P. Alivisatos. Synthesis and properties of biocompatible water-soluble silica-coated cdse/zns semiconductor quantum dots. *Journal of Physical Chemistry B*, 105(37):8861–8871, 2001.

[41] Daniele Gerion, Fabien Pinaud, Shara C. Williams, Wolfgang J. Parak, Daniela Zanchet, Shimon Weiss, and A. Paul Alivisatos. Synthesis and properties of biocompatible water-soluble silica-coated CdSe/ZnS semiconductor quantum dots. *Journal of Physical Chemistry B*, 105(37):8861–8871, 2001.

[42] R. Gill, I. Willner, I. Shweky, and U. Banin. Fluorescence resonance energy transfer in cdse/zns-dna conjugates: Probing hybridization and dna cleavage. *Journal of Physical Chemistry B*, 109(49):23715–23719, 2005.

[43] Ron Gill, Itamar Willner, Itzhak Shweky, and Uri Banin. Fluorescence resonance energy transfer in CdSe/ZnS-DNA conjugates: Probing hybridization and DNA cleavage. *Journal of Physical Chemistry B*, 109(49):23715–23719, 2005.

[44] E. R. Goldman, A. R. Clapp, G. P. Anderson, H. T. Uyeda, J. M. Mauro, I. L. Medintz, and H. Mattoussi. Multiplexed toxin analysis using four colors of quantum dot fluororeagents. *Analytical Chemistry*, 76(3):684–688, 2004.

[45] Zuguang Gu, Roland Eils, and Matthias Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, 2016.

[46] Piyush B. Gupta, Christine M. Fillmore, Guozhi Jiang, Sagi D. Shapira, Kai Tao, Charlotte Kuperwasser, and Eric S. Lander. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644, 2011.

[47] Brian Haas, Alexander Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas Doak, Natalie Pochet, Jing Sun, Catherine Wu, Thomas Gingeras, and Aviv Regev. Star-fusion: Fast and accurate fusion transcript detection from rna-seq. *bioRxiv*, (120295), 2017.

[48] M. Y. Han, X. H. Gao, J. Z. Su, and S. Nie. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nature Biotechnology*, 19(7):631–635, 2001.

[49] Che Ming J. Hu, Ronnie H. Fang, Kuei Chun Wang, Brian T. Luk, Soracha Thamphiwatana, Diana Dehaini, Phu Nguyen, Pavimol Angsantikul, Cindy H. Wen, Ashley V. Kroll, Cody Carpenter, Manikantan Ramesh, Vivian Qu, Sherrina H. Patel, Jie Zhu, William Shi, Florence M. Hofman, Thomas C. Chen, Weiwei Gao, Kang Zhang, Shu Chien, and Liangfang Zhang. Nanoparticle biointerfacing by platelet membrane cloaking. *Nature*, 526(7571):118–121, 2015.

[50] Linping Hu, Kun Ru, Li Zhang, Yuting Huang, Xiaofan Zhu, Hanzhi Liu, Anders Zetterberg, Tao Cheng, and Weimin Miao. Fluorescence in situ hybridization (FISH): An increasingly demanded tool for biomarker research and personalized medicine. *Biomarker Research*, 2(1):1–13, 2014.

[51] X. Hu, Q. Wang, M. Tang, F. Barthel, S. Amin, K. Yoshihara, F. M. Lang, E. Martinez-Ledesma, S. H. Lee, S. Zheng, and R. G. W. Verhaak. Tumorfusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res*, 46(D1):D1144–D1149, 2018.

[52] Norman Huang, Zhangming Yan, Weixin Wu, Weizhong Chen, Yiqun Jiang, Jingyao Chen, Xuerui Huang, Xingzhao Wen, Jie Xu, Qiushi Jin, Kang Zhang, Zhen Chen, Shu Chien, and Sheng Zhong. Genome-wide colocalization of RNADNA interactions and fusion RNA pairs. *Proceedings of the National Academy of Sciences of the United States of America*, 116(8):3328–3337, 2019.

[53] Florent Hubé and Claire Francastel. Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Frontiers in genetics*, 9:140, apr 2018.

