DEVELOPMENT AND EXPLORATION OF TECHNIQUES TO STUDY THE SPATIAL ORGANIZATION OF THE HUMAN GENOME

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Max Garrett Kushner

August 2021

DEVELOPMENT AND EXPLORATION OF TECHNIQUES TO STUDY THE SPATIAL ORGANIZATION OF THE HUMAN GENOME

Max Garrett Kushner, Ph D.

Cornell University 2021

Gene expression is regulated by a number of different mechanisms, but of particular interest recently is the regulation of gene expression through spatial organization of the genome. Various techniques have been developed and continue to be developed to further characterize and study the dependence of transcriptional regulation on genome organization. In this dissertation, I will discuss a series of projects aimed at designing a novel method to examine the relationship between transcriptional activity and genomic contacts.

I will discuss efforts to characterize photoactivatable compounds including members of the psoralen family for use as a tool to examine genomic contacts. I will explain the determination of many different photochemical properties of psoralen compounds. I will also mention optimization and the use of psoralen compounds with two-photon excitation.

Here I present a technique we call Femto-Seq. This technique utilizes a photoactivatable DNA crosslinking compound with an affinity tag to take a snapshot of the DNA sequences spatially around a genomic locus of interest. After allowing the photoactivatable compound to intercalate, two-photon excitation is used to covalently bind sequences in a nuclear volume of interest (around a fluorescently labeled gene for example). The affinity tag on the crosslinker is used to enrich for sequences from the

irradiated volume which are then analyzed through sequencing. I will present pilot experiments designed to examine the efficacy of such a method by looking at the enrichment of a targeted transgene. We report a 15-fold enrichment of the transgene matching expected enrichment based on the nuclear volume of interest. I will describe potential applications and extensions of such a technique.

Many parts of the Femto-Seq method can be further optimized. Since the technique involves fluorescently labeling a genomic locus of interest, fluorescently labeled engineerable DNA binding proteins are particularly valuable. While many exist, we were particularly interested in Transcription Activator Like Effectors (TALEs). In order to understand the mechanisms of TALE binding we utilized a variety of techniques including DNA curtains, single molecule co-localization and Fluorescent Correlation Spectroscopy. I will also discuss the potential use of microfluidic platforms to increase throughput of the pulldown and cleanup processes.

Biographical Sketch

Max Garrett Kushner was born and grew up in Reisterstown, Maryland where his father works as a physician and his mother as a nurse. Max graduated from George Washington Carver Center for Arts and Technology where he participated in the Computer Science magnet program. Upon graduation, Max studied Physics at the University of Maryland, College Park. While there he participated in biophysics related research. Under the mentorship of Dr. Arthur La Porta he utilized an optical trap to examine the folding dynamics of G-Quadruplex structures under molecularly crowded conditions. After graduating with departmental honors, Max sought to further his knowledge by pursuing a graduate degree in biophysics at Cornell University. At Cornell, Max worked under the mentorship of Warren Zipfel. He was able to pursue several projects that intersected molecular biology, optics, nanofabrication, chemistry and physics. Max worked on a project funded by the NIH 4D Nucleome Common Fund. He worked on the development of a novel technique to study genomic contacts using a photoactivatable crosslinking compound. Max received a NYSTEM Grant training fellowship in 2018. Upon receiving his Ph.D., Max will look for private industry jobs which will allow him to apply skills learned in his graduate program.

Acknowledgements

I'd first like to thank Warren Zipfel. His mentorship and assistance throughout my stay at Cornell cannot be matched. Many members of the Zipfel Lab have helped me over the years. Avtar Singh, Alex Van Slyke, and Juan Wang provided friendship and assistance as I was just joining the lab. All were quick to help with any issues I had when I was beginning here. Alex deserves extra acknowledgement for providing guidance with the early single molecule TALE experiments. I want to give special thanks to Juan Wang for her help on the Femto-Seq project. Her help with the early experiments and the cell line creation was critical to get the project working. New members of the lab that joined while I was here also deserve thanks. Thomas Ciavatti provided feedback and offered discussion on my work. Jack Crowley helped with extensions of FemtoSeq including applying it to the nucleoplasmic reticulum. Of particular importance, Thomas Roberts and I worked closely on the Femto-Seq project. His help in particular helped move that project to success.

I'd like to thank my committee as well, John Lis and Michelle Wang both provided input and resources needed to complete numerous aspects of my project. Members of the Lis Lab deserve special credit as well. Specifically, Abdullah Ozer and Judhajeet Ray provided extremely helpful advice and support on the Femto-Seq project. Without their molecular biology expertise, the project would have found little success.

I'm grateful to all the core facilities at Cornell University. The imaging core led by Rebecca Williams provided all equipment necessary to complete most of these projects.

The NanoScale Science and Technology Facility was instrumental in the creation of the microfluidic projects mentioned here.

I of course must thank my funding sources. Both the NIH and the Empire State Stem Cell fund provided the financial support I needed to work on these projects. Without them none of this would be possible.

Finally, I would like to thank my family. I would not have been able to make it through this program without their endless support.

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Overview

The human genome contains all the information needed to control all cellular functions in our bodies. This information can be found in both the coding and non-coding regions of our genome. On the surface, the coding sequences do the majority of the work. They encode proteins and RNA which carry out most biological processes. Despite their importance, coding sequences only make up approximately 2% of all genomic sequences (Figure 1.1). The remaining 98% of genomic sequences are non-coding but still vital [1]. Non-coding sequences include telomeres [2], centromeres [3], and origins of replication [4]. The former is important for DNA repair while the latter are important for DNA replication processes. Non-coding sequences also include control elements such as promoters and enhancers [5]. While these sequences do not actually code anything, they are still vitally important for controlling gene expression and thus controlling downstream biological processes.

Regulation of gene expression is important for a number of different reasons. Cell differentiation processes require specific genes to be activated or repressed [6]. Additionally, gene expression needs to change in order to respond to environmental stimuli [7]. Control of gene expression comes in many different forms. Gene expression can be controlled at the transcription [8], mRNA processing [9] and translation [10] levels. Recently, transcriptional control caused by spatial organization within the

