A versatile protein platform for sequence-specific nucleic acid imaging and detection

Ruihong Wang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Raymond Monnat Jr., Chair

Patrick Stayton

James Carothers

Alshakim Nelson

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

A versatile protein platform for sequence-specific nucleic acid imaging and detection

Ruihong Wang

Chair of the Supervisory Committee:
Prof. Raymond Monnat
Department of Laboratory Medicine and Pathology

Detection and quantification of genetic material is fundamental for many research and diagnostic applications. Many different methods have been developed towards this objective, where each is typically tailored to detect a specific types of nucleic acid target under defined conditions. We have devised a platform technology based on a single-protein probe architecture compatible with detection of all types of genetic materials (RNA, ssDNA, dsDNA), to achieve design flexibility and compatibility with existing biological assays, ease of mass production and quality control for clinical applications and establish direct link between sequence recognition and signal generation.

In this dissertation, I describe the development of a universal platform for sequence-specific nucleic acid imaging and detection based on Transcription Activator-Like Effector (TALE) proteins with versatile functional domains, characterization of the platform, and its application in live cells, fixed cellular specimens, and cell-free context.

In fixed and live cell specimens, I used TALE probes to robustly label individual genome loci under most native nucleic acid condition in the cell environment. The engagement of TALEs with

genomic DNA is minimally disruptive to chromatin and is analogous to the interaction of endogenous transcription factors with their target sites. Furthermore, because TALEs are pure, single-component protein probes, they can be easily engineered to incorporate any protein modality used for visualization. These properties render TALEs an attractive and minimally invasive platform for revealing native genomic structure and function by imaging.

TALE architecture can also be used for *ex vivo* nucleic acid detection in cell-free context. Nucleic acid detection methods have facilitated the development of biological basic research and served as a core technique in diagnostics for pathogens in various biological samples. I deployed TALE probes with various reporter units to implement sequence-specific nucleic acid detection in biological samples. The detection is highly sensitive and comparable detection limit with contemporary assays with robust expansion to multiplexing targeting.

Thus, the TALE platform provides a useful and versatile new way to understand the spatial and temporal aspects of nucleic acid in cell environment, and a corresponding new platform and technology for nucleic acid detection in clinical or other biological samples.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Throughout the years of my graduate studies, I received massive supports and helps from my mentors, colleagues, friends, and families, who provide unconditional help with my graduate studies and life.

I very much appreciate Dr. John Stamatoyannopoulos for being a tremendous support to my graduate study in many resources. Your advisory on my work in scientific research lets me explore further in the area of study.

Over the past years of study, I have special thanks to my previous and current advisors and committee members for all the mentoring and teaching on my research and how to be a good researcher. Especially, I thank Dr. Shreeram Akilesh, Dr. Fyodor Urnov, and Dr. Pavel Zrazhevskiy for my direct mentor and supervisor over my thesis project and pushing me to expand my project and horizons. You passed on a lot of good advice and experiences so I can be a better researcher. Also, my past and current committee members, Dr. Raymond Monnat, Dr. James Carothers, Dr. Andy Scharenberg, Dr. Michael Jensen, Dr. Patrick Stayton, and Dr. Alshakim Nelson, offer consistent support.

My former and current colleagues in the lab have provided numerous resources I need. I would like to thank Bill Kervin, Vivek Nandakumar, Tobias Ragoczy, Sanna Sullivan, and Tanya Kutyavin while I developed my thesis project. Also thank Cell biology team, TALE assembly team, HSDU, and computational team for their supportive work in data production.

My family and friends are always supportive to me. I express my greatest gratitude for continuous help and encouragement for the past years. My family has been supportive since I decided to come to US for college and graduate studies. The unconditional love has motivated me to pursue my dream. My grandmother Dongmei passed away during the pandemic period and

unfortunately, I cannot make it back during this hard time. I owed her deeply and hope she would be proud of me. I would like to thank all my friends from college, graduate school, and many other encounters for all the joys you have brought me.

Lastly, I thank my funding source from Centers of Excellence in Genome Science (CEGS) grant from National Human Genome Research Institute (NHGRI). Molecular Engineering and Sciences Institute and PhD program at University of Washington provided consistent help and support in my initiation and continuation of my graduate study.

Chapter 1 Introduction

Technology has been developed for decades in molecular and cell biology. New techniques have been extensively renovated to improve the assay efficacy, turnaround time, and efficiency. These new advances participate in the scientific research to a faster process and higher throughput. On the other hand, improved basic science is serving as the scientific fundamentals for the technology innovation. The positive feedback loop helps us continuously make the technology more accessible to research works and real life.

1.1 Detection of nucleic acids in biological samples

Detection of nucleic acids is an important technique enabling detection of specific nucleic acid targets in various biological samples. Polymerase chain reaction (PCR) is an easy and well-established method to amplify a single target to millions of copies to be read out. We can use this protocol to evaluate the existence and quantity of one or more specific nucleic acid targets. For example, clinical laboratories use PCR or other amplification methods to detect pathogen existence in clinical samples (*1*).

To start with, the entire procedure relies on two parts: detection module and readout module. The detection module for the nucleic acid is a probe, mostly sequence specific, binding towards a unique region of the genetic material in the specimens. The detection module can be nucleic acid, protein, or chemical. There are other featured biosensors developed for the nucleic acid detection by electronics as well (*2*). Nucleic acid detection modules commonly used are nucleic acid probes, such as fluorescence or isotope labeled nucleic acid or analog probes (*3*). In use of selected quantitative-PCR and southern blotting, processed samples are detected with chemical dye labeled nucleic acid probes whose sequences are complementary to the target of interest (*4*). Another

commonly used modules is chemical dye. In the less specific assay, such as traditional agarose gel electrophoresis, intercalating dyes bind to the major or minor groove of DNA double helix structure (*5*). Each method has its own advantages and disadvantages which can be chosen based on the applications.

After the nucleic acid targets are probed with detection modules, specific or unspecific to the sequences, we need a readout from the assay. The readout modules can be direct or indirect. Direct readout could be radioactive (*6*), fluorescence (*7, 8*), or luminescence (*9, 10*). These signals are generated directly through the conjugation functional group on the detection modules. Others could be evaluated through procedure-based results. In the colorimetric LAMP reaction (*1*), the colorimetric change in the reaction color does not come from any of the probes and primers in the reaction. Instead, the color change is from a pH change in the reaction due to the active process of target amplification. Some other probes are functionalized with a conjugate group to which another directly labeled complementary could specifically bind, such as streptavidin-biotin pairing (*11*).

The output of the binding form can be finally quantified by either imaging instrumentation or direct readout machines which can read absorbance or fluorescence intensities. The specific usage of instrumentation is based on the readout module of the probing system and the nature of samples themselves. If we are to investigate the nucleic acid in cell context, especially the spatial distribution of genomic elements, the ideal measurement ought to be microscope. If the samples are cell-free context of nucleic acid in a free lysis form, the detection could be purely in laboratory plates or tubes and would be easier to handle in plate reader.

1.2 Current technologies and their limitations

Common technologies have two major sequence specific detection modules: nucleic acid- and protein-based.

DNA probes are usually used in non-live cell environments for nucleic acid detection. Some recent technologies, such as Flow FISH (*12*), are also implemented in flow cytometry for cell transcriptome expression quantification. DNA probes bind to the specific targeted sequences and are attached with direct fluorescence labeling or indirect readouts modules. Southern blotting (*13*) and TaqMan assays (*14*) use the specific nucleic acid pairing to reveal the targeted nucleic acid of interest. Fluorescence *in situ* Hybridization (FISH) is a traditional method to detect single nucleotides locus in the cells and could be easily expanded to tissue sections (*15, 16*). Based on the Watson-Crick base pairing, these probes are highly specific and unique to their respective binding region. Fixed specimens staining has intensive and rigorous washing steps with various temperatures and salt concentrations, so the samples have good signal to noise ratio (SNR). But this also becomes a limitation to the sample preparation. The FISH protocol typically takes hours for probes binding to the target nucleotides sequence. And then it takes another few hours for washing redundant probes away from samples. The protocol supplies good SNR but is limited to low throughput. The procedure also involves global denaturation which damages the chromatin-DNA binding relations and further causes inability of detecting chromatin accessibility (*17*).

FISH also expands to detection of RNA at single molecule level in the cells, particularly in cytoplasm (*16*). With similar assay to DNA-FISH but without the global denaturation step, small oligo nucleotides can probe to the targeted molecules to achieve enrichment in a small region.

With discovery of modifiable DNA binding proteins with or without facilitation of modified ligand, these proteins were used for the genome locus detection. They are zinc finger protein (ZF), Transcription Activator-like Effector (TALE), and Clustered Regularly Interspaced Palindromic

Repeats (CRISPR) with its associated protein family. All these proteins have high affinity to the desired DNA sequences, some of which require additional component for binding.

Zinc finger protein was used before by tagging with Green Fluorescence protein (GFP) for repetitive sequence in microsatellites (*18*). It was also reported that the Zinc Finger protein fusion with nuclease domains to implement dimerized double strand breaking (DSB) (*19*). But zinc finger protein is not subject to easy synthesis procedure and presents challenges for an ordinary lab to adopt it.

The CRISPR-associated protein was engineered with various functional domains and binds to specific DNA or RNA targets with help of guide RNA (gRNA). Cas9, one of the Cas family proteins, was engineered to remove the nuclease function (dCas9) and became nucleic acid binding protein with facilitation of guide RNA. With certain limitations such as protospacer adjacent motif (PAM) sequences, the dCas9 protein can be used as a probe both in living cells and fixed cells (*20-23*). Recently, in the trend of developing new assays for pandemic, CRISPR system was also implemented with Cas 12a, 13, and 14 to explore the point of care applications for fast testing by cleavage (*24*). Indirect readout modules are traditionally more sensitive to low copy input molecules and able to boost the signals. However indirect readouts also introduced extra operation procedures and potential background signals. CRISPR systems present applications towards nucleic acid imaging in both live cells and fixed specimens. Two-component system, Cas family protein and guide RNA, poses an uncertainty towards the binding efficacy.

The current technologies pose tailored optimization towards their own areas and scopes. The customized features demonstrated drawbacks for a universal nucleic acid detection platform. The platform we need needs to address the following challenges: 1) ability to bind nucleic acid in its native form; 2) ability to bind with short time and high affinity; 3) ability to achieve single

molecule sensitivity; 4) being easy to program and probe to nearly possible sequences; 5) convenient delivery and adaptation to different assay format.

1.3 TALE as a protein-based nucleic acid detection module

TALE consists of multiple repeats binding to the major groove of double strand DNA. TALE is originally a bacteria-oriented activator by *Xanthomonas (25, 26)*. It is found that TALE is composed of highly conserved repeats with 33-34 amino acids. The 12th and 13th amino acids are divergent depending on the targeted DNA sequence. But these two specific positions are conserved towards specific nucleotides, or more precisely, deoxyribonucleotides. The earlier literature has reported the di-residues of amino acids corresponding to each of four nucleotides (*25*). These di-residues are called Repeat Variable Di-residues (RVDs). Through Golden Gate assembly, it is possible to synthesize multiple coding sequences in a high-throughput method (*27, 28*). TALE has high specificity (binding up to 34-nt DNA) and no design limitation (***Figure 1.1***).

Traditionally TALEs were fusion with cleavage domain of FokI to dimerization leading to double strand break in genome DNA (*25, 28-35*). The dimerization required by the FokI enhances the specificity because it requires longer binding sequence and proper sequence spacing between two halves determines the orthogonality of effective region.

But one thing should be noticed. Since the base pair binding is very specific and strong, the binding free energy is around -40 kcal/mol. The dCas9 protein binding to the sgRNA is around -6 kcal/mol (*36*). So, the affinity determining step is the dCas9-sgRNA binding. TALE binding to DNA sequence is simply DNA-protein interaction. The dissociation constant is around 3.3 nM (*37, 38*), which is like the value for the dCas9-sgRNA (2.7 nM). Therefore, TALE could be more sensitive in DNA binding than CRISPR/dCas9. The K_d value for transcription factor-DNA is at 1

μM magnitude and the bonding strength is lower than TALE-DNA interaction (*39*). CRISPR (*40*) and TALE binding to the DNA might be competitive to the native chromatin status. The many other advantages of TALE binding still render TALE as a good candidate for nucleic acid detection in cell environment (***Table 1.1***)

Table 1.1 Comparison between different DNA-binding probes to date (adapted from (41))

Probe	DNA-FISH	CRISPR-Cas9	TALE-based system
Difficulty of probe generation	Reliable	Variable	Difficult
Immediate readout	No	Yes	Yes
Visualize intact DNA	No	Yes	Yes
Fixed cell imaging	Yes	Yes	Yes
Live cell imaging	No	Yes	Yes
Sensitivity to chromatin accessibility	No	Possible	Possible

Due to the specific DNA binding features and less design limitations compared to other similar binding domains, such as CRISPR and Zinc Finger, TALEs can conjugate with multiple functional domains for the applications in genome and epigenome editing. The most common application is the direct conjugation to FokI cleavage domain, which is part of restriction enzyme FokI but without the DNA binding region. The cleavage domain of FokI needs to be dimerized to be functional so two TALE Nucleases (TALENs) must be close enough to allow the proper spacing and orientation (*42, 43*).

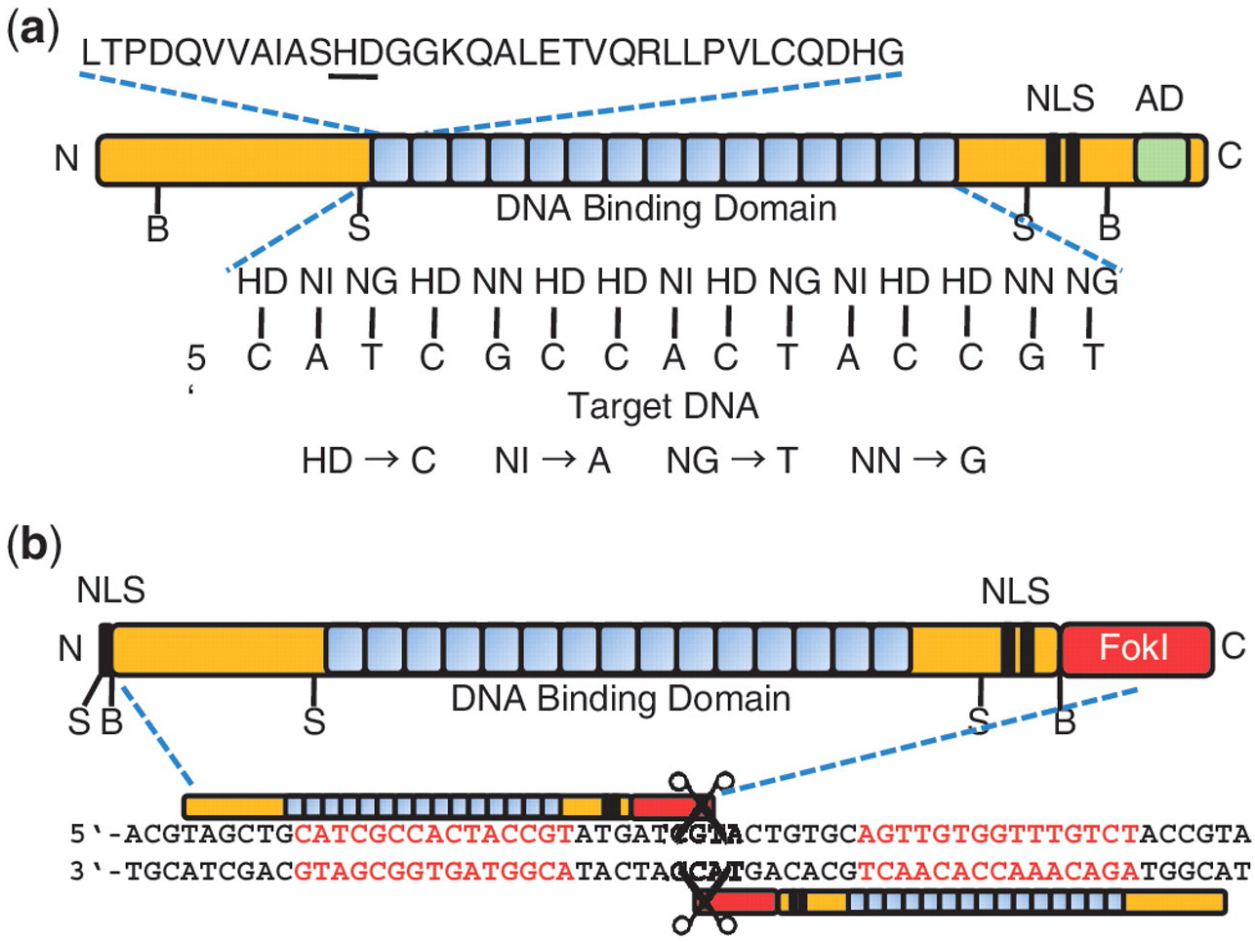

Figure 1.1 TALE structure and RVDs

Figure was reproduced from Cermak et al, 2011 with license © 2011 Oxford Academic, in which (**a**) indicated the unique RVD binding correlations between two amino acid residues and nucleic acid; (**b**) showed a sample of mechanism of TALEN in introducing DSB to DNA sequence.

TALE can also be functionalized with epigenetic domains to exert gene up- and down-regulating by blocking the gene transcription activity. As reported, TALE-KRAB was used in mice hematopoietic stem cells to regulate miR-302/367 (*44*). Compared to siRNA strategy, TALE-KRAB has more flexibility towards the promoter or enhancer region. This flexibility allows more design possibilities and screening opportunities. Compared to traditional gene knockout by DSB, epigenetic regulations only require a single half of TALEs and could be deployed in multiplexing. There are other efforts in silencing gene element by methyltransferase DNMT3A and DNMT3L or hydroxylase TET1 (*45, 46*).

TALE was used by several groups for repetitive sequence targeting, mostly telomere (*47-50*). A stable labeling against telomeres in embryonic stem cells were proposed to track the movement of telomeres during mitosis and differentiation in living cells.

1.4 Universal strategy to detect nucleic acids in complex setting

To explore a universal platform for the nucleic acid detection in various biological samples, the protocol series need to be adequately robust to support TALE probes implemented in complex settings and be flexible in the probe production to adapt different input requirements (***Figure 1.2***).

There are two sides of the story to dissect: the target of interest and the detection module adaptations. Targets of interest are nucleic acids present in live cells, fixed specimens, and cell-free context. These targets are choices of dsDNA, ssDNA and ssRNA. It was reported that TALE protein can bind to dsDNA and DNA-RNA heteroduplex. To make designated targets accessible to TALE protein binding, targets of interest need to be pre-processed. Single stranded DNA and RNA can be complemented with matching ssDNA oligo. The direct complementary method translates the targeted molecules 1:1 towards the desired forms. Sometimes the desired molecules

are in scarcity. To detect that scarcity, amplification of target molecules might be necessary. Traditional amplification technology, such as PCR or LAMP, increases the final input molecule number and transforms them towards double stranded structure.

Complex settings might need us to alter the choice of readout modules. Direct labeling is straight-forward and less time-consuming but has limitations in the signal's robustness. Due to the molecule number from individual detection modules are most likely to be few, the required probe number and input detection targets need to be increased. Indirect labeling requires extra steps to conjugate direct readout modules with the detection modules, such as Immunofluorescence (IF) or Tyramide Signal Amplification (TSA) (*51*). Because the signal boost strategies change the number of readout modules attached to a single detection module, the necessary detection probes and input target molecules number can be titer down.

Depending on the environment where the samples are presented and estimation of molecules abundance, the readout modules can be adjusted to either direct or indirect labeling. Also, the input nucleic acids can be adapted to single copy transformation to multi-copy amplification.

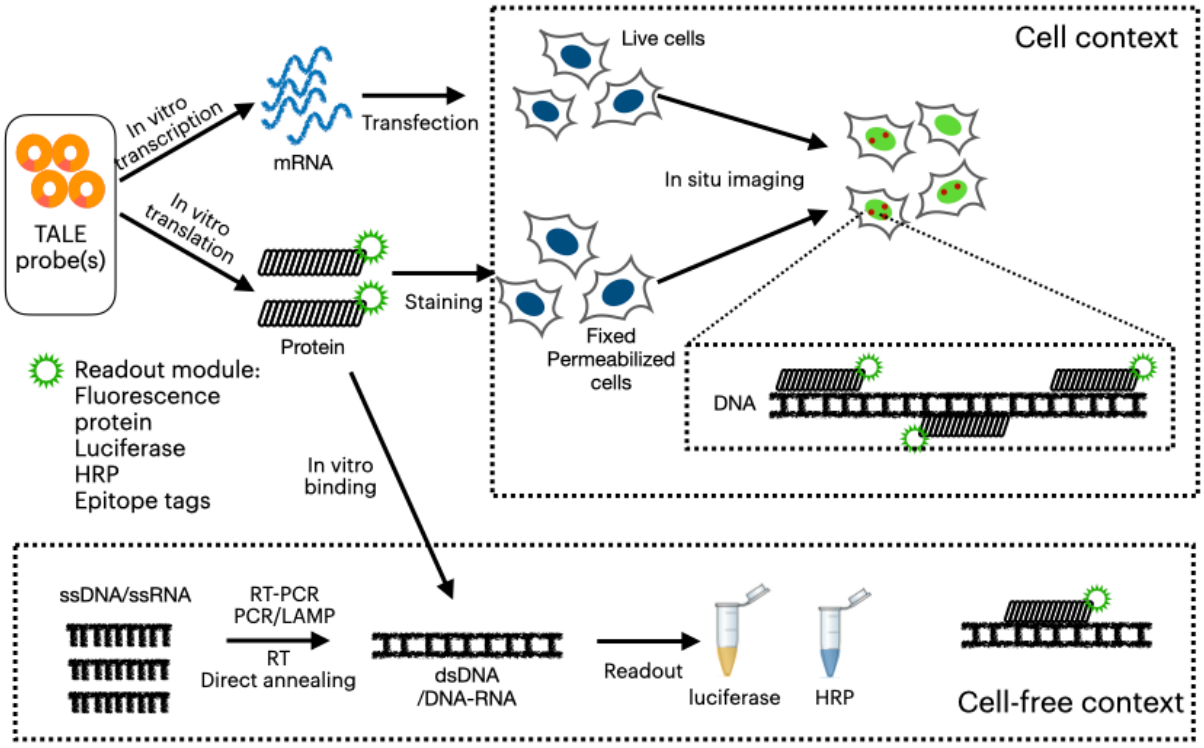

Figure 1.2 Overview of TALE platform in nucleic acid detection

TALE probes were synthesized in plasmid vectors and introduced into both sample contexts. Cell context could have live cell and fixed cells samples, where we deployed two different strategies to add TALE probes. The final spots in imaging would be multi-copy recruitment of probes on designated detection loci. For cell-free context, we processed the raw samples to complementary double strand structure before adding TALE protein cocktail. The direct readout could be luminescence by luciferase or colorimetry by Horseradish Peroxidase activities.

Chapter 2 Methodologies and Rationale

TALE platform for nucleic acids detection in different cellular and cell-free contexts is robust and consists of various efforts input from molecular biology to sequencing. This chapter introduces the rationale of all methods used in the technology development procedure in Chapter 3 to 5. This will include probe designs, the characterization and quality control procedures following design and assembly, experiment designs, rationales, and methodologies.

2.1 Methodology overview

The specificity of the TALEs is crucial to the general nucleic acid detection efficacy and performance. One benefit of the TALE designs in genome wide context is that TALE designs only limit by the initial thymine binding from N-terminus, which enables massive designing possibilities for tiling and saturating multiple probes into a single region. However, it also leads to increased complexity of screening hundreds of designs.

Previous studies showed estimated fluorophores needed in the context of live cell imaging (*20*). Yet, we still need to determine that number empirically, which varied from the reported value from literature. As a reference, the number of the fluorophore will be limited to c.a. 100, while in a single region of interest (e.g., 1 kb), there could be more than five-fold possible designs. We need to implement several procedures to evaluate the input designs and output quality control.

2.2 TALE probes design rationale

2.2.1 TALE DNA binding domains

We summarized some design rules through previous research experiences in TALEN editing and related research works from literature (*52-54*). The preliminary selection criteria were based on the length of the DNA binding domain to be 17 to 21. In common situations, shorter binding sequence results in lower specificities and increased events of off-target binding in genome-wide. On the other hand, longer target binding sequences increased the difficulty in assembly and lowered down binding efficiency due to the less favorable binding enthalpy.

As reported in other literature as well, TALE protein has favorable specificity towards thymine and cytosine due to possibly strong hydrogen bonds (*52*). Therefore, binding sequences possessing higher combinatory thymine and cytosine percentages could be potentially more specific.

There are more prior research indicating more design rules for specificity improvement (*30, 53*). However, the described two criteria have limited down the design possibilities, and we designed more methods in quality control to screen.

2.2.2 TALE vector modular designs and cloning

TALE probes mainly consist of five sections: pre-N-terminus domain, N-terminus, modular DNA binding domains, C-terminus, and post-C-terminus functional domain. Among these, N- and C-terminus tend to be highly conservative so traditionally there is no modification towards them. For the modulated DNA binding domains, the variability is fairly limited to the di-residue in the middle of each 34 amino acid DNA-binding single module. To functionalize the TALE, functional domains need to be added to outside of TALE main coding frame (***Figure 2.1***).

Detection of nucleic acids in cell contexts needs direct imaging of fluorophore attachment or any tags which are able to form fluorophore conjugation. In the current live cell setup, fluorophores are popular to direct labeling the target with single or more fluorophore on one probe, such as

mNeonGreen and conjugation tags such as HALO-tag or SNAP-tag. N-terminus could also be functionalized with detection tags. However, the native N-terminus tends to be conservative in the binding events (*55*). Additional coding sequence added to the head of proteins might destabilize the binding event.

More fluorophore recruitment to single probe can reduce the probe number by the same fold. Tandem arrays of epitope were applied to either N- or C-terminus to enable co-labeling of single region. Moreover, tandem epitope array is smaller in the coding sequence and helps reduce the side products forming as truncation. Smaller functional domains pose less perturbation to the DNA-binding domain stability.

Instead of fluorophore and immune-tags, functional domains were cloned with enzymes to further diversify the applications of TALE probes towards multiple combinations between samples and targets. The direct enzymatic reaction from TALE probes makes the experimental flow faster and reduce turnaround time.