[54] D. Huber, L. Voith von Voithenberg, and G. V. Kaigala. Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro and Nano Engineering*, 1:15–24, 2018.

[55] J. K. Jaiswal, H. Mattoussi, J. M. Mauro, and S. M. Simon. Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nature Biotechnology*, 21(1):47–51, 2003.

[56] Tanja Jakoš, Anja Pišlar, Anahid Jewett, and Janko Kos. Cysteine Cathepsins in Tumor-Associated Immune Cells. *Frontiers in immunology*, 10:2037, aug 2019.

[57] Tadeusz Janas, Teresa Janas, and Michael Yarus. Specific RNA binding to ordered phospholipid bilayers. *Nucleic Acids Research*, 34(7):2128–2136, 2006.

[58] Tadeusz Janas and Michael Yarus. Visualization of membrane RNAs. *Rna*, 9(11):1353–1361, 2003.

[59] Teresa Janas, Tadeusz Janas, and Michael Yarus. A membrane transporter for tryptophan composed of RNA. *RNA (New York, N.Y.)*, 10(10):1541–1549, oct 2004.

[60] Teresa Janas, Tadeusz Janas, and Michael Yarus. Human tRNA(Sec) associates with HeLa membranes, cell lipid liposomes, and synthetic lipid bilayers. *RNA (New York, N.Y.)*, 18(12):2260–2268, dec 2012.

[61] Damien Jeandard, Anna Smirnova, Ivan Tarassov, Eric Barrey, Alexandre Smirnov, and Nina Entelis. Import of Non-Coding RNAs into Human Mitochondria: A Critical Review and Emerging Approaches. *Cells*, 8(3):286, mar 2019.

[62] D. Kampa, J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana, and T. R. Gingeras. Novel rnas identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res*, 14(3):331–42, 2004.

[63] Neda R. Kasim, Kateina Kuželová, Aleš Holoubek, and Michael A. Model. Live fluorescence and transmission-through-dye microscopic study of actinomycin D-induced apoptosis and apoptotic volume decrease. *Apoptosis*, 18(4):521–532, 2013.

[64] A Khvorova, Y G Kwak, M Tamkun, I Majerfeld, and M Yarus. RNAs that bind and change the permeability of phospholipid membranes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(19):10649–10654, sep 1999.

[65] H. Kirshner, F. Aguet, D. Sage, and M. Unser. 3-d psf fitting for fluorescence microscopy: implementation and localization application. *Journal of Microscopy*, 249(1):13–25, 2013.

[66] C. Kumar-Sinha, S. Kalyana-Sundaram, and A. M. Chinnaiyan. Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med*, 7:129, 2015.

[67] Sunjong Kwon. Single-molecule fluorescence in situ hybridization: Quantitative imaging of single RNA molecules. *BMB Reports*, 46(2):65–72, 2013.

[68] Suzanne L. Lababidi, Mariana Pelts, Moumita Moitra, Laura G. Leff, and Michael A. Model. Measurement of bacterial volume by transmission-through-dye imaging. *Journal of Microbiological Methods*, 87(3):375–377, 2011.

[69] J. Lai, J. An, I. Seim, C. Walpole, A. Hoffman, L. Moya, S. Srinivasan, J. L. Perry-Keene, Bioresource Australian Prostate Cancer, C. Wang, M. L. Lehman, C. C. Nelson, J. A. Clements, and J. Batra. Fusion transcript loci share many genomic features with non-fusion loci. *BMC Genomics*, 16:1021, 2015.

[70] Pavel P. Laktionov, Svetlana N. Tamkovich, Elena Yu Rykova, Olga E. Bryzgunova, Andrey V. Starikov, Nina P. Kuznetsova, and Valentin V. Vlassov. Cell-surface-bound nucleic acids: Free and cell-surface-bound nucleic acids in blood of healthy donors and breast cancer patients. *Annals of the New York Academy of Sciences*, 1022:221–227, 2004.

[71] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[72] Chatarina Larsson, Ida Grundberg, Ola Söderberg, and Mats Nilsson. In situ detection and genotyping of individual mRNA molecules. *Nature Methods*, 7(5):395–397, 2010.