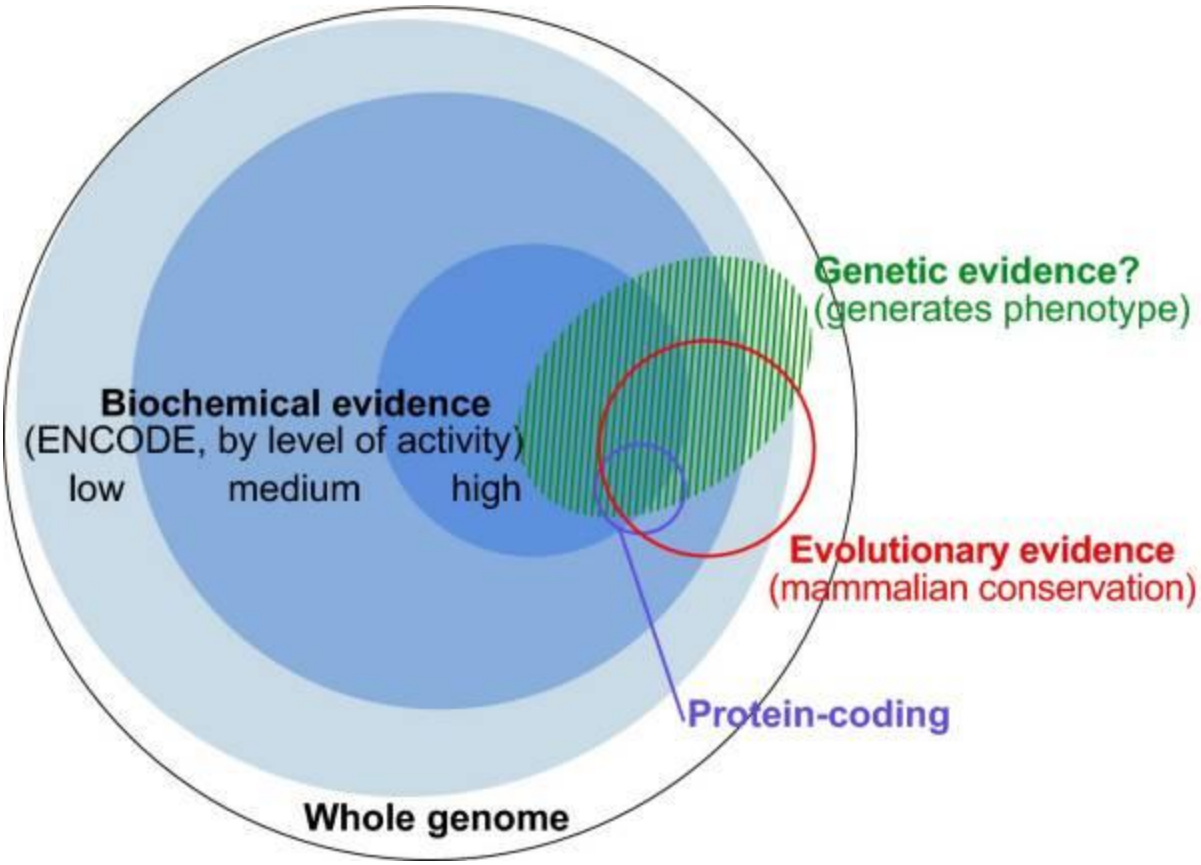


Figure 1.1. Overview of functional DNA elements in the human genome.

A very small fraction of the human genome (encompassing circle) corresponds to protein coding regions (small purple circle). Despite the limited amount of protein coding regions, there is evidence many non-coding regions have important regulatory roles. The red circle represents the fraction of regions believe to serve a functional role based on evolutionary conservation between mammals. The green shaded area conceptually represents non-coding regions which produce phenotypical changes upon modification (as true estimates are limited). The blue circles represent DNA regions with biochemical activity as detected in the ENCODE project (such as histone modification patterns, DNA methylation and transcription factor binding). Despite protein coding regions making up a small portion of the genome, there is much evidence that other non-coding regions serve important regulatory functions. Adapted from [11].

genome has become particularly interesting. Understanding how changes in the physical location of sequences, proteins and nuclear bodies affect gene expression is important to fully understand developmental processes.

1.2 Methods to examine Spatial Genome Organization:

A thorough discussion transcription, regulatory elements and methods to determine and examine regulatory sequences can be found in Appendix A. There are a variety of methods that have been utilized to examine spatial organization of the genome. Still today, much work goes into developing new and novel methods to examine such interactions.

1.2.1 Chromosome Conformation Capture Techniques:

Chromosome Conformation Capture techniques have been used to look at chromosome organization for close to two decades [12] (Figure 1.2). It started with the development of 3C technology, but this sort of technology has been adapted and utilized to create a slew of methods to examine chromosome conformation. The initial steps in all of these techniques are as follows. First, chromatin is fixed, usually using formaldehyde. The fixed chromatin is then cut with a restriction enzyme. The sticky ends of the chromatin fragments are then ligated together in a dilute environment. This ensures that only intramolecular crosslinks between crosslinked sequences are ligated together. Once the crosslinks are reversed all that remains is several linear DNA strands composed of segments that were spatially close together in the nucleus even if they weren't close together in the genome [12]. Once a set of linear strands are generated, they can be analyzed through a variety of methods.

3C:

The original 3C utilized semi-quantitative qPCR to analyze whether two genomic contacts were crosslinked and thus spatially close to each other [12]. This technique was originally used to show that yeast chromosome III forms a ring structure. It was later used to show that the upstream LCR looped back to interact with the β-globin locus forming an ~50kb chromatin loop using qPCR [13].

4C:

4C creates small DNA circles via another ligation. Primers designed to examine a locus of interest are used to perform an inverse PCR [14]. The resulting strands are then analyzed using a microarray [15]. This provides a snapshot of all the sequences around a specific locus of interest determined by the primers utilized in the inverse PCR reaction. 4C experiments were first used to examine the DNA interactions of tissue specific genes that were embedded in an inactive region [15]. It was also used to examine the contacts of a housekeeping gene in a very active region of the genome [15].

5C:

5C allows the examination of interactions between multiple distinct sequences compared to 3C which looks at the interaction of two specific loci and 4C which looks at all the interactions by a single locus [16]. Following the reversal of the crosslinking reaction, the DNA template is hybridized to a set of oligos designed to bind to restriction enzyme binding sites with universal primers on the 5’ end. The DNA is then amplified using these universal primers creating a set of genomic contact templates which can be

analyzed using a microarray or sequencing [16]. 5C has been used to look at the β-globin gene locus where it shows similar results to previous 3C studies [16].

Hi-C:

Hi-C was the first all versus all method brought about by the emergence of next generation sequencing [17]. In Hi-C protocols, prior to the ligation step, the DNA fragments have their ends filled in with biotinylated nucleotides. After the ligation, a biotin pulldown is performed to isolate only the ligated sequences. After sequencing, one is left with a set of reads only corresponding to ligated sequences. When a set of sequences come from different restriction fragments they are treated as an interaction between those sets of sequences. Resulting from all the reads, a matrix of ligation frequencies can be created [17]. The original resolution of Hi-C was ~1Mb based on approximately 10 million paired end reads [17]. Since Hi-C measures all versus all, resolution increases require a quadratically scaling increase in the amount of sequencing. Hi-C confirmed the presence of spatial separation of active and inactive regions [17]. While these were seen in previous studies, Hi-C allowed researchers to see this behavior on a genome wide scale.

Drawbacks and Potential Issues:

The theoretical limit to the resolution of these techniques is dependent on the DNA cutter being used [18]. Cutters that fragment DNA into smaller pieces have a lower theoretical limit. The presence of repetitive DNA sequences can make the analysis of certain regions more difficult, but this can be mitigated with increased sequencing depth [18]. All of these techniques are dependent on the capture probability between genomic