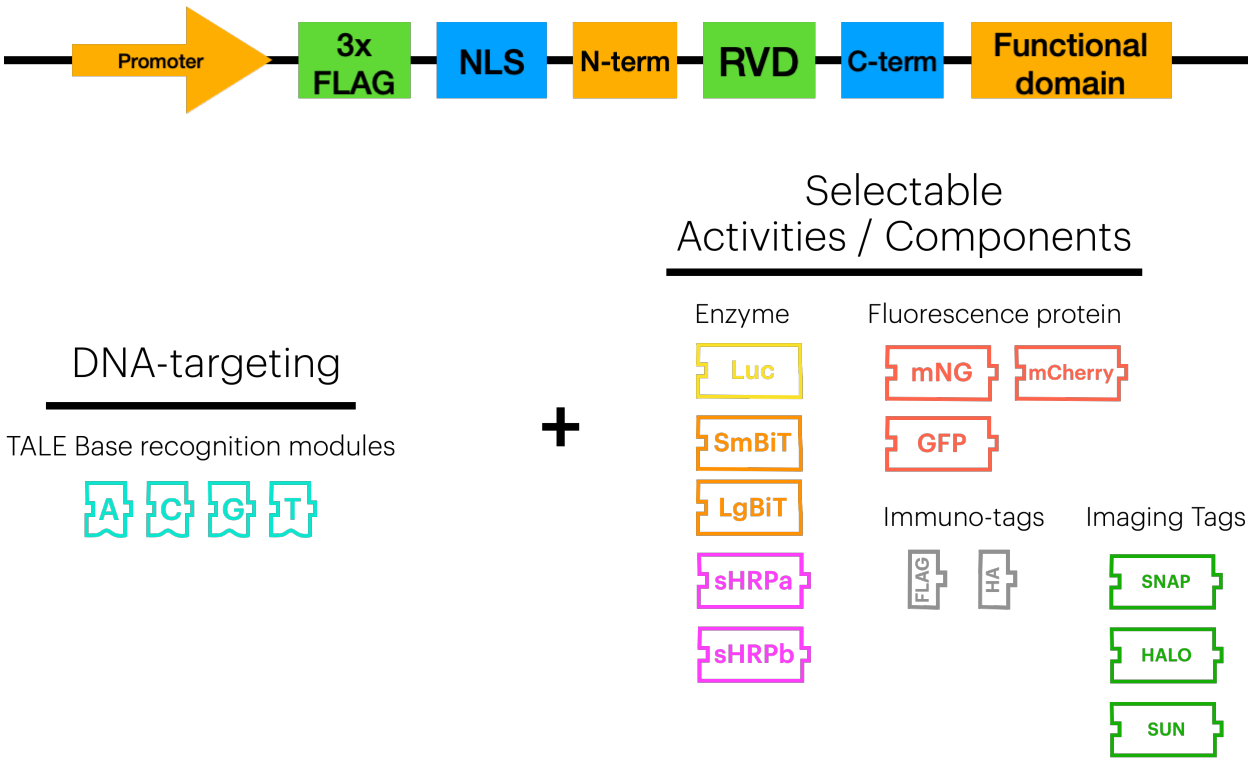

Figure 2.1 Illustration of sample expression cassette design

The expression cassette included a double promoter, CMV and T7, in which T7 promoter is crucial for mRNA and protein *in vitro* production. 3xFLAG tag and functional domains are adjustable depending on the applications. From NLS to C-terminal will be conservative. Functional domain can be easily diversified through the design of active components vectors.

2.2.3 Functionalization with epitopes, enzymes, and luminescence units

We created a large library of destination vectors for the assembly of TALE probes to diversify the applications and assay format. The applications of each vector were adapted for the convenience of assay format and instrumentations.

In the live cell imaging section, C-terminus of TALE protein was functionalized with monomeric NeonGreen, which is one kind of GFP derived from Branchiostoma lanceolatum (*56*). A single Glycine-Serine peptide linker was introduced between C-terminus and the functional domain. Traditional superfolder GFP and monomeric Cherry fluorescence proteins were also attached to the C-terminus for potential applications in multiplexed labeling. For high resolution imaging, we also prepared compatible vector library with SNAP and HALO-tag.

In the cell imaging *in vitro* section, C-terminus of TALE protein was functionalized with either 10 tandem copies of GCN4 epitope (*57*) or 4 tandem copies of HA epitope. The N-terminus of TALE protein was functionalized with 3 tandem copies of FLAG epitope or single copy of V5 epitope. GS linkers were inserted between domains and between each tandem copy to guarantee the freedom of the epitope array and binding of antibodies towards each epitope. The purpose of the multiple tandem copy introduction is mainly to amplify the signals from fewer probes necessary.

Detection of nucleic acids in the non-cell contexts reused the epitope-containing TALE constructs from the *in vitro* cell imaging section. TALE-ELISA relied on these epitopes to add secondary antibodies conjugates with Horseradish Peroxidase (HRP). In the future applications, liquid phase detection of RNA needs the N- and C-terminus of the TALE proteins to be functionalized with split HRP subunit a and b (*58*) or split nano Luciferase LgBiT/SmBiT (*59*).

2.2.4 Affinity purification of TALE probes

The production of TALE protein relies on *in vitro* transcription/translation. The commercialized kit used in our research used rabbit reticulocyte lysates as a source of protein and post-translational modification environment (*60*). One potential drawback of this assay is their low yield (*61*), but the limit of detection is reasonably low. The developed assays have low demands of protein and nucleic acid molecules number.

The final products of the *in vitro* transcription and translation kit could be diluted in the known buffers. Due to the big dilution factor (usually lower than 1:30) and low molecules needs, the original reaction system would be highly diluted and not contribute to the perturbation of desired salt concentrations in destination buffer system. Also, the majority of inactive ingredients are rabbit hemoglobin, which is essentially inert in imaging and TALE-ELISA protocol. Therefore, unpurified IVTT products are suitable with our described assays and research purposes.

However, to standardize the protocol for future extension towards point-of-care, we have developed a protein purification protocol in high-throughput scale for the low-yield input/output. This protocol is based on my research efforts in Chapter 5. TALE proteins are able to bind to their targeted DNA sequences in various buffer environment. This fact enables the isolation of desired proteins from liquid phase and anchoring towards solid phase such as magnetic beads or plates. On the other hand, salt concentrations are critical to the stability of DNA-binding proteins with their corresponding ligands (*62*). Higher salt concentrations could perturbate the hydrogen bonding and/or ionic bonding. Short-time surge of salt concentrations sabotages the DNA-protein binding events but does not influence the re-folding of protein to its functional state. Upon the spiking of salt in the binding system, the proteins are released from the solid phase and resolubilized into the

liquid phase (***Figure 2.2***). This methodology is similar to other phase exchange methods for DNA, RNA, and protein purifications.

This method is particularly easy to use in high-throughput and automation system. However, it limits the input of desired protein to be low yield. Thus, the purification protocol could be operated at maximal binding capacities.

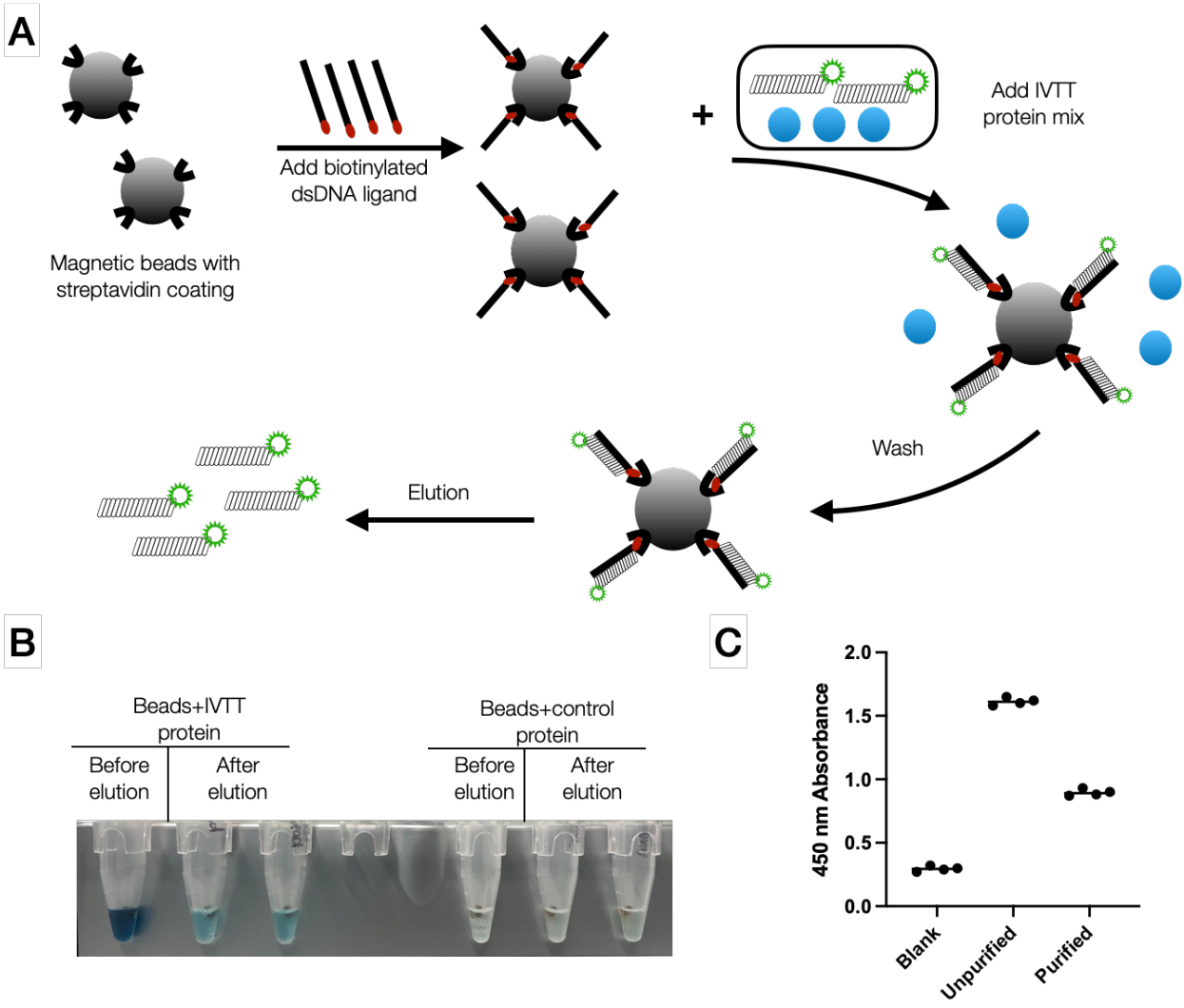

Figure 2.2 TALE-ELISA based low quantity protein purification

(**A**) is the illustration of TALE-ELISA based protein purification protocol. (**B**) is evaluation of IVTT protein specificity on the beads surface. Left three samples contained beads with dsDNA ligand and the right three did not. Before and after protein elution step, the beads were saved for residue protein fouling test. After added with TMB substrate, the secondary antibody conjugate with HRP binding to the FLAG tag on TALEs would show the protein abundance left on the beads surface. The darker the color is, the more protein residues are there. This showed the TALE protein fouling towards the beads surface is minimal compared to the beads with ligand immobilization. (**C**) is functional assay of TR242-ST protein purification. Protein samples were added assuming 100% recovery efficiency and the sample volumes were adjusted. The true recovery rate was around 50-60% depending on the calculation method.

2.3 Quality control and characterizations

The output control from assembly pipeline was conducted in serial logic steps to evaluate the TALE protein size, functional group intactness, on-target binding on designated dsDNA ligand, total off-target binding profile in fixed specimens, and genome-wide live cell binding profile (***Figure 2.3***).

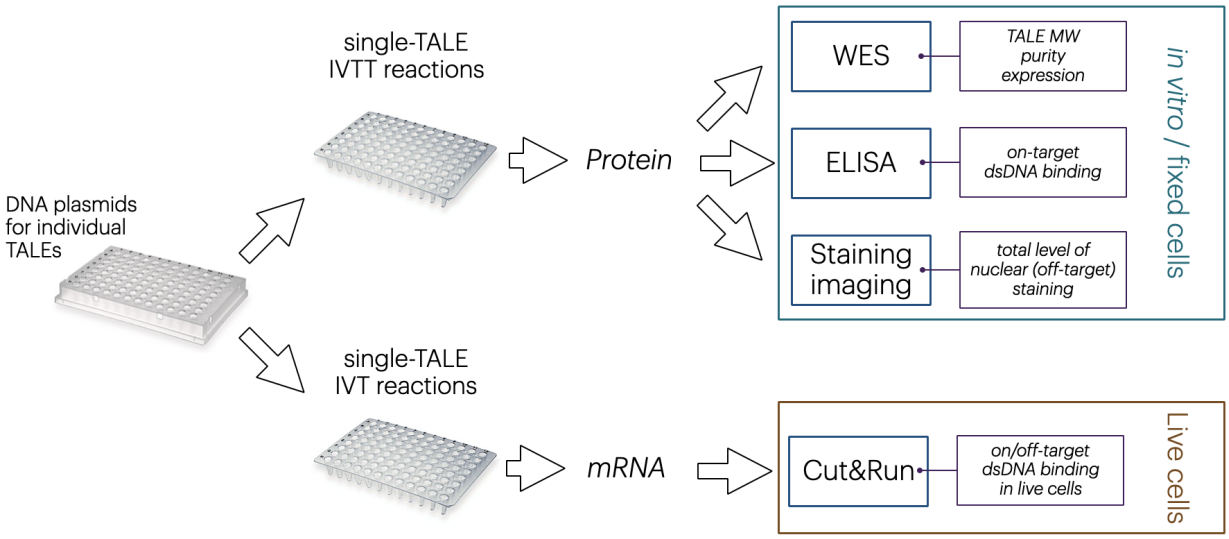

Figure 2.3 TALE probes QC and characterization

TALE probes were screened after production QC. *In vitro* translated protein probes were evaluated on their quality, on-target binding affinity, and global off-target binding profile in fixed cells. CUT&RUN assay evaluated the on-/ off-target binding events in genome in a live cell setup.

2.3.1 *In vitro* translation expression and yield quantification

In vitro transcription and translation kit (Promega) did not explicitly demonstrate yield of targeted protein. Some other available kits from other vendors suggested inconsistent values depending on properties of proteins themselves and the expression vectors where the protein coding sequences were inserted.

Expression of the protein was extensively verified by Western blotting and total protein assay (Protein Simple). Size and area under curve of protein curves in the digitized graph appear similar. We believed that all the yields of different TALE proteins in IVTT reactions were in a proximity range. Traditionally, DNA or RNA use either absorbance at 230 nm or 260 nm wavelength to determine concentrations based on the proportional relationship. RNA concentrations can be read by fragment analyzer and conventional plate reader. Two different reading methods provided two reading values, and it was not difficult to assume that the values derived from two methods were deviated from the true values. I proposed that the concentration of the protein would be in a formulated relationship with the Area under Curve (AUC) or the peak intensity. However, the distribution of a single protein was wide at high concentration input. Sole interpretation of peak value did not reflect the concentration change in the input samples.

We adopted a standardized 3xFLAG peptide from vendor (Sigma) with 2 mg/mL stock concentration. The control proteins in serial dilution of 1:10 were analyzed through Compass software. The normalized AUC values were plotted against the known concentrations. It appears that the plot has an obvious sigmoidal curve. The semi-log plot indicates a linear relationship between the value of AUC and the concentrations with statistical significance (***Figure 2.4***).

$$Area\ under\ curve(AUC) = 9 * 10^6 * \log_2(Protein\ conc.\ in\ IVTT) - 2 * 10^7$$

AUC is the direct value computed from Compass for Wes (Simple Wes), and protein concentration is in unit of ng/μL.

TR242-SunTag protein was also performed Wes with series dilutions. The AUC value reduction folds reflect the same dilution factor because in the low concentration range, the correlation between AUC and concentration is close to a linear relationship. It further explained the AUC value is more suitable with the concentration determination using these reference methods. The value of the AUC was input into the formula and a concentration value was calculated to be 150 ng/μL, which is approximately 1.5 pmol/μL (TALE proteins are around 100 kDa and exact sizes depend on the functional domain and RVD length).

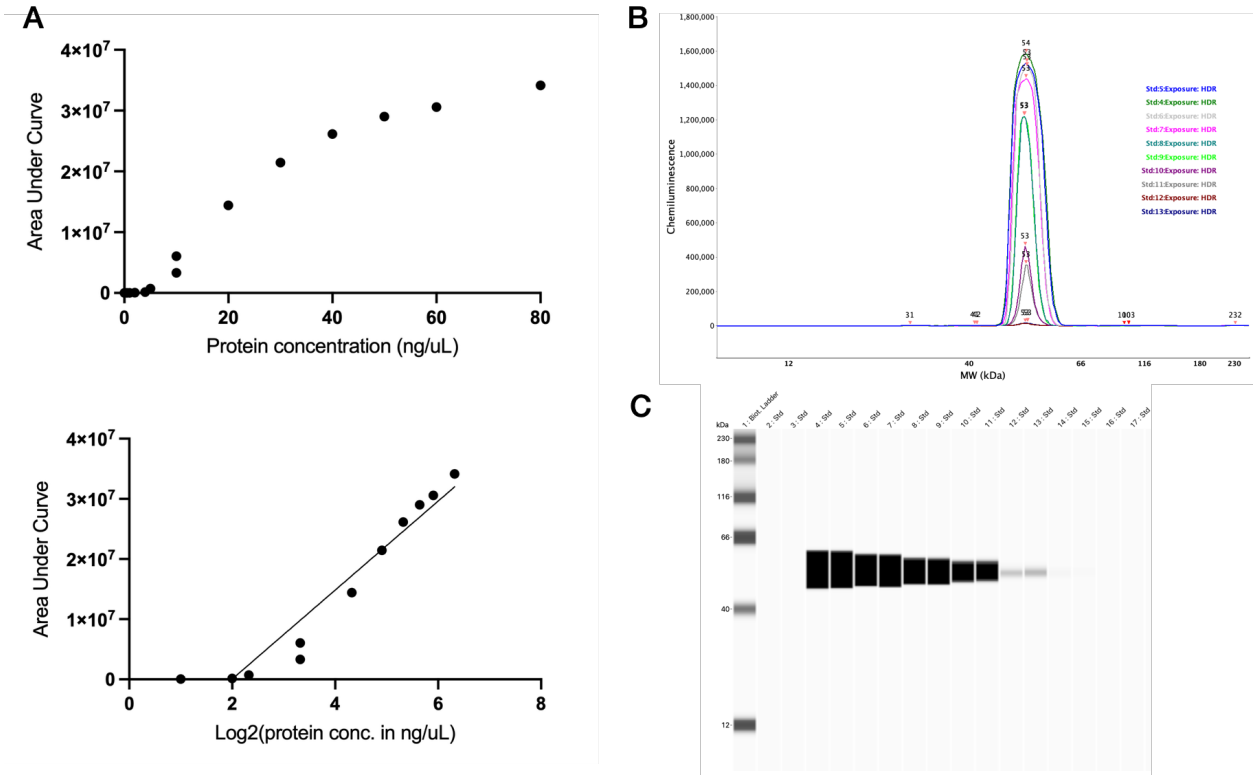

Figure 2.4 IVTT yield quantification and Western blot

(**A**) showed relationship between area under curve vs. FLAG peptide normalized concentration standard curve. The direct figure was a sigmoid curve indicating linear relationship at semi-log graph. (**B**) and (**C**) was example of Wes result of 3xFLAG peptide with anti-FLAG (M2, Sigma, 1:100) as chemiluminescence intensity graph and Western blot gel image. (**C**) had protein concentration gradient from left to right (where concentration were indicated to the left). More QC Western Blot gels were listed in Figure S1.

2.3.2 Specificity evaluation

Although the specificities of the TALE probes were evaluated by alignment with the whole genome, there are two major issues associated. Firstly, BLAST is a computational method to align target sequences to the reference genome. The processed alignment number heavily relies on the accuracy of the reference genome. For the repetitive locus, the repeats number and length varied across cell lines and individual cells. Reference genomes typically use shot-gun whole genomes and the library built consists of small fragments in the range of 150-200 base pairs including adapters. Repetitive sequence information might be compromised in the indistinguishably library segments (*63*). Secondly, it is difficult to screen partial binding events of probes towards imperfect nucleic acid. Strict screening criteria with all possible mismatch's intolerance might exclude excess probes and loose screening criteria tolerating potential mismatches was not functioning in low specificity probes selection.

ChIP-seq is able to evaluate the TALE probes binding events in the genome wide profile (*64, 65*). However, the protocol is trivial for large scale of samples and procedures. Cut under target and Release under nuclease (CUT&RUN) is an improved assay having similar function as ChIP-seq (*66*). This method provides insights of *in vivo* binding profile of one or more TALEs. Recently it was proposed to have higher throughput of binding events profile (*67*). By identifying the peaks, the assay draws out the genome wide binding landscape; by normalizing the fragment counts at base pair resolutions, the binding strength could be evaluated at different time points. In TALE detection in cell contexts, the pooled probes were evaluated on time series binding affinity and off-target binding profile. In CUT&RUN assay, we evaluated the binding profile of TALE probes pool binding to the designated target (***Figure 2.5***). The probes shown in the figures are the ones used in the Chapter 3 and 4 for genomic loci imaging.

In Figure 2.5, (B) showed designed 60 probes around two DHSs in the proximity of PPP1R12C promoter, which is also known as AAVS1 safe harbor (*68*). From the total pool transfection, B and D set of AAVS1 probes showed good binding on the designated sequence and showed a Gaussian distribution peak near the center of binding sequences. The wave-like peak indicated the combinational base pair resolution reads count of multiple individual peak overlapping. C set had pretty bad binding possibly due to the heterochromatin status between promoter DHS and intron DHS region. This proved the possibility of TALE proteins not penetrating the native chromatin-DNA status but also is because the probes binding to this region were not strong enough

In Figure 2.5, (C) showed 50 LCR probes distributed to four DHSs in HS1-4 in beta globin LCR region. This LCR region is one of symbolic control region using multiple DHS to control upstream coding sequence in developmental biology (*69-71*). Individual probe set on each DHS and combinational transfection showed good binding. We noticed in HS1 transfection, despite of HS1 peak there is a minor peak shown in HS2 site. Rather than side products or off-target binding events from the assay, we tend to believe that it is a sign of active chromatin hub in the local region (*70, 72, 73*), where the region in proximity was also under protein of MNase digestion strength.

However, in each probe pool, the designed probe number is too expensive to build a single probe binding profile library. We developed another assay to evaluate the off-target binding profile in the cell-context, exclusively in *in vitro* context though. Through the stringent binding and washing steps, any single probe binding in the imaging sample should have three behaviors: on-target binding on DNA, off-target binding on DNA, and off-target binding in global cell context. Entrapment of the TALE imaging probes lead to background level obscuring the signals. To establish the baseline for the probe specificity on the similar DNA targets and cell environment influences. Implementing the single probe staining (SPS) assay on the individual TALE probe

could directly demonstrate the off-target binding events. This assay might have sensitivity issues with few off-target binding in DNA and cell context. But the point was to identify the major contributor to the overwhelming background (***Figure 2.6***).

In the example figure of SPS mean intensity quantification, LCR probe set has consistent signals at low value, which indicates either no or few binding events in the cell context. To exclude out non-binders, the next subsection will talk about evaluation of on-target binding between TALEs and their respective dsDNA ligand. High intensity single probe potentially highlights the possibility of obscuring the true intensity. At further investigations of single molecule labeling, the signals could be very low because of low number binding probes and even minor background sabotages correct spots identification.

In a separate investigation to explore the TALEN specificity by p53BP1 assay and SPS score (data not shown), the SPS intensity showed negative correlation with the specificity score from TALEN editing, which indicated high SPS intensity was reflecting the low specificity in TALEN editing.

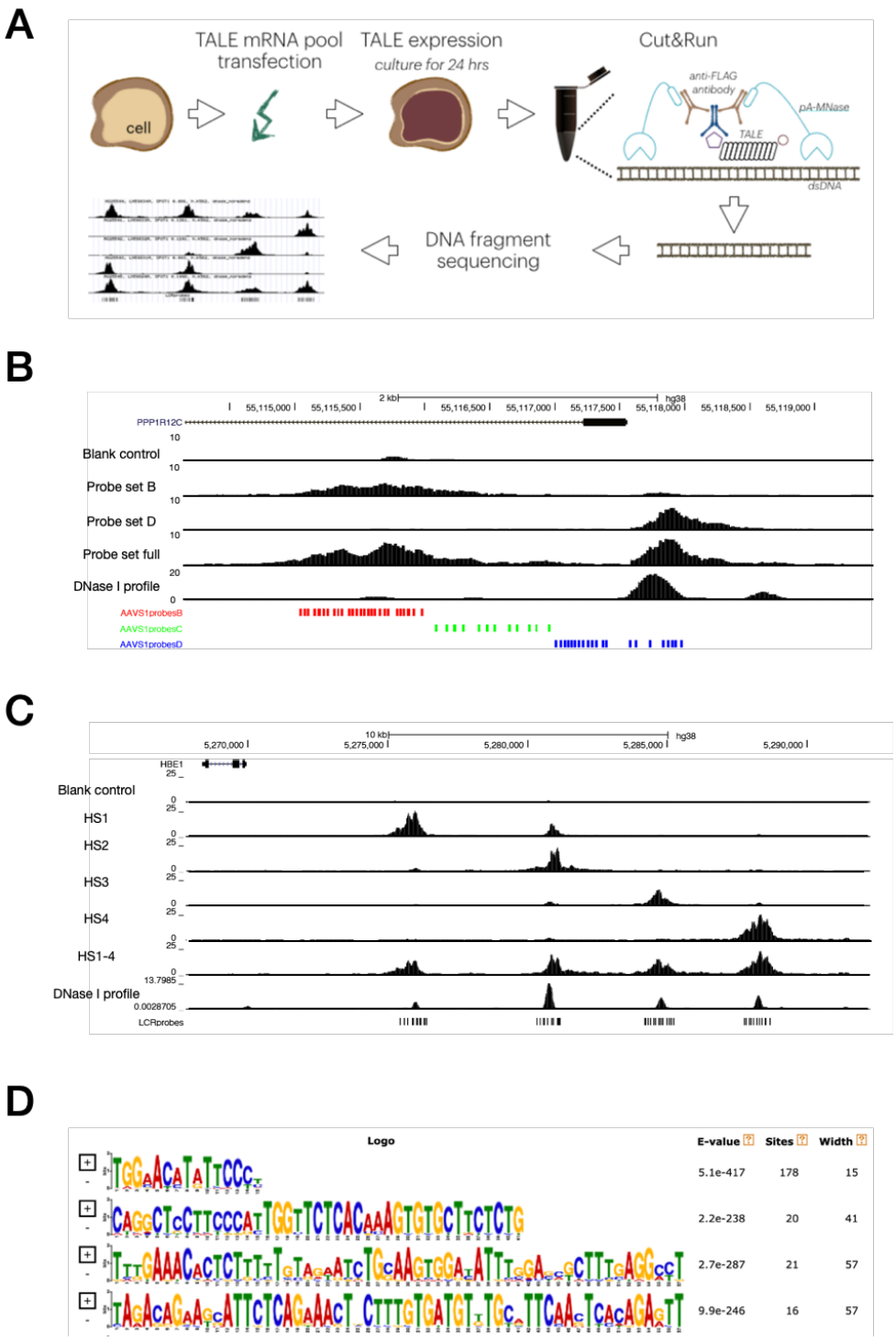

Figure 2.5 Pipeline for evaluation TALE binding specificities in vivo with TALE probe pools by CUT&RUN

(**A**) is diagram of how CUT&RUN assay works. (**B**) and (**C**) were samples of AAVS1 and LCR probe set binding events. Cells were collected at 24-hrs. Y axis indicated the normalized density for the total sequencing run but values were not normalized to the intrinsic *E. coli* genome

alignment. From top to bottom were blank control, individual probe sub pool, and total pool results. Then we include DNase I profile in the corresponding cell line and the probe location in the genome. **(D)** is a sample of off-target motif search for the AAVS1- B probe set. The top motif corresponding to one of TALE probe partial binding site.