[73] Michael Lawrence, Wolfgang Huber, Herv Pags, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin T. Morgan, and Vincent J. Carey. Software for computing and annotating genomic ranges. *PLOS Computational Biology*, 9(8):e1003118, 2013.

[74] C. Lee and M. Roy. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biology*, 5(7), 2004.

[75] Christopher Lee and Meenakshi Roy. Analysis of alternative splicing with microarrays: Successes and challenges. *Genome Biology*, 5(7):5–8, 2004.

[76] K. Lee, Y. Cui, L. P. Lee, and J. Irudayaraj. Quantitative imaging of single mrna splice variants in living cells. *Nat Nanotechnol*, 9(6):474–80, 2014.

[77] L. Y. Lee, S. L. Ong, J. Y. Hu, W. J. Ng, Y. Y. Feng, X. L. Tan, and S. W. Wong. Use of semiconductor quantum dots for photostable immunofluorescence labeling of cryptosporidium parvum. *Applied and Environmental Microbiology*, 70(10):5732–5736, 2004.

[78] Tineke L. Lenstra, Joseph Rodriguez, Huimin Chen, and Daniel R. Larson. Transcription Dynamics in Living Cells. *Annual Review of Biophysics*, 45(1):25–47, 2016.

[79] H. Li, J. Wang, G. Mor, and J. Sklar. A neoplastic gene fusion mimics trans-splicing of rnas in normal human cells. *Science*, 321(5894):1357–61, 2008.

[80] Aifu Lin, Qingsong Hu, Chunlai Li, Zhen Xing, Guolin Ma, Cheng Wang, Jun Li, Yin Ye, Jun Yao, Ke Liang, Shouyu Wang, Peter K. Park, Jeffrey R. Marks, Yan Zhou, Jianwei Zhou, Mien Chie Hung, Han Liang, Zhibin Hu, Hongbing Shen, David H. Hawke, Leng Han, Yubin Zhou, Chunru Lin, and Liuqing Yang. The LINK-A lncRNA interacts with PtdIns(3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nature Cell Biology*, 19(3):238–251, 2017.

[81] Shawn C. Little, Mikhail Tikhonov, and Thomas Gregor. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*, 154(4):789–800, 2013.

[82] Eric Lubeck and Long Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods*, 9(7):743–748, 2012.

[83] Eric Lubeck, Ahmet F. Coskun, Timur Zhiyentayev, Mubhij Ahmad, and Long Cai. Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods*, 11(4):360–361, 2014.

[84] ATL Lun, M Perry, and E Ing-Simmons. Infrastructure for genomic interactions: Bioconductor classes for hi-c, chia-pet and related experiments [version 2; referees: 2 approved]. *F1000Research*, 5(950), 2016.

[85] A. Lyubimova, S. Itzkovitz, J. P. Junker, Z. P. Fan, X. Wu, and A. van Oudenaarden. Single-molecule mrna detection and counting in mammalian tissue. *Nat Protoc*, 8(9):1743–58, 2013.

[86] F. B. Markey, W. Ruezinsky, S. Tyagi, and M. Batish. Fusion fish imaging: single-molecule detection of gene fusion transcripts in situ. *PLoS One*, 9(3):e93488, 2014.

[87] M. P. Martelli, G. Sozzi, L. Hernandez, V. Pettirossi, A. Navarro, D. Conte, P. Gasparini, F. Perrone, P. Modena, U. Pastorino, A. Carbone, A. Fabbri, A. Sidoni, S. Nakamura, M. Gambacorta, P. L. Fernandez, J. Ramirez, J. K. Chan, W. F. Grigioni, E. Campo, S. A. Pileri, and B. Falini. Eml4-alk rearrangement in non-small cell lung cancer and non-tumor lung tissues. *Am J Pathol*, 174(2):661–70, 2009.

[88] H. Mattoussi, J. Matthew Mauro, E. R. Goldman, G. P. Anderson, V. C. Sundar, F. V. Mikulec, and M. G. Bawendi. Self-assembly of CdSe-ZnS quantum dot bioconjugates using an engineered recombinant protein. *Journal of the American Chemical Society*, 122(49):12142–12150, 2000.