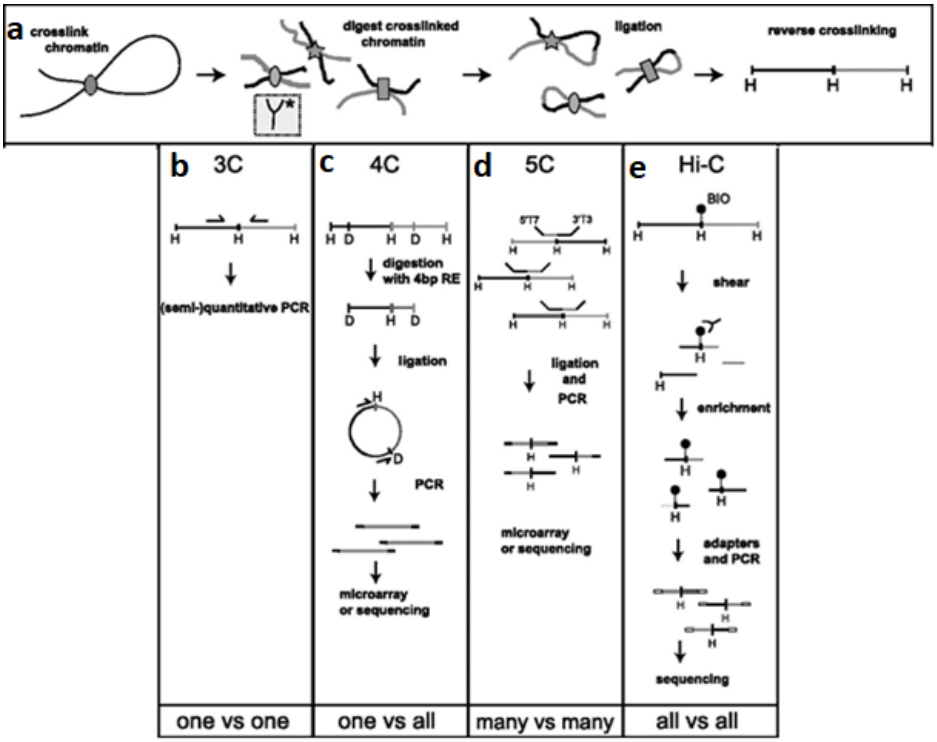


Figure 1.2. Chromatin Conformation Capture Technologies.

(a) Most traditional chromatin capture technologies follow this experimental pathway. Chromatin is crosslinked usually using formaldehyde. Chromatin is then digested and ligated in a dilute solution to ensure ligation only occurs on individual chromatin chunks and not between chromatin chunks. Crosslinks are then reversed and the resulting fragments are analyzed. (b) 3C utilizes qPCR to examine contact frequency between two individual loci. A primer from each locus is chosen to allow quantification of fragments that contain both loci. (c) 4C utilizes inverse PCR and microarray or sequencing analysis to look at all the contacts of a single locus of interest. (d) 5C is an extension of 3C technology to look at the interactions of several different loci with each other. (e) Hi-C is a genome wide application of 3C technology. The ligation step includes inserting a biotinylated nucleotide which can be enriched. This is followed by sequencing to generate a map of all genomic contacts in the genome. Adapted from [49].

loci. As loci get farther apart on the genome their capture probability sharply decreases. In order to detect long range interactions, a large number of reads are required. Thus, long range interactions are often described statistically as loci which are in contact with each other with a higher probability than other regions the same linear distance from each other [18]. It has been shown that spatial organization of the genome is not static in time and can vary between cells in a given population [19]. Finally, Chromosome Conformation Capture techniques are traditionally limited to studying pairwise interactions [17]. Interactions from multiple loci that form a cluster would not be easily extractable from the data.

1.2.2 Microscopy based Techniques:

While the previous mentioned methods have been biochemical-based a whole suite of microscopy-based techniques have also been developed to look at spatial organization of the genome. These techniques all revolve around labeling genomic loci utilizing a variety of methods some of which are described below.

Fluorescence in Situ Hybridization:

Fluorescence in Situ Hybridization (FISH) utilizes fluorescently labeled oligonucleotides to visualize genomic loci (Figure 1.3). In this technique, cells are fixed and their chromatin is denatured. A fluorescent probe that is designed to bind a specific sequence is then allowed to hybridize to the denatured chromatin. The probe is then visualized using microscopy [20]. This technique has been used to look at very repetitive DNA such as centromeres [21] and telomeres [22]. It has also been used to look at the location of multicopy genes such as ribosomal DNA [23]. Chromosome

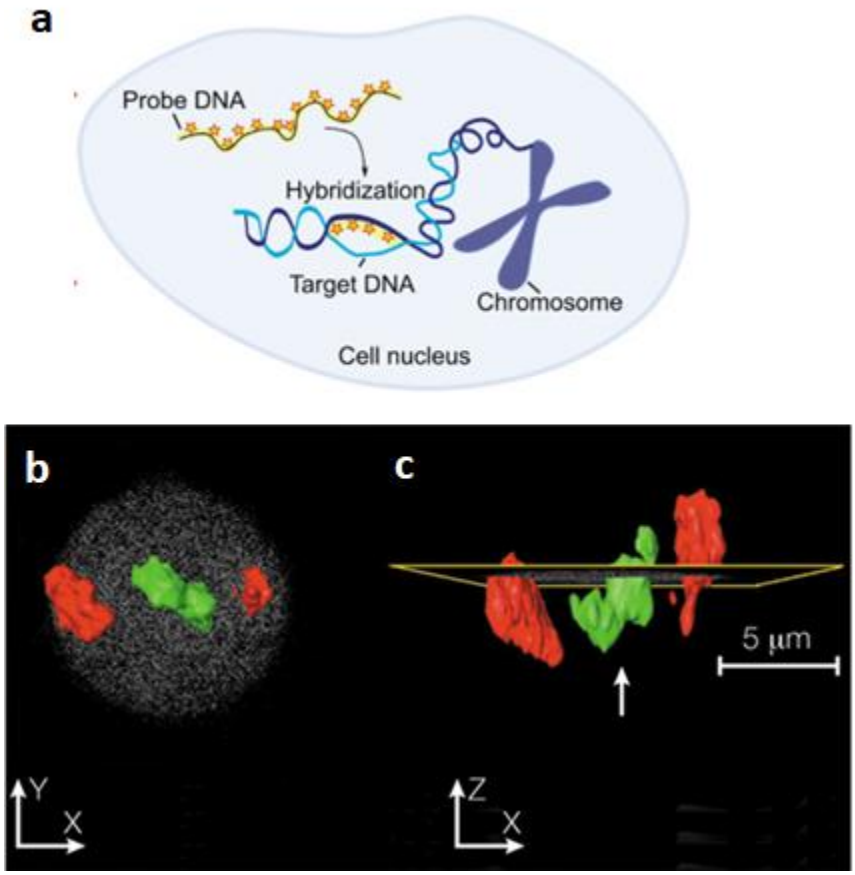


Figure 1.3. Fluorescence in Situ Hybridization.

(a) Fluorescence in Situ Hybridization protocol. Nuclei are permeabilized to allow fluorescently labeled probe DNA sequences into the nucleus. Chromatin is denatured and the probe DNA is allowed to bind its complimentary sequence. Nuclei are then fixed and imaged. (b) FISH reconstruction of chromosome 18 (red) and 19 (green). Chromosome 19 localizes towards the center of the nucleus while 18 localizes towards to the periphery. (c) 3D reconstruction. Adapted from [50] and [24].

paint studies have been used to show that chromosome occupy distinct chromosome territories within the nucleus [24]. While the data from FISH experiments has been extremely useful, there are some shortcomings of such protocols. The efficiency of the binding of FISH probes is dependent on a variety of factors including the probe itself, the denaturation and hybridization method. Another drawback of FISH is that the fixation protocol can result in the loss of very fine DNA spatial organizational structure. Repeated examination of genomic loci has been used to illustrate the transcriptional activity of chromosomes [25].

1.3 Global Nuclear Architecture:

The previously described techniques have been used to identify and study many different scopes of spatial genome organization. These layers of genome organization include compartments, topologically associated domains, chromosome territories, and Lamina Associated Domains.

Compartments:

At the largest level nuclear organization can be broken down into two compartments. These compartments, A/B, were first examined by early Hi-C experiments [17]. Stronger interaction has been shown to occur within a compartment rather than between compartments [17]. A/B compartments have many other additional features which distinguish them. The A compartment is composed of mainly open active chromatin. Compartment A correlates with known genes, transcriptional activity and DNase I hypersensitivity [17]. Compartment A also correlates with histone marks demarking open chromatin. The B compartment is the opposite, it is composed of

mainly closed and silent chromatin [17]. Pairs of loci in Compartment B tended to have higher interaction frequency further suggesting that Compartment B consists of heterochromatin [17]. The A compartment tends to localize to the center of the nucleus in eukaryotes and the B compartment tends to localize to the periphery. Compartments are cell type specific, regions that are in the A compartment in one cell may be found in the B compartment in a different cell [26]. Despite the cell type specificity, the open and closed chromatin distinction is maintained.