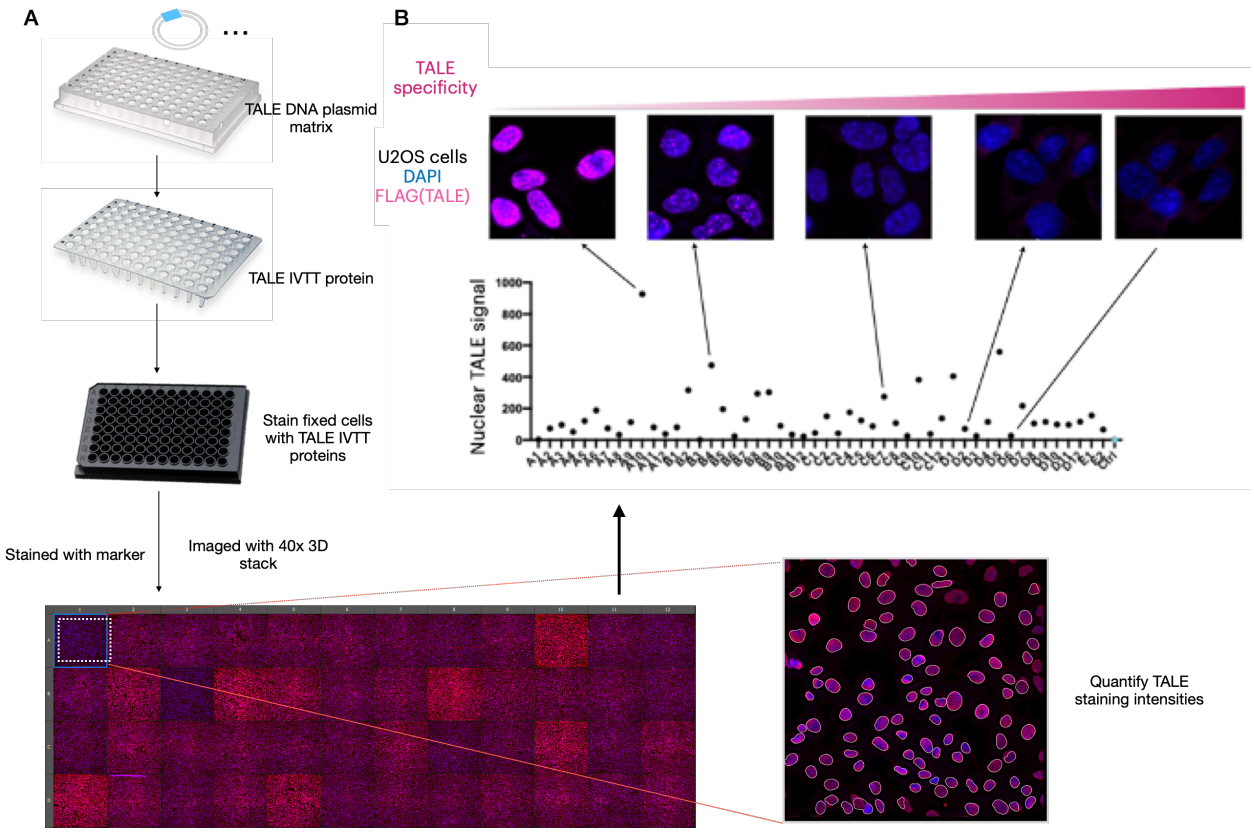

Figure 2.6 SPS evaluation of specificity in single TALE probe staining

(**A**) is the flow of Single probe staining (SPS) assay. TALEs DNA and IVTT were prepared in 96-well format. Therefore, the protocol is highly streamline and scaled up. The acquired images were processed to quantify the TALE staining intensity inside nuclei, which showed as mean value in (**B**). Single TALE probe entrapment by either cell context or unspecific binding towards off-target binding can be evaluated through staining intensities. Through quantification of FLAG intensities in the nuclei, the extreme outliers can be identified as major contributor to the pooled cell staining background.

2.3.3 Binding strength evaluation

In the previous section, I introduced using CUT & RUN to evaluate on- and off-target binding profiles of single TALE probes in the live cell context. By normalizing the count number of existing *E. coli* genome introduced by the protein A-MNase native DNA, the normalized count number and frequency at base pair resolution can be compared across multiple samples and experiments. Yet, the single probe CUT & RUN assay is lengthy and bulky in library production. For fixed specimens, live cell environment binding profile does not reflect the same on- and off-target targeting events.

TALE-ELISA is a newly developed method to assess the binding affinity of TALE probes towards their respective DNA target oligos. Nucleic acid binding strength could be assessed through high-throughput functional assay (*74*). Firstly, a certain amount of oligos were immobilized onto the solid phase. Then the protein molecules were incubated with the functionalized surface. If the input protein molecule number is less than the oligos molecule number, stronger binding affinity TALE probes will have higher retaining rate, which results in higher colorimetric readouts. Stronger binding affinity leads to a longer residence time for TALE protein sitting on ligands. Transcription factor (TF) residence time is related to its effect on the local gene regulation (*75*). TALE, as a similar product as TF, has similar effects potentially.

The absorbance of colorimetric readout was measured at λ= 450 nm, but the absolute value cannot compare across different batches of experiment. The consumed time for the color development and cease varied each time so the values were not generated under exactly same conditions. However, the SNR value is relatively accurate. The control groups were also under identical conditions and experimental setup with rest of respective experimental groups. So, SNR values are reasonable to compare for binding strength.

In ***Figure 2.7 D***, the scatter plots correlated the SNR values of two probe sets with same DBD with different functional domains. The strong linear correlations showed the binding strength was solely dependent on the binding domain instead of the functional domains. In the TALE imaging and detection platform, the functional domains should not interfere with the binding domain. In the future experiment, the screening of one specific probe only needs to be conducted once to determine the DBD binding affinity. After the probe is made with different entry vector, the probe binding affinity can be trusted unless there are mistakes during assembly.

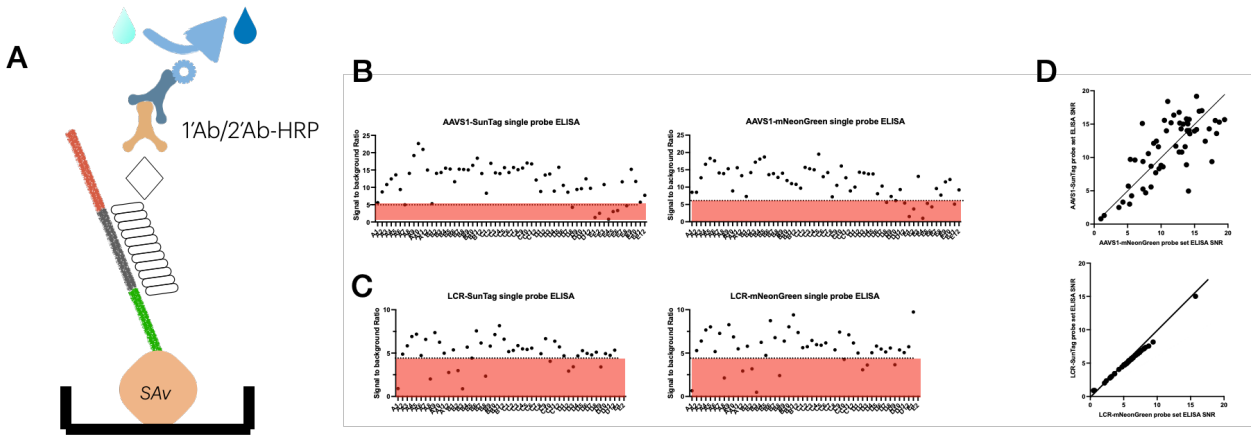

Figure 2.7 TALE binding affinity evaluation for probe set

(**A**) is an illustration of TALE-ELISA setup. The designated DNA targets were amplified with biotinylated primers and mounted to immobilized streptavidin surface. If tested TALE protein probe binds to the target DNA, the signal to noise ratio will be high after readout step with HRP. (**B**) and (**C**) are figures of AAVS1 and LCR probe set. mNeonGreen was for living cell imaging and SunTag was for fixed cell imaging. Dotted line indicated one standard deviation below mean value and shaded region highlighted the probe with weak binding affinity. (**D**) correlated the SNR values from two different functional domains. There was a linear relationship between the SNRs of same DNA binding domain, which further demonstrated the DNA binding domain binding strength is independent from the functional domain.

2.4 Summary

This chapter described the overall strategies on the quality control and evaluation methodologies over the TALE probes designs and characterization post-assembly. I also highlighted some key QC results on the probes used in Chapter 3 and 4 for nucleic acid imaging in cells. The major contributors to the potential off-target binding or non-binding were picked out and subject to exclusion. The assays described above can be applied further to expanded targets and probe sets with potential to possess comparable results with similar protocols.

2.5 Materials and Methods

Vector cloning

For N-terminal functionalization, plasmid pVax with four different last repeats were digested with BstXI and SacII restriction enzymes (New England Biolab) for 1 hour at 37 ^{o}C in the CutSmart buffer. The digested products were separated by 1% agarose-TAE gel. The expected band was 6 kb and purified through Nucleospin PCR and gel purification kit (Macherey-Nagel).

For C-terminal functionalization, plasmid pVax with four different last repeats were digested with BamHI-HF and XhoI restriction enzymes (New England Biolab) for 1 hour at 37 ^{o}C in the CutSmart buffer. The digested products were separated by 1% agarose-TAE gel. The expected band was 5 kb and purified through Nucleospin PCR and gel purification kit (Macherey-Nagel).

Functional domain coding sequences were adapted from Addgene repository sequence. The ordered sequences contained 15 bp of homologous arms to the fitting destination vectors on both 5'- and 3'-ends. The ordered ultramers and gBlock (IDT) were amplified with high-fidelity polymerase to minimize error prone (Phusion, Invitrogen). The digested backbone and PCR inserts were mixed at molar ratio of 1:5 (insert size >300 bp) or 1:10 (insert size <200 bp) with InFusion

(Takara) or NEBuilder enzyme mix (New England Biolab). The assembly reactions were then transformed with 10-beta competent cells (New England Biolab) and selected on Kanamycin plates (Teknova). The cloning was verified by Sanger sequencing (Genewiz) and digestion (when necessary). Large preparation of the destination vectors was conducted through Plasmid Plus Midiprep (Qiagen) or Midiprep Xtra Plus (Macherey-Nagel).

In vitro translation

Production of *in vitro* coupled transcription/translation (IVTT) protein followed the manufacturer's instruction (Promega). DNA plasmid from the TALE assembly product was first normalized to 100 ng/μL concentration before mixed to a single cocktail. For one standard reaction (50 μL), 9 μL DNA, 1 μL 1mM Methionine, and 40 μL Master Mix were added and incubated at 30 ^{o}C for 90 minutes. After mixed with an equal volume of 50% glycerol, the final IVTT can be stored in -20 ^{o}C for 1 month without losing binding activity.

SPS imaging

U2OS cells (ATCC) were seeded into 96-well glass bottom plate (ThermoFisher Scientific) and cultured in McCoy 5A medium (Corning) with 10% FBS (Cytiva) and 1% Penicillin/ Streptomycin (Corning) for 24-hours. Cells were washed with 1X PBS (Corning) twice and fixed with freshly prepared Methanol: Acetic Acid mixture (3:1 v/v; Sigma). Fixed specimens were then treated with RNase A (Sigma) and washed with 1X PBS. IVTT protein was diluted in TALE dilution buffer (1xPBS, 0.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 5 mM TCEP, 0.1% BSA, 0.02 ng/μL poly(dI-dC)) in 1:50 and blocked for 10 minutes before applying to blocking buffer treated fixed cells. After 2 hours incubation at room temperatures, the cells were washed and subject to corresponding

antibody staining by default diluted in antibody dilution buffer (1xPBS, 2% BSA, 0.05% Tween-20). The plate was later stained with DAPI and imaged by Inverted Nikon Ti microscope with a 40x air objective with PFS.

Protein purification

T1 beads (Invitrogen) were used to anchor 1 pmol of amplified biotinylated target DNA ligands in 1x B&W buffer (10 mM Tris-HCl, 1M NaCl, 0.1% Tween-20). Protein(s) were added to the ligand-conjugated beads and incubated in the rotator for 1 hour at room temperature. After 20 mM Tris-HCl wash steps, the immobilized ligand-protein-beads were incubated with a release buffer (20 mM Tris-HCl, 500 mM NaCl, 2.7 mM KCl, 0.1% Tween-20) for 10 minutes. Save the elution and repeat steps for another twice. All liquid fractions were collected and analyzed by NuPAGE and Wes for verification. Elutes were also used for TALE-ELISA to confirm active binding after the purification step.

CUT&RUN

TALE DNA was transcribed to mRNAs by Cellscript mRNA production or mMessage T7 kit (Invitrogen) and transfected into K562 cells by BTX (Harvard Apparatus). The transfected cells were harvested at 24- and 48-hr time points. The harvested samples were processed following Henikoff protocol (*66*) and amplified into library for NGS.

TALE-ELISA

General screening of TALE binding strength was performed in solid phase detection. Pre-blocked streptavidin coated 8-strips or plate (ThermoFisher Scientific) were used depending on

the scale of experiment. The well would be pre-equilibrated with 100 µL BB-2 (1xPBS, 0.5 mM $MgCl_2$, 10 µM $ZnCl_2$, 0.1% Tween-20, 0.1% Bovine Serum Albumin) at room temperature for 5 minutes or longer. 2 µL IVTT products were mixed with 25 µL BB-6 (based BB-2, add additional 5 mM TCEP and 0.02 ng/µL poly(dI-dC)) and incubated at room temperature for 10 minutes. PCR products would not not required to purify if the input molecule number is lower than the well binding capacity (4 pmol). Mix IVTT and 2 µL PCR products and transfer the mixture into equilibrated and decanted wells Incubate at room temperature for 60 minutes and wash 5 times with BB-2. Primary antibodies (anti-FLAG, mouse, Sigma) diluted 1:1000 in BB-1 (1xPBS, 0.05% Tween-20, 2% Bovine Serum Albumin) were added after the washing step. Primary antibodies were incubated for 60 minutes and washed 5 times with WB-1 (1xPBS, 0.05% Tween-20) afterwards. Secondary antibodies (Goat anti-mouse-HRP, monoclonal, Abcam) diluted 1:1000 in BB-1 were added after the washing step. Secondary antibodies were incubated for 30 minutes in dark and washed 5 times with WB-1 afterwards. TMB (ThermoFisher Scientific) substrate was pre-warmed to room temperature prior to addition. 100 µL TMB substrate was added into wells and a blue color would be developed in 2-10 minutes. 50 µL of 450 nm STOP solution (Abcam) was added to stop the conversion and final yellow color could be read by Spectra M2 plate readers (Molecular Devices).

Chapter 3 Detection of cellular nucleic acids *in vitro*

3.1 Introduction

Human health relies on the medical development. Precision medicine was tailored to monitor and treat individual being. The difference between human individuals results from gene expression differences. To understand this difference, it is important to investigate in a bottom-up approach to profile unique genomic signatures. Single cell DHS profiling is possible to analyze how the active genes and regulation behave in the cell composite for individuals. Through the current high-throughput imaging instrumentation, this process will be completed in fast turnaround time.

Global understanding of the genomic locus spatial and temporal distribution in the cell context facilitates the investigations into the chromosome conformation and the interaction between regulatory elements. Fixed specimens, like frozen tissue sections or cell culture, are easy to treat with proper detection module after proper treatments such as fixation and permeabilization (*76*). Upon fixation, the relative positioning of locus of interest are anchored to the proximity of native locations in the living cell state. Compared to the dynamic status of the genomic locus in live cell imaging, the fixed cell imaging captures a snapshot of thousands of cells at different cell states. The current high-throughput imaging instrumentation and technologies enable fast readouts of multiple specimens and substantial amounts of cells in a timely manner.

3.1.1 Overview of genome locus imaging in fixed specimens

The regulation of epigenomes in human cells plays a significant role in maintaining normal metabolism and functioning of cellular activities (*77*). The 3-dimensional organization of the genome is one of the keys in maintaining the epigenetic inheritance and genome stability (*78*). Comprehensive understanding of chromatin structure and genome arrangement will help us on the

expansion of genome engineering and further applications. One way to actively learn the structure and arrangement is to localize the genome loci.

Chromatin is the basic structure of the eukaryotic chromosome. It consists of DNA and protein. When it packages the DNA into dense state, histones form strong electrostatic forces between negatively charged DNA and multiple lysine amino acid residues on histone. This force makes the DNA inaccessible to transcription factor binding and further inhibits the gene expression or other gene activity. When the DNA is treated with DNase I, the DNase I digests the unprotected sites, which typically are the genetically active sites in the genome (*79*). Sequencing the leftover will give us the map of regulatory DNA and active genes at current cell state. The ENCODE project has developed high resolution to base pair level (*80*). These chromatin states tell us more about transcription factor binding kinetics and possible connection in gene variation. It further connects to the common disease and functional phenotyping.

These active regions which are highly susceptible to DNase I treatment are called DNase I hypersensitive sites (DHSs). It contains various *cis*-regulatory elements, ranging from enhancer, promoter, silencer, insulator, to locus control regions. Certain features are noticed. Active genes and regulatory elements in different cell identities are very selective and exclusive. And cancer cells may have different DHS activations than the normal, original cells (*81*).

When we construct such a big library for different cell types and their DHS profiling, the question coming next is what we should do with them. These DHSs are subject to transcription factor binding and should be also accessible for visualization probes binding. Several applications could be considered. Since different cell identities have different DHSs, certain signature DHS could be labeled and barcoded as cell identity symbols. If the probe can bind to the cell of interest exclusively, we can identify the specific cell from multiple cell mixtures or even tissue sections.

Similar thoughts can be applied into abnormal cell identification in cancer cells or somatic mutated cells. By detecting the abnormal DHSs, it may be developed into a diagnostics method. Incorporating with microarray, this detection can be run in a high-throughput manner.

Imaging, as a detection tool, could resolve the profile to single cell level. DNase I-seq provides an overall picture of how a group of cells behave and how their DHSs demonstrate. However, this limits the resolution to thousands of cells and most likely in a single cell line mixture. This fails to give us the information of individual cell behavior in its own pool or complexed mixture (*82*). To address the problem, development of single cell DHS functional labeling could enable us to understand: 1) how single cell DHSs are in single cell line pools; 2) how single cell DHSs behave responding to the surrounding cells, especially different cell lines; and 3) how sensitive the single cell DHSs respond to intracellular and extracellular signals.

To achieve the goals of understanding DHS biology through imaging tools, firstly we need to implement a robust strategy to label the single genomic locus.

3.1.2 Current technologies and their limitations

The nucleic acid detection study via imaging tools needs to start with a robust technique which can label a single region of genome locus in fixed specimens.

Fluorescence *in situ* Hybridization (FISH) is a traditional way to detect single nucleotides loci. Other viable targets in the cells could be easily expanded to tissue section (*83*). Based on the Watson-Crick base pairing, these protocols have high specificity. Fixed sample staining has a lot of steps of washing to reduce the background noise. So, they also have good signal to noise ratio (SNR). But this also becomes a limitation in the sample preparation. The FISH protocol typically takes several hours for probe binding to the target nucleotide sequences. And then it takes another few hours for washing additional probes away from samples. The protocol provides good SNR but

is limited to low turnout. The procedure also involves a global denaturation which damages the chromatin-DNA binding relations and further causes inability of detecting chromatin accessibility.

Cas9, the CRISPR-associated protein, was engineered to remove the nuclease function (dCas9) and became nucleic acid binding protein with facilitation of guide RNA (gRNA). With certain limitations such as protospacer adjacent motif (PAM) sequences, the dCas9 protein can be used as a probe both in living cells and fixed cells. This two prior research demonstrated the robustness of labeling repetitive DNA locus and mRNA transcriptome in cytoplasm CASFISH used fluorescence dye conjugated nucleic acids on the gRNA (*22*). RCasFISH used indirect labeling by MS2-PP7 bridging to amplify the mRNA binding towards HER2 mRNA expression in cytoplasm (*84*). Both these technologies showed strong signals and high signals to noise ratio in the single dots. They required pre-assembly of Cas protein with the gRNA or gRNA conjugates, which left room for unspecific binding due to unreacted halves in incomplete assembly. As Chapter 2 described, CRISPR system has more favorable free energy in binding to dsDNA structure, so it is able to further the off-target binding at less specific loci binding.

3.1.3 Overview of TALE imaging in fixed specimens

TALE protein-based imaging, on the other hand, inherits the benefits of single component detection module in FISH and of unnecessary need for the global denaturation procedure in Cas-based fixed cell imaging. When specimens were fixed with certain reagents, the native protein-DNA interaction could be preserved as their relative position was snapshot and available for detection (*85*).

TALE imaging in fixed specimens is like immunofluorescence protocol. Conventional IF protocol uses a primary antibody to specially bind to a protein of interest in the specimens (*86*). The primary antibody could be directly conjugated to a readout module, like Alexa Fluor® dye,

or need a secondary antibody with the readout module piggy-back on the primary antibody. Since the secondary antibody is binding to the Fc domain in the primary, the copy number of available secondary ones is more than single one. Thus, the signals from few copies of protein in the specimens are amplified.

TALE imaging translates the nucleic acid target to a protein surrogate via direct protein-DNA interactions. Protein-based assay is well established with IF or TSA, which are gentle to the samples and do not introduce additional perturbation towards the sample integration. When TALE probes translate the designated target DNAs to a protein output, this procedure adds two hours before the traditional IF protocol. Also due to the direct conjugates of various epitopes on the N- and C-terminal of protein, the target detection can be multiplexed with multi-color labeling in different probe sets. This could save tediousness like in the CASFISH protocol by sequential binding procedure (*22*).

We could further explore the rationale for transforming single copy target to multi-copy *in situ*. The available sequence can be amplified by localized multimerization of targeted sequence (*87*). Detailed methods for this protocol need more optimization.

3.2 Protocol development

The protocol was generally like the IF protocol as described. The IF staining procedure is well-established if there are specific antibodies towards the epitope conjugated to the TALE protein probes. For the multiplexed labeling, two or more species primary antibodies are also required. The key points are to address the new protocol between sample preparation and IF.

3.2.1 Probe selection

TALE probes are the key components in the assay. Their specificities and binding strengths largely determine whether the nucleic acid targets can be detected with high specificity and

robustness. In chapter 2, we introduced a series of rules of designing and QC steps for evaluating individual probes properties. These rules and QC procedures were implemented in each probe set.

3.2.2 Sample treatments

Preliminary sample treatment was developed based on the traditional IF and FISH protocol with consideration of minimal impact to the native sample preservation. The sample treatments prior to TALE probe input are fixation, RNase A digestion, and blocking. However, we found that our antibody anti-GCN4 had unspecific staining towards the cytoplasm region, which is likely mitochondria, so we decided to use Pepsin to remove the cytoplasm section. This is to prevent the obscurement of staining on mitochondria interfering with spots in the nuclei.

3.2.3 Blocking and binding conditions

The traditional IF protocol used skim milk or Bovine serum albumin to serve as inert proteins binding to all potential unspecific protein-protein binding sites in the samples. In the TALE-DNA binding process, the TALE probe binds to the unspecific sites due to partial binding or global searching mechanism of protein in the cell context. In the fixed cell environment, the cell state is static but the molecules interaction intracellular are also static due to the fixation step. Fixation method introduces extra electrostatic or covalent bonds to anchor the internal relative locations. TALE searching mechanism inside live cell context is a mixture of global searching and local search, as reported (*88*). Fixation has extra hindrance to the TALE probes finding their corresponding targets.

To address the blocking efforts in protein-DNA interactions, we decided to look at similar assays which evaluate DNA-protein interactions. In SELEX (*89*) and EMSA assay (*90*), a DNA

analog called poly(deoxyinosinic-deoxycytidylic), a.k.a. poly(dI-dC) was used to serve as dsDNA analog to block the nonspecific protein binding towards nucleic acid.

As for other components like salts and surfactants, we took similar assays as our references as most protocols evaluating TALE protein performance also require integrity and native form of protein during assay procedure.

3.2.4 Wash conditions

The wash condition is critical for immunoassay. Partial binding of TALE or DNA probes towards their targets favors a lower Gibbs free energy. The surfactants and salts sabotage the partially binding probes off by weakening electrostatic forces and hydrogen bonding. However, the ingredients cannot be so strong to destabilize the on-target binding as well. In the protein purification assay in Chapter 2, we evaluated salt and thermal effects on the protein-DNA interactions. High concentrations of Magnesium or sodium can totally wipe out the binding events (*91*). We used the MUC4 probes to evaluate the background removal performance as a standardization. We took the high concentrations and titrated down the salt gradient to see the flip point between retaining the stronger signals for on-target binding and maximum effect on the off-targets. Selected results are listed in Chapter 5. With 100 mM $MgCl_2$ to fully destroy the TALE ligand binding events, the background level showed a reducing trend at 0-60 mM input in standardized probe data.

3.3 Results and Discussions

3.3.1 TALE probes robustly label low complexity targets

The level of complexity indicates the copy number of individual TALE probe can bind to. Low complexity target contains simpler repeats which single TALE probe is able to recruit enough

fluorophores to a single locus. High complexity target is the locus where every single TALE probe only binds to single copy of target sequence in the reference genome, which leads to more complicated TALE probe design and multiplexed introduction.

For the complexity, we defined the low complexity corresponding to 10^3 magnitude targeted sequences (e.g., telomere) and medium complexity corresponding to 10^2 magnitude targeted sequences (e.g., subtelomeric and repetitive coding sequences).

