UNIVERSITIY OF CALIFORNIA SAN DIEGO

Visualizing the Dynamics of 4D Nuclear Architecture during Enhancer-Promoter Transcription Activation Using Live Imaging Techniques

A thesis submitted in partial satisfaction of the requirements
for the degree of Master of Science

in

Biology

by

Yeeun Leah Kim

Committee in charge:

Professor Michael Geoffrey Rosenfeld, Chair
Professor Cornelis Murre, Co-Chair
Professor James Kadonaga

2020

The Thesis of Yeeun Leah Kim is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____
Co-chair

_____
Chair

University of California San Diego

2020

DEDICATIONS

I would like to dedicate my thesis to my parents, who have been unconditionally supporting me throughout my life. Without them, I could have not pursued my education this far, especially away from my home country. I would also like to dedicate my thesis to my friends, who always cheered me up and helped me stand up again whenever I encountered hard times.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

Polymerase II	Pol II
Homology-directed repair	HDR
Liquid-liquid phase separation	LLPS
PCP	PP7 coat proteins
MCP	MS2 coat proteins
dCas9	Nuclease-deactivated Cas9
PRO-Seq	Precision nuclear run-on sequencing
4-C	Circular chromosome conformation capture
TFF1e	TFF1 enhancer
c-MYCe	c-MYC enhancer
HPV	Human Papillomavirus

ACKNOWLEDGMENTS

I would like to acknowledge Professor Michael Geoff Rosenfeld for his full support as my committee chair. Your advice and vision made me contribute my life in academia, pursuing to be a professor like you who has a lifelong passion and dedication in research as well as in education.

I would like to acknowledge Professor Cornelis Kees Murre for his support as my committee co-chair. Your kind words made me feel that I belong in academia and motivated me to push harder beyond my limit.

I would like to acknowledge Professor James Kadonaga for his support as my last but not least committee member. You not only sparked my interest in molecular biology via BIMM112 class, but also taught me how to direct my love of molecular biology as a successful scientist via BGGN202 class.

I would like to acknowledge Susan Wang and Dr. Thomas Suter for mentoring me since I started working in this lab back in almost three years ago. Your patience and mentoring have shaped me who I am as a scientist as well as a person right now. I will never forget your unconditional support and would love to pass your legacy to my future students. Because of you two, I grew up not only as a scientist but also as a person. I cannot thank you two more than enough.

I would like to acknowledge Dr. Soohwan Oh for inspiring me with a great master's project. I was very lucky to have a person with a same nationality as me so that I could feel a small part of our lab like home. Also, thank you for answering my questions regarding benchwork especially when I first started working in this lab.

I would like to acknowledge Kenny Ohgi, our lab manager, who always kept our lab clean, organized, and functional. Thank you for being very generous and friendly to me especially when I first started working in this lab.

I would like to acknowledge Rachel Pardee, our lab administrator, who always warmly welcomed me whenever I had to ask for administrative requests. Thank you for inputting all my grades and managing Ph.D. admission related requests in a timely manner.

I would like to thank to all the lab members in the Rosenfeld lab for creating a friendly and supportive environment. I could have never learned this much without you all.

This thesis is coauthored with Kim, Yeeun (Leah); Rosenfeld, Michael Geoffrey; Wang, Susan; Suter, Thomas; Oh, Soohwan. The thesis author was the primary author of this paper.

ABSTRACT OF THE THESIS

Visualizing the Dynamics of 4D Nuclear Architecture during Enhancer-Promoter Transcription Activation Using Live Imaging Techniques

by

Yeeun Leah Kim

Master of Science in Biology

University of California San Diego, 2020

Professor Michael Geoffrey Rosenfeld, Chair
Professor Cornelis Murre, Co-Chair

The transcription machinery requires spatiotemporal coordination of its regulatory elements for precise gene expression. It is believed that the dynamics in the nuclear architecture like enhancer:promoter looping interaction is essential for the regulation of the transcription machinery. However, the detail of the change in 4D nuclear architecture during transcription is poorly understood. Here we seek to visualize a pair of enhancer and its associated target gene based on a real time, live cell imaging that uses

optimized CRISPR-Cas9 and MS2/PP7 system. Our prior work on measuring TFF1 transcription bursting in MCF7 cells confirmed the efficiency and sensitivity of our imaging techniques. In this study, we aim to visualize c-MYC enhancer(c-MYCe) and c-MYC gene in live HeLa cells based on our preliminary data proving that c-MYC enhancer is a functional enhancer. We plan to measure the relative timing of signal, the duration of activity, and the distribution of intensity over the course of activity. The spatial kinetics between c-MYCe and c-MYC over time will be measured, and their motion relative to transcriptional activity (on/off signal) will be examined. Collectively, we hope to advance the understanding of the precise 4D dynamics of the events underlying enhancer-mediated transcription regulation.

INTRODUCTION

For all mammalian cells to undergo the correct differentiation, the precise regulation of transcription machinery initiated by the upstream signals is vital. If such regulation is disrupted, the human body is susceptible to metabolic disorders, diabetes, immune disorders, and cancers. Enhancers are one of the cis-acting DNA elements that mainly regulate transcription. Several models of how enhancers regulate transcription has been proposed. The looping model is one of the prevalent models where the chromatin makes a loop to place enhancer in the proximity of the promoter, facilitating the activation of the enhancer-associated target genes. The tracking model proposes that Polymerase II (Pol II) and the relevant transcription factors will be bound on enhancer while Pol II moves along chromatin (tracking) towards the enhancer-associated target gene [1]. Thus, enhancers are brought near to the promoter of the target genes. Linking/chaining model emphasizes the "chain" that transcription factors form by binding to each other. The "chain" of transcription factors extends from enhancer to promoter, allowing communication between enhancer and its associated target gene [1]. A more contemporary model, like the hub and condensate model addresses how multiple enhancers activate different associated target genes in a limited space. When enhancers are bound to transcription factors and mediators, the movement of its associated target gene is limited. This cluster of transcription factors and mediators allow enhancers and their associated target genes to be in proximity [1]. The hub and condensate model explains the formation of this cluster by liquid-liquid phase separation (LLPS) [1, 2]. LLPS is defined as the demixing of liquid with different