[89] H. Mattoussi, J. M. Mauro, E. R. Goldman, G. P. Anderson, V. C. Sundar, F. V. Mikulec, and M. G. Bawendi. Self-assembly of cdse-zns quantum dot bioconjugates using an engineered recombinant protein. *Journal of the American Chemical Society*, 122(49):12142–12150, 2000.

[90] I. L. Medintz, H. T. Uyeda, E. R. Goldman, and H. Mattoussi. Quantum dot bioconjugates for imaging, labelling and sensing. *Nature Materials*, 4(6):435–446, 2005.

[91] F. Mertens, B. Johansson, T. Fioretos, and F. Mitelman. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer*, 15(6):371–81, 2015.

[92] G. P. Mitchell, C. A. Mirkin, and R. L. Letsinger. Programmed assembly of dna functionalized quantum dots. *Journal of the American Chemical Society*, 121(35):8122–8123, 1999.

[93] Gregory P. Mitchell, Chad A. Mirkin, and Robert L. Letsinger. Programmed assembly of DNA functionalized quantum dots [10]. *Journal of the American Chemical Society*, 121(35):8122–8123, 1999.

[94] EVGENIY S MOROZKIN, PAVEL P LAKTIONOV, ELENA Y RYKOVA, and VALENTIN V VLASSOV. Extracellular Nucleic Acids in Cultures of Long-Term Cultivated Eukaryotic Cells. *Annals of the New York Academy of Sciences*, 1022(1):244–249, jun 2004.

[95] T. C. Nguyen, X. Cao, P. Yu, S. Xiao, J. Lu, F. H. Biase, B. Sridhar, N. Huang, K. Zhang, and S. Zhong. Mapping rna-rna interactome and rna structure in vivo by mario. *Nat Commun*, 7:12023, 2016.

[96] Tan A Nguyen, Blake R C Smith, Kirstin D Elgass, Sarah J Creed, Shane Cheung, Michelle D Tate, Gabrielle T Belz, Ian P Wicks, Seth L Masters, and Ken C Pang. SIDT1 Localizes to Endolysosomes and Mediates Double-Stranded RNA Transport into the Cytoplasm. *Journal of immunology (Baltimore, Md. : 1950)*, 202(12):3483–3492, jun 2019.

[97] W. J. Parak, D. Gerion, D. Zanchet, A. S. Woerz, T. Pellegrino, C. Micheel, S. C. Williams, M. Seitz, R. E. Bruehl, Z. Bryant, C. Bustamante, C. R. Bertozzi, and A. P. Alivisatos. Conjugation of dna to silanized colloidal semiconductor nanocrystalline quantum dots. *Chemistry of Materials*, 14(5):2113–2119, 2002.

[98] Wolfgang J. Parak, Daniele Gerion, Daniela Zanchet, Anke S. Woerz, Teresa Pellegrino, Christine Micheel, Shara C. Williams, Markus Seitz, Richard E. Bruehl, Zev Bryant, Carlos Bustamante, Carolyn R. Bertozzi, and A. Paul Alivisatos. Conjugation of DNA to silanized colloidal semiconductor nanocrystalline quantum dots. *Chemistry of Materials*, 14(5):2113–2119, 2002.

[99] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[100] Catherine Rabouille. Pathways of Unconventional Protein Secretion. *Trends in cell biology*, 27(3):230–240, mar 2017.

[101] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. Stochastic mrna synthesis in mammalian cells. *Plos Biology*, 4(10):1707–1719, 2006.

[102] Arjun Raj, Patrick van den Bogaard, Scott A. Rifkin, Alexander van Oudenaarden, and Sanjay Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10):877–879, 2008.

[103] Arjun Raj and Alexander van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008.

[104] X. Ren, R. Deng, K. Zhang, Y. Sun, X. Teng, and J. Li. Splicerca: in situ single-cell analysis of mrna splicing variants. *ACS Cent Sci*, 4(6):680–687, 2018.

[105] U. Resch-Genger, M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke, and T. Nann. Quantum dots versus organic dyes as fluorescent labels. *Nature Methods*, 5(9):763–775, 2008.