Telomere sequence is known for its long repetitive region and simple repeat. It is an ideal pilot target for assay initiation. TeloS1 probe contains the TTAGGG repeats binding sequence and can label around 70 foci in the U2OS cells (***Figure 3.1***) (*48*), which is consistent with the estimated number based on the telomere length. In K562 cells, the TeloS1 probe did not present good performance against the telomere sequence. The length of telomere and TRF2 protein abundance contributed mostly to the efficacy discrepancy. After we stained K562 cells with Telomere FISH probes followed by anti-TRF2 antibody, we could see that the signals from these two assays were complementary (shown in Appendix II). When the FISH signal was strong in one focus, the antibody staining spots were weaker in intensity. TRF2 protein is sheltering protein around telomere sequence (*92*). When there is more shelter protein protecting the telomere sequence, the available sequence in the telomeric region decreases where TeloS1 TALE probes bind to. The fact that TeloS1 did not penetrate the TRF2 protein sheltering around DNA sequences showed that the TALE probes are more sensitive to the local environment and existing protein-DNA interactions in the fixed specimens.

3.3.2 TALE probes robustly label medium complexity targets

It was reported that the chromosome 13q has a repetitive section in the subtelomeric region (*48*). With informatics analysis, we designed a series of TALE probes with various repeat numbers in

the selected region, ranging from 9-250 copies based on the hg38 reference genome. The true copy number is yet to be determined due to the limitation of whole genome sequencing in the repetitive sequences.

We assembled all selected probes and evaluated their performance by TALE-ELISA. They all showed strong binding affinity to designated target DNA sequences. To assess our antibody efficacy and reduce the probe number for applications in small genomic regions, the N- and C-terminals of the TALE backbone were attached with 3xFLAG tag and 10xGCN4 tag respectively.

TR series labeled fixed cells with high efficiency and low background. U2OS cells are hypertriploid on chromosome 13 (*93*), which is consistent with the spots calling pattern in TR series labeling. Spots calling mode resided in the 3 and 4 spots per cell (***Figure 3.2***).

TR series numbering indicated the copy number based on analysis of the reference genome. We also would like to use it as reference to estimate the number of probes or fluorophores necessary for one single locus imaging. It was reported in live cell CRISPR imaging and RCasFISH which estimated the number of probes input to be 30-40 (*20, 84*). However, both environment and probes are different from TALE imaging. The data showed that the TR series have guaranteed spots counting of hypertriploid in copy number 170 and above (***Figure 3.3***). Copy numbers less than 171 will have problems detecting any spots or fitting estimation. It is interesting to find out the TR19 also possessed good spots calling in the cells, considering the expected copy is way lower than the threshold. The designs might not be consistent with the true copy number of sequences in the targeted cells. Rest of probes between TR19 and TR171 did not give convincing spots while passing binding specificity and affinity QC.

There is slight variation between 171 and 252 copy spots intensities. This probably resulted from the saturation of probe enrichment in a single locus. As for TR19, the signal intensity from

single spots did not deviate by 10-fold as the copy number fold difference. Therefore, we believe that the TR19 successful labeling of chromosome 13 is because of the irregular subtelomeric region in chromosome 13 having extended length or repeat patterns.

On the other hand, we also tested probe MUC4-E2 in the coding sequence of MUC4 exon 2 repeats. This region had estimated 100 or more copies in the exon 2 within the MUC4 coding sequence. When we stained cells with this probe, the spots count mode is in 3, which is consistent with the expected chromosome 3 copy number in U2OS cells. The spots intensity and size were reduced compared to TR series spots. MUC4-E2 probe labeling region and estimated copy number are both lower than successful TR series probes. We noticed that the cell staining had more background staining inside nuclei due to unspecific binding of protein towards cell context entrapment and of DNA in off-target sequences.

3.3.3 Medium complexity target used as a standard to optimize experimental conditions

The MUC4-E2 probe we tested in medium complexity is able to label the targeted loci with high efficiency and present nice spot counting. However, the background resulting from the staining protocol was solely coming from the TALE probe staining.

There are three events where TALE probes enter a cellular environment: 1) TALE binds to the targeted DNA sequence; 2) TALE binds to non-targeted DNA sequences; 3) TALE binds to cell context such as intracellular proteins. We only favor the first scenario. Nonetheless, the other two binding events are inevitable. Unspecific binding events of TALE probes in DNA and proteins lead to higher background and harder distinguishment of expected signals in the imaging. To reduce the background events, the procedure before and after TALE probe staining needs to be improved. The procedure and results were listed in the methods development section.

3.3.4 High complexity targets can be labeled with tandem array of TALE probes

PPP1R12C, as known as AAVS1 safe harbor, is traditionally used as a knock-in locus for the homologous directed recombination (HDR). Beta globin LCR is a tandem five DHS peaks in a 10 kb region responsible for the control and expression of upstream hemoglobin coding sequences. We adopted PPP1R12C promoter DHS region and beta globin LCR loci as our primary targets in high complexity probe staining. Each probe used in the pool was screened genome-wide to verify the orthogonality of sequences. In PPP1R12C, 3 kb regions in the proximity of promoter DHS were clustered with 60 probes. In LCR HS1 to HS4, there were 50 probes distributed in each 1 kb flank region around DHSs. All the probes were designed around DHS because we want to minimize the effect of nucleosome-DNA interaction and potential methylation status in CpG loci.

We stained U2OS cells with the PPP1R12C probe set and K562 cells with both sites. PPP1R12C is a housekeeping gene but it is traditionally used as safe harbor. In U2OS cells, the TALE imaging can capture the binding events of the probe set. Spots count was lower than expected value, but the intensity stood out from the background (***Figure 3.4***). As for the cell population spot distribution, approximately 40 percent of cells harbor at least one spot.

There are more experimental conditions which could be improved to increase the detection efficiencies. The harsh wash conditions might sabotage some on-target binding events, but milder wash conditions promote the high background in the context.

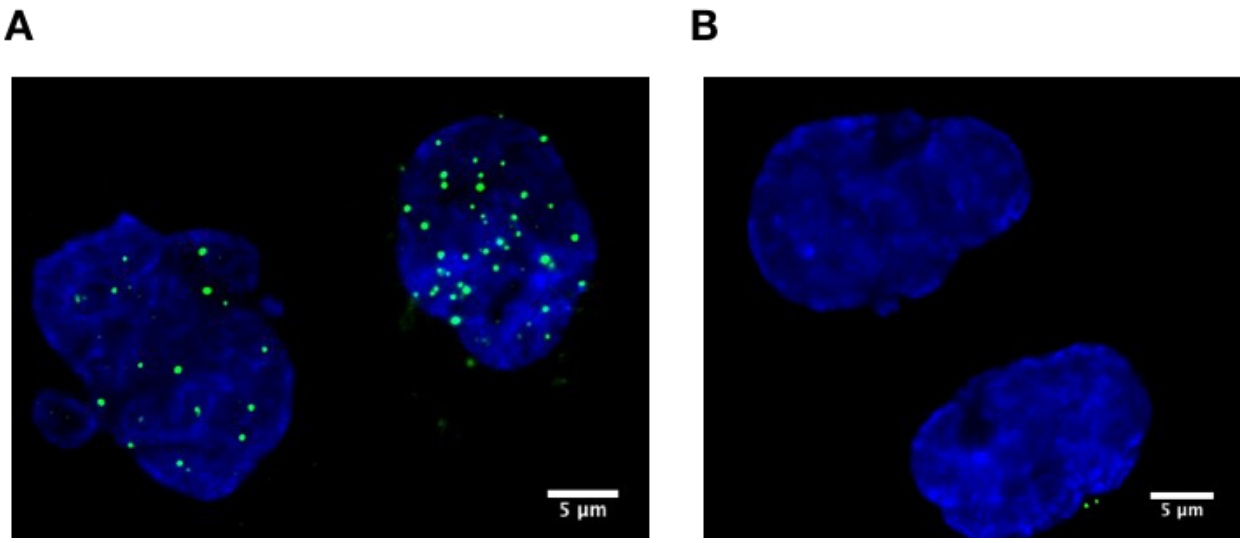

Figure 3.1 In vitro system telomere imaging

IVTT probes stained U2OS cells with targets of telomere system. (**A**) TeloS1 probes and primary conjugates. (**B**) no TALE probes.

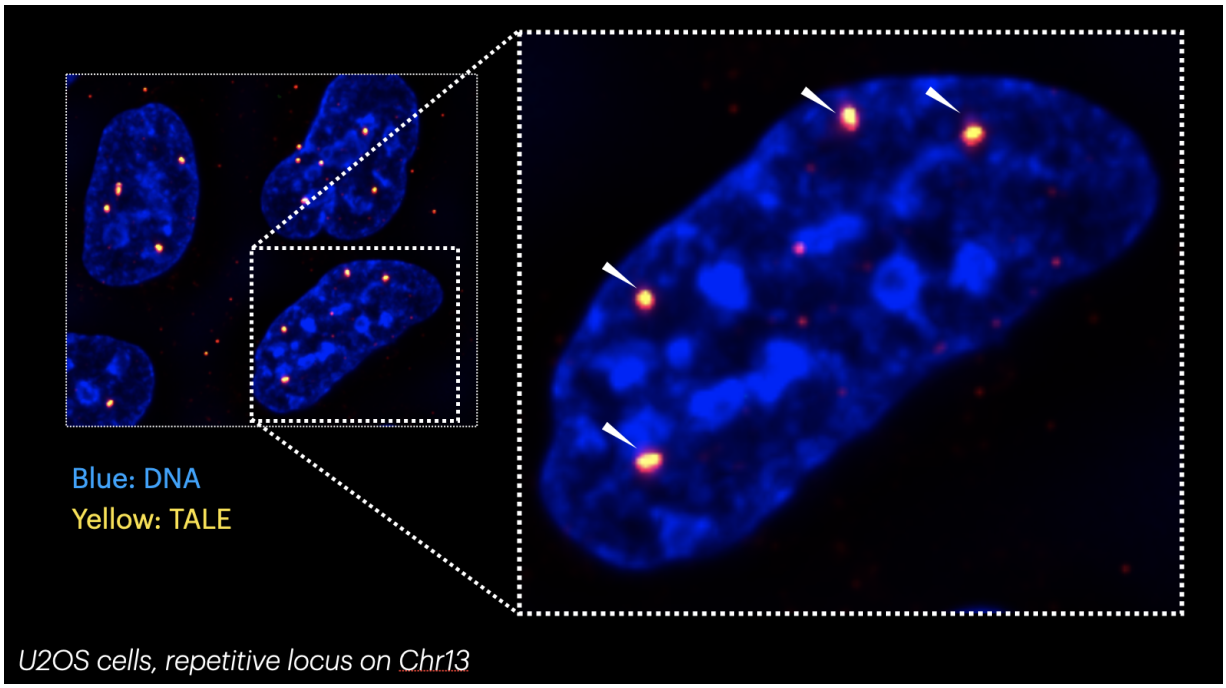

Figure 3.2 TALE in vitro labeling of subtelomeric region of Chromosome 13q

Shown is U2OS cells labeled with TR242-SunTag probe. This cell line has four copy of chromosome 13 and targeted region is on the long arm. There were partial off-target staining in the cytoplasm due to the unspecific binding of anti-GCN4 to the mitochondria, which can be further eliminated by pepsin treatment.

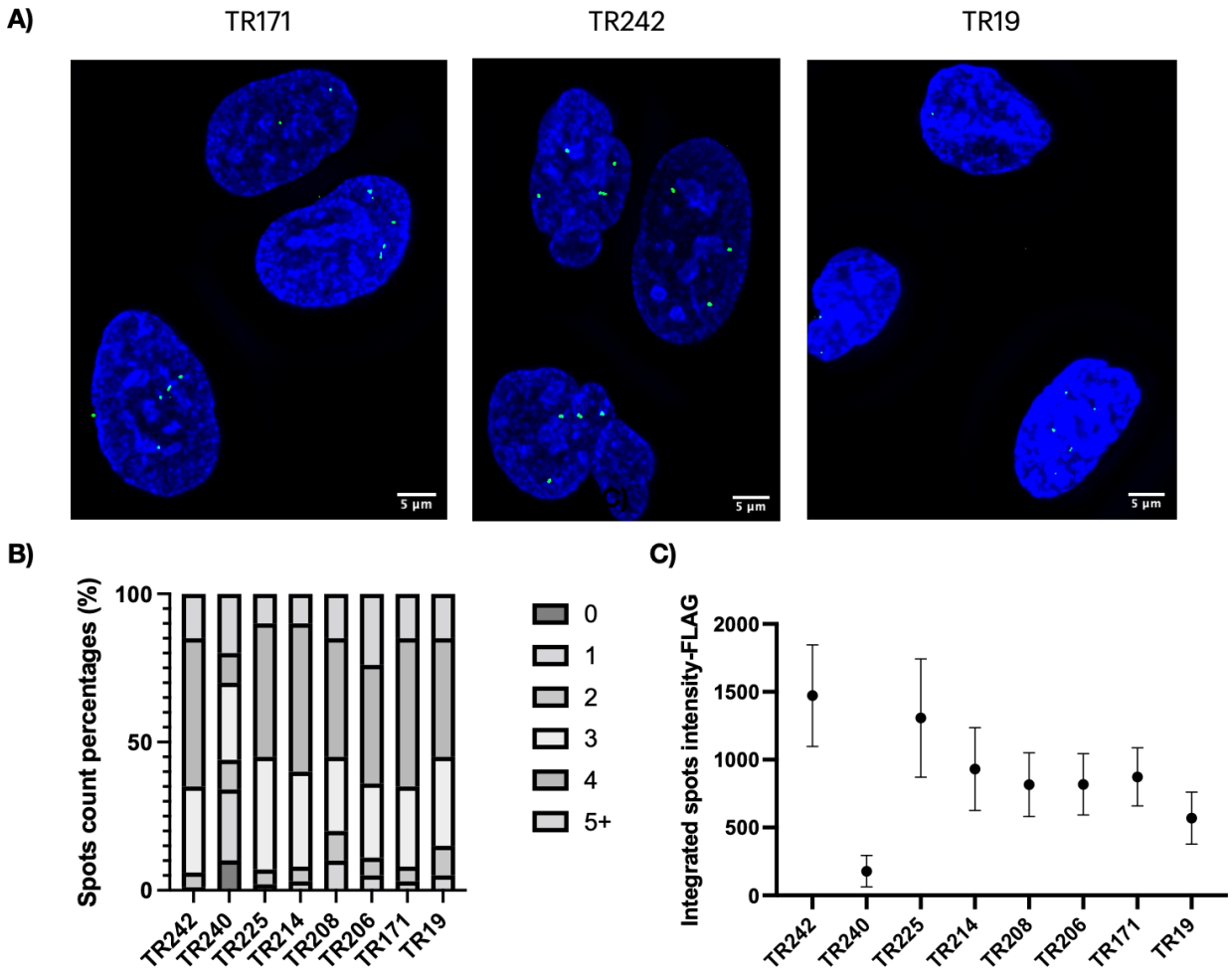

Figure 3.3 Comparison between different copy number probes in same region

(**A**) are sample images of TR series for least visible copy number 171, most copy number 242, and mysterious copy number 19. Scale bar = 5 µm. (**B**) is the spots counts in each TR series which have visible spots. As most of them fit the expected value of residing 3-4 spots per cell. (**C**) is the spots intensity normalized to the background. As data shown, the mean spots intensity was decreasing with copy number reduction. Except for TR240 and TR19, theoretical repeat copy number was related to the spots intensities.

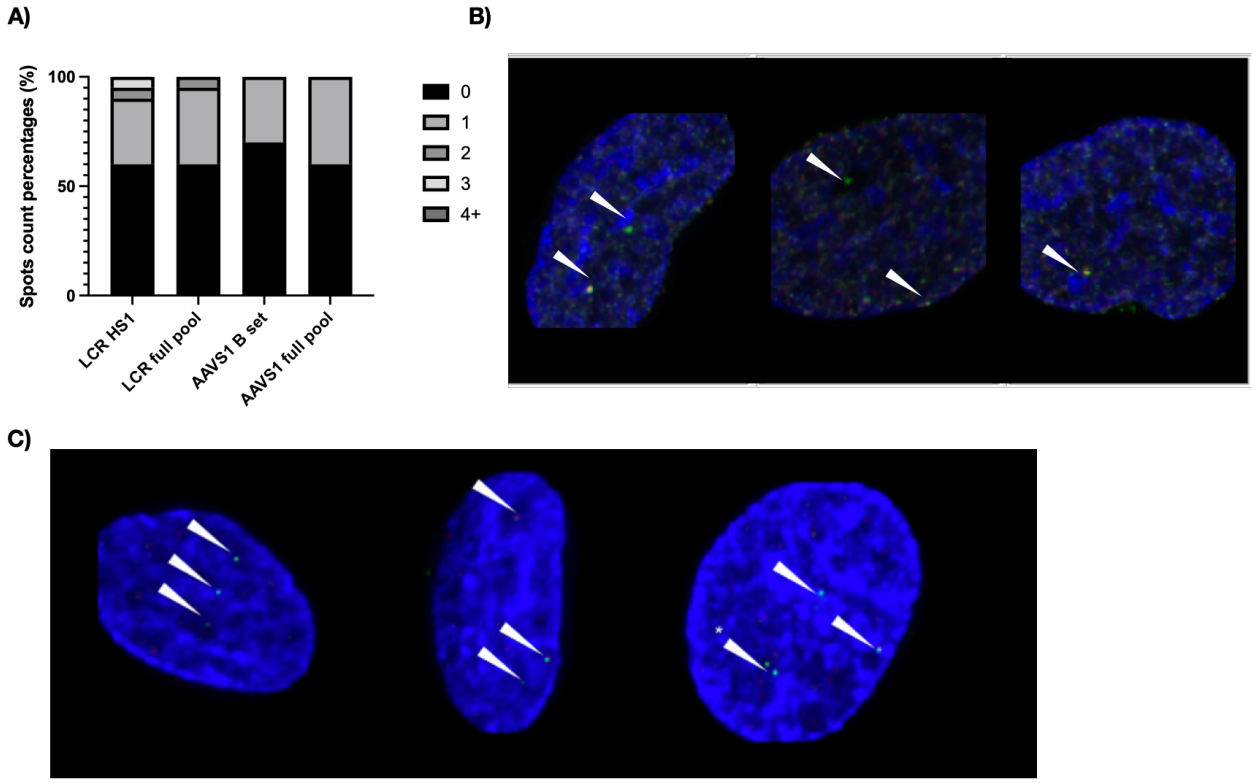

Figure 3.4 AAVS1 and beta-globin LCR labeling with multiple tandem TALE probes in vitro

(**A**) are spots count for selected subpool and full pool probe sets staining. (**B**) are example images for AAVS1 full set TALE labeling. (**C**) are example images for beta globin LCR full set TALE labeling. The asterisk highlighted was indication for replication doublet.

3.4 Summary

In this chapter, I described the nucleic acid imaging in fixed cells. In the low complexity target, I showed the telomere target could be labeled in various cell lines and demonstrated the telomere biology with the TRF2 protein abundance. TR series and MUC4 targets, as a higher complexity target, started to reflect potential background contributors and also demonstrated the minimal number of fluorophores necessary to visualize single locus. But TALE probes were able to label the few kilobase region robustly. In the example of promoter and enhancer gene loci, TALE showed labeling of spots in the fixed cells, with reduced efficiency and spots count. This might be due to the lower probe number enriched in single region.

3.5 Materials and Methods

TALE fixed cell imaging

Targeted cells (U2OS, K562) were fixed with Methanol: Acetic acid at ratio of 3:1 in 24-well plate (#1.5 High performance glass, CellVis). Samples were treated with 25 ng/μL RNase A in 1xPBS at 37 ºC for 30 minutes and 0.1% Pepsin in 0.01M HCl for 5 minutes to remove RNA and cytoplasm. TALE probes were translated to protein as described and diluted in the TALE dilution buffer for 15 minutes. Samples were blocked with TALE dilution buffer with extra 2% BSA. TALE probes were incubated with fixed cell samples for 2 hours.

We washed cells with 1xPBST with 60 mM $MgCl_2$ for three times and then primary antibodies were added 1:1000 in the antibody dilution buffer. Secondary antibodies were added afterwards at 1:2000 dilution. Samples were further stained with Hoechst 33342 (Life technologies).

Microscope imaging

The samples were mounted with 12 mm coverslips (Fisherband). We used a Nikon Ti inverted with a 60x oil immersion objective. The samples were imaged with High Content Acquisition and processed by commercially available deconvolution software (microvolution).

Chapter 4 Detection of nucleic acid targets in living cells

4.1 Introduction

Dynamic status of one or more genome loci could narrate a time-series process of regulatory elements and the corresponding gene interactions. There are several efforts on high sensitivity and long-time live cell imaging of protein motion (*94*) and genomic locus tracking. Given that the hardware and software aspects of single molecules visualization are well-established, the only limit is how we should probe the elements of interest. The live cell setup is more straightforward to see the real-time interactions between regulatory elements. With the TALE imaging system developed in Chapter 3, we can peek at how this system works in the vigorous live cell environment.

4.1.1 Challenges of nucleic acid imaging in live cells

Live cell imaging setup, compared to the fixed cell context, is more complicated and intriguing. Despite the challenges and efforts other teams have already input, the imaging assay still has several things needed to conquer.

To deliver our TALE imaging system into a live cell, the transfectants are either DNA, RNA, or protein. Technology involved in delivery of a few nucleotides sequence and protein are mature, such as electroporation and chemical induction. However, when transfectants are to enter a single cell all at the same time, there are chances where the single cell will not receive all the probes at the same time. To ensure one single locus to be localized by a tandem array of probes, all transfectants must enter the same cell. The traditional mRNA transfection is usually able to achieve efficiency of over 90% in non-primary cell lines. Yet, the combinational efficiency for all probe successful introduction to one single cell is much lower. That is also one of the reasons why multiplexed genome editing is so difficult.

Therefore, it is challenging to achieve simultaneous multiplexed delivery and we expect to have low spots counting among the cell population due to the difficulty.

4.1.2 Current technologies and limitations

Cas9, the CRISPR-associated protein, was engineered to remove the nuclease function (dCas9) and became nucleic acid binding protein with facilitation of guide RNA (gRNA). With certain limitations such as protospacer adjacent motif (PAM) sequences, the dCas9 protein can be used as a probe both in living cells and fixed cells. One strategy is to attach fluorescence protein on the terminal of dCas9 protein (*20, 21, 23, 50, 57, 95-103*). One of the advantages of this system is that, for living cells, only one protein is needed for the transfection. Fewer numbers of foreign DNA introduced into the living system could reduce the cytotoxicity and elongate the viability. However, CRISPR has been recently reported to possess considerable off-target effect (*104*). And the gRNA forms an RNA-dsDNA triplex conformation, which could perturb the normal gene function in living cells if directly probing to active DNA.

In the recent paper, a non-repetitive enhancer and promoter labeling probe set was reported to label loci in mouse embryonic stem cells through multiplexed gRNA probes array (*105*). However, the CARGO array used in the proximal location of Fgf5 is in DNase I non-peak region. RNA-seq data also indicated the probe sets in the living cell did not interfere with the transcription functions. The dCas9 protein apparently possesses the ability to compete with nucleosomes and access directly to DNA sequences. Although several studies showed that the Cas9 cutting efficiency could be influenced by the chromatin accessibility, the binding efficiency of dCas9 was not well studied. TALEs are directly binding to the DNA sequences and dCas9 distorts the DNA to form triplexes. Then TALEs have potential of less competition with nucleosomes near the targeted region.

4.1.3 Overview of TALE-based imaging in live cells

A TALE-based imaging system in live cells was implemented in two tracks. One track is to use direct labeling of mNeonGreen. By direct attachment of fluorescence protein on the C-terminal of TALE probes, we expect to see less spots count and signal intensities because recruitment of fluorescence molecules reduces in given location. This way allows us to have direct observation of signals in the binding events.

The other track is to use indirect labeling of the SunTag system (***Figure 4.1***). This approach uses the two components system where the TALE probes carry the binding domain and epitope array available for reporter protein binding to it. The reporter protein contains the domain to bridge the epitope array and the single chain variable fragments. There is also a GFP tag and GB1 peptide to increase solubility of reporter protein in live cell environment. The constructs were adapted from previous literature (*57*). The recruitment events have two stages where the targeted genome loci recruit the designed probes and tagged GFP binds to the epitope array. The two-step enrichment could potentially solve the issue with reduced probe transfection efficiency in the cell pools.

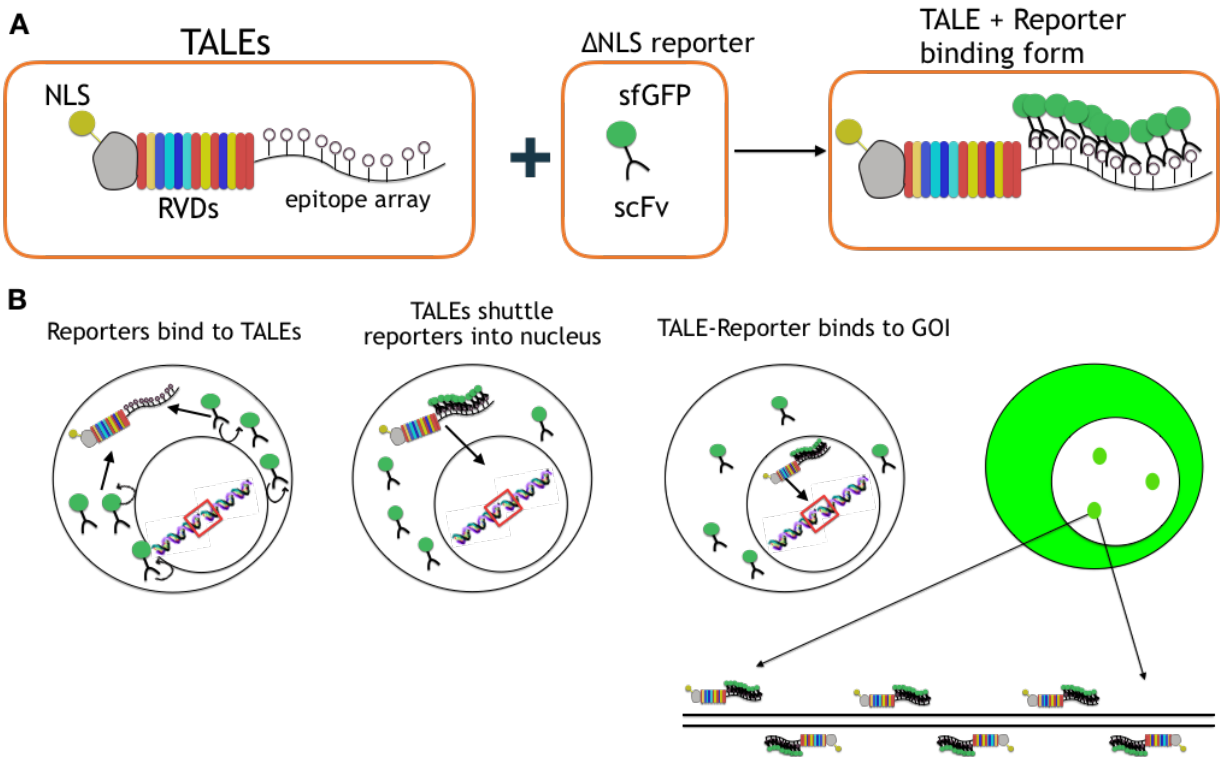

Figure 4.1 Mechanism of TALE-based imaging system with two components

(**A**) described the binding form between TALE-based DNA Binding Domain and optimized reporter system and created a multi-copy fluorescence molecule recruited form. (**B**) illustrated the mechanism of how the system was optimized and works. The reporter and TALE protein were both translated in cytoplasm and reporter could not enter nucleus without NLS. TALE recruited reporter and carried them into nucleus. Then the binding form approached the targeted locus by tandem multiple copies of TALEs in narrow range. Then the genome loci could be localized.