Topologically Associated Domains:

Topologically Associated Domains (TADs) are megabase size domains that show increased interaction within the domain over sequences outside the domain [27]. It has been shown that TADs function as spatial insulators, allowing enhancer promoter interaction within the TAD, but preventing crosstalk between promoters and enhancers within different TADs [28]. Alterations in the location of TAD boundaries have been shown to lead to changes in gene expression profiles leading to various pathologies [29]. Boundaries of TADs have increased levels of housekeeping genes [30]. In mammals, most TAD boundaries are marked by the presence of CCCTC-binding factor (CTCF) sites and structural maintenance of chromosomes (SMC) cohesion complexes [30]. Increased contact frequency between the edges of TADs (seen as corners on a Hi-C map) may indicate the presence of CTCF loops [31]. CTCF shows association with insulators as well, further suggesting its importance in defining TAD boundaries. Depletion of CTCF has been shown to cause increased interaction between TADs. Alterations or removal of CTCF sites has been shown to change the position of TAD boundaries [32]. This behavior is not universal though. In Drosophila, depletion of CTCF

is not shown to broadly effect genome architecture, despite binding to the same sequence [33]. It has been suspected that TAD organization in Drosophila is affected by other insulator proteins such as BEAF-32 [34]. Other species such as plants do not have CTCF, but still show TAD-like behavior [35].

Chromosome Territories:

Chromosomes themselves have been shown to occupy distinct territories of the nucleus [36]. Positioning of chromosome territories is a tissue specific phenomenon and conserved between different mammalian species [37]. Changes in the position of chromosome territories can occur over minutes and in response to external stimuli [38]. The volume of chromosome territories depends heavily on the density of active genes rather than the size of the chromosome itself [39]. Chromosome territories are not strict as much intermingling between chromosomes is noticed [39]. It has been shown that this intermingling occurs as transient movement of chromosomes increases on both a genome wide and gene specific scale [39]. Despite the noticeable presence of intermingling, regions that intermingle have been shown to not be enriched for euchromatin or actively transcribing genes [39].

Lamina Associating Domains:

The nucleus is surrounded by a membrane bilayer spotted with nuclear pores, complexes which allow transportation of molecules across the nuclear membrane [40]. The inner membrane is coated with a large mesh-like structure called the nuclear lamina [41]. The nuclear lamina is composed of a variety of Lamin proteins and other membrane associated proteins [42]. Depletion of Lamin proteins has been shown to be

associated with nucleus deformation suggesting that Lamin proteins play a vital role in maintaining nuclear structure [41]. Lamin proteins are absent from plants and fungi [42]. In many cell types it has been shown that heterochromatin localizes to the periphery of the nucleus while euchromatin localizes to the center of the nucleus [43, 44]. Examining interactions between Lamin-B and chromatin has led to the discovery of large genomic regions termed Lamin Associated Domains (LADs) [45]. LADs can span up to 10kb and show enrichment for repressive histone marks [45]. Some LADs termed facultative LADs show differentiation with cell type, while other LADs termed constitutive LADs show remarkable conservation between species and increased A-T nucleotide density [46]. LADs generally show low gene density and the genes contained in LADs show little transcriptional activity [45]. However, gene repression at the nuclear periphery is not a universal phenomenon. Several studies have indicated subregions of increased gene expression at the nuclear periphery specifically localized at nuclear pore complexes [47, 48].

1.4 Development of a Novel Method to Examine Genomic Contacts:

With these previously discussed techniques and nuclear architecture in mind, we sought to develop a novel method to examine genomic contacts. Our method, which we call Femto-Seq, is a novel method to look at genomic sequences in a specific volume of the nucleus. Similar to 4C, this method can look at genomic contacts around a specific genomic locus; however, it is not limited to that. Femto-Seq is able to look at any nuclear volume that can be fluorescently labeled. Femto-Seq does not require formaldehyde crosslinking of chromatin fragments. Instead, it utilizes photoactivatable compounds which can provide spatial resolution. Femto-seq is performed using a two-

photon microscope which allows examination of specific sub-populations that may be missed in ensemble-based measurements.

The experimental process of Femto-Seq is as follows (Figure 1.4a). Prior to the experiment the nuclear volume of interest must be fluorescently labeled (Figure 1.4b). Many methods can accomplish this some of which will be discussed in Chapter 3. After labeling a nuclear volume of interest, nuclei are incubated with a psoralen based crosslinker with an affinity tag. After allowing this compound to intercalate, nuclei are imaged on a two-photon microscope. The photoactivatable crosslinker is selectively activated in the region of interest with two-photon excitation using the fluorescent label as a target. This process covalently bonds the photoactivatable crosslinker with an affinity tag to only the DNA located in the irradiated volume. Chromatin is then extracted and sheared. Chromatin from the nuclear volume of interest is enriched by performing a bead pulldown using the affinity tag attached to the crosslinker. The crosslinking compound is removed from the chromatin and the chromatin is then sequenced effectively providing a snapshot of the sequences located in the nuclear volume of interest..

1.5 Structure of Thesis:

In this chapter, I have briefly summarized the importance of studying chromatin organization as it relates to gene expression regulation. I have discussed numerous previously developed techniques designed to examine transcriptional regulation. I have also introduced Femto-Seq as a novel method we have developed to examine genomic contacts.

In Chapter 2, I will discuss how photoactivatable compounds have been used to provide spatial resolution in a variety of contexts. I will discuss exploration of photoactivatable compounds for use in Femto-Seq experiments. I will then discuss determination of various photochemical properties of psoralen-based compounds for the purpose of being used in a Femto-Seq experiment.

In Chapter 3, I will discuss summarize developed techniques to study chromatin organization including novel methods and improvements of older techniques. I will discuss the results of a pilot Femto-Seq experiment and suggest future improvements and applications of such a method.

In Chapter 4, I will mention smaller projects I have worked on during my study. These projects while not directly related to Femto-Seq could provide the basis for further improvements of the method.