properties. Many transcription factors and mediators have intrinsically disordered regions prone to undergo LLPS during enhancer:promoter communication. Thus, enhancer, promoter, transcription factors, and mediators are thought to be in a liquid droplet or a condensation during active transcription. However, since transcription is episodic and the liquid droplets are not permanent, the synchronization of enhancer and promoter has been studied under the name of "transcriptional bursting."

A number of models were proposed to explain the change in the nuclear architecture for the recruitment and regulation of the transcription machinery. However, there has not been a consensus on which model correctly explains the spatiotemporal rearrangement in the nuclear structure during the activation of the enhancer-associated target genes. The only way to capture these spatiotemporal events in the nucleus is by live imaging. To aid in understanding the dynamics of 4D nuclear architecture, we aim to visualize pairs of enhancers and their associated target genes using real-time, live-cell imaging techniques.

We used a combination of modern technologies for successful visualization in live cells. The development of CRISPR-Cas9 technology enabled the precise, stable integration of the visualization cassette. Multiple sgRNA candidates that are complementary to the target region of interest is designed. The most efficient sgRNA is cloned into the appropriate Cas9-expressing vector. Upon transfection and proper expression, sgRNA and Cas9 will form a sgRNA-Cas9 complex. While sgRNA binds to its complementary region of the genome, Cas9 will create a double-strand break. Then the donor plasmid repairs this double-strand break via homology-directed repair (HDR). For the efficient and precise repair, optimizing donor plasmid is vital. The length of

each homology arm, the sgRNA location within the donor plasmid for linearization upon Cas9 cut, and the size of the stably integrated part must be considered.

As our live cell visualization cassette, we chose to use MS2/PP7 system. MS2 system was firstly developed in Singer lab [3]. The MS2 technology adapted the high binding affinity between the MS2 bacteriophage RNA stem-loops and the MS2 RNA stem-loop binding coat proteins (MCP). To obtain readout of active transcription, MS2 stem-loop sequence needs to be precisely inserted to a known, actively transcribed region by CRISPR-Cas9. MCP fused to a fluorescent protein needs to be inserted into a genome as well for color-labelling active transcription. Upon transcription, the RNA transcripts attached to MS2 RNA stem-loops, which are bound to MCP fused to a fluorescent protein, are produced. By using MS2 technology, Singer lab visualized ASH1 mRNA for the first time and revealed its function in budding yeast [3]. Also, they visualized β-actin gene's mRNA in transgenic mouse and recorded its transcriptional bursting upon serum addition in real time [4].

PP7 technology was adapted from MS2 system, where bacteriophage PP7 RNA stem-loops and its specific PP7 RNA stem-loop binding coat proteins (PCP) are used instead. The color-labelling mechanism of PP7 system is the same as that of MS2 system [5]. Simultaneous two-color imaging is possible if both MS2 and PP7 system are used in a single cell, while fusing different fluorescent protein to MCP and PCP, respectively. For example, the enhancer in interest can be tagged with PP7-PCP-Turquoise and enhancer-associated target gene can be tagged with MS2-MCP-YFP. By using MS2 and PP7 system, Levine lab was first successful showing the transcriptional bursting in *Drosophila* embryo [6]. They found that the strength of the enhancer

correlates with the bursting frequency [6]. Also, they challenged the conventional looping models, where the shared enhancer interacts with one gene at a time by random chance. Instead, they observed co-bursting of two genes, suggesting that the shared enhancer simultaneously interacts with two genes [6]. This confirmed that live cell imaging is essential when delineating the intrinsically dynamic nature of transcription.

To replicate this in mammalian system, Larson lab used MS2 system to visualize TFF1 gene in live MCF7 breast cancer cell in real time [7]. They found that upon estradiol (E_2) treatment, the duration of TFF1 inactive transcription is much longer than TFF1 active transcription. Also, inactive period is highly variable among cell population [7]. They observed that TFF1 alleles are cross-correlated to each other where one allele is prone to burst if the other allele already bursted within past two hours [7].

Prior data published by our lab showed that the most robust enhancers adjacent to E_2-upregulated coding genes cause increased enhancer-promoter looping in mammary cell regulatory transcriptional programs [8]. Nuclear Estrogen receptor α (ERα) dimerizes upon binding of its natural ligand, 17β-estradiol (E_2), and translocates to the nucleus. The ERα-bound enhancers are active, which consequentially activate their associated target genes. This suggests that ERα acts as a potent regulator of gene expression. TFF1 enhancer (TFF1e), being one of the most robustly activated ERα-bound enhancers, and its target gene TFF1 are thus chosen in this study for further investigation.

However, since TFF1 and TFF1e are not resolvable by distance (10kb apart) via current live imaging techniques, we could only record the rate and pattern of TFF1 enhancer and gene activation without seeing the spatiotemporal rearrangement of

enhancer:promoter (Figure 1). PP7-PCP-Turquoise was used to visualize eRNA of TFF1 enhancer. Upon E2 treatment, TFF1e transcription was turned "ON" for 35 minutes while it was turned "OFF" for more than one hour. The transcription "OFF" time was highly variable among the cells compared to the "ON" time. Our result aligned with the similar work done by Larson lab [7].