[106] Blanka Robesova, Monika Bajerova, Kvetoslava Liskova, Jana Skrickova, Marcela Tomiskova, Sarka Pospisilova, Jiri Mayer, and Dana Dvorakova. TaqMan based real time PCR assay targeting EML4-ALK fusion transcripts in NSCLC. *Lung Cancer*, 85(1):25–30, 2014.

[107] N. Roch, S. Florens, V. Bouchiat, W. Wernsdorfer, and F. Balestro. Quantum phase transition in a single-molecule quantum dot. *Nature*, 453(7195):633–U3, 2008.

[108] Livia A.Casciola Rosen, Grant Anhah, and Antony Rosen. Autoantigens targeted in systemic lupus erythematmus are clustered in two populations of surface structures on apoptotic keratinocytes. *Journal of Experimental Medicine*, 179(4):1317–1330, 1994.

[109] Jeffrey M. Rothenstein and N. Chooback. ALK inhibitors, resistance development, clinical trials. *Current Oncology*, 25(June):S59–S67, 2018.

[110] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser. Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. *Ieee Transactions on Image Processing*, 14(9):1372–1383, 2005.

[111] Daniel Sage, Franck R. Neumann, Florence Hediger, Susan M. Gasser, and Michael Unser. Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. *IEEE Transactions on Image Processing*, 14(9):1372–1383, 2005.

[112] Wes Sanders and Alain Laederach. Membrane RNAs in bacteria. *Molecular microbiology*, 79(1):1–2, 2011.

[113] D. Sapkota, A. M. Lake, W. Yang, C. Yang, H. Wesseling, A. Guise, C. Uncu, J. S. Dalal, A. W. Kraft, J. M. Lee, M. S. Sands, J. A. Steen, and J. D. Dougherty. Cell-type-specific profiling of alternative translation identifies regulated protein isoform variation in the mouse brain. *Cell Rep*, 26(3):594–607 e7, 2019.

[114] Takaaki Sasaki, Scott J. Rodig, Lucian R. Chirieac, and Pasi A. Jänne. The biology and treatment of EML4-ALK non-small cell lung cancer. *European Journal of Cancer*, 46(10):1773–1780, 2010.

[115] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene-expression patterns with a complementary-dna microarray. *Science*, 270(5235):467–470, 1995.

[116] Mark Schena, Dari Shalon, Ronald W Davis, and Patrick O Brown. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, 270(5235):467 LP – 470, oct 1995.

[117] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nat Methods*, 9(7):671–5, 2012.

[118] S. Semrau, N. Crosetto, M. Bienko, M. Boni, P. Bernasconi, R. Chiarle, and A. van Oudenaarden. Fusefish: robust detection of transcribed gene fusions in single cells. *Cell Rep*, 6(1):18–23, 2014.

[119] Bingdong Sha and Ming Luo. Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nature Structural Biology*, 4(3):239–244, 1997.

[120] Sydney M. Shaffer, Min Tzu Wu, Marshall J. Levesque, and Arjun Raj. Turbo FISH: A Method for Rapid Single Molecule RNA FISH. *PLoS ONE*, 8(9), 2013.

[121] A. M. Smith and S. M. Nie. Compact quantum dots for single-molecule imaging. *Jove-Journal of Visualized Experiments*, (68), 2012.

[122] Y. Song, O. B. Botvinnik, M. T. Lovci, B. Kakaradov, P. Liu, J. L. Xu, and G. W. Yeo. Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol Cell*, 67(1):148–161 e5, 2017.

[123] R Srinivasan, E Ozhegov, Y W van den Berg, B J Aronow, R S Franco, M B Palascak, J T Fallon, W Ruf, H H Versteeg, and V Y Bogdanov. Splice variants of tissue factor promote monocyte-endothelial interactions by triggering the expression of cell adhesion molecules via integrin-mediated signaling. *Journal of thrombosis and haemostasis : JTH*, 9(10):2087–2096, oct 2011.

[124] A. Stahlberg and M. Kubista. The workflow of single-cell expression profiling using quantitative real-time pcr. *Expert Review of Molecular Diagnostics*, 14(3):323–331, 2014.