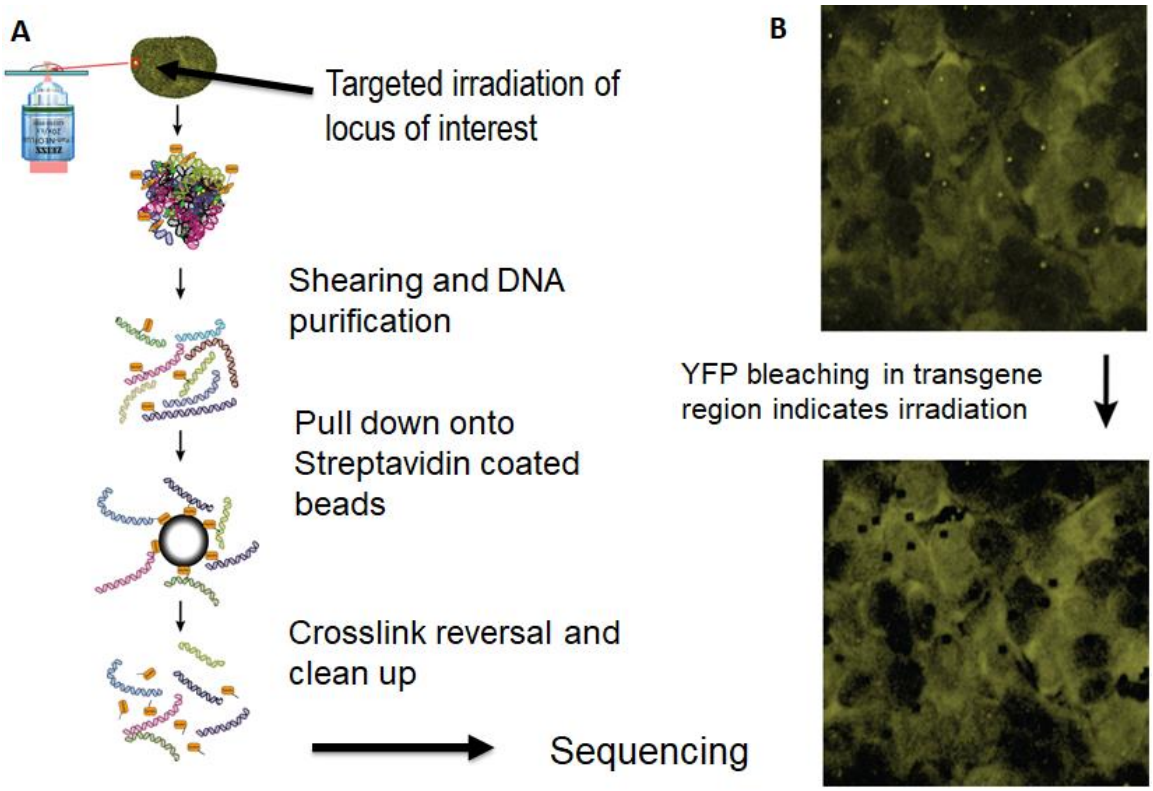


Figure 1.4: Overview of Femto-Seq.

(a) Overview of the Femto-Seq technique. A genomic volume of interest is fluorescently labeled. A photoactivatable crosslinker with a biotin tag is then added. 3D localized two-photon excitation is used to excite the crosslinker around the volume of interest. DNA is purified and the crosslinked DNA is enriched via streptavidin bead pulldown. After crosslink reversal and clean up, sequencing is performed. Since we enrich for the chromatin from the volume of interest, a snapshot of sequences from the volume of interest is obtained. (b) Example images before and after targeted irradiation. A transgene is marked by punctate YFP spots and after targeted two-photon excitation clear bleaching patterns at the loci are seen.

1.6 References:

[1] Lander ES, et al. Initial sequencing and analysis of the human genome. Nature 409: (2001);860-921.

[2] Greider CW. Telomeres. Curr Opin Cell Biol 3: (1991);444-51.

[3] Vig BK. Sequence of centromere separation: role of centromeric heterochromatin. Genetics 102: (1982);795-806.

[4] Leonard AC, Méchali M. DNA replication origins. Cold Spring Harb Perspect Biol 5: (2013);a010116.

[5] Levine M. Transcriptional enhancers in animal development and evolution. Curr Biol 20: (2010);R754-63.

[6] Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell 144: (2011);327-39.

[7] Bartlett JB, Dredge K, Dalgleish AG. The evolution of thalidomide and its IMiD derivatives as anticancer agents. Nat Rev Cancer 4: (2004);314-22.

[8] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: (1961);318-56.

[9] Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15: (2014);163-75.

[10] Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Hartsch T, Wimberly BT, Ramakrishnan V. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. Science 291: (2001);498-501.

[11] Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RC. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A 111: (2014);6131-8.

[12] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 295: (2002);1306-11.

[13] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10: (2002);1453-65.

[14] Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38: (2006);1341-7.

[15] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38: (2006);1348-54.

[16] Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16: (2006);1299-309.

[17] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: (2009);289-93.

[18] Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 43: (2011);1059-65.

[19] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502: (2013);59-64.

[20] Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A 79: (1982);4381-5.

[21] Amakawa G, Ikemoto K, Ito H, Furuya T, Sasaki K. Quantitative analysis of centromeric FISH spots during the cell cycle by image cytometry. J Histochem Cytochem 61: (2013);699-705.

[22] Meeker AK, Gage WR, Hicks JL, Simon I, Coffman JR, Platz EA, March GE, De Marzo AM. Telomere length assessment in human archival tissues: combined telomere fluorescence in situ hybridization and immunostaining. Am J Pathol 160: (2002);1259-68.

[23] Vitturi R, Libertini A, Armetta F, Sparacino L, Colomba MS. Chromosome analysis and FISH mapping of ribosomal DNA (rDNA), telomeric (TTAGGG)n and (GATA)n

repeats in the leech Haemopis sanguisuga (L.) (Annelida: Hirudinea). Genetica 115: (2002);189-94.

[24] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2: (2001); 292–301.

[25] Su JH, Zheng P, Kinrot SS, Bintu B, Zhuang X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. Cell 182: (2020);1641-1659.e26.

[26] Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol 16: (2015);180.

[27] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: (2012);376-80.

[28] Kim YJ, Cecchini KR, Kim TH. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. Proc Natl Acad Sci U S A 108: (2011);7391-6.

[29] van Bemmel JG, Galupa R, Gard C, Servant N, Picard C, Davies J, Szempruch AJ, Zhan Y, Żylicz JJ, Nora EP, Lameiras S, de Wit E, Gentien D, Baulande S, Giorgetti L, Guttman M, Hughes JR, Higgs DR, Gribnau J, Heard E. The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. Nat Genet 51: (2019);1024-34.

[30] Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A, Gelfand MS, Shevelyov YY, Razin SV.

Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res 26: (2016);70-84.

[31] Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG, Flicek P, Dekker J, Merkenschlager M. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res 23: (2013);2066-77.

[32] Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A 111: (2014);996-1001.

[33] Kaushal A, Mohana G, Dorier J, Özdemir I, Omer A, Cousin P, Semenova A, Taschner M, Dergai O, Marzetta F, Iseli C, Eliaz Y, Weisz D, Shamim MS, Guex N, Lieberman Aiden E, Gambetta MC. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat Commun 12: (2021);1011.

[34] Wang Q, Sun Q, Czajkowsky DM, Shao Z. Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. Nat Commun 9: (2018);188.

[35] Liu C, Cheng YJ, Wang JW *et al.* Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. Nature Plants 3: (2017); 742–748.

[36] Parada LA, McQueen PG, Misteli T. Tissue-specific spatial organization of genomes. Genome Biol 5: (2004);R44.

[37] Tanabe H, Müller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci U S A 99: (2002);4424-9.

[38] Bridger JM. Chromobility: the rapid movement of chromosomes in interphase nuclei. Biochem Soc Trans 39: (2011);1747-51.

[39] Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 4: (2006);e138.

[40] Strambio-De-Castillia C, Niepel M, Rout MP. The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11: (2010);490-501.

[41] Houben F, Willems CH, Declercq IL, Hochstenbach K, Kamps MA, Snoeckx LH, Ramaekers FC, Broers JL. Disturbed nuclear orientation and cellular migration in A-type lamin deficient cells. Biochim Biophys Acta 1793: (2009);312-24.

[42] Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD. Nuclear lamins. Cold Spring Harb Perspect Biol 2: (2010);a000547.

[43] Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B. Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38: (2006);1005-14.

[44] Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38: (2010);603-13.

[45] Amendola M, van Steensel B. Mechanisms and dynamics of nuclear lamina-genome interactions. Curr Opin Cell Biol 28: (2014);61-8.

[46] Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L, van Steensel B. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res 23: (2013);270-80.