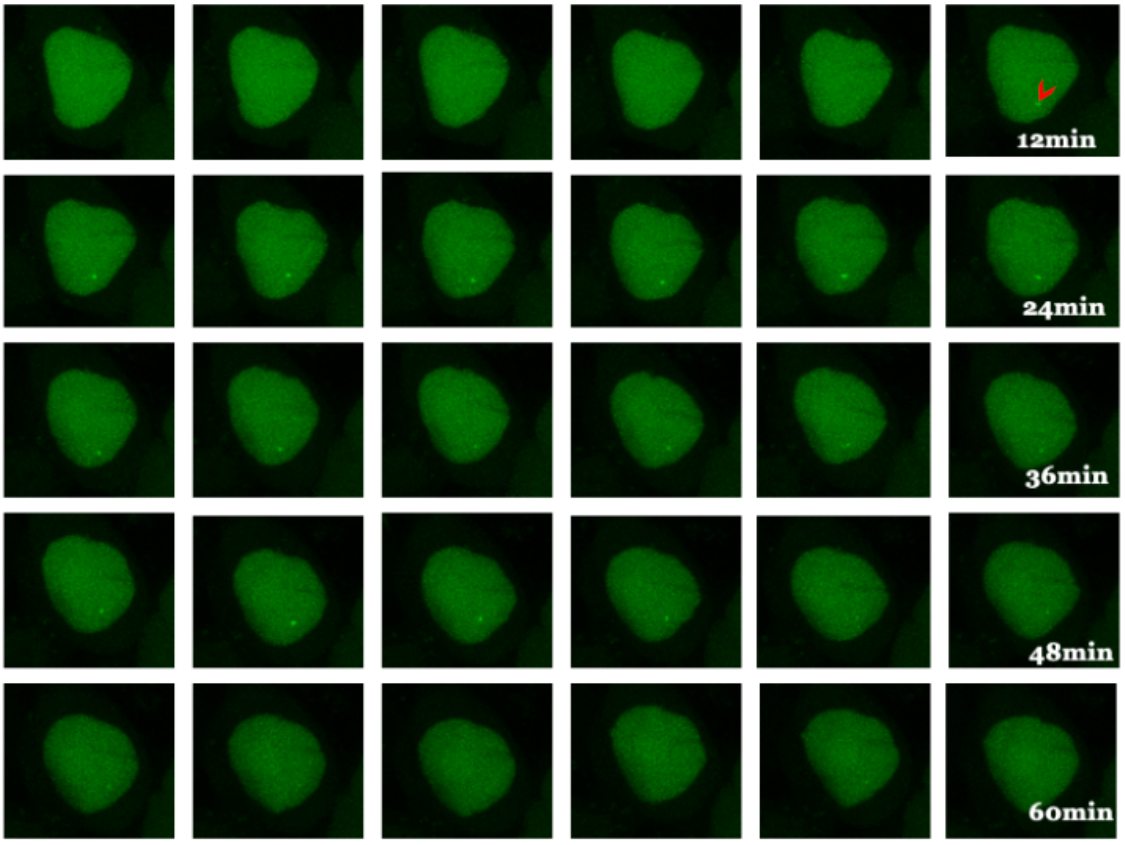

Figure 1: Live cell imaging of TFF1 gene 3' UTR 24xPP7-PCP-Turquoise in MCF7 cell. A 2 hour-movie of one of the positive clones that successfully integrated 24x PP7 repeats and PP7 coat proteins fused with Turquoise. The green foci (red arrow) indicates a sufficient accumulation of fluorescent signal during TFF1 RNA transcript production. The transcription ON period was around 35 minutes while the transcription OFF period was highly variable among the cells, averaging around 1 hour.

Because TFF1 enhancer and TFF1 gene distance is not resolvable via current live imaging techniques, we selected c-MYC enhancer:gene pair to additionally examine the 4D dynamics of the subnuclear events underlying enhancer-mediated transcription regulation. c-MYC enhancer (c-MYCe) is 500kb away from c-MYC gene.

c-MYCe includes the integrated HPV DNA fragment that is suspected to be regulating proto-oncogene c-MYC as its putative enhancer.

Human Papillomavirus (HPV) infection is known to be the leading cause of cervical cancer. Despite the existence of HPV vaccination, cervical cancer remains the fourth most common cancer among women worldwide [9]. The malignant transformation of cervical cells is caused by the viral oncoproteins E6 and E7, mainly produced by high-risk HPV types like HPV-16 and HPV-18. Both E6 and E7 inhibits tumor suppressors such as p53 and pRb, which is essential in controlling cell cycle and repairing DNA [10]. Recently, rather than the direct effects of viral oncoproteins, different viewpoints of the oncogenicity of cervical cells have been reported. The dysregulation of c-MYC is thought to be the main candidate due to its dominant role in cell growth and proliferation [11]. In *Drosophila*, c-MYC's upregulation led to uncontrollable cell growth [12]. Interestingly, HPV-DNA integration site is near a single copy of c-MYC at chromosome 8q24, exclusively in HeLa cervical cancer cells [13]. On top of c-MYC's proximity near the HPV-DNA integration site in HeLa cells, Gimenes and colleagues showed a significant correlation between frequent c-MYC activation and HPV-DNA integration in HeLa cells [14]. Recently, it was found that HPV integration site, c-MYC, and 8q24.22 site spatially coordinate in HeLa cells [15]. Therefore, it is worth investigating the regulatory relationship between integrated HPV-DNA and c-MYC, specifically in HeLa cells.