[125] N. Stransky, E. Cerami, S. Schalm, J. L. Kim, and C. Lengauer. The landscape of kinase fusions in cancer. *Nat Commun*, 5:4846, 2014.

[126] D. Z. Sun and O. Gang. Dna-functionalized quantum dots: Fabrication, structural, and physicochemical properties. *Langmuir*, 29(23):7038–7046, 2013.

[127] Dazhi Sun and Oleg Gang. DNA-functionalized quantum dots: Fabrication, structural, and physicochemical properties. *Langmuir*, 29(23):7038–7046, 2013.

[128] Masayuki Takahashi, Viorica Raluca Contu, Chihana Kabuta, Katsunori Hase, Yuuki Fujiwara, Keiji Wada, and Tomohiro Kabuta. SIDT2 mediates gymnosis, the uptake of naked single-stranded oligonucleotides into living cells. *RNA biology*, 14(11):1534–1543, nov 2017.

[129] Toshiya Takahashi, Nikhil Nitin Kulkarni, Ernest Y. Lee, Ling Juan Zhang, Gerard C.L. Wong, and Richard L. Gallo. Cathelicidin promotes inflammation by enabling binding of self-RNA to cell surface scavenger receptors. *Scientific Reports*, 8(1):1–13, 2018.

[130] Yuqi Tan and Patrick Cahan. SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Systems*, 9(2):207–213.e2, 2019.

[131] J. Tazi, N. Bakkour, and S. Stamm. Alternative splicing and disease. *Biochim Biophys Acta*, 1792(1):14–26, 2009.

[132] W. Torres-Garcia, S. Zheng, A. Sivachenko, R. Vegesna, Q. Wang, R. Yao, M. F. Berger, J. N. Weinstein, G. Getz, and R. G. Verhaak. Prada: pipeline for rna sequencing data analysis. *Bioinformatics*, 30(15):2224–6, 2014.

[133] W. Valleron, E. Laprevotte, E. F. Gautier, C. Quelen, C. Demur, E. Delabesse, X. Agirre, F. Prosper, T. Kiss, and P. Brousset. Specific small nucleolar rna expression profiles in acute leukemia. *Leukemia*, 26(9):2052–2060, 2012.

[134] II Vandenbroucke, J. Vandesompele, A. D. Paepe, and L. Messiaen. Quantification of splice variants using real-time pcr. *Nucleic Acids Res*, 29(13):E68–8, 2001.

[135] K. E. Varley, J. Gertz, B. S. Roberts, N. S. Davis, K. M. Bowling, M. K. Kirby, A. S. Nesmith, P. G. Oliver, W. E. Grizzle, A. Forero, D. J. Buchsbaum, A. F. LoBuglio, and R. M. Myers. Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat*, 146(2):287–97, 2014.

[136] J. A. Vendrell, S. Taviaux, B. Beganton, S. Godreuil, P. Audran, D. Grand, E. Clermont, I. Serre, V. Szablewski, P. Coopman, J. Mazieres, V. Costes, J. L. Pujol, P. Brousset, I. Rouquette, and J. Solassol. Detection of known and novel alk fusion transcripts in lung cancer patients using next-generation sequencing approaches. *Sci Rep*, 7(1):12510, 2017.

[137] Maria Vera, Jeetayu Biswas, Adrien Senecal, Robert H. Singer, and Hye Yoon Park. Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. *Annual Review of Genetics*, 50(1):267–291, 2016.

[138] C. Vonesch and M. Unser. A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution. *Ieee Transactions on Image Processing*, 17(4):539–549, 2008.

[139] L Wakefield and G G Brownlee. RNA-binding properties of influenza A virus matrix protein M1. *Nucleic acids research*, 17(21):8569–8580, nov 1989.

[140] Kevin C. Wang and Howard Y. Chang. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, 43(6):904–914, 2011.

[141] Y. Wang, J. Liu, B. O. Huang, Y. M. Xu, J. Li, L. F. Huang, J. Lin, J. Zhang, Q. H. Min, W. M. Yang, and X. Z. Wang. Mechanism of alternative splicing and its regulation. *Biomed Rep*, 3(2):152–158, 2015.