[47] Brickner JH, Walter P. Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol 2: (2004);e342.

[48] Casolari JM, Brown CR, Komili S, West J, Hieronymus H, Silver PA. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. Cell 117: (2004);427-39.

[49] de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev 26: (2012);11-24.

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

Chapter 2

Examination and Characterization of Photoactivatable DNA Binding Compounds for Use as a Tool to Examine Genomic Contacts

2.1 Overview:

Photoactivatable compounds have been combined with various illumination techniques in the past to study and interact with a variety of biological phenomenon [1]. As far back as the late 1970s, researchers were altering molecules such as ATP by adding photochemical protecting groups. These caged ATP molecules were unable to be metabolized until liberation of the ATP molecule via 340nm light [2]. Creating caged ATP generally involves altering the γ phosphate by adding a photoactivatable group, usually a 2-nitrophenylehtyl group that gets cleaved away yielding functional ATP (Figure 2.1a) [2]. Caged ATP has been used to study excitation-contraction coupling mechanisms in rabbit muscle fibers [3]. While caged ATP uses cleavage of a molecule to photoactivate it, other mechanisms have been used. Photoisomerization of azobenzenes has been used to transform its conformation from cis to trans or vice versa [4]. Caging of cations (CA2+ and Mg2+) has been achieved by introducing photoactivatable groups into molecules such as BAPTA [5] or EDTA [6]. The “nitr” series of molecules is based on BAPTA (Figure 2.1b). The second approach based on EDTA has been called “DM-nitrophen” (Figure 2.1c). Both these compounds work similarly. Upon irradiation with 300nm light a cleavage reaction occurs releasing a cation as well as some biologically inert side products (Figure 2.1b-c). Caged cations have primarily been utilized to study ion channels [7] and cation transport mechanisms [8]. Caged cations provide a couple of unique advantages which are shared with other

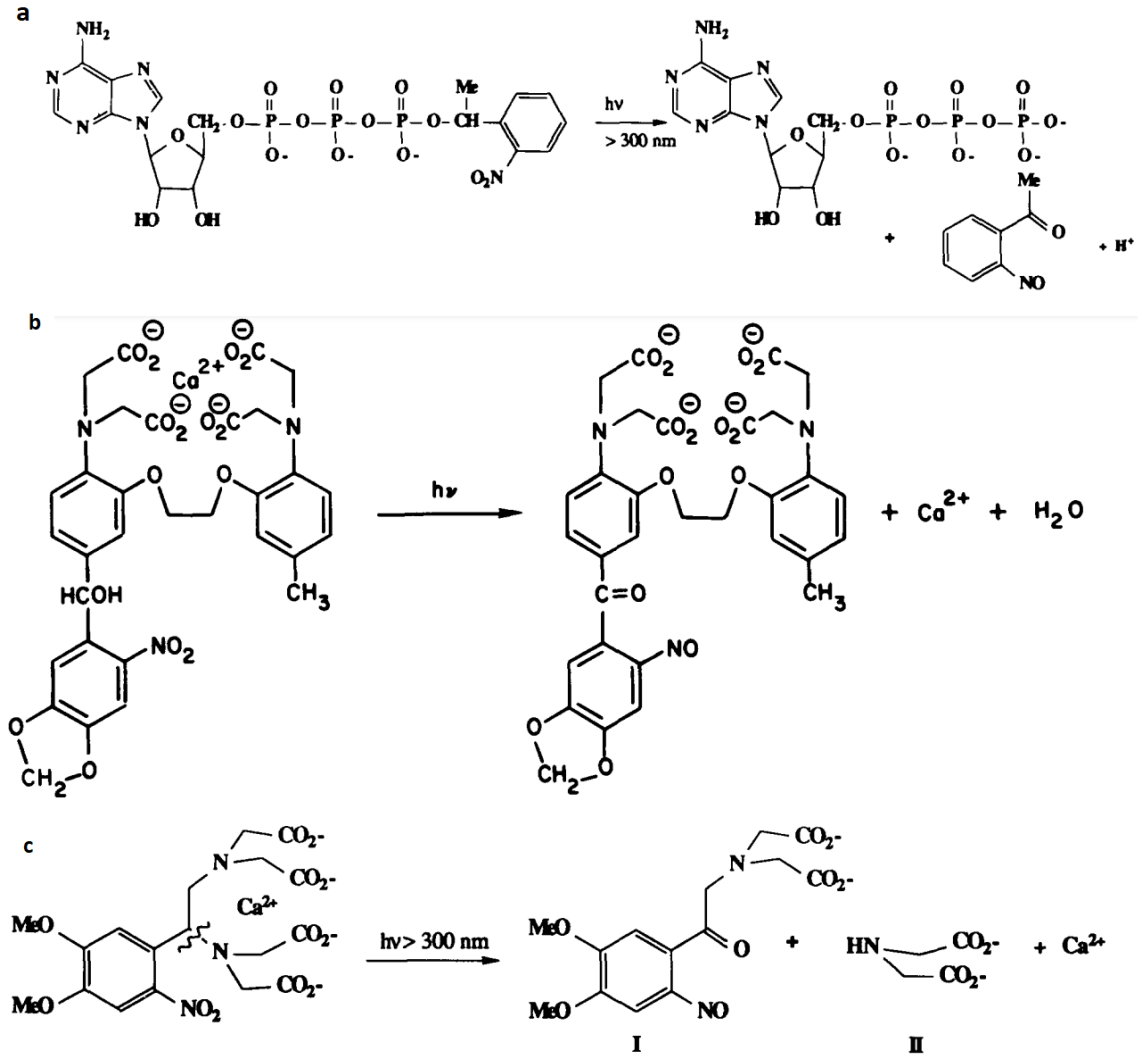


Figure 2.1. Examples of photoactivatable compounds: Caged-ATP, NITR-5 and DM-Nitrophen.

(a) Photochemical reaction of Caged-ATP. The γ phosphate is altered making the compound inactive until a light caused cleavage occurs at the 2-nitrophenylethly group leaving behind an active ATP molecule. (b) Photochemical reaction of NITR-5, after irradiation the Ca2+ cation is released. (c) Similar photoreaction to (b) but utilizes DM-Nitrophen as a photoactivatable compound for releasing Ca2+. Adapted from [9].

photoactivatable compounds. As photoactivatable compounds are inert until activated with light, they tend to have strong temporal resolution when compared to other methods that are commonly used to introduce small molecules and ions [9]. While at the time of their initial discovery the spatial resolution did not see widespread use, it became an advantage that would be utilized as illumination techniques continued to improve [9].

Similar to caged molecules, photoactivatable fluorescent proteins have seen extensive use due to their spatial and temporal resolution. Photoactivatable fluorescent proteins can be broken down into 3 categories based on their response to light stimuli. Photoactivatable fluorescent proteins can either switch from a dark state to a bright state, convert to a different fluorescent color, or reversibly photoconvert [10]. The first class includes proteins such as PA-GFP [11] (Figure 2.2a) and PA-mRFP [12]. Both of these compounds remain dark until activated with UV light, after which they brightly fluoresce green or red respectively (approximately 70-fold increase in fluorescence) [11,12]. Proteins that switch colors upon activation include EosFP [13], Dendra2 [14], and Kaeda [15] (Figure 2.2b). These proteins fluoresce green but transition to red emission upon UV activation [10]. The primary advantage of these compounds is that unlike PA-GFP they are easily visualizable prior to photoswitching which can be useful in targeting specific volumes for photoactivation. Finally, reversible highlighters such as Dronpa [15] and Kindling-FP [16] (Figure 2.2c) demonstrate a reversible change between two different fluorescent colors.