MYC, the protein product of proto-oncogene c-MYC, is a well-known transcription factor that upregulates and downregulates certain gene sets [16]. While c-MYC produces a transcription factor as its protein product, the c-MYC gene itself could

also be transcriptionally regulated by cis DNA regulatory elements like enhancers. The c-MYC specific enhancers are not universally confirmed, but inactivating enhancer-docking site prevents specific cancer related-super-enhancers in proximity to c-MYC [17]. This shows that a putative enhancer that is site-specific to cervical cancer might exist. Furthermore, the knock-out experiment of the integrated HPV sequence dramatically decreased c-MYC expression, specifically in HeLa cells [15]. Similarly, one RNA-Seq experiment showed that c-MYC is overexpressed only in HPV-18 integrated HeLa cells [13]. The preliminary data of precision nuclear run-on sequencing (PRO-Seq) from our lab also showed that the knock-down of c-MYC eRNA expression by dCas9-KRAB resulted in the decreased c-MYC mRNA expression (Figure 2). Circularized Chromosome Conformation Capture (4-C) data from our lab also revealed that c-MYC and c-MYCe come in direct contact (Figure 3). All these findings support the functional importance of the HPV-associated c-MYC enhancer in HeLa cell.

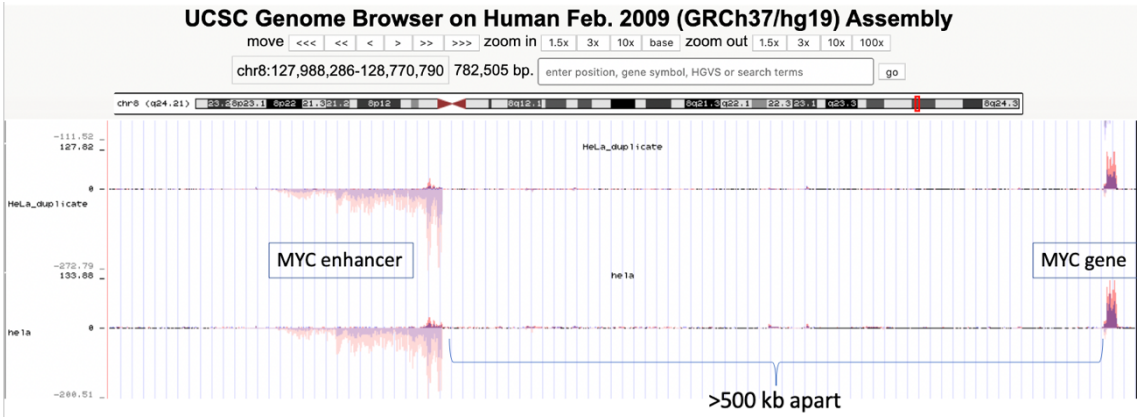

Figure 2: Decreased mRNA expression of c-MYC as eRNA expression was knocked down by dCas9-KRAB. According to our lab's preliminary PRO-Seq data, c-MYC mRNA level decreased as c-MYC eRNA level was knocked down by dCas9-KRAB.

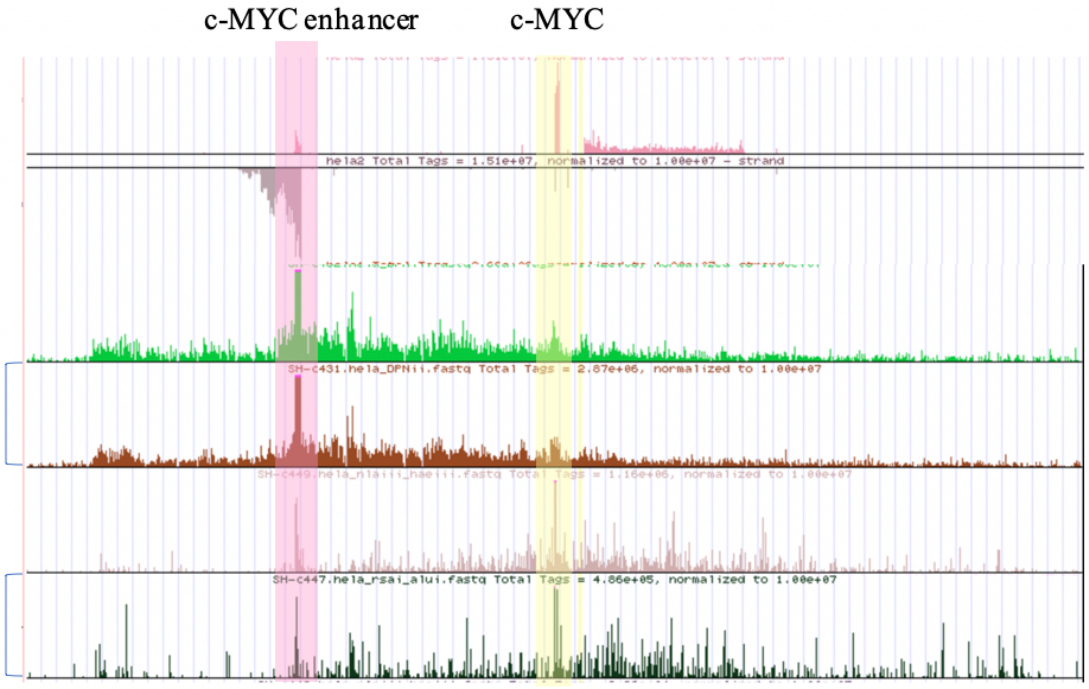

Figure 3: Interaction between c-MYC enhancer and c-MYC via Circularized Chromosome Conformation Capture (4-C). Via 4-C, the interaction between c-MYC enhancer and c-MYC was viewed reciprocally. The first two graphs (green and brown) demonstrates a viewpoint from c-MYC enhancer and the bottom two graphs (pink and black) demonstrates a viewpoint from c-MYC. The high peaks at the location of c-MYC enhancer (vertically highlighted region in pink) and the high peaks at the location of c-MYC (vertically highlighted region in yellow) means high frequency of physical interaction between c-MYC enhancer and c-MYC.