4.2 Results and discussion- direct labeling

4.2.1 RNA as transfectant improved the signal consistency

There are many ways for delivery of targeted protein. Apart from traditional DNA nucleofection, many other novel delivery methods were invented for easier adaptation and higher efficiency (*32, 34, 35, 106*). Compared to the DNA transfection, mRNA as a transfectant has a shorter turnover rate but more consistent signals. This could be a potential issue in long time imaging but at the current status, it is not necessary to consider time-lapse tracking. From the 7-day cytometry diagrams, mRNA transfection showed more uniform signals in the GFP expression while the plasmid transfection lasted longer and had a wide range of expression levels (***Figure 4.2***). For imaging, to avoid the outstanding signals such as aggregation treated as false positive signals, the reporter expression needs more consistency and lower intensity. In the temporal imaging, the mRNAs might be modified to elongate the half-life or transition to continuous feeding of protein probes.

4.2.2 TALE labeled the subtelomeric targets robustly

Series titration of 40 ng, 80 ng, and 160 ng of TR242-mNeonGreen were transfected in pre-seeded U2OS cells and it was imaged at 8- and 24-hour time points. U2OS cells were imaged with confocal units for spots calling at z-stack (***Figure 4.3***). It showed that most cells captured 4 spots in the maximum intensity projection images. The puncta were distinct and stood out from the background. Cells were later fixed with formaldehyde to confirm the spots with its own FLAG tag on N-terminus.

Dosage titrations can be easily shown by the intensity of positive cells which contains bright GFP signals. However, some of cells possessing low signals overall in the nuclei still had strong

spots labeling with expected counting. It appeared that cells only need small amount of transfectant to express adequate protein probes for efficiency binding events. However, the intensities of spots did not reflect too much on the dosage gradient. The local saturation of fluorophore recruitment resulted in reaching upper bound of spots intensities.

4.2.3 AAVS1- and LCR- mNeonGreen probes label spots in living cells

Two pools of probes were transfected in multiple subpools and reduced dosage titration to evaluate the performance of probe sets in responding to low probe number and low transfectants. We saw a decreasing trend of spots calling from single subpool to the full pool despite the same dosages. The reduced probe number answers to the lower efficacies in single locus labeling due to the lower efficiency of multiplexing transfection (***Figure 4.4***).

Lower dosage did not respond to the signal intensity variation but more to the lower transfection efficiencies. There were fewer cells with background signals in the nuclei but most of them had obvious clustered aggregation near the nucleolus. Nucleolus had outstanding signals of GFP compared to the rest of nuclei and cytoplasm. The higher intensity may come from the fouling of GFP-tagged TALE probes binding to the nucleolus outer surface.

AAVS1 probe set has higher labeling events than LCR. As described in Chapter 3, PPP1R12C is a housekeeping gene and expressed in many other cell lines. The promoter DHS is believed to be active in U2OS cells. U2OS cells only have single active HS2 and might not be able to harness many designed probes. The heterochromatin status potentially blocked the other TALE probes binding event.

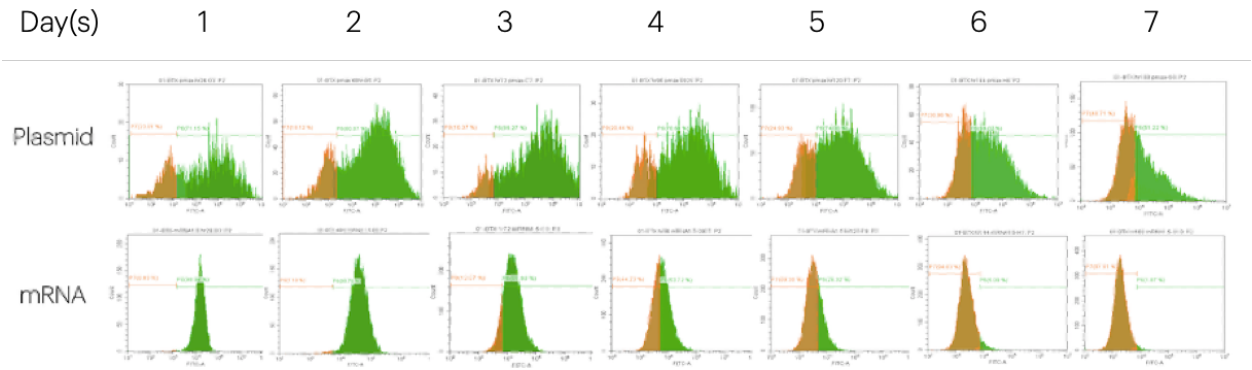

Figure 4.2 Plasmid vs. mRNA post-transfection time-course fluorescence reading

K562 cells were transfected with plasmid DNA and mRNA produced from the plasmid. mRNA transfection produced more uniform and consistent signals but DNA transfection last longer.

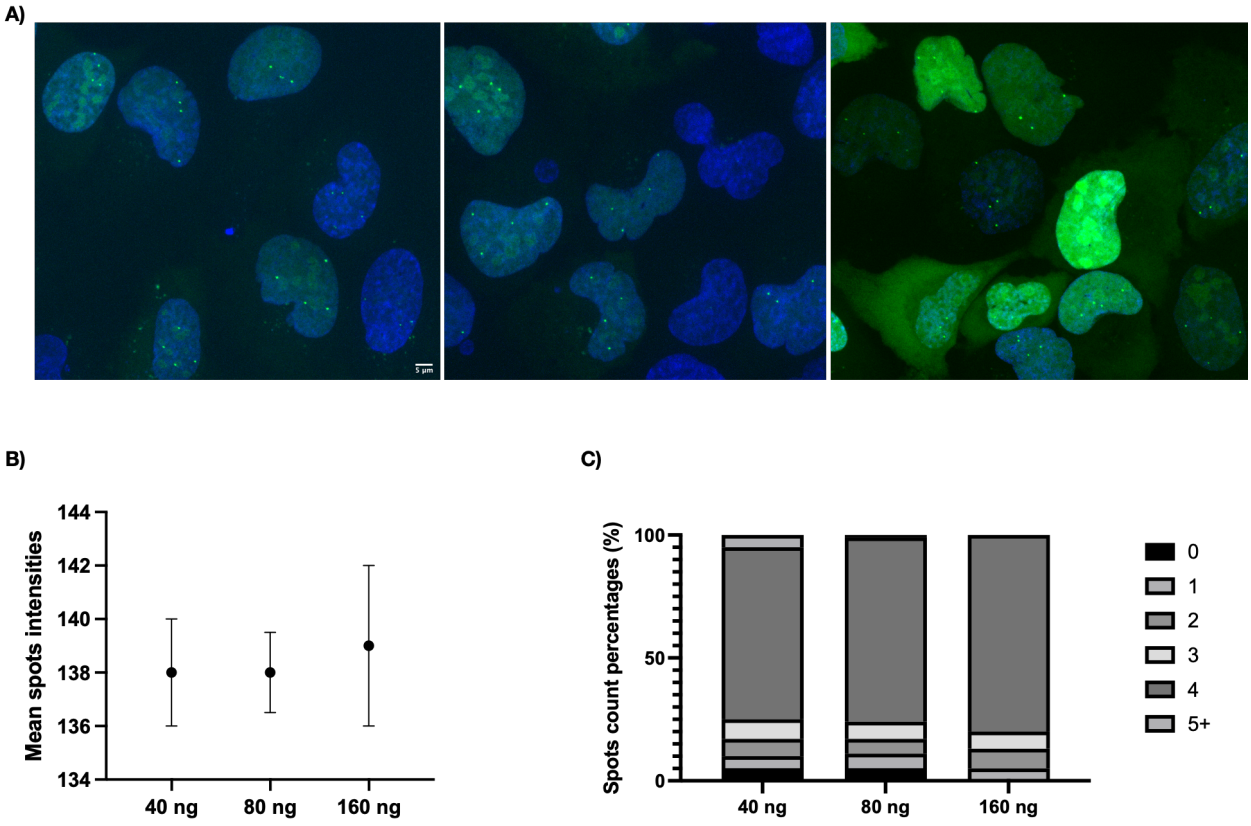

Figure 4.3 TR242 dosage titration in live cell at 24-hr

(**A**) is TR242 dosage titration example images. From left to right is 40 ng, 80 ng, and 160 ng. Imaged at 24-hr after Hoechst staining in DPBS. Scale bar = 5 μm. (**B**) is the intensity distribution for same batch of experiment. (**C**) is spots count distribution.

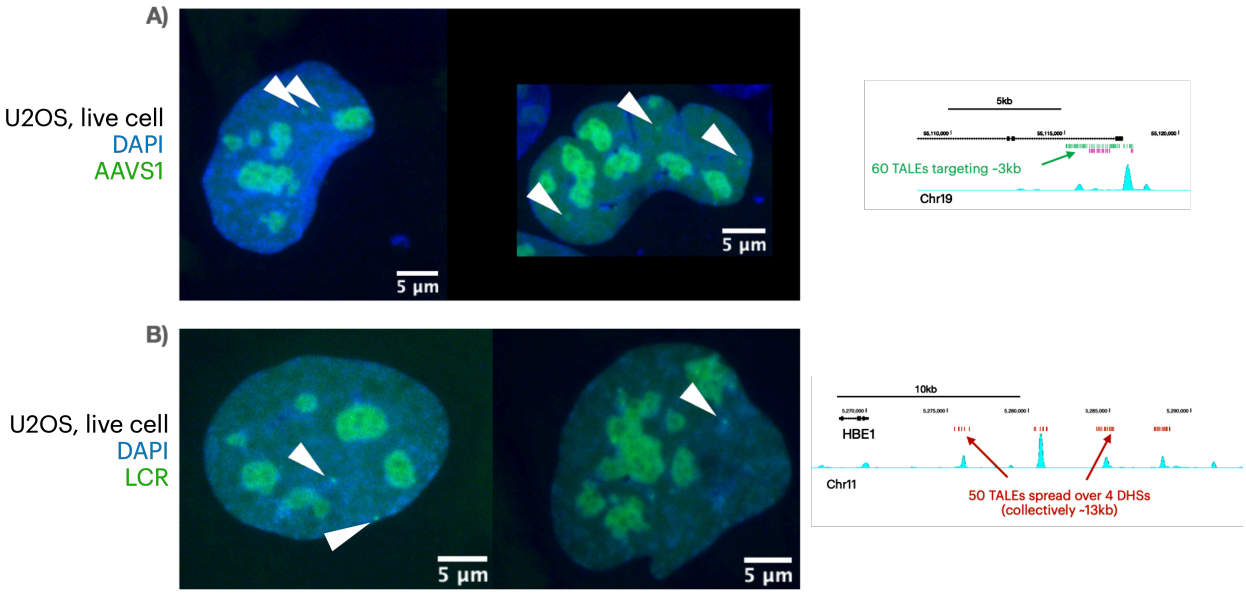

Figure 4.4 Beta globin LCR and AAVS1 genomic loci imaging in live cell

(**A**) are example images for AAVS1 full set transfection. (**B**) are example images for beta globin LCR full set transfection. All transfected with 200 ng and imaged at 24-hr. Scale bar = 5 µm.

4.3 Results and discussions- two components labeling

Results from 4.2 indicates that the high number of TALE probes co-transfection resulted in low successful entrance of all viable probes. The chance of all viable probes entering single cell simultaneously was low and can be even lower to increase probe number. To increase the labeling efficiency, there are two methods. One is to increase the probe number by over-saturating the targeted region with additional probes. Therefore, if some probes cannot enter the single cell at the same time, the transfected probes should be sufficient to produce enrichment in single region. However, this method will reduce each transfectant concentration in the reaction, but higher dosage might also contribute to higher background and increased translation burden for the cells.

The other method is to reduce the probe number but to increase fluorophore attached to single probe (*57, 107, 108*). By recruiting multiple copies of fluorescence protein to single probe, the required probe number can also be reduced by several folds. This is also the methodologies in fixed cell imaging by probing to single molecule sensitivity.

4.3.1 Probe numbers enhanced overall signal intensity

When the reporter dosage was fixed, increasing the probe number raises the total intensity due to the recruitment. The binding kinetics between reporter and probe epitope array affected the recruited copy number. Therefore, the signal intensity increased folds are not proportional to the probe number. But this still indicates that the probe number- probe dosage increase enhanced the overall signal intensities.

4.3.2 Dosage titration reduced the background but retained the desired signals

The standard transfection dosage is 1 ug and marked as 1X. 0.2X and 0.04X are equivalent to 200 ng and 40 ng respectively. In ***Figure 4.5 B*** and ***C***, the dosage titration down reduced the overall

staining intensity to very low level but remained significant from blank baseline. Considering the number of probes in the system, the lower dosage is better for the imaging purpose.

4.3.3 Removal of nuclear localization signal improved SNR in nucleus

When 8 probes were used in the transfection, even 0.04X dosage of the original reporter construct yielded a high background in the nucleus. By removing the Nuclear Localization Signal (NLS) from reporter constructs, reporter protein translated in the cytoplasm would stay until the probe epitope array recruits them and carries them into the nucleus. In ***Figure 4.6***, using TeloS1 as a standard, the low dosage original reporter had more background than the modified ΔNLS reporter.

Meanwhile, the reporter proteins were easily susceptible to proteasome degradation in the cytoplasm. Imaging time point was changed to 24-hr for the following imaging.

4.3.4 ΔNLS reporter proteins were more susceptible to degradation

With drug Bortezomib treatment and inhibiting 26S proteasome activity, the reporter signals sustained (***Figure 4.6 B*** and ***Figure 4.7 B***) but also resulted in a large scale of cell deaths (***Figure 4.7 C***). When reporter protein stayed in cytoplasm and the majority of protein degradation pathways was also there, the proteins were easily degraded. The time point of 48 hours was too late to image without a degradation inhibitor but considering the toxicity of the drug, it is better to change imaging time to 24 hours post transfection.

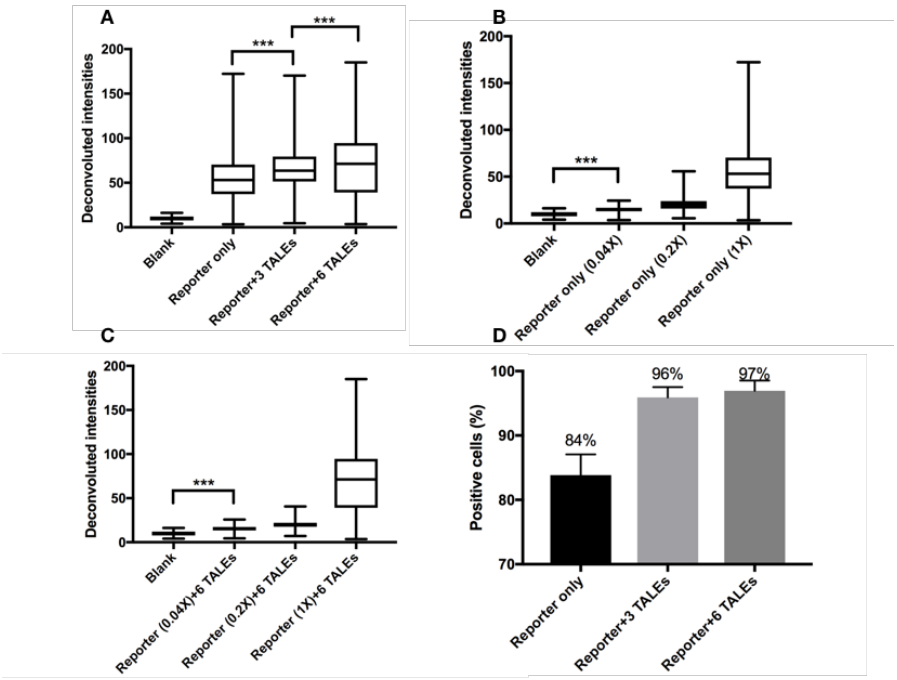

Figure 4.5 Probe number and reporter titration test

(**A**) When reporter dosage is fixed, increasing probe number enhanced the overall intensity. (**B**) and (**C**) Reporter titrated down would not compromise signals but also reduced background. (**D**) Increasing probe numbers promoted positive cells ratio.

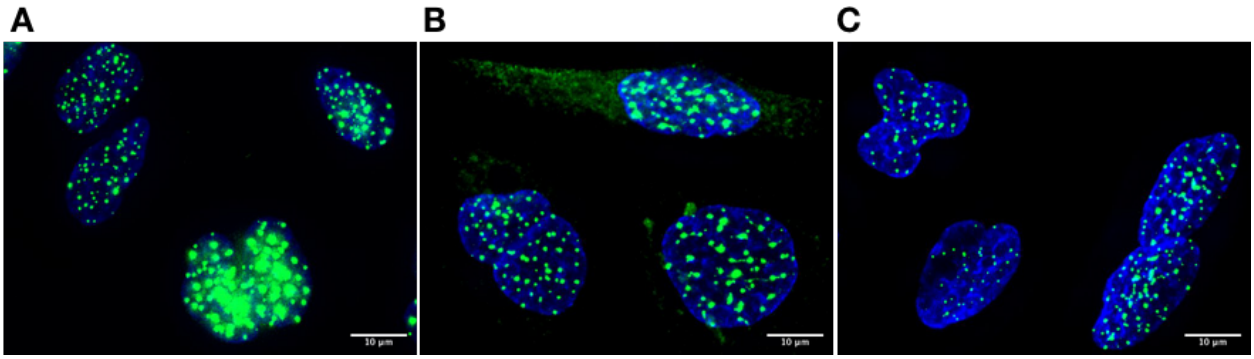

Figure 4.6 Comparison between different reporter constructs and inhibitor treatment.

The three images are TeloS1 transfected with (**A**) 0.04X original reporter, (**B**) and (**C**) ΔNLS reporter in U2OS cells. (**B**) had additional 5 nM Bortezomib treatment while (**C**) did not. The signals from (**A**) and (**B**) were more redundant than (**C**).

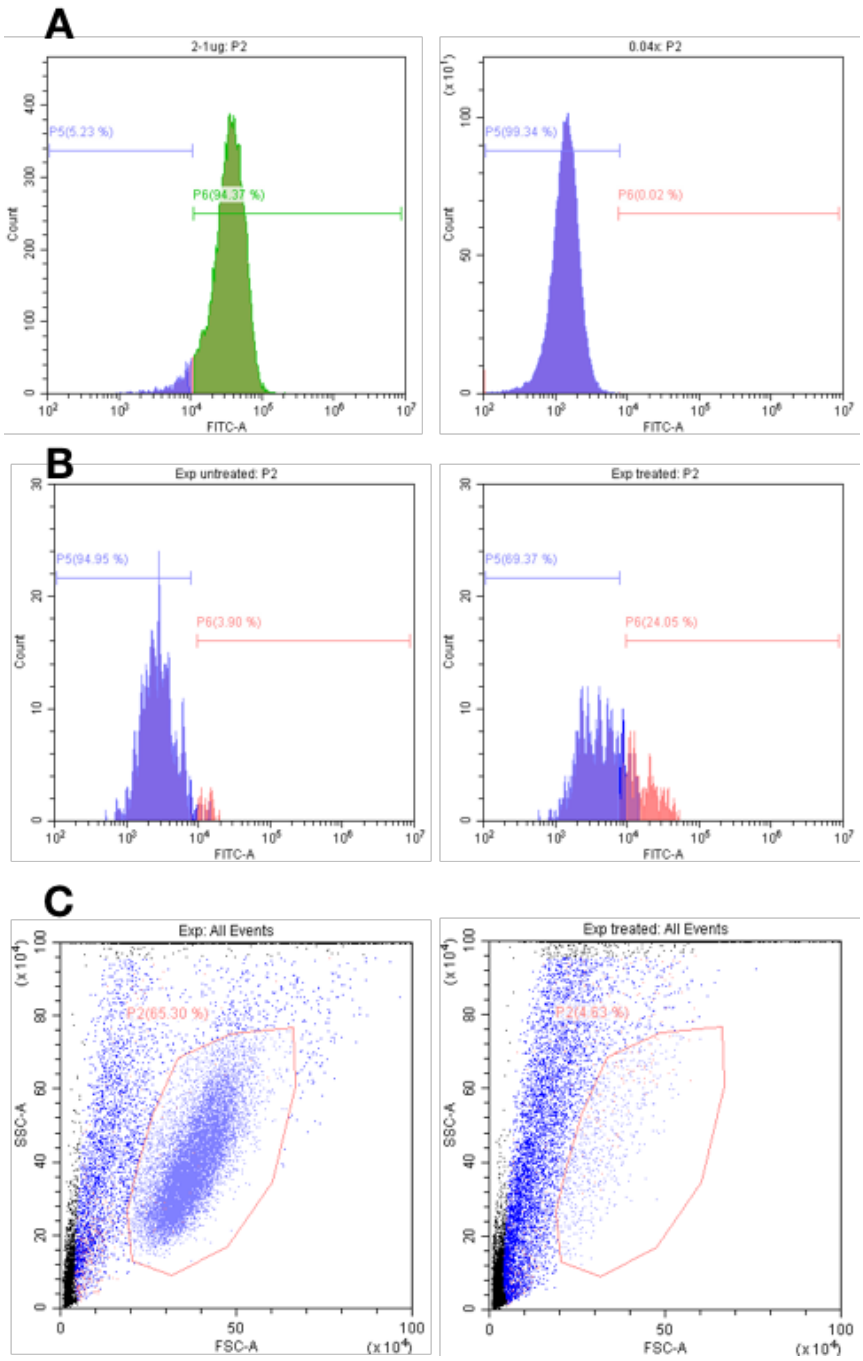

Figure 4.7 Reporter transfection efficiency and ΔNLS reporter inhibition assay

(**A**) was original reporter transfection efficiency at 1 ug (1X, left) and 40 ng (0.04X, right). The signals at low dosage were too weak to detect. (**B**) showed ΔNLS reporter transfection (1 ug) with inhibitor treatment increased the positive signals overall at 48-hr (Left, no inhibitor treatment; right, with inhibitor treatment). (**C**) showed the inhibitor treatment was killing the living cells at 48-hr time point.

4.3.5 Telomere was visualized *in vivo* in post-fixed cells

TeloS1 and ΔNLS reporter were transfected into three targeted cell lines K562, A549, and U2OS. Also, TRF2 antibody was used to verify the loci.

The TeloS1 showed reasonable amounts of spots in U2OS cells but low abundance in K562 and A549. Among three tested cell lines, U2OS has the longest telomere. So, it makes sense that U2OS had both a pretty amount of TRF2 and TeloS1 loci (***Figure 4.8***). In other two cell lines, the TRF2 loci was inversely related to TeloS1 and TeloFISH loci amount. The phenomenon is because the TRF2 formed shelterin complexes around telomere sequences to prevent shortening. Where TRF2 is more abundant most likely has fewer available telomere sequences. TeloS1 and TeloFISH both rely on availability of telomere sequences while the TRF2 protein complex hinders the exposure.

4.3.6 Telomere was visualized *in vivo* in living cells

U2OS cells were directly imaged after Hoechst nuclei staining in PBS (***Figure 4.9***). The imaging pattern in the GFP channel showed fixed cell staining images. Also, one trial included the original reporter in 0.04X dosage and it showed more background in the nucleus. So, for the short time point, the ΔNLS reporter performs better than low dosage original reporter. For longer time living cell imaging, it needs either inhibitor addition or even lower dosage of the original reporter.

One thing also noticed is that the autofluorescence was strong in living cells, which might be the result from autofluorescence of cell endogenous proteins and out-of-focus light from other planes.

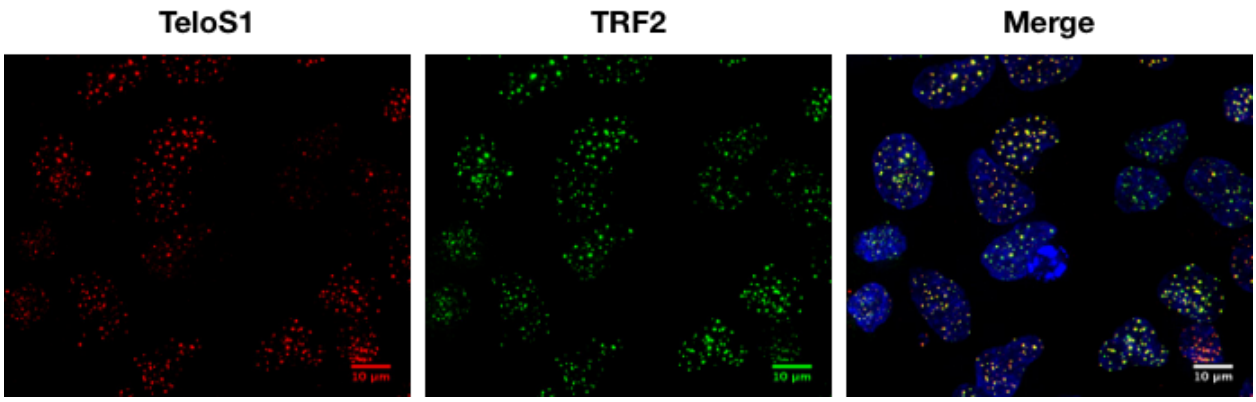

Figure 4.8 TeloS1 system transfection and post-transfection IF in U2OS cells

Left is TeloS1 and reporter transfection followed by GFP staining. Middle is TRF2 protein IF staining used to verify spots from transfection. Right is merged image. Blue is nuclei and yellow is colocalization signal.

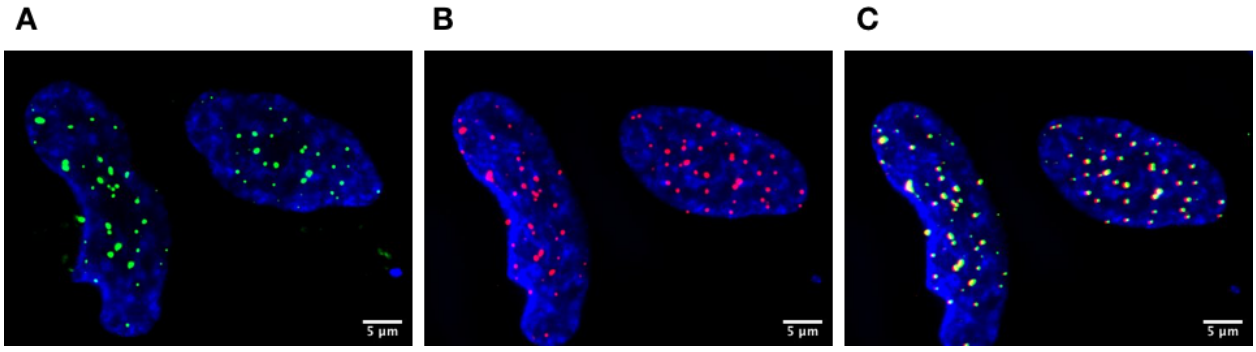

Figure 4.9 Live cells imaging of telomeres in U2OS

(**A**) was living cell imaging of telomeres. (**B**) was the exact same sample imaged in (**A**) and stained with TRF2 antibody. (**C**) was the merged images using nucleus boundaries as cell registry matching.