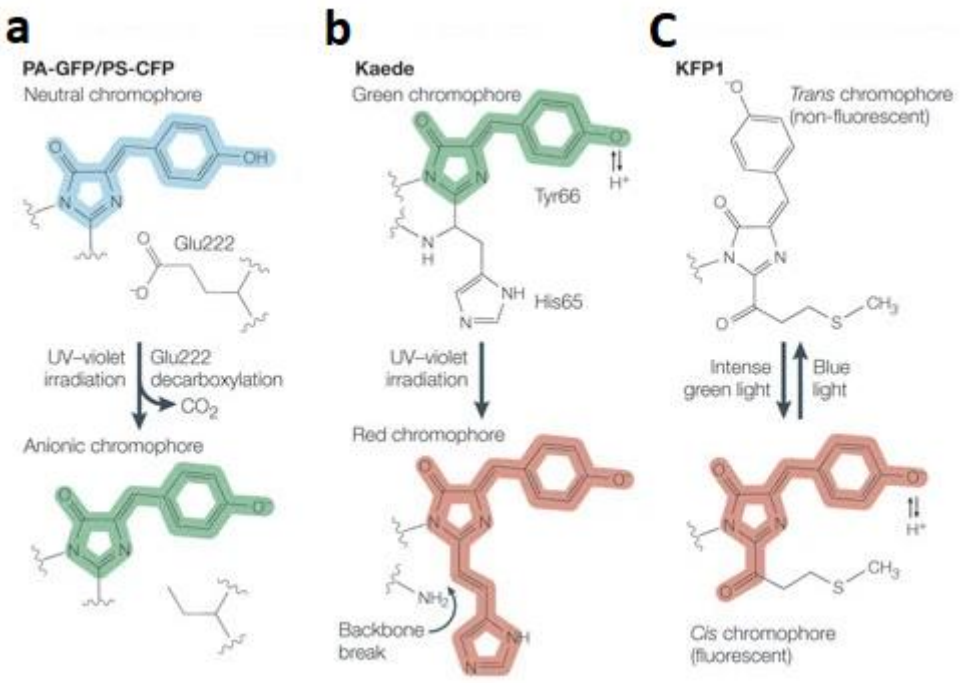


Figure 2.2 Example photochemical reaction of common photoactivatable fluorescent proteins.

(a) Photochemical reaction of PA-GFP. After UV irradiation a green chromophore is left behind and abled to be imaged. Unactivated chromophores are not fluorescent at that wavelength. This process is irreversible. (b) Photoconversion of Kaede from a green fluorescent protein to a red fluorescent protein. This irreversible process results in a backbone break to change the fluorescence color of the molecule. (c) Reversible process for activation of KFP-1. Green light changes the chromophore conformation from trans to cis making it fluorescent. This process can be reversed by using blue light. Adapted from [23].

As stated earlier, photoactivatable fluorescent proteins have a variety of uses because of their spatial and temporal resolution. Photoactivatable fluorescent proteins allowed researchers to do photo-chase experiments [10]. In a given cellular compartment the total protein concentration may remain constant as gain and loss are balanced, preventing visualization of protein trafficking. Photoactivatable fluorescent proteins can be used to track protein movement by selectively activating a small population of proteins and tracking their movement throughout the cell. This has been used to track the origins of peroxisomes [17], tubulin behavior [18], and actin movement [19].

Similar to how photoactivatable fluorescent proteins can be used for tracking proteins in cells, they can also be used to track cell movement and development in tissues. By having embryonic cells express a photoactivatable fluorescent protein like tdEosFP researchers were able to study cell movement by selectively activating tdEosFP in a subpopulation of cells and tracking them over time [20]. Lastly, photoactivatable fluorescent proteins have been used to examine protein binding kinetics, in a method similar to Fluorescence Recovery After Photobleaching (FRAP). Rather than bleach out a volume of fluorescent proteins, a subset of photoactivatable fluorescent proteins is activated and tracked as they diffuse [10].

Perhaps one of the most interesting uses of photoactivatable proteins is their use in super resolution imaging. While many techniques have been developed to break the diffraction limit of conventional microscopy, photoactivatable fluorescent proteins have been used for Photoactivated Localization Microscopy (PALM) [21]. Stochastic optical reconstruction microscopy utilizes photoswitchable organic dyes to achieve the same goal [22].

Lastly, photoactivatable compounds are particularly interesting for therapeutic purposes. Photodynamic therapy has been used to treat many different types of cancers, including breast cancer [25], pancreatic cancer [26], and colorectal cancer [27]. Photodynamic therapy works by introducing a photosensitizing agent to cancer cells and applying light [24]. As with the previously discussed methods, the spatial resolution provided by utilizing light in this way allows specific targeting of tumor cells over other benign tissue.

Perhaps most pertinent to the rest of this chapter, phototherapy has been used in the treatment of psoriasis, a skin condition resulting from increased skin cell production. Psoralen Ultraviolet A is a form of photochemotherapy used to treat certain psoriasis variants. Psoralen which is photosensitive is introduced to the site either topically or through an injection. Ultraviolet A light is then used to activate the psoralen and induce apoptosis [28]. Psoralen and similar compounds operate by first intercalating into DNA at sites where it can form monoadducts or interstrand crosslinks. Upon absorption of a UV photon, psoralen forms a pair of covalent bonds at a DNA residue (preferably at Thymine residues). Activation by a second UV photon causes it to form a second pair of covalent bonds with a neighboring Thymine on the opposite strand forming an interstrand diadduct [33] (Figure 2.3a,b). Many compounds behave similarly to psoralen, but differ in other characteristics such as binding affinity, quantum yield, and solubility [34] (Figure 2.3c).