As mentioned above, we decided to additionally investigate c-MYCe:c-MYC interaction on top of TFF1e:TFF1 interaction to observe a different looping dynamics in a different transcription machinery. The long distance between c-MYCe and c-MYC (500kb apart) will allow us to visualize the enhancer:gene communication in full scope. We will record transcriptional bursting as well as the precise kinetics of each element during active transcription in real time. Collectively, we hope to reveal the correct enhancer:promoter communication model used by mammalian cells by using 4D live imaging techniques.

RESULTS

The preliminary data (Figure 1) showed that the transcription of TFF1 was in active state for 35 minutes in average. The transcription repressive state was around more than one hour but highly variable among the cells. Even if did not carry out the whole experiment, I contributed establishing live imaging technology prior to my master's program.

With the optimized imaging strategy that I helped establish, I successfully constructed the All-in-one CRISPR-Cas9 pX459 and pC1-mCherry donor plasmid for visualization of c-MYC enhancer (Figure 4).

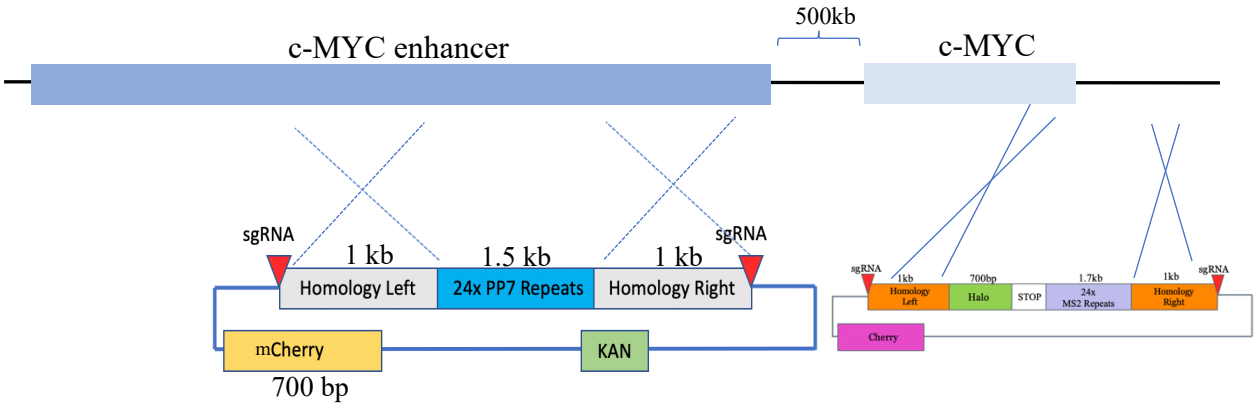

Figure 4: Schematic of 24x PP7/MS2 repeats knock-in donor plasmid for c-MYC enhancer and c-MYC, not in scale: Schematic diagram of the donor plasmid for knock-in of 24x PP7 repeats into c-MYC enhancer and c-MYC, not in scale. The same sgRNA that is cloned into All-in-one CRISPR Cas9 pX459 was placed near both of each homology arm to increase the homology-directed repair (HDR) efficiency by linearizing the donor plasmid. For c-MYC enhancer, 24x PP7 repeats will be integrated where the eRNA signal is present (checked by PRO-Seq, Figure 2), but at the location where the signal is low enough that knock-in will not affect the expression level of c-MYC enhancer itself. For c-MYC, 24x MS2 repeats will be integrated at 3'UTR. Both donor plasmid constructs as well as All-in-one CRISPR Cas9 pX459 will allow RNA imaging of c-MYC and its enhancer. While the construction of the donor plasmid for c-MYC enhancer is complete, note that construction of the donor plasmid for c-MYC is still in progress. Thus, further experiments are only done with c-MYC enhancer.

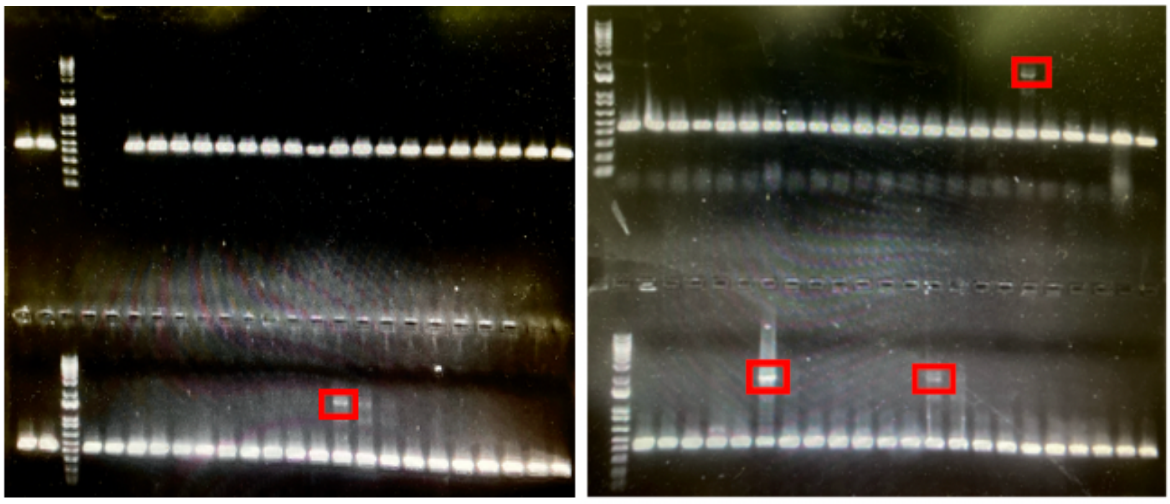

Figure 5: PCR validation of 200 HeLa cells that stably integrated 24x PP7 repeats in c-MYC enhancer region. The PCR primers that flank the expected 24x PP7 repeats knock-in c-MYC enhancer region were designed to validate the stable integration of 24x PP7 repeats. The single clones that have a successful 24x PP7 repeats integration showed around 2kb band. Since there are two copies of MYCe, the other allele that does not have a 24x PP7 integration showed 400bp band in 1% agarose gel. Four red boxes indicate a successful knock-in of 24x PP7 repeats in 4 cells out of 200 cells, meaning the knock-in efficiency of 24x PP7 repeats is 2%. The PCR validation results of the other cells that all had only 400bp band are not shown.