4.3.7 Single DHS in LCR was visualized *in vivo*

HS2 8-probe system and ΔNLS reporter were transfected into U2OS and A549 cells for living cell imaging. K562 cells were not included because suspension cells will detach from the surface if the imaging time is too long, and the suspension cells might move to different focus planes during imaging and acquisition. Some spots could be seen in living cells but the autofluorescence also formed lots of spots in whole cells. Also, the out-of-focus light introduced more background on the detected plane. To verify the spots from living cell imaging, a post-fixation was performed immediately after imaging and the GFP reporter was stained. The imaging coordinates were recorded and used in the fixed cell imaging. Later a cell registry would be matched and overlapped for channel colocalization.

Fixed cell imaging after transfection had two parts. One was, as described above, using cell registry to match living cell imaging and check whether the living cells spots were true GFP loci. The other was collected at 24-hr time point and did not go through living cell imaging. But this would use LCR FISH to colocalize the GFP spots. Idea here is that, since FISH cannot be performed in living cells, FISH can colocalize with GFP reporter spots and then fixed cell spots can colocalize with living cells signals.

Unfortunately, in the cell registry, there was no spot colocalization, but the cell morphology had changed. The texture of chromatin became distorted and contained aggregates. Also, the size of nuclei was generally larger. The change was majorly coming from living cell imaging time and media. PBS lacks Magnesium and Calcium ions which are important for cell attachment. Also, the osmosis between cells and environment caused cell morphology change. These can be solved by using specialty living cell imaging serum. Post-transfection IF also did not show spots colocalization and diffusion of reporter signals could disturb the signal patterns.

In the post-transfection FISH-IF, three cell lines were all tested. LCR FISH was confirmed to be working in K562 but not yet in U2OS and A549. From the DHS profile, the A549 and U2OS have active HS2 and HS5. So, the LCR FISH should still probe to target, perhaps at lower efficiency (***Figure 4.10***). However, the LCR FISH still did not perform as well as in K562 cells. The current data is not enough to support the colocalization spots in the GFP channel.

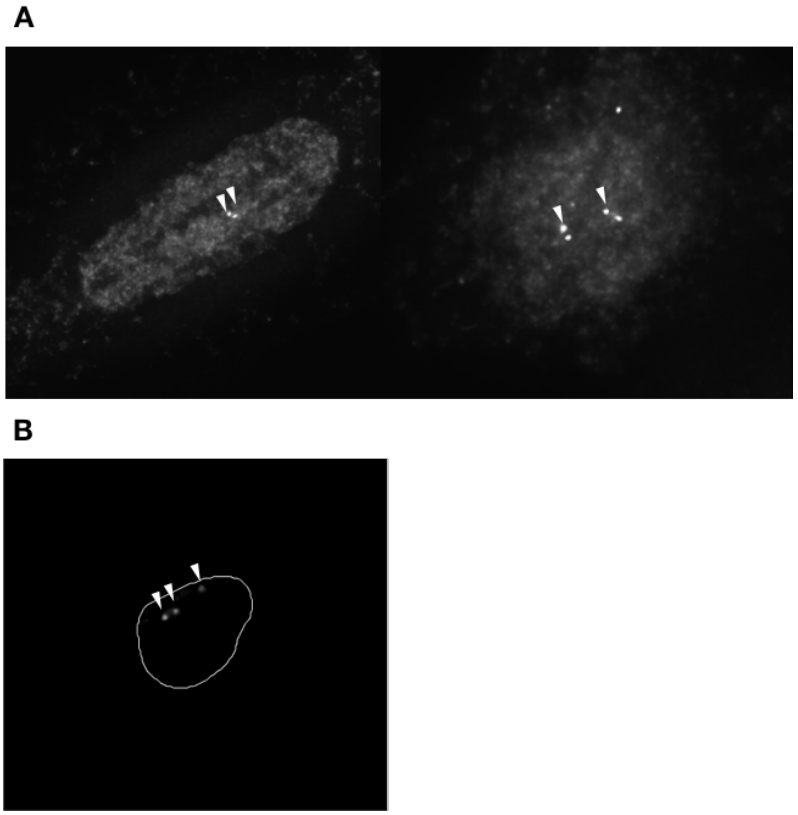

Figure 4.10 Detection of potential HS2 loci in U2OS and A549 cells

(**A**) U2OS cells and (**B**) A549 cells, 24-hr post-transfection fixation, stained with anti-GFP (Abcam) and DAPI. (**B**) used white outline to highlight nucleic boundary after segmentation.

4.4 Summary

In this chapter, I took the scope from fixed cell imaging by TALE platform towards more dynamic environment in the live cell setup. Here the TALE imaging can label the telomere and TR series targets at high efficiency. At even lower concentration of transfectants, the cells can still express adequate protein for target region labeling. This lays a foundation for multiple probe transfection in live cell since individual probe transfectants concentration is lower. I showed that the TALE labeling of non-repetitive loci by two methods: direct labeling and indirect labeling. Direct labeling is easier to handle but requires more probe number in the transfection, which further leads to lower efficiency of simultaneous entrance of multiplexed TALE probes. Indirect labeling requires two components system. The extra recruitment of multiple copy of fluorophores onto single TALE probe can drastically reduce the necessary number of probes. The imaging results also showed that the local chromatin status in different cell lines influenced potential spots calling and TALE probes binding.

4.5 Materials and methods

U2OS cells transfections

U2OS cells (ATCC) were cultured in McCoy 5A medium (Corning) with 10% FBS (Cytiva) and Penicillin/Streptomycin (Corning). 50,000 cells were seeded into a 24-well glass-bottom plate (CellVis) and culture overnight in a 37 ºC incubator. Desired TALE probes (and reporter RNA) were pre-aliquoted in 1.7 mL tubes with 25 µL of Opti-MEM (Gibco). 1.5 µL of RNAiMAX lipofection reagents (Invitrogen) was mixed with 25 µL of Opti-MEM. Two components were mixed up and incubated for 5 minutes at room temperature before added into pre-seed cells. The cells were incubated in 30 C for 8-hour before live cell imaging.

Protease Inhibitor assay

26S protease inhibitor PS-341 (Selleckchem) was diluted 1:10000 into DMSO. The diluted samples were aliquoted and stored in -20 ^{o}C. Upon treatment, the pre-seed cells in 6-well plate (Corning) were treated with recommended dosage by the manufacturer and control group was treated with 0.1% DMSO to the final concentration.

Flow cytometry

Cells to be tested were washed with 1xPBS (Corning) and incubated with Ghost dye 780 for 30 minutes. The cells were evaluated through Cytoflex S (Beckman Coulter) by FITC and PE channels.

Microscope imaging

Cells at each time point were washed with DPBS (with Mg^{2+} and Ca^{2+}) and incubated with Hoechst for 10 minutes in the dark. The images were acquired by Nikon Ti2 with inverted 60x water immersion objective and spinning disk confocal units.

Chapter 5- Nucleic acid detection in cell-free context

5.1 Introduction

In Chapter 3 and 4, TALE, has been demonstrated as strong probes to detect DNA targets in the cell samples both *in vivo* and *in vitro*.

TALE-ELISA was used to determine the specificity and strength for the TALE protein probe to their respective DNA targets. From the previous results, data showed the TALE proteins can distinguish the DNA targets with a certain level of specificity. Normally low to no signals indicates low to no binding between protein and nucleic acids.

5.1.1 Diagnostic needs for the detection of cell-free nucleic acids

Infectious diseases account for more than one quarter of death per year worldwide (*109*). The existing market has huge amounts of assays for assessing the existence of certain pathogens. Most infectious diseases are detectable by analyzing body fluid from humans, such as saliva, mucus, or blood (*110-113*). Easy access to these specimens makes the biological assay feasible to deploy in a basic laboratory setting or point of care in an even less accessible environment.

Detection and identification of the pathogens for infectious diseases are very important for prevention of spreading and mortality in a global heath staging. Especially in an area lacking basic healthcare infrastructure and laboratory setup, an assay with low cost, fast turnover rate, and moderate sensitivity would be very critical (*114*).

However, despite numerous assays being produced or being developed, new communicable diseases have come every year and easily become a global pandemic. There are several efforts in describing new methodologies with CRISPR system (*24, 113, 115, 116*). However, as mentioned, the design limitations render it harder to target all sequences, especially those in need. A rapid

design scheme for a gene specific detection assay will facilitate the development progress and make the applications of assay in a timely manner.

5.1.2 Universal strategy for cell-free nucleic acid detection with TALE platform

In the TALE platform of nucleic acid detection, the methods could be categorized to two types: solid phase and liquid phase. Solid phase detection is asking for anchoring nucleic acid of interest (NAOI) to a solid phase through streptavidin-biotin reaction. The capture oligos are immobilized to a solid phase where low-fouling surface was pre-functionalized with streptavidin protein at certain concentrations, ranging from 4 to 20 pmol. Normally the NAOI can be captured by biotinylated oligos in polymerase reaction, reverse transcription, or direct annealing. This method used the "sandwich" structure as in traditional ELISA methodology and inherited its benefits as well. Multiple steps of capture, binding, and antibody reaction enable enhancement of specific binding and low background, resulting in high signal to noise ratio (SNR).

Liquid phase detection gives up the solid phase as in magnetic beads or well plates and redirects to a one-pot reaction containing all NAOI and TALE detection modules. Since this method has no background reduction step, such as a wash buffer, the specificity needs to be improved. Also, the final readout has to be a pairwise enzymatic reaction or fluorescence system instead of direct readouts from a single reporter system. TALEs were cloned by switching the domain on C-terminus to split enzymes. Currently there are two systems we adopted: split nano luciferase (sNLuc) and split horseradish peroxidase (sHRP). Both split enzyme systems require two proteins located in proximity range. In a regular liquid environment, due to the Brownian motion, two split systems are difficult to be in close range without facilitation. NAOI provided one molecule carrying both landing pads for TALEs with two enzyme halves. With a proper spacing design, two enzyme halves ought to reconstitute to a fully functional enzyme.

5.1.3 Overview of TALE-based detection methodologies for RNA, ssDNA, and dsDNA

Previous literature suggested TALE could bind to single strand RNA (ssRNA) molecules through DNA-RNA heteroduplex (*117*). TALE would bind to the complementary strand of ssRNA. The DNA-RNA heteroduplex structure could be formed through reverse transcription (RT) or direct annealing of antisense ssDNA. In a non-cell context, TALE binding to nucleic acid targets will be improved due to a less complex environment which is free of an excess amount of potential off-target binding targets (***Figure 5.1***).

Here we demonstrated the functionality of TALE proteins as a detection module for nucleic acids and characterization of the probes in sensitivity and specificity.

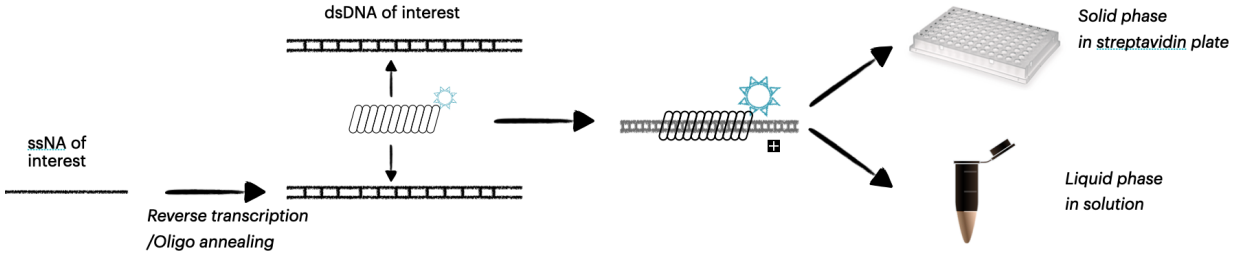

Figure 5.1 Illustration of nucleic acid detection in cell free context

Single stranded nucleic acid was transformed to double strand structure first before TALE protein probing. The transformation was usually conducted by polymerase or reverse transcriptase. Depending on the molecule abundance in the sample, the nucleic acid of interest might be amplified by PCR or RT-PCR or RT-LAMP. For high molecule number, the samples were complemented by matching oligos or RT. When TALE protein probes bind to the desired targets, the protein-NA binding form can be detected by either immobilized streptavidin surface or liquid phase in solution.

5.2 Results and Discussion

5.2.1 TALE probes tolerate multiple mismatches but have different sensitivities towards truncations from N- or C- terminus

The assay sensitivity is important for robustness and performance of probes under different ligands conditions. Better knowledge of assay tolerance to mismatches and truncations in ligands will help us understand the limitations and what is needed for optimization.

TR242 used in imaging context was designed to be exact 242 copies in subtelomeric region in chromosome 13. There is potentially higher copy number due to the volatile variation in the subtelomeric region. Partial binding of the probe was plausible and yet to be determined for future applications. Due to the searching mechanisms of TALE binding to DNA and nature that TALE protein is more considered conservative towards N-terminus of DNA binding domain, TALE proteins might be less sensitive or specific to the C-terminus mismatch.

To evaluate the specificities of TALE probes, deliberation of mismatch was introduced into the target binding oligos. SNR value showed a decreasing trend when the mismatched targets went towards the N-terminus conservative region. TR242 TALE probes showed apparent binding to C-terminal truncations when the mismatch can be up to 9 bps under both binding buffer conditions. N-terminus was considered as a more conservative region in binding. The TR242 probe was more sensitive towards the truncation from the N-terminus. After 3 bp mismatch from initial T, the SNR value reduced drastically. In this probe and specific base pair switching pattern, the sensitivity of mismatch is higher in N-terminus compared to C-terminus. But other possibility of base pair switching, or probe might possess different paradigm of sensitivity bias.

We noticed that the mismatch tolerance is also varied depending on the probe itself. Some probes showed more sensitivity to the truncation mismatch on the C-terminus. TALE probe

AAVS1-B2 binding to PPP1R12C promoter DHS region could tolerate up to 3 bp mismatch and showed a cliff drop in SNR from 1 bp to 3 bp mismatch.

The sensitivity of C-terminus mismatch is more dependent upon the protein itself instead of the assay conditions. From previous data and design principles, we showed that the higher TC ratio resulted in more specific probes because the thymine and cytosine have stronger binding strength.

We noticed that, in TR242, the single mismatch was not sensitive on neither N- nor C- terminus truncation designs. SNR ratio showed constant value along the single base pair mismatch in the targeted DNA binding oligos.

This mismatch tolerance might come from the distorted DNA conformation in the produced ligands. Although the ligand mismatch deliberation was intended to be a sequence where first N base pairs are still identical, the remaining changed sequence was not supposed to form any secondary or partial mismatch structure to allow protein binding. The unorthodox DNA conformation is able to contribute to additional binding free energy (*118*).

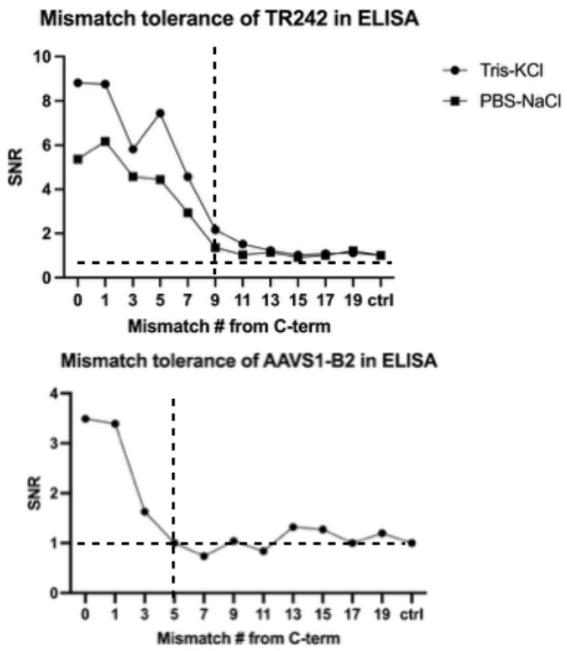

Figure 5.2 Comparison of two TALE probes mismatch tolerance on the C-terminus

Upper is the TR242 probe mismatch tolerance evaluated by TALE-ELISA with two different buffer system. The truncations by deliberated mismatch from C-terminus had consistent reduction trend in SNR. PBS-NaCl and Tris-KCl system did not showed significant difference in the SNR reflection at certain truncation length. One to nine bp mismatches in the dsDNA ligand still had SNR value larger than 1. Lower one is the AAVS1-B2 probe C-terminus mismatch tolerance with same design scheme as upper one. 1 and 3 bp mismatch were the only two demonstrating significant changes compared to the controls.

5.2.2 TALE-ELISA is sensitive towards CpG to TpG substitutions in target DNA

DNA methylation is important in the epigenetic control mechanism of gene expression in cells (*119*). Previous literature reported that the TALE proteins were sensitive towards DNA methylation status (*120-123*). Also, it was reported that the DNA methylation caused a biased TALE binding towards thymine when cytosine was methylated to 5mC (*122*), which probably resulted from the side group modification on fifth position on the carbon ring. As described in Chapter 2, one of the design criteria in the TALE sequence design was to select based on high TC ratio in the DNA binding domain. Because thymine and cytosine are more selective on the respective TALE RVDs, they have potentially stronger and more specific behaviors towards binding TALE protein.

We performed a CpG to TpG substitution in the targeted DNA oligos. 84 probes possess 0 to 3 CpG points in the binding site. In the probes possessing CpG positions, SNR showed reduced value for binding strength of the respective TALE probes towards the corresponding CpG->TpG DNA binding oligos (***Figure 5.3***). Sensitivities towards the substitution increased when DNA binding oligos possessed more substitutions. On the base pair positions, sensitivities increased when the positions of change were moving towards 5' of binding oligos (or to say, N-terminus of TALE proteins). It matched the conservative binding behavior of N-terminus of the TALE protein. TpG substitution was supposed to tolerate the methylation status on the cytosine side chain (*124*) but we failed to see TpG TALE probes binding efficiently to the methylated DNA ligands.

The sensitivity could be applied towards TALE detection of methylated samples in cell context (*125*) but needs more tests on the scale of robustness in the detection.

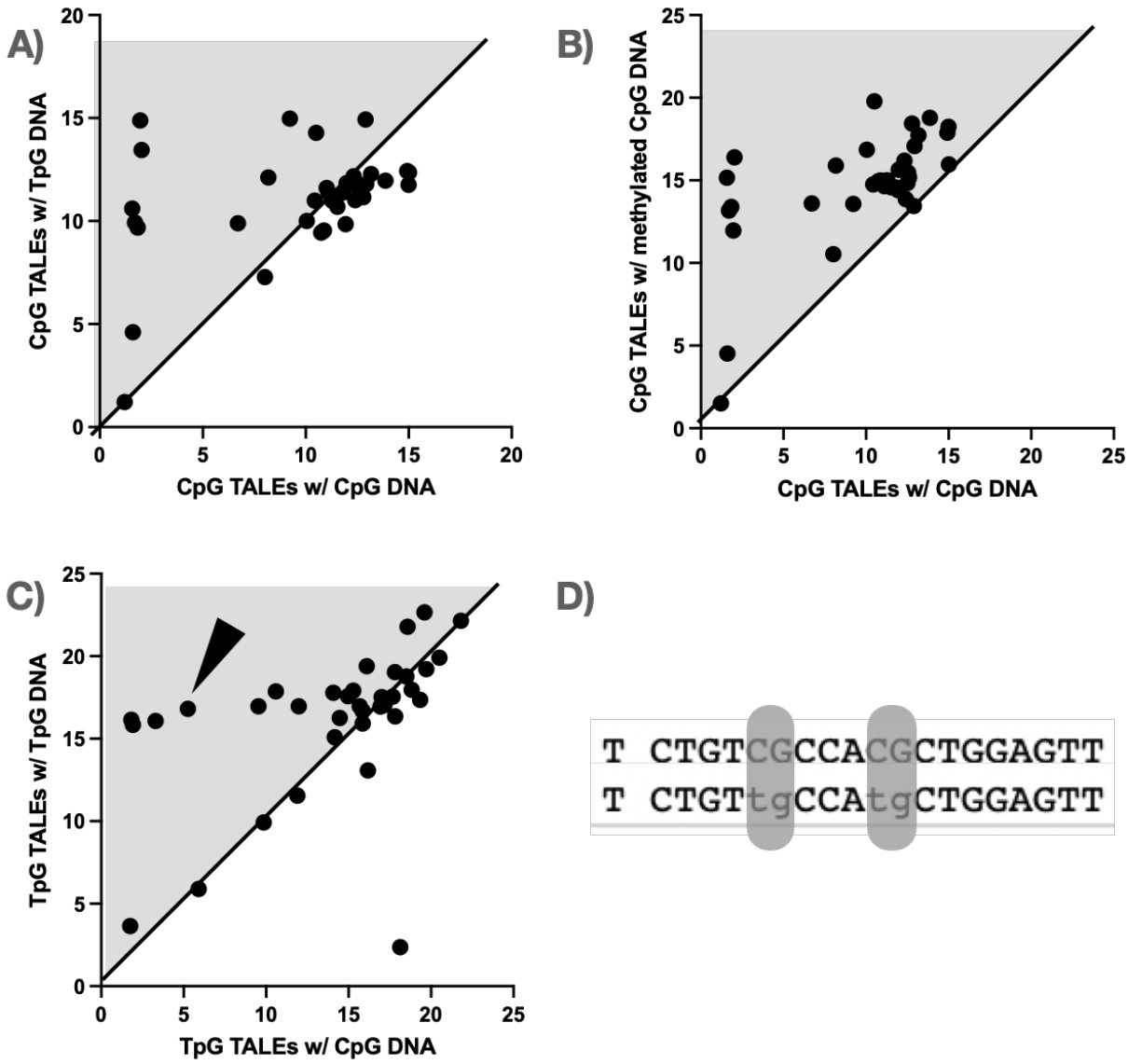

Figure 5.3 AAVS1 probe set sensitivity towards SNP and DNA methylation status

(**A**), (**B**), and (**C**) were comparisons between TALEs with different ligands, which contain 0-3 SNP mismatches introduced by deliberation of base pair switching or M. SSSI methylation on CpG di-residue. If TALE probe is not sensitive to the ligand change, the dots will fall on the middle line. Shadow area highlighted the sensitive region where TALE proteins had reduced SNR in the ligands in the x-axis than y-axis. (**D**) was one of probes in AAVS1-D set which was sensitive to the CpG methylation in two of binding bases.

5.2.3 TALE is able to detect ssDNA and ssRNA with ssDNA facilitation

Single stranded DNA can be detected through TALE probes if the samples are processed with complementary ssDNA oligos. Traditionally the oligo annealing is processed by high temperature denature followed by graduate cooling step to precipitate annealing. Or the samples can be treated with ambient temperature but for longer time. However, the complementary step is not only limited to temperature. As we know, restriction enzyme typically needs a header sequence before its binding region. This header sequence is not necessarily sequence specific but must retain certain length so the protein can have a searching and binding motion. TALE protein also possesses similar searching mechanism in cells (*126, 127*). The local searching mechanism indicates TALE protein might need more than its own binding sequence but more recognition dsDNA structure on the 5' or 3'.

Here we implemented a landing pad identified by one of designs for RNA binding and created multiple length of complementary ssDNA oligos for full length oligo used in TALE-ELISA protocol. The result showed that the addition of sequence beyond protein-binding sequence is necessary for TALE protein binding to its target landing pad. At least five base pairs were necessary to anchor protein to address higher than background signals. Longer sequence on either 5' and 3' demonstrated higher binding affinity but still lower than the PCR products, which was used as same molarity as direct annealing oligos (***Figure 5.4 A***). The efficiency of direct annealing still has lower efficiency and needs further optimization.

Previous literature reported TALE binding to the DNA-RNA heteroduplex by probing the DNA strand side (*117*). To detect ssRNA in the *ex vivo* setting, the targets had to be reformed to either dsDNA by RT-PCR or DNA-RNA duplex by RT or ssDNA annealing. The choice of method depends on the input molecule number. As mentioned above, the current LoD is 10^9 molecules. If

the ssRNA molecule number in the sample of interest is above the threshold, the complement method could be simply reverse transcription or direct annealing by single strand DNA oligos to form a DNA-RNA heteroduplex. If the input molecule number is below the threshold, we must amplify the NAOI to reach the LoD by RT-PCR or RT-LAMP.

TALE proteins contain repetitive RVDs in the DNA binding region which contains 34 amino acids each unit binding to a single DNA base. Except for two amino acids specific for four types of DNA single unit, the 32 amino acids are largely similar. Due to the assembly method, the DNA coding sequences for the homogenous section varied at small scales, but most coding sequences remained the same. Several HBG TALEN mRNAs were taken for an initial study.

17 TALEs were designed targeting the example sequences of a full map TALEN plasmid. 3 were identified to be strong binders during general TALE-ELISA screening. In the RT method, these TALEs demonstrated good binding towards the cDNA-mRNA heteroduplex products and exhibited low background in either mRNA only or blank protein control (***Figure 5.4 B***).

However, the limit of detection (LoD) was determined to be 10 fmol, which is about 10^9 magnitudes of molecule number. This limit also applies to the sensitivity of agarose gel electrophoresis and general qPCR after 30 cycles from a single copy template. The RT method does not increase molecule number from single molecule. The prior trial with TALEN mRNA required 30 ng, which is equivalent to 10^9 limits. If the NAOI is low concentration in the sample, without amplification strategies, there will be a lower chance of detection by the TALE module.

This method could be implemented in the RNA detection in living cell or fixed specimens, compared to the current experiments using MS2-PP7 in CRISPR system (*128, 129*). Living cell mRNA imaging might be more challenging by TALE due to the two-component system.