[142] Yan Wang, Jun Zhang, Guanghui Gao, Xuefei Li, Chao Zhao, Yayi He, Chunxia Su, Shijia Zhang, Xiaoxia Chen, Jie Zhang, Wei Li, Bing Li, Jing Zhao, Likun Hou, Chunyan Wu, Shengxiang Ren, and Caicun Zhou. EML4-ALK fusion detected by RT-PCR confers similar response to crizotinib as detected by FISH in patients with advanced non-small-cell lung cancer. *Journal of Thoracic Oncology*, 10(11):1546–1552, 2015.

[143] Weixin Wu, Zhangming Yan, Tri C Nguyen, Zhen Bouman Chen, Shu Chien, and Sheng Zhong. Mapping RNA-chromatin interactions by sequencing with iMARGI. *Nature protocols*, 2019.

[144] X. Wu, H. Liu, J. Liu, K. N. Haley, J. A. Treadway, J. P. Larson, N. Ge, F. Peale, and M. P. Bruchez. Immunofluorescent labeling of cancer marker her2 and other cellular targets with semiconductor quantum dots. *Nat Biotechnol*, 21(1):41–6, 2003.

[145] Xingyong Wu, Hongjian Liu, Jianquan Liu, Kari N. Haley, Joseph A. Treadway, J. Peter Larson, Nianfeng Ge, Frank Peale, and Marcel P. Bruchez. Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nature Biotechnology*, 21(1):41–46, 2003.

[146] J. W. Wynne, S. Todd, V. Boyd, M. Tachedjian, R. Klein, B. Shiell, M. Dearnley, A. J. McAuley, A. P. Woon, A. W. Purcell, G. A. Marsh, and M. L. Baker. Comparative transcriptomics highlights the role of the activator protein 1 transcription factor in the host response to ebolavirus. *J Virol*, 91(23), 2017.

[147] J. W. Wynne, S. Todd, V. Boyd, M. Tachedjian, R. Klein, B. Shiell, M. Dearnley, A. J. McAuley, A. P. Woon, A. W. Purcell, G. A. Marsh, and M. L. Baker. Comparative transcriptomics highlights the role of the activator protein 1 transcription factor in the host response to ebolavirus. *J Virol*, 91(23), 2017.

[148] J. Yao, D. R. Larson, H. D. Vishwasrao, W. R. Zipfel, and W. W. Webb. Blinking and nonradiant dark fraction of water-soluble quantum dots in aqueous solution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14284–14289, 2005.

[149] K. Yoshihara, Q. Wang, W. Torres-Garcia, S. Zheng, R. Vegesna, H. Kim, and R. G. Verhaak. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, 34(37):4845–54, 2015.

[150] Grace X Y Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, Mark T Gregory, Joe Shuga, Luz Montesclaros, Jason G Underwood, Donald A Masquelier, Stefanie Y Nishimura, Michael Schnall-Levin, Paul W Wyatt, Christopher M Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D Ness, Lan W Beppu, H Joachim Deeg, Christopher McFarland, Keith R Loeb, William J Valente, Nolan G Ericson, Emily A Stevens, Jerald P Radich, Tarjei S Mikkelsen, Benjamin J Hindson, and Jason H Bielas. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, 2017.

[151] D. J. Zhou, J. D. Piper, C. Abell, D. Klenerman, D. J. Kang, and L. M. Ying. Fluorescence resonance energy transfer between a quantum dot donor and a dye acceptor attached to dna. *Chemical Communications*, (38):4807–4809, 2005.

[152] Dejian Zhou, Joe D. Piper, Chris Abell, David Klenerman, Dae Joon Kang, and Liming Ying. Fluorescence resonance energy transfer between a quantum dot donor and a dye acceptor attached to DNA. *Chemical Communications*, (38):4807–4809, 2005.

[153] Joseph X. Zhou and Sui Huang. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in Genetics*, 27(2):55–62, 2011.

[154] P. Zrazhevskiy and X. H. Gao. Quantum dot imaging platform for single-cell molecular profiling. *Nature Communications*, 4, 2013.