Both pX459 and the optimized donor plasmid were validated by Sanger Sequencing, which showed all the necessary components for knock-in of 24x PP7 repeats. According to the flow cytometry result that sorted for mCherry of donor plasmid, the transfection efficiency of HeLa cells was around 10%. After PCR validation, only 4 out of 200 single clones (2%) successfully integrated 24x PP7 repeats, which showed both 2kb and 400bp band (Figure 5). 196 out of 200 single clones that did not integrate the repeats only showed 400bp bands after PCR validation. Initially, 800 single clones were expected to be sorted, but some wells had more than one cell, so the knock-in efficiency was calculated based on 200 single clones rather than 800 single clones.

After PP7 coat protein-Turquoise transfection, all four single clones survived after G418 Sulfate treatment (Figure 5).

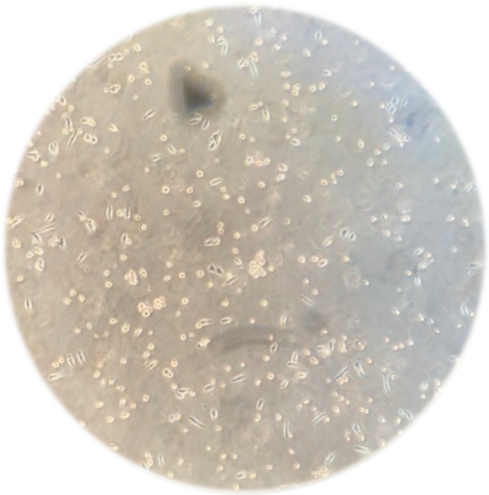

Figure 6: 3 days of G418 Sulfate treatment of PCP-Turquoise transfected HeLa cells. 24x PP7 stably integrated HeLa cells were transfected with 0.6ug of PCP-Turquoise. After 3 days of G418 Sulfate treatment, those cells that only integrated PCP-Turquoise survived.

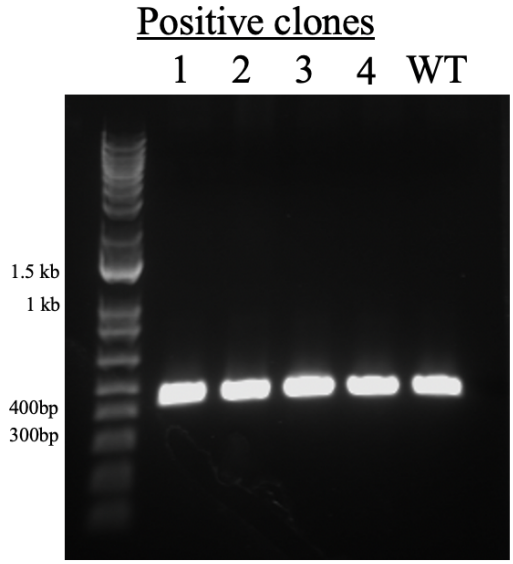

Figure 7 : Re-PCR validation of four single clones that stably integrated 24x PP7 repeats. Four single clones that were initially validated to have 24x PP7 repeats were re-validated with the same PCR validation primers after 2 weeks of passaging. All four single clone's PCR amplicon showed 400bp instead of 2kb band in 1% agarose gel.

However, when PCR validation was repeated after three weeks, all four single clones did not show 2kb bands (Figure 6). Even though the stable integration of PCP-

Turq was successful, we could not proceed to live imaging of c-MYCe since 24x PP7 repeats were lost.

While troubleshooting for knock-in of 24x PP7 repeats to c-MYCe, the construction of pC1-mCherry donor plasmid for knock-in of 24x MS2[2] repeats into c-MYC gene was in progress. The Gibson "insert" including left Gibson overlap sequence from pC1-mCherry plasmid, left homology arm, Halo®, stop codon, multi cloning site (for cloning in 24x MS2 repeats), right homology arm, and right Gibson overlap sequence from pC1-mCherry plasmid was constructed. Also, the Gibson "backbone" which is a digested backbone with EcoRI and BamHI was completed as well. However, the Gibson assembly of the previously mentioned Gibson "insert" and Gibson "backbone" and its troubleshooting are still in progress.

DISCUSSION

The preliminary data such as PRO-Seq (Figure 2) and 4-C (Figure 3) experiments supported that c-MYCe is a functional enhancer regulating c-MYC gene. These data were sufficient for proceeding to live imaging experiments. Although initial PCR validation showed that the stable integration of 24x PP7 repeats at c-MYCe was successful, further PCR validation (three weeks after initial validation) showed that the repeats were lost. One possible reason was that since c-MYC is one of the highly transcribed regions, the repeats could have been lost after three weeks of passaging. Also, considering the transfection efficiency of HeLa cell was only 14%, which is much lower than the expected transfection efficiency, we suspect that this specific strain of HeLa cell is not stable, thus, incapable of holding such a large integration.