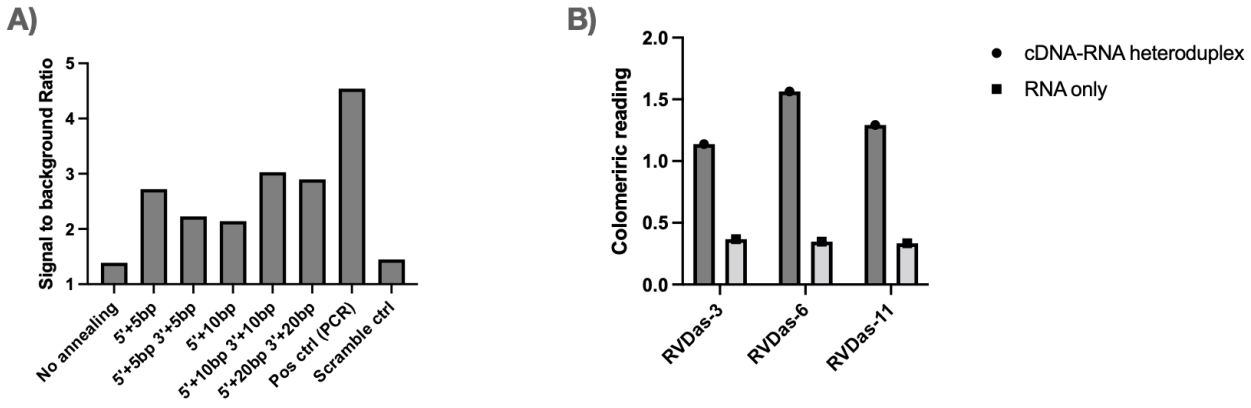

Figure 5.4 TALE binding to the ssDNA and ssRNA with facilitation

TALE RNAas-6 was identified as strong binder towards RVD in TALE sequences. For ssDNA landing pad containing the RNAas-6 binding sequences, multiple oligos were tested about 5' and 3'- sequences to allow TALE binding. (**A**) showed that the 5' end need at least 5 bp and 3' end need at least 10 bp to be efficient binding compared to positive controls. For ssRNA detection, reverse transcription with commercial kit can generate enough template for TALE binding and detection. (**B**) showed the top three binders in TALE mRNA anti-sense strand.

5.2.4 TALE-ELISA could detect viral RNA in low-cycle amplification protocol

TALE-ELISA was further developed to identify long RNA fragment in SARS-CoV-2 viral RNA controls. We designed 16 TALEs for the ORF1a region, 14 TALEs for N protein, and 14 TALEs for RNase P control in the human transcriptome. Produced TALEs were evaluated with standard QC methods described for binding specificity towards targeted dsDNA ligand and binding strength.

Screened TALEs were identified and proceeded with detection assay with low-cycle amplified RT-PCR products from SARS-CoV-2 viral control RNA. So far, we reached 10^3 copy input from RT step and 10^2 copy input in the PCR steps as the limit of detection testing with single TALE probe binding. The agarose gel shown had close-to-sensitivity faint band, but TALE-ELISA still demonstrated strong signal to background ratio. Tandem array of multiple binding probes was able to reduce the total number of cycles in the amplification step and input molecule number in the assay.

To shorten the assay time further, instead of using the TALE protein with efficient labeling tags, we also deployed a direct labeling system aiming at proximate LoD. By fusion with Nano Luciferase peptide to the C-terminal functional domain, the TALE had abilities to implement direct labeling with substrate. We here showed a result of triple replicates of the TALE-NLuc single probe applying to the low cycle RT-PCR products (***Figure 5.5***). It was noticed that the high background of the absolute luminescence unit lies in the blank control group. The lowest signals from 10^2 copy input are still 30% higher than the background. Judging by the experience with traditional TALE-ELISA, this SNR value is not ideal but replicate data showed a consistent trend of linear relationship between assay input molecule number and absolute luminescence unit.

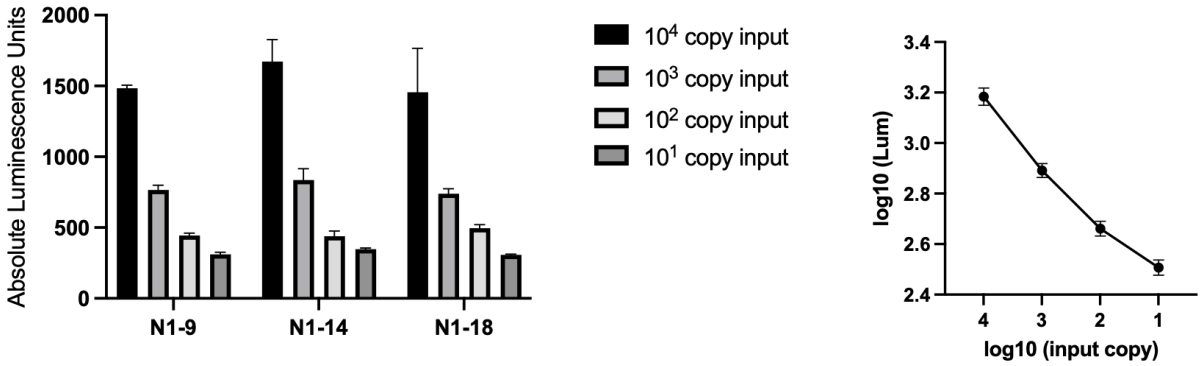

Figure 5.5 Single TALE detection system in viral RNA low cycle amplified products

Shown is three TALE probes designed on N1 flank region in SARS-CoV-2. These three probes were functionalized with Nano Luciferase on the C-terminus. The luminescence signals were directly proportional to the input copy number. Due to the large fold change in input molecule number, log-log plot was easier to visualize the linear correlations.

5.3 Summary

In this chapter I showed an expansion of TALE platform in cell-free context with high sensitivity and specificity in complex environment. TALE-ELISA protocol demonstrated robust response of signals triggered by different functional domains. Although there are certain limitations in the assay specificity towards probe, I have optimized the protocol steps to make the assay more resistant to the unspecific binding events. In the viral RNA detection, TALE-ELISA protocol showed high sensitivity and specificity towards the low-cycle amplified viral targets.

5.4 Materials and Methods

ssDNA design

For the general TALE binding strength ELISA test, the design method was listed in Chapter 2.

PCR

JumpStart enzyme (Sigma) was used to amplify the 60-mer ssDNA targets (IDT) following manufacturer's instruction. The annealing temperature was 62 ºC and elongation time was 15 secs. PCR products were verified by 1.5% agarose gel electrophoresis.

In vitro transcription and translation (IVTT)

TALEs were assembled in pVax vectors (Altius Institute for Biomedical Sciences). The vector contains a T7 promoter on the upstream of the TALE coding sequences. The synthesized plasmids were normalized to 110 ng/μL with molecular biology grade water prior to use. The IVTT reactions were assembled based on manufacturer's instructions (Promega). Depending on the experiment scale, reaction volume can be adjusted as low as 5 μL. The proteins were ready to use

post-reaction or could be used later by mixing with equal volume of 50% glycerol (final glycerol concentration 25%, stored at -20 ^{o}C, one month). Products were verified by Western Blotting and SDS-PAGE.

In vitro transcription of TALEN mRNA

TALEN plasmids were linearized by EagI-HF (New England Biolabs) 500 bp downstream of the poly-A sequence. The linearized DNAs were used as template to produce mRNA by mMachine mMessage T7 kit following manufacturer's instruction (Invitrogen). The products were purified by RNAXP beads (Ampure). The purified mRNAs were analyzed on the Agilent Bioanalyzer to determine product sizes and concentrations.

TALE-ELISA

General screening of TALE binding strength was performed in solid phase detection. Pre-blocked streptavidin coated 8-strips or plate (ThermoFisher Scientific) were used depending on the scale of experiment. The well would be pre-equilibrated with 100 µL BB-2 (1xPBS, 0.5 mM MgCl2, 10 µM ZnCl2, 0.1% Tween-20, 0.1% Bovine Serum Albumin) at room temperature for 5 minutes or longer. 2 µL IVTT products were mixed with 25 µL BB-6 (based BB-2, add additional 5 mM TCEP and 0.02 ng/µL poly(dI-dC)) and incubated at room temperature for 10 minutes. PCR products would not be required to purify if the input molecule number is lower than the well binding capacity (4 pmol). Mix IVTT and 2 µL PCR products and transfer the mixture into equilibrated and decanted wells Incubate at room temperature for 60 minutes and wash 5 times with BB-2. Primary antibodies (anti-FLAG, mouse, Sigma) diluted 1:1000 in BB-1 (1xPBS, 0.05% Tween-20, 2% Bovine Serum Albumin) were added after the washing step. Primary antibodies

were incubated for 60 minutes and washed 5 times with WB-1 (1xPBS, 0.05% Tween-20) afterwards. Secondary antibodies (Goat anti-mouse-HRP, monoclonal, Abcam) diluted 1:1000 in BB-1 were added after the washing step. Secondary antibodies were incubated for 30 minutes in dark and washed 5 times with WB-1 afterwards. TMB (ThermoFisher Scientific) substrate was pre-warmed to room temperature prior to addition. 100 μL TMB substrate was added into wells and a blue color would be developed in 2-10 minutes. 50 μL of 14 mM 450 nm STOP solution (Abcam) was added to stop the conversion and final yellow color could be read by Spectra M2 plate readers (Molecular Devices).

Reverse transcription (RT)

mRNA and sequence-specific biotinylated primer were added with reverse transcription kit components (Revert-Aid cDNA synthesis kit, ThermoFisher Scientific). The cDNA synthesis followed the manufacturer's instructions. The RT products were verified by agarose gel electrophoresis with no template control and RNA only control.

RT-PCR

RT was performed as described above. The products were used in the RT-PCR reaction as templates by no more than 1/10 reaction volume. Primers were either biotinylated or not depending on the experimental group. PCR was performed on manufactures' instruction (Phusion II, ThermoFisher Scientific). Products were verified by agarose gel electrophoresis with proper controls.

Chapter 6 Summary and future directions

In the previous chapters, I described the three applications of TALE platform in nucleic acid detection. Cellular and cell-free contexts gave two edges of investigating into the nucleic acid detection in their natural form. With complementary methodologies, diverse nucleic acid targets were translated into viable target for TALE detection. Generally, TALE platform enabled us with enough foundation for universal methodologies for nucleic acid labeling with considerable efficiency and robustness.

Moreover, the characterization and screening platform is crucial for massive screening of TALE probes. Unlike genome and epigenome editing, single detection probe cannot be easily assessed by its functional assay, such as indel rate or histone marker modification. TALE-ELISA platform showed that the DNA binding domain binding affinity towards designated targets is solely dependent on the DNA-binding domain itself and independent from the functional domain. This indicates that the screening of every single probe is one time work and if the QC-passed probe is switching functional domain, the probe binding affinity will still be trustworthy and no longer need repetitive screening process, assuming assembly protocol has no problem.

All protocols, especially fixed cell imaging, demonstrated one key feature as in reduced time in sample processing. Conventional FISH requires overnight hybridization of oligos towards its target in complex cell environment crowded with macromolecular crowing agents. As shown in Chapter 5, the oligo annealing in a pure ssDNA reaction took one hour to reach half binding molecules. In the TALE fixed cell imaging, the TALE searching, and binding events happened within two hours with possibility to reduce the processing time. The rest of steps is standard immunofluorescence protocol with variance in assay time. This enhanced protocol of DNA imaging sample preparation can drastically increase the efficacy of experimental enablement.

Reduced time, ambient temperature, less disruption to the samples, and cheap probe production render the TALE fixed cell imaging to be a good candidate for future high-throughput applications.

TALE platform is a pure protein-based system for DNA/RNA binding, which means the two ends of protein can functionalized with any possible protein tags and reporters. Currently there are tens of commercial tags at dispose. Moreover, nanobody and single chain variable fragment development has been mature and implemented to create new pairing. The TALE platform can be easily multiplexed and scaled up with multicolor labeling to enhance the scale of simultaneous detection of more gene fragment or transcriptomes in cell samples.

TALE imaging of nucleic acid will be further explored with the improvement of efficacy. The robust labeling of non-repetitive genomic loci needs further optimization in protocol and experimental conditions for high throughput single cell genotyping. Upon the high labeling efficiency of DHSs in the cell context, *in vivo* or *in vitro*, we could use the spots count or spots intensity to evaluate the DHS status in single cell level. TALE platform generates a broad scope in profiling single genome locus in the cellular context. Previous discussion talked about sensitivity of TALE towards heterochromatin in editing and binding events. Contrast to the traditional DNase-I seq or ATAC-seq (*130*), TALE functional labeling of DHS with sensitivity could demonstrate the DHS status at single cell level. Dynamic status of DHS in spatial distribution reflects how the intracellular signal interaction influences the hierarchy of gene regulation in cell population. For example, in a solid tumor, gene perturbation from a drug can affect the peripheral cells versus inside cells shielding from external surface. Through either live cell tracking of DHS labeling or temporal section of tissue demonstrated the regional gene response. Instead of investigating enhancer-promoter interactions, loss and gain of DHS in gene regulation indicates the real-time local chromatin behavior in small gene regions (c.a. 1kb). Single cell level genetic

signature with dynamic status tracking will be easily accessible through common lab instrumentations.

Recent findings revealed more possibilities in the pair-wise signal amplification and specificty improvement by split-HRP and Proximity Ligation Assay (PLA) (*131*). Similar to ZFN and TALEN, pair-wise interactions require a dimerization between two DBDs in a close proximity range. Like FRET (*132*) and BRET (*133*), the signals have to be generated when two probes sit on a small region. We have generated a preliminary trials by PLA to prove the pilot result of single pair of TALE labeling HS2 in LCR (not shown). The multi-fold signal amplification results in higher sensitivity to lower copy of nucleic acid in the biological specimens. Single element resolution needs smaller probe set enrichment in a targeted region, PLA or split HRP methods are compatible with TALE platform and could titrate the resolution to 100 bp in theory. TALE platform in fixed cell imaging can start label trace interactions among regulatory elements in close proximity. With the current super resolution imaging tool, the labeling is highly functional and will offer insight into spatial distribution of *trans*- and *cis*- regulatory elements profiling.

To extensively explore the motion of TALE protein kinetics in the fixed specimens, we could deploy similar methodology (*88, 97*). The motion dynamics of protein will be significant for us to understand the kinetics of protein searching mechanism in the targeted regions and potentially comprehension of context-driven TALE binding profile, especially near heterochromatin. Enhanced instrumentation, such as Light Sheet Fluorescence Microscopy (LSFM) (*134*), enables us to dissect the movement of the protein molecules in a higher resolution and better imaging quality.

There are massive potential in exploration of nucleic acid imaging and detection by TALEs in various applications, which can be beneficial towards molecular and cellular biology, with further

expansion towards single cell or single element profiling of genomic elements in cellular contexts. This will largely advance our technology development and fundamental research in genetics and clinical applications.

Literature cited

1. A. Ganguli *et al.*, Reverse Transcription Loop-Mediated Isothermal Amplification Assay for Ultrasensitive Detection of SARS-CoV-2 in Saliva and Viral Transport Medium Clinical Samples. *Anal Chem* **93**, 7797-7807 (2021).
2. M. T. Hwang *et al.*, Ultrasensitive detection of nucleic acids using deformed graphene channel field effect biosensors. *Nat Commun* **11**, 1543 (2020).
3. K. L. Taneja, Localization of trinucleotide repeat sequences in myotonic dystrophy cells using a single fluorochrome-labeled PNA probe. *Biotechniques* **24**, 472-476 (1998).
4. O. Kalinina, I. Lebedeva, J. Brown, J. Silver, Nanoliter scale PCR with TaqMan detection. *Nucleic Acids Research* **25**, 1999-2004 (1997).
5. G. P. McDermott *et al.*, Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR. *Anal Chem* **85**, 11619-11627 (2013).
6. D. B. Schowalter, S. S. Sommer, The generation of radiolabeled DNA and RNA probes with polymerase chain reaction. *Analytical Biochemistry* **177**, 90-94 (1989).
7. J. G. Wetmur, DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol* **26**, 227-259 (1991).
8. P. Zrazhevskiy, X. Gao, Quantum dot imaging platform for single-cell molecular profiling. *Nat Commun* **4**, 1619 (2013).
9. T. Ozawa, A. Kaihara, M. Sato, K. Tachihara, Y. Umezawa, Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal Chem* **73**, 2516-2521 (2001).
10. J. C. Chien *et al.*, A multiplexed bioluminescent reporter for sensitive and non-invasive tracking of DNA double strand break repair dynamics *in vitro* and *in vivo*. *Nucleic Acids Res* **48**, e100 (2020).
11. A. Raap *et al.*, Ultra-sensitive FISH using peroxidase-mediated deposition of biotin-or fluorochrome tyramides. *Human Molecular Genetics* **4**, 529-534 (1995).
12. N. Rufer, W. Dragowska, G. Thornbury, E. Roosnek, P. M. Lansdorp, Telomere length dynamics in human lymphocyte subpopulations measured by flow cytometry. *Nat Biotechnol* **16**, 743-747 (1998).
13. E. M. Southern, Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517 (1975).
14. P. M. Holland, R. D. Abramson, R. Watson, D. H. Gelfand, Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* **88**, 7276-7280 (1991).
15. P. R. Langer-Safer, M. Levine, D. C. Ward, Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A* **79**, 4381-4385 (1982).
16. A. F. Savulescu *et al.*, Interrogating RNA and protein spatial subcellular distribution in smFISH data with DypFISH. *Cell Reports Methods* **1**, (2021).
17. J. M. Levsky, R. H. Singer, Fluorescence *in situ* hybridization: past, present and future. *Journal of Cell Science* **116**, 2833-2838 (2003).
18. B. I. Lindhout *et al.*, Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res* **35**, e107 (2007).
19. J. C. Miller *et al.*, An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-785 (2007).

20. B. Chen *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).
21. B. Chen *et al.*, Expanding the CRISPR imaging toolset with Staphylococcus aureus Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res* **44**, e75 (2016).
22. W. Deng, X. Shi, R. Tjian, T. Lionnet, R. H. Singer, CASFISH: CRISPR/Cas9-mediated *in situ* labeling of genomic loci in fixed cells. *Proc Natl Acad Sci U S A* **112**, 11870-11875 (2015).
23. Y. Fu *et al.*, CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat Commun* **7**, 11707 (2016).
24. R. Aman, A. Mahas, M. Mahfouz, Nucleic Acid Detection Using CRISPR/Cas Biosensing Technologies. *ACS Synth Biol* **9**, 1226-1233 (2020).
25. T. Cermak *et al.*, Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).
26. J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436 (2010).
27. F. Zhang *et al.*, Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011).
28. T. Sakuma *et al.*, Efficient TALEN construction and evaluation methods for human cell and animal applications. *Genes Cells* **18**, 315-326 (2013).
29. S. Boissel *et al.*, megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res* **42**, 2591-2601 (2014).
30. D. Reyon *et al.*, FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460-465 (2012).
31. T. Gaj, C. A. Gersbach, C. F. Barbas, 3rd, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397-405 (2013).
32. J. Liu, T. Gaj, J. T. Patterson, S. J. Sirk, C. F. Barbas, 3rd, Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. *PLoS One* **9**, e85755 (2014).
33. L. Poirot *et al.*, Multiplex Genome-Edited T-cell Manufacturing Platform for "Off-the-Shelf" Adoptive T-cell Immunotherapies. *Cancer Res* **75**, 3853-3864 (2015).
34. U. Mock *et al.*, mRNA transfection of a novel TAL effector nuclease (TALEN) facilitates efficient knockout of HIV co-receptor CCR5. *Nucleic Acids Res* **43**, 5560-5571 (2015).
35. J. Liu *et al.*, Efficient delivery of nuclease proteins for genome editing in human stem cells and primary cells. *Nat Protoc* **10**, 1842-1859 (2015).
36. E. A. Josephs *et al.*, Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **43**, 8924-8941 (2015).
37. H. Wan, J. P. Hu, K. S. Li, X. H. Tian, S. Chang, Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PLoS One* **8**, e76045 (2013).
38. J. F. Meckler *et al.*, Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118-4128 (2013).
39. S. Belikov, O. G. Berg, O. Wrange, Quantification of transcription factor-DNA binding affinity in a living cell. *Nucleic Acids Res* **44**, 3045-3058 (2016).
40. M. I. E. Uusi-Makela *et al.*, Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLoS One* **13**, e0196238 (2018).
41. B. Chen, J. Guan, B. Huang, Imaging Specific Genomic DNA in Living Cells. *Annu Rev Biophys* **45**, 1-23 (2016).

42. M. Christian *et al.*, Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761 (2010).
43. J. C. Miller *et al.*, A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).
44. Z. Zhang, E. Wu, Z. Qian, W. S. Wu, A multicolor panel of TALE-KRAB based transcriptional repressor vectors enabling knockdown of multiple gene targets. *Sci Rep* **4**, 7338 (2014).
45. A. Amabile *et al.*, Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167**, 219-232 e214 (2016).
46. M. L. Maeder *et al.*, Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137-1142 (2013).
47. K. Thanisch *et al.*, Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Res* **42**, e38 (2014).
48. H. Ma, P. Reyes-Gutierrez, T. Pederson, Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc Natl Acad Sci U S A* **110**, 21048-21053 (2013).
49. R. Ren *et al.*, Visualization of aging-associated chromatin alterations with an engineered TALE system. *Cell Res* **27**, 483-504 (2017).
50. Y. Geng, A. Pertsinidis, Simple and versatile imaging of genomic loci in live mammalian cells and early pre-implantation embryos using CAS-LiveFISH. *Sci Rep* **11**, 12220 (2021).
51. G. Wang, C. L. Achim, R. L. Hamilton, C. A. Wiley, V. Soontornniyomkij, Tyramide signal amplification method in multiple-label immunofluorescence confocal microscopy. *Methods* **18**, 459-464 (1999).
52. J. Streubel, C. Blucher, A. Landgraf, J. Boch, TAL effector RVD specificities and efficiencies. *Nat Biotechnol* **30**, 593-595 (2012).
53. L. Cong, R. Zhou, Y. C. Kuo, M. Cunniff, F. Zhang, Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968 (2012).
54. F. C. Rinaldi, L. A. Doyle, B. L. Stoddard, A. J. Bogdanove, The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Res* **45**, 6960-6970 (2017).
55. B. M. Lamb, A. C. Mercer, C. F. Barbas, 3rd, Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res* **41**, 9779-9785 (2013).
56. N. C. Shaner *et al.*, A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods* **10**, 407-409 (2013).
57. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646 (2014).
58. J. D. Martell *et al.*, A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat Biotechnol* **34**, 774-780 (2016).
59. A. S. Dixon *et al.* (Google Patents, 2017).
60. W. Hoerz, K. S. McCarty, Initiation of protein synthesis in a rabbit reticulocyte lysate system. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **228**, 526-535 (1971).

61. M. Anastasina, I. Terenin, S. J. Butcher, D. E. Kainov, A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *Biotechniques* **56**, 36-39 (2014).
62. R. O'Brien, B. DeDecker, K. G. Fleming, P. B. Sigler, J. E. Ladbury, The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic archaeon. *J Mol Biol* **279**, 117-125 (1998).
63. T. J. Treangen, S. L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**, 36-46 (2011).
64. A. Mukhopadhyay, B. Deplancke, A. J. Walhout, H. A. Tissenbaum, Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. *Nat Protoc* **3**, 698-709 (2008).
65. T. H. Kim, J. Dekker, ChIP-Quantitative Polymerase Chain Reaction (ChIP-qPCR). *Cold Spring Harb Protoc* **2018**, (2018).
66. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *Elife* **8**, (2019).
67. D. H. Janssens *et al.*, Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat Genet* **53**, 1586-1596 (2021).
68. P. Ward, C. E. Walsh, Targeted integration of a rAAV vector into the AAVS1 region. *Virology* **433**, 356-366 (2012).
69. Q. Li, K. R. Peterson, X. Fang, G. Stamatoyannopoulos, Locus control regions. *Blood* **100**, 3077-3086 (2002).
70. J. Kooren *et al.*, Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J Biol Chem* **282**, 16544-16552 (2007).
71. D. Noordermeer, W. de Laat, Joining the loops: beta-globin gene regulation. *IUBMB Life* **60**, 824-833 (2008).
72. W. de Laat, F. Grosveld, Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447-459 (2003).
73. D. A. Jackson, J. C. McDowell, A. Dean, Beta-globin locus control region HS2 and HS3 interact structurally and functionally. *Nucleic Acids Res* **31**, 1180-1190 (2003).
74. S. Mukherjee *et al.*, Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat Genet* **36**, 1331-1339 (2004).
75. K. Clauss *et al.*, DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res* **45**, 11121-11130 (2017).
76. J. R. Mansfield, Imaging in cancer immunology: phenotyping immune cell subsets in situ in FFPE tissue sections. *MLO Med Lab Obs* **46**, 12-13 (2014).
77. A. Murrell, V. K. Rakyan, S. Beck, From genome to epigenome. *Human Molecular Genetics* **14**, R3-R10 (2005).
78. D. Michieletto, E. Orlandini, D. Marenduzzo, Polymer model with Epigenetic Recoloring Reveals a Pathway for the de novo Establishment and 3D Organization of Chromatin Domains. *Physical Review X* **6**, 041047 (2016).
79. A. Wolffe, *Chromatin: structure and function*. (Academic press, 1998).
80. E. P. Consortium *et al.*, Perspectives on ENCODE. *Nature* **583**, 693-698 (2020).
81. G. Felsenfeld, J. Boyes, J. Chung, D. Clark, V. Studitsky, Chromatin structure and gene expression. *Proc Natl Acad Sci U S A* **93**, 9384-9388 (1996).
82. W. Jin *et al.*, Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142-146 (2015).

83. S. Bradley, L. Zamechek, J. Aurich-Costa, in *Fluorescence In Situ Hybridization (FISH) — Application Guide,* T. Liehr, Ed. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), pp. 67-73.
84. M. Wang *et al.*, RCasFISH: CRISPR/dCas9-Mediated *in Situ* Imaging of mRNA Transcripts in Fixed Cells and Tissues. *Anal Chem* **92**, 2468-2475 (2020).
85. E. A. Hoffman, B. L. Frey, L. M. Smith, D. T. Auble, Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* **290**, 26404-26411 (2015).
86. D. M. Miller, D. C. Shakes, Immunofluorescence microscopy. *Methods in cell biology* **48**, 365-394 (1995).
87. S. Jitrakorn *et al.*, *In situ* DIG-labeling, loop-mediated DNA Amplification (ISDL) for highly sensitive detection of infectious hypodermal and hematopoietic necrosis virus (IHHNV). *Aquaculture* **456**, 36-43 (2016).
88. S. Jain *et al.*, TALEN outperforms Cas9 in editing heterochromatin target sites. *Nat Commun* **12**, 606 (2021).
89. N. Ogawa, M. D. Biggin, in *Gene Regulatory Networks*. (Springer, 2012), pp. 51-63.
90. K. Larouche, M. J. Bergeron, S. Leclerc, S. L. Guerin, Optimization of competitor poly(dI-dC).poly(dI-dC) levels is advised in DNA-protein interaction studies involving enriched nuclear proteins. *Biotechniques* **20**, 439-444 (1996).
91. L. Cuculis *et al.*, Divalent cations promote TALE DNA-binding specificity. *Nucleic Acids Res* **48**, 1406-1422 (2020).
92. K. Ancelin, C. Brun, E. Gilson, Role of the telomeric DNA‐binding protein TRF2 in the stability of human chromosome ends. *Bioessays* **20**, 879-883 (1998).
93. A. Janssen, R. H. Medema, Genetic instability: tipping the balance. *Oncogene* **32**, 4459-4470 (2013).
94. M. A. Caporizzo, B. L. Prosser, The microtubule cytoskeleton in cardiac mechanics and heart failure. *Nature Reviews Cardiology*, 1-15 (2022).
95. D. M. Shechner, E. Hacisuleyman, S. T. Younger, J. L. Rinn, Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* **12**, 664-670 (2015).
96. H. Ma *et al.*, Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* **112**, 3002-3007 (2015).
97. S. C. Knight *et al.*, Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823-826 (2015).
98. H. Ma *et al.*, Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* **34**, 528-530 (2016).
99. S. Wang, J. H. Su, F. Zhang, X. Zhuang, An RNA-aptamer-based two-color CRISPR labeling system. *Sci Rep* **6**, 26857 (2016).
100. S. Shao *et al.*, Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res* **44**, e86 (2016).
101. P. Qin *et al.*, Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat Commun* **8**, 14725 (2017).
102. Y. Takei, S. Shah, S. Harvey, L. S. Qi, L. Cai, Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. *Biophys J* **112**, 1773-1776 (2017).
103. H. Wang *et al.*, CRISPR-mediated live imaging of genome editing and transcription. *Science* **365**, 1301-1305 (2019).