The construction of c-MYC donor plasmid has been partially successful. The Gibson assembly of the "insert" that involves the necessary components for knock-in of 24x MS2 repeats has been successful. However, the final Gibson assembly of the "insert" and digested pC1-mCherry backbone has been unsuccessful. It is possible that the Gibson overlap sequence was not long enough for an efficient Gibson assembly, or the digestion of pC1-mCherry backbone could have been incomplete so that there were less digested plasmids for a successful Gibson assembly.

Overall, this study sets the foundation of visualizing the 4D kinetics between HPV-associated c-MYC enhancer, and c-MYC by live imaging techniques. Adapted from the optimized live imaging techniques for TFF1 gene, all the steps just before imaging with microscopy were completed. Also, Halo sequence integration in the donor

plasmid far more improved the cell sorting efficiency during flow cytometry (From TFF1 gene donor plasmid, data not shown in this paper).

Further experiments, including the repeating knock-in of 24x PP7 repeats into c-MYC enhancer, completing the construction of c-MYC donor plasmid with 24x MS2 repeats, and the consequent round of stable integration of PCP-Turquoise and MCP-YFP will be done to complete the stable line generation for visualization under super-resolution microscopy. Although we could not see either the spatiotemporal rearrangement or RNA bursting pattern of c-MYCe and c-MYC in this study, we successfully prepared and optimized the donor plasmid for knock-in of the repeats, which is the most crucial initial step of the visualization in live cells. If the stable lines are successfully generated for both c-MYCe and c-MYC, we plan to measure the relative timing of signal, including the start and stop of activity, the duration of activity, and the distribution of intensity over the course of activity. Additionally, the spatial kinetics between the enhancer and gene over time will be measured and their motion relative to transcriptional activity (on/off signal) will be examined. We hope our work contributed in investigating the kinetics and functional significance of the enhancer:promoter association in 4D genome to better understand diseases induced by the dysregulation of the transcription machinery.

Stable line generation of c-MYCe-24xPP7-PCP

By using CRISPR-Cas9 technology, we tagged c-MYCe with 24x PP7 repeats. sgRNA was chosen based on the previously mentioned PRO-Seq data, where the knock-in of the 24x PP7 repeats would not destroy c-MYC's transcription machinery (Figure 6). Then sgRNA was ligated into All-in-one CRISPR Cas9 pX459 plasmid.

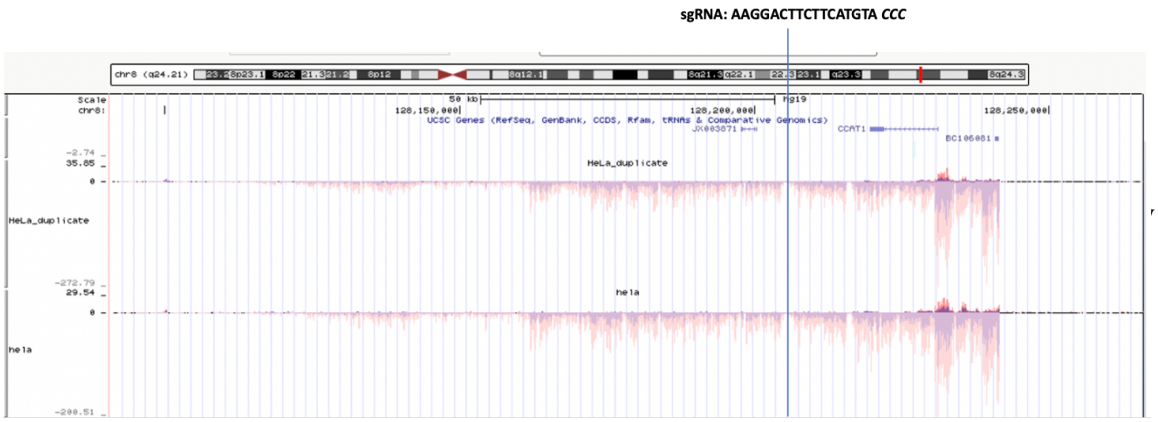

Figure 8: sgRNA selection strategy for c-MYC enhancer tagging. sgRNA was selected where the eRNA signal is low enough so that the eRNA expression of c-MYC enhancer is not altered.

Upon sgRNA selection, the donor plasmid using pC1-mCherry plasmid that facilitates knock-in of the repeats by homology-directed repair was carefully optimized (Figure 4). The same sgRNA that was cloned into All-in-one CRISPR Cas9 px459 was placed near both of each homology arm to increase the homology-directed repair efficiency by linearizing the donor plasmid. The optimized donor plasmid will express the left and right homology arms flanking the 24x PP7 repeats. When transcribed, those repeats form a hairpin structure and recruit PP7 coat proteins (PCP) that are fused to Turquoise.

Table 1: Primer sets for generation of c-MYCe donor plasmid of 24x PP7 repeats knock-in. After careful optimization, sgRNA was selected and cloned into both pX459 and pC1-mCherry donor plasmid. Below are the sequences that form each left and right arm of the donor plasmid of 24x PP7 repeats knock-in.

sgRNA and *PAM*	5' – AAGGACTTCTTCATGTACCC *CGG* – 3'
Left arm F	5' – GAAGGAGGGCTGATCTGAGCATTCAGGTTTCAGAATGC – 3'
Left arm R:	5' – GGTACCTGCGCGGCCGCTGCACTAGTCCCCGGGCCAGA TTCACAGCATCTGGG – 3'
Right arm F:	5' – ACTAGTGCAGCGGCCGCGCAGGTACCTACATGAAGAAG TCCTTAATTGCAGTCATTTACATGGTAGATTCTCTATAA TCATTTAATTTGC – 3'
Right arm R:	5' – GGGCTGGCCAGGTCAGTGCAACTTCAAAGTCG– 3'

Transfection and knock-in validation protocol

1 day before transfection, HeLa cells were plated at 70% confluency in 24-well plate. Then, each 0.3ug of All-in-one CRISPR Cas9 pX459 plasmid and pC1-mCherry donor plasmid were co-transfected using Lipofectamine 2000 (Thermo Fisher). After 6hr of incubation, the cells were washed and cultured in DMEM media with 10% fetal bovine serum (FBS). The cells were moved to 6-well plate when 100% confluent.

Flow cytometry was done after 24-36 hours to allow maximal protein expression. The cells were lifted by Accutase (Innovative Cell Technologies) and isolated into single cell by Falcon Cell Strainers (Corning). Then the cells were spun down at x 300g and resuspended in FBS with 10% PBS. The cells that had mCherry signal was sorted into 96-well plate.

Further validation of the cells that positively integrated 24x PP7 repeats was done by PCR. DNA of the cells were extracted by QuickExtract DNA Extraction Solution (Lucigen). PCR primers that flanked 24x PP7 repeats and expected to give 2kb amplicon were designed. The wild type HeLa cells and those cells that did not integrate the repeats were expected to give 400bp amplicon. The validated cells were further passaged in 6-well plate.

0.6ug of PP7 coat protein fused to Turquoise (PCP-Turq) was transfected to the PCR-validated cells using the same transfection protocol explained above. After allowing maximal protein expression, the cells were treated with G418 Sulfate (Gemini) for selection for three days. The cells that positively integrated both the 24x PP7 repeats and PCP would be visible under green channel of Keyence microscope.

This thesis is coauthored with Kim, Yeeun (Leah); Rosenfeld, Michael Geoffrey; Wang, Susan; Suter, Thomas; Oh, Soohwan. The thesis author was the primary author of this paper.

REFERENCES

[1] Furlong, Eileen & Levine, Michael. (2018). Developmental enhancers and chromosome topology. Science (New York, N.Y.). 361. 1341-1345.

[2] Nair, S. J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M. J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., Ma, Q., Zhang, J., Hu, Y., Tan, Y., Ohgi, K. A., Jayani, R. S., Banerjee, P. R., Aggarwal, A. K., & Rosenfeld, M. G. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. Nature structural & molecular biology, 26(3), 193–203.

[3] Bertrand, Edouard & Chartrand, Pascal & Schaefer, Matthias & Shenoy, Shailesh & Singer, Robert & Long, Roy. (1998). Localization of ASH1 mRNA particles in living yeast. Mol Cell. 2. 437-45.

[4] Wu, B., Chao, J. A., & Singer, R. H. (2012). Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. Biophysical journal, 102(12), 2936–2944.

[5] Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A., & Singer, R. H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science (New York, N.Y.), 332(6028), 475–478.

[6] Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell, 166(2), 358–368.

[7] Rodriguez, Joseph & Day, Christopher & Chow, Carson & Larson, Daniel. (2018). Intrinsic Dynamics of an Endogenous Human Gene Reveal the Basis of Expression Heterogeneity. SSRN Electronic Journal.

[8] Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H. S., Glass, C. K., & Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature, 498(7455), 516–520.

[9] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018), Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68: 394-424.

[10] Kraus, I., Molden, T., Holm, R., Lie, A. K., Karlsen, F., Kristensen, G. B., & Skomedal, H. (2006). Presence of E6 and E7 mRNA from human papillomavirus types 16, 18, 31, 33, and 45 in the majority of cervical carcinomas. Journal of clinical microbiology, 44(4), 1310–1317.

---

[11] Eilers, M., & Eisenman, R. N. (2008). Myc's broad reach. Genes & development, 22(20), 2755–2766.

[12] Orian, A., Grewal, S. S., Knoepfler, P. S., Edgar, B. A., Parkhurst, S. M., & Eisenman, R. N. (2005). Genomic binding and transcriptional regulation by the Drosophila Myc and Mnt transcription factors. Cold Spring Harbor symposia on quantitative biology, 70, 299–307.

[13] Adey, A., Burton, J. N., Kitzman, J. O., Hiatt, J. B., Lewis, A. P., Martin, B. K., Qiu, R., Lee, C., & Shendure, J. (2013). The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. Nature, 500(7461), 207–211.

[14] Gimenes, F., Souza, R. P., de Abreu, A. L., Pereira, M. W., Consolaro, M. E., & da Silva, V. R. (2016). Simultaneous detection of human papillomavirus integration and c-MYC gene amplification in cervical lesions: an emerging marker for the risk to progression. Archives of gynecology and obstetrics, 293(4), 857–863.

[15] Shen, C., Liu, Y., Shi, S., Zhang, R., Zhang, T., Xu, Q., Zhu, P., Chen, X., & Lu, F. (2017). Long-distance interaction of the integrated HPV fragment with MYC gene and 8q24.22 region upregulating the allele-specific MYC expression in HeLa cells. International journal of cancer, 141(3), 540–548.

[16] Sabò, A., Kress, T. R., Pelizzola, M., de Pretis, S., Gorski, M. M., Tesi, A., Morelli, M. J., Bora, P., Doni, M., Verrecchia, A., Tonelli, C., Fagà, G., Bianchi, V., Ronchi, A., Low, D., Müller, H., Guccione, E., Campaner, S., & Amati, B. (2014). Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. Nature, 511(7510), 488–492. https://doi.org/10.1038/nature13537

[17] Schuijers, J., Manteiga, J. C., Weintraub, A. S., Day, D. S., Zamudio, A. V., Hnisz, D., Lee, T. I., & Young, R. A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. Cell reports, 23(2), 349–360.