104. X. H. Zhang, L. Y. Tee, X. G. Wang, Q. S. Huang, S. H. Yang, Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
105. B. Gu *et al.*, Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050-1055 (2018).
106. J. A. Zuris *et al.*, Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat Biotechnol* **33**, 73-80 (2015).
107. Y. H. Huang *et al.*, DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. *Genome Biol* **18**, 176 (2017).
108. D. Virant *et al.*, A peptide tag-specific nanobody enables high-quality labeling for dSTORM imaging. *Nat Commun* **9**, 930 (2018).
109. WHO. (World Health Organization, 2020).
110. D. G. Bausch *et al.*, Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. *The Journal of infectious diseases* **196**, S142-S147 (2007).
111. R. H. Yolken, Enzyme immunoassays for the detection of infectious antigens in body fluids: current limitations and future prospects. *Reviews of infectious diseases* **4**, 35-68 (1982).
112. H. Lee, S. I. Kim, Review of Liquid Chromatography-Mass Spectrometry-Based Proteomic Analyses of Body Fluids to Diagnose Infectious Diseases. *International Journal of Molecular Sciences* **23**, 2187 (2022).
113. B. Casati *et al.*, ADESSO: a rapid, adaptable and sensitive Cas13-based COVID-19 diagnostic platform. *medRxiv*, 2021.2006.2017.21258371 (2021).
114. H. Chen, K. Liu, Z. Li, P. Wang, Point of care testing for infectious diseases. *Clinica chimica acta* **493**, 138-147 (2019).
115. T. Y. Liu *et al.*, Accelerated RNA detection using tandem CRISPR nucleases. *Nat Chem Biol* **17**, 982-988 (2021).
116. L. Zhou, R. Peng, R. Zhang, J. Li, The applications of CRISPR/Cas system in molecular detection. *J Cell Mol Med* **22**, 5807-5815 (2018).
117. P. Yin *et al.*, Specific DNA-RNA hybrid recognition by TAL effectors. *Cell Rep* **2**, 707-713 (2012).
118. A. Afek *et al.*, DNA mismatches reveal conformational penalties in protein-DNA recognition. *Nature* **587**, 291-296 (2020).
119. T. I. Yusufaly, Y. Li, W. K. Olson, 5-Methylation of cytosine in CG:CG base-pair steps: a physicochemical mechanism for the epigenetic control of DNA nanomechanics. *J Phys Chem B* **117**, 16436-16442 (2013).
120. D. Deng *et al.*, Recognition of methylated DNA by TAL effectors. *Cell Res* **22**, 1502-1504 (2012).
121. H. Kaya, H. Numa, A. Nishizawa-Yokoi, S. Toki, Y. Habu, DNA Methylation Affects the Efficiency of Transcription Activator-Like Effector Nucleases-Mediated Genome Editing in Rice. *Front Plant Sci* **8**, 302 (2017).
122. S. Tsuji, K. Shinoda, S. Futaki, M. Imanishi, Sequence-specific 5mC detection in live cells based on the TALE-split luciferase complementation system. *Analyst* **143**, 3793-3797 (2018).
123. G. Kubik, M. J. Schmidt, J. E. Penner, D. Summerer, Programmable and highly resolved *in vitro* detection of 5-methylcytosine by TALEs. *Angew Chem Int Ed Engl* **53**, 6002-6006 (2014).
124. J. Valton *et al.*, Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* **287**, 38427-38432 (2012).

125. K. L. Kim *et al.*, Systematic detection of m(6)A-modified transcripts at single-molecule and single-cell resolution. *Cell Rep Methods* **1**, (2021).
126. L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat Commun* **6**, 7277 (2015).
127. L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, TALE proteins search DNA using a rotationally decoupled mechanism. *Nat Chem Biol* **12**, 831-837 (2016).
128. B. Wu, J. Chen, R. H. Singer, Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci Rep* **4**, 3615 (2014).
129. J. Wu *et al.*, Live imaging of mRNA using RNA-stabilized fluorogenic proteins. *Nat Methods* **16**, 862-865 (2019).
130. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
131. K. Zhang *et al.*, Direct Visualization of Single-Nucleotide Variation in mtDNA Using a CRISPR/Cas9-Mediated Proximity Ligation Assay. *J Am Chem Soc* **140**, 11293-11301 (2018).
132. R. B. Sekar, A. Periasamy, Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *The Journal of cell biology* **160**, 629 (2003).
133. Y. Xu, D. W. Piston, C. H. Johnson, A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proceedings of the National Academy of Sciences* **96**, 151-156 (1999).
134. J. Lim, H. K. Lee, W. Yu, S. Ahmed, Light sheet fluorescence microscopy (LSFM): past, present and future. *Analyst* **139**, 4758-4768 (2014).

Appendices

Appendix I: Acronyms

AUC: Area under Curve

BLAST: Basic Local Alignment Search Tool

ChIP: Chromatin Immuno-Precipitation

CRISPR: Clustered Regularly Interspaced Palindromic Repeats

CUT&RUN: Cut under target and Release under nuclease

DHS: DNase I Hypersensitive Site

DSB: Double Strand Break

ELISA: Enzyme-linked Immunosorbent Assay

EMSA: Electrophoretic Mobility Shift Assay

FISH: Fluorescence *in situ* Hybridization

GFP: Green Fluorescence Protein

gRNA: guide RNA

HDR: Homologous Directed Recombination

HRP: Horseradish Peroxidase

IF: Immuno-Fluorescence

IVT: *In vitro* transcription

IVTT: *In vitro* transcription/ translation

LAMP: Loop-Mediated Isothermal Amplification

LCR: Locus Control Region

LoD: Limit of Detection

NAOI: Nucleic acid of Interest

NLS: Nuclear Localization Signals

NLuc: Nano Luciferase

PAM: Protospacer Adjacent Motif

PCR: Polymerase Chain Reaction

RT: Reverse Transcription

RVD: Repeat Variable Di-residues

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

SNR: Signal to Noise Ratio

SPS: Single Probe Staining

TALE: Transcription Activator-like Effector

TALEN: Transcription Activator-like Effector Nuclease

TSA: Tyramide Signal Amplification

Appendix II: QC data of TALE screening and detection platform

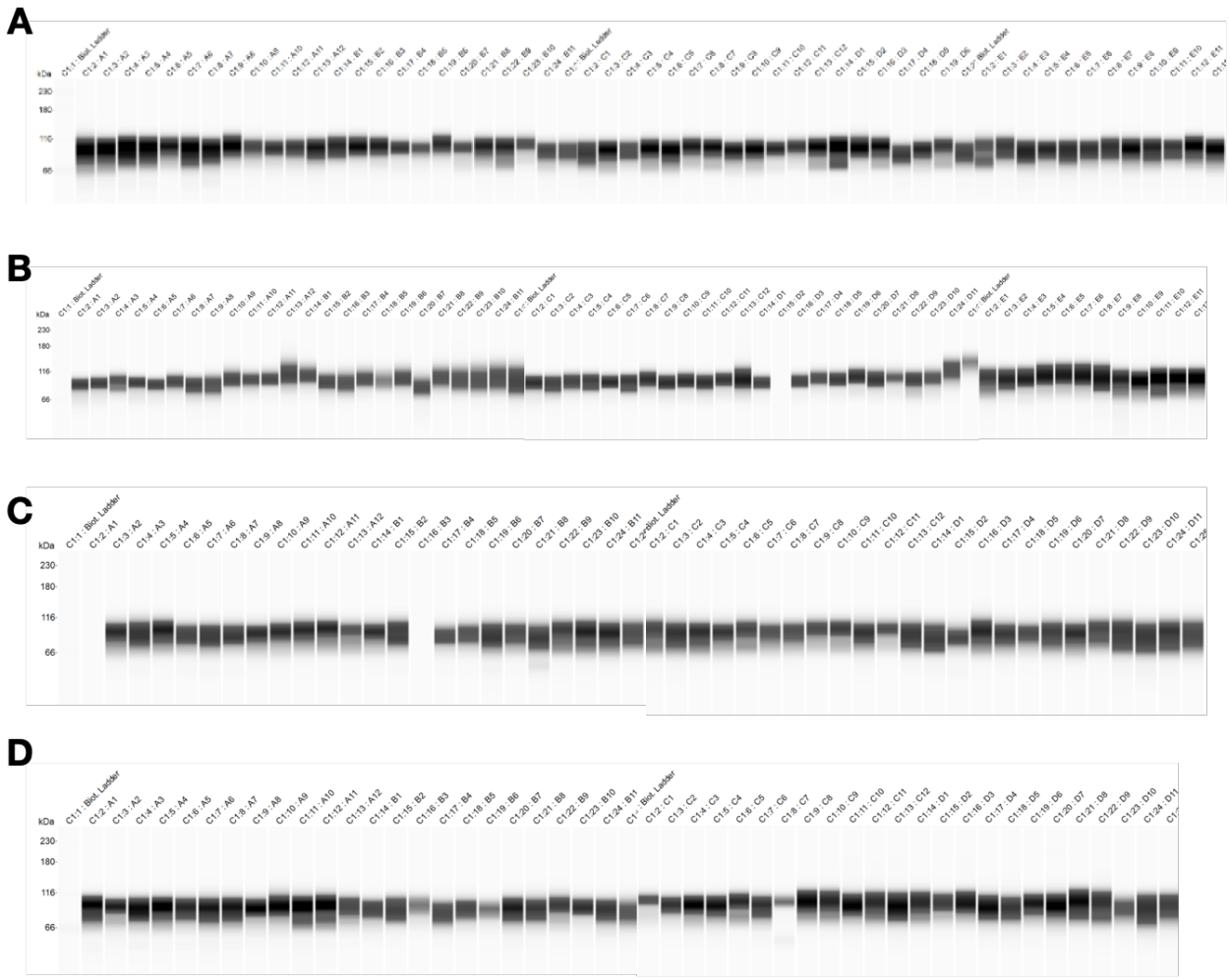

Figure S1. Western Blot QC gels of TALE probes used in cellular context high complexity nucleic acid targets detection

Shown are (**A**) AAVS1-SunTag probes, (**B**) AAVS1-mNeonGreen probes, (**C**) LCR-SunTag probes, and (**D**) LCR-mNeonGreen probes. Reference size for the TALE proteins falls in range of 85 to 110 kDa, depending on the DBD length and functional domain sizes.

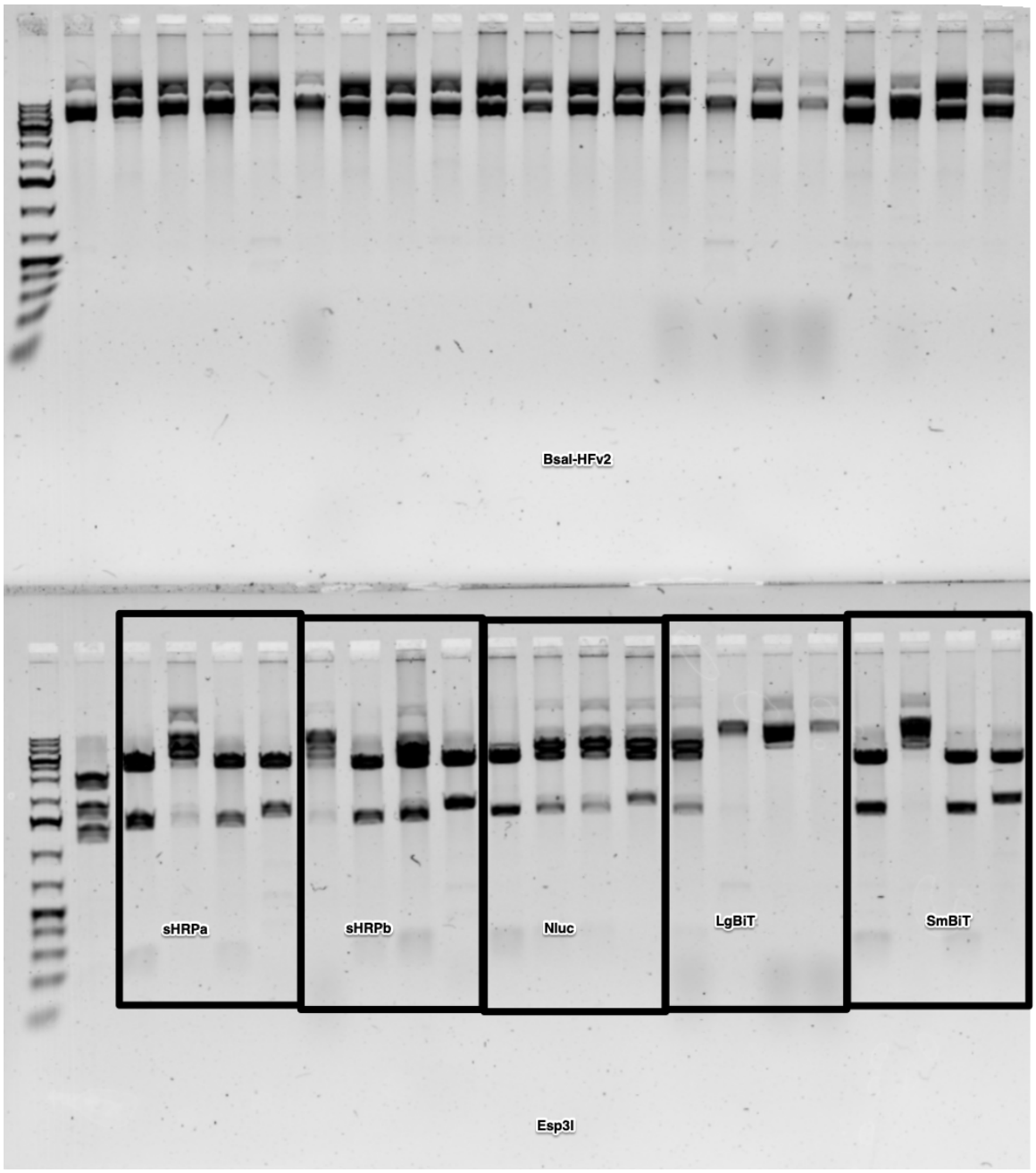

Figure S2. Verification of selected TALE probe assembly destination vectors

Shown ladder on the left is Generuler 1kb plus (ThermoFisher Scientific). The digestion was to verify the size of functional domain and successful cloning. Within each black box are different colonies for same constructs.

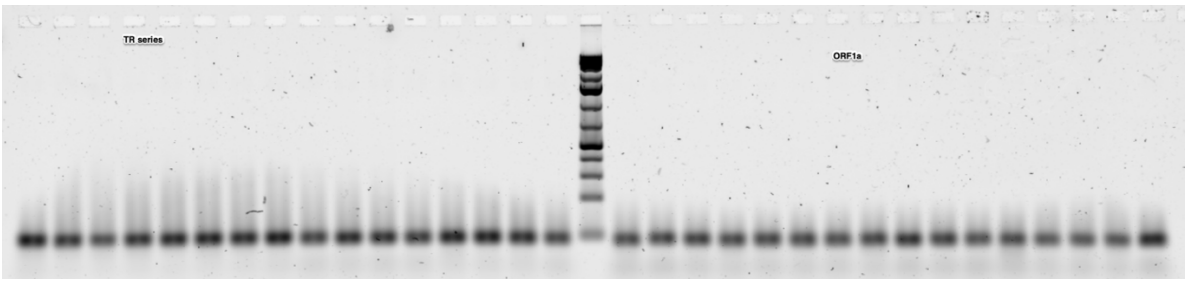

Figure S3. Sample QC gel for TALE-ELISA oligos PCR

Shown ladder is Generuler 1kb plus (ThermoFisher Scientific). The left half were oligos for TR series and right half are oligos for SARS-CoV-2 viral RNA ORF1α. Expected amplicon size is 60 bp.

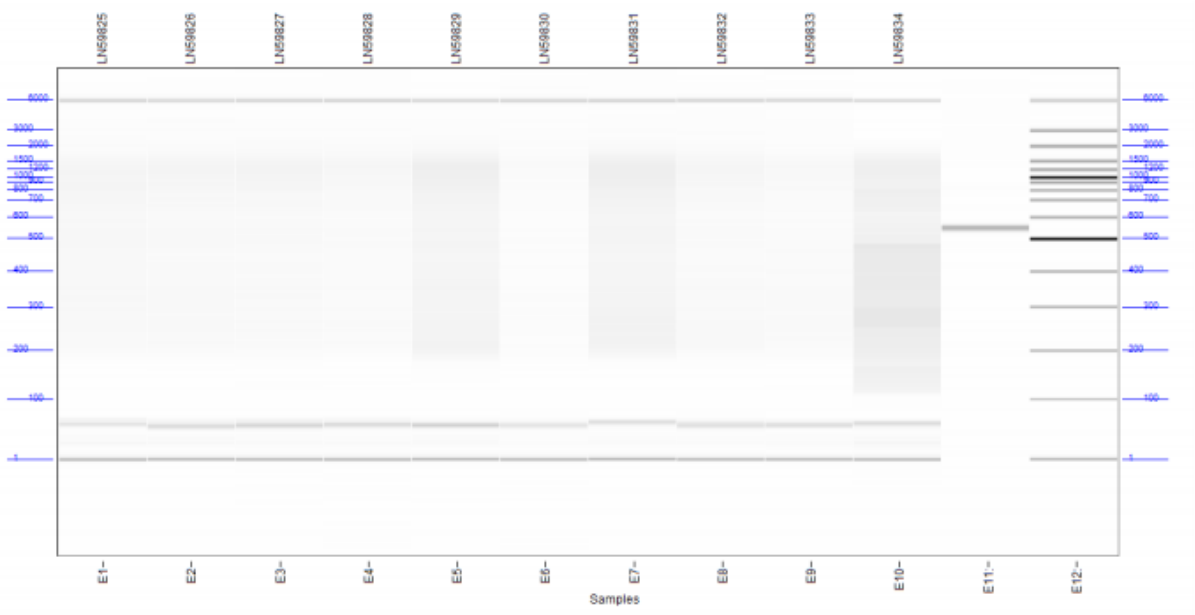

Figure S4. Sample QC for CUT&RUN assay trace gel for NGS library

Shown is trace gel for CUT&RUN fragment released from assay before first round amplification and barcoding. Products were around 70 bp.

Appendix III: Supplementary Figures

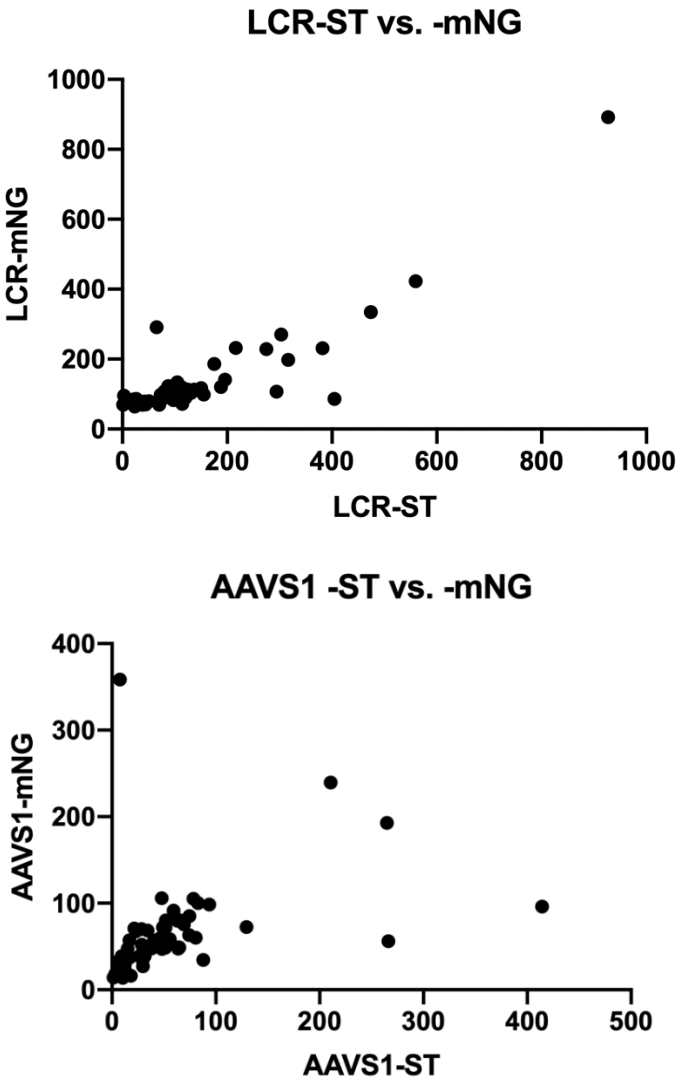

Figure S5. Correlation between SPS intensities from same probe set with two different functional domains

The pattern showed a linear correlation between two sets with same DNA binding domains. However, the intensities can also be influenced by cell contexts and off-target entrapment. When most of probes are specific, the data points are mostly clustered around ground zero.

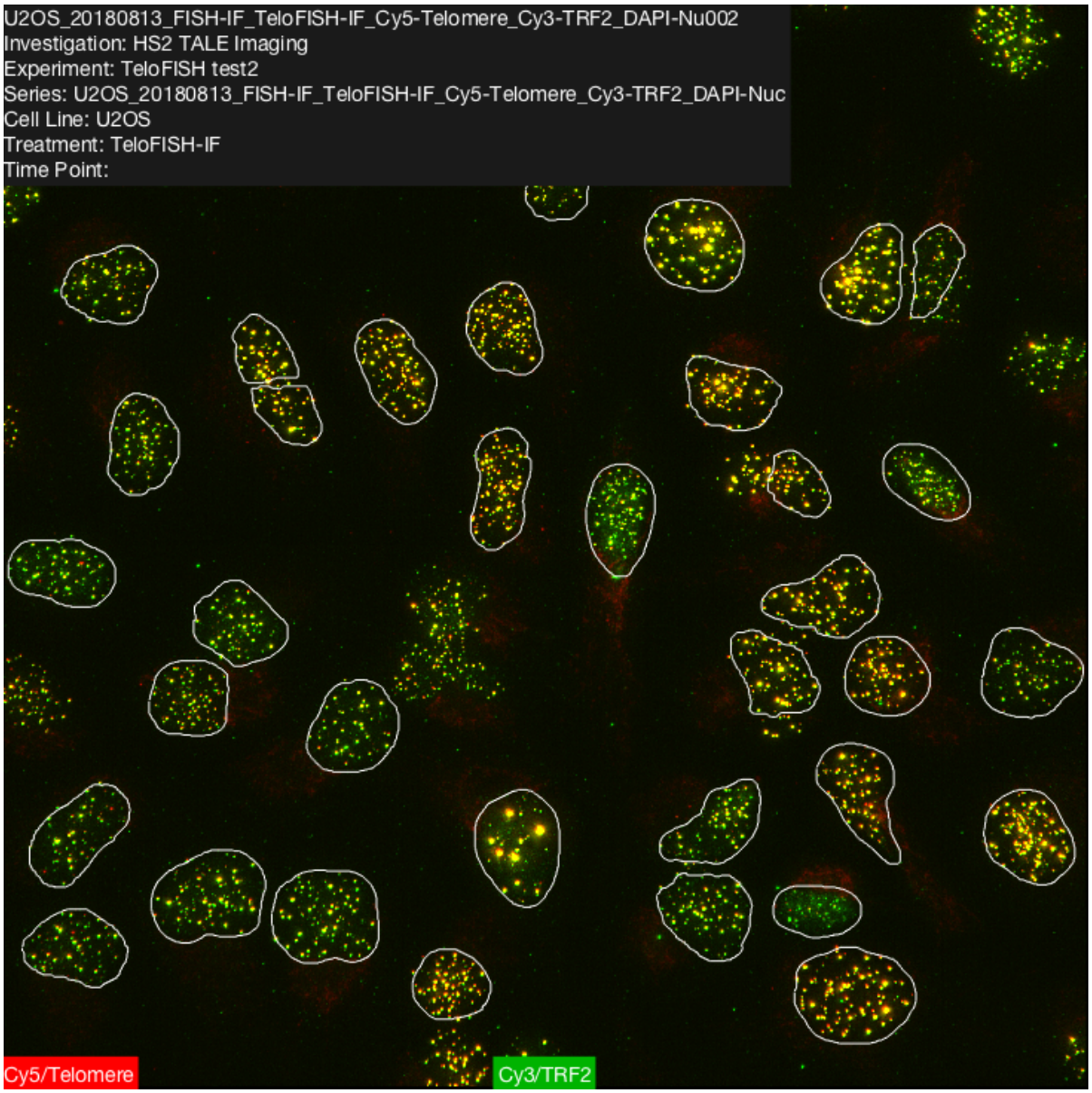

Figure S6. Telomere FISH colocalization with TRF2 protein

Shown is colocalization test between Telomere FISH (Cambio) and TRF2 antibody (Santa Cruz Biotechnology). Red color is telomere FISH stained with streptavidin-Cy5 conjugates and green color is TRF2 staining with anti-mouse-Cy3 conjugates. Yellow is colocalized spots. White outlines are nuclei boundaries after segmentation. TRF2 protein showed the telomere spots and served as surrogate to colocalize with TeloS1 TALE probes used in live cell imaging.

VITA

Ruihong “Redd” Wang took his English name from one of his favorite movies, *The Shawshank Redemption*, as being fond of the character Ellis Redding.

He was born in Changsha and spent his first 18 years in China. Starting college in US, he has stayed in the Emerald city for almost 12 years. He really likes the backcountry of this state and tried to make use of the nature for the spare time.