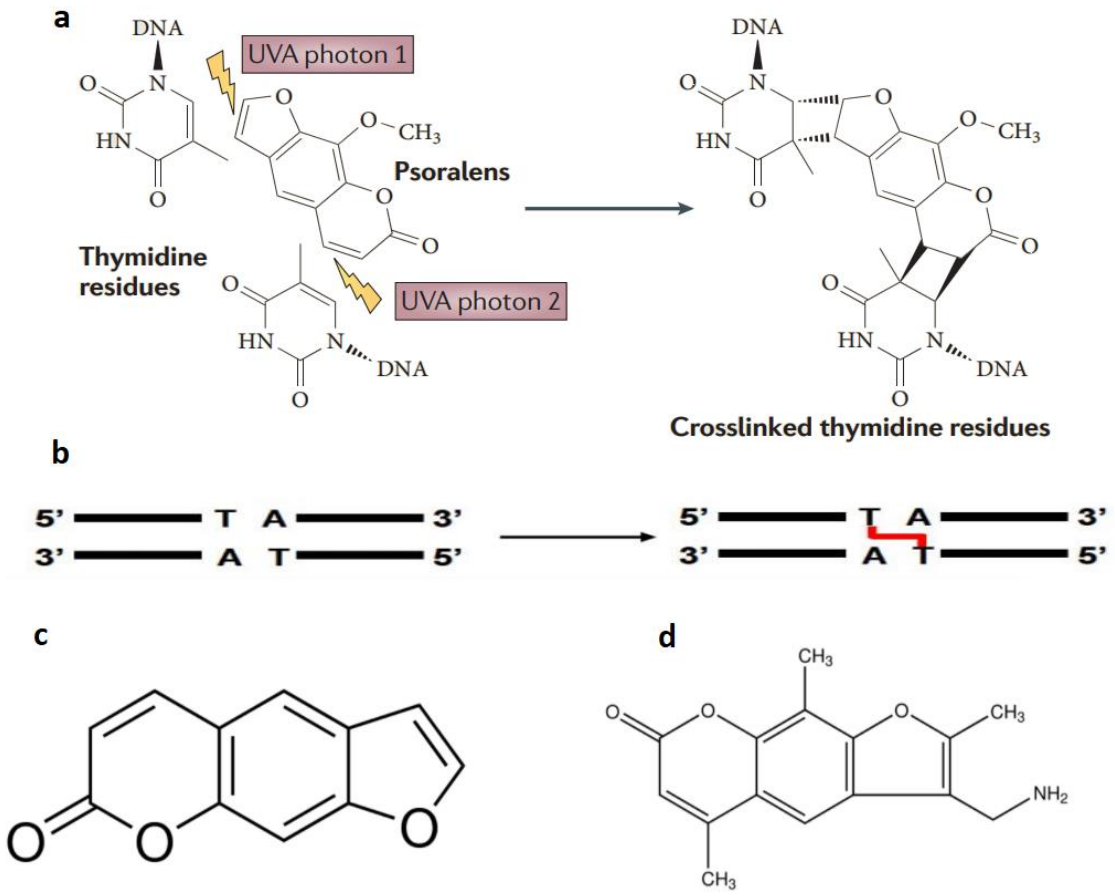


Figure 2.3 Overview of Psoralen Compounds and Photochemistry.

(a) Example schematic of a psoralen crosslinking reaction. Covalent bonds between psoralen like compounds and thymidine residues occur in two separate light activated reactions. Adapted from [33]. (b) Visualization of the interstrand crosslink formed by a psoralen molecule. The crosslink binds both strands of DNA together across thymidine residues. (c) Psoralen molecule. (d) 4'-aminomethyl trioxsalen. This compound behaves similar to Psoralen with slightly different crosslinking efficiency and intercalation properties.

Compounds that have not been used as photoactive agents in the past are being altered today to have photoactivatable properties. Cisplatin has been used extensively as a chemotherapy drug to treat a variety of cancers including lung, ovarian, and breast cancer [29]. These compounds tend to be inactive in blood, but in the cytoplasm with lower chloride anion concentration undergo hydrolysis of the chloride ligands to create a species that forms monoadducts with the N7 position of guanine or adenine [30] (Figure 2.4). These adducts prevent replication and transcription of cellular DNA leading to eventual cell death (Figure 2.4). Unfortunately, cisplatin-based compounds tend to have severe side effects brought on by off targeting of nonmalignant cells [31]. In order to reduce this off-target cell toxicity, researchers sought to introduce photoactivatable groups to take advantage of the spatially targeted activation [32].

As discussed, photoactivatable compounds offer unique spatial and temporal resolution. While these compounds have seen a variety of uses for a plethora of different purposes as described previously, they were of particular interest to us as a platform to study genomic interactions. As described in the previous chapter, spatial genomic organization is important to understand from a disease and development related perspective. In this chapter, I will discuss our efforts to investigate several photoactivatable compounds we identified as potential candidates for use in methods to study genomic DNA interactions. Photoactivatable DNA binding compounds, as discussed previously, were of particular interest to us.

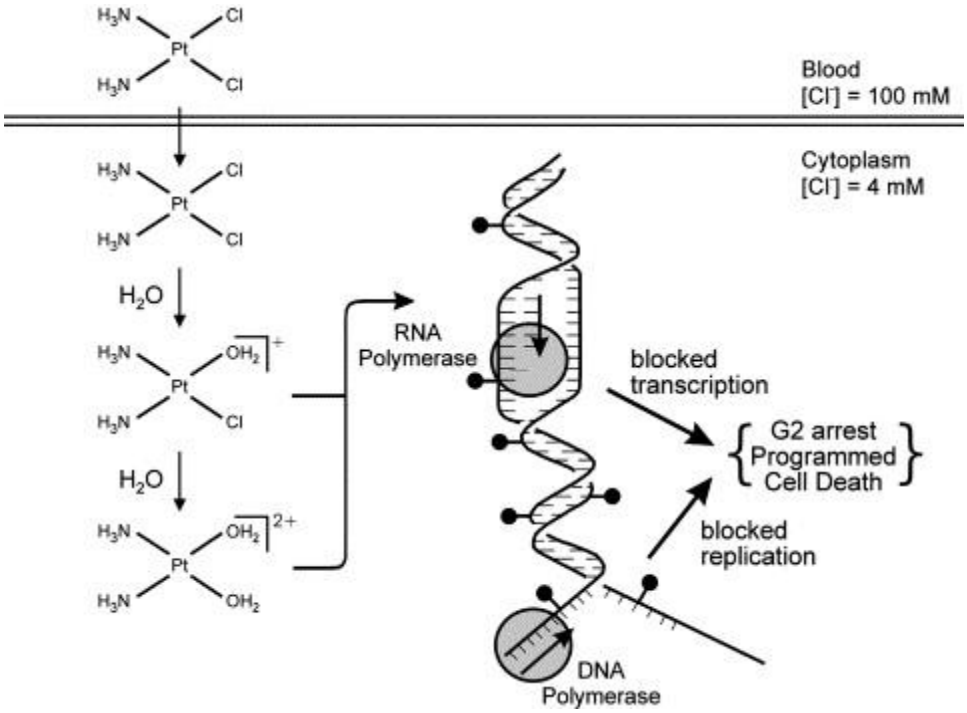


Figure 2.4 Overview of Cisplatin activity to cause cell apoptosis.

Cisplatin remains in an inactive form in the blood where Chloride anion concentration remains relatively high. Upon entering the cell cytoplasm where Chloride anion concentration is lower, the chloride ligands undergo hydrolysis. This reaction allows cisplatin to bind numerous biologically relevant molecules including cellular DNA. While the efficiency with which it binds different molecules can vary substantially, it tends to bind the N7 position of Guanine or Adenine. Adapted from [32]

2.2 Exploration of Rhenium Based Compounds:

Similar to cisplatin, other metallo-organic compounds show promise as photoactivatable DNA binding compounds. The lab of Justin Wilson at Cornell University brought a rhenium-based compound, [Re(CO)2(Phen)(DAPTA)(Cl)] (abbreviated SCM-I-230) [35] (Figure 2.5a) to our attention. This compound has several features which are worth mentioning. Similar to Cisplatin, where a Platinum atom is the central atom, Rhenium, another heavy metal element, is the central atom. Another similarity to cisplatin is the use of carboxyl groups. These functional groups form the basis of the compounds binding potential. After irradiation, the CO group is released leaving the compound with an active binding site for other organic molecules [36]. Other functional groups may serve as better ligands for these purposes. The DAPTA ligand is to make the compound water soluble [37]. The Phosphine group gives the compound its photoactivity and phosphorescence, as similar compounds without a phosphine group have very limited photoactivity and phosphorescence [38].

In order to test the photoactivatable binding potential of SCM-I-230 to DNA, we measured the phosphorescence lifetime of SCM-I-230 in the presence and absence of DNA and before and after UV irradiation. Calf Thymus DNA and SCM-I-230 were incubated together briefly after which it underwent irradiation using a 365nm LED. Irradiation level varied from no irradiation (Figure 2.5b), a low power irradiation for 30 minutes (Figure 2.5c) and a high-power irradiation for a shorter time (Figure 2.5d). Dialysis was then performed removing the unbound SCM-I-230 and phosphorescence lifetime curves were acquired determine if any SCM-I-230 remained. We noticed a strong phosphorescence signal in the absence of dialysis as the unbound SCM-I-230 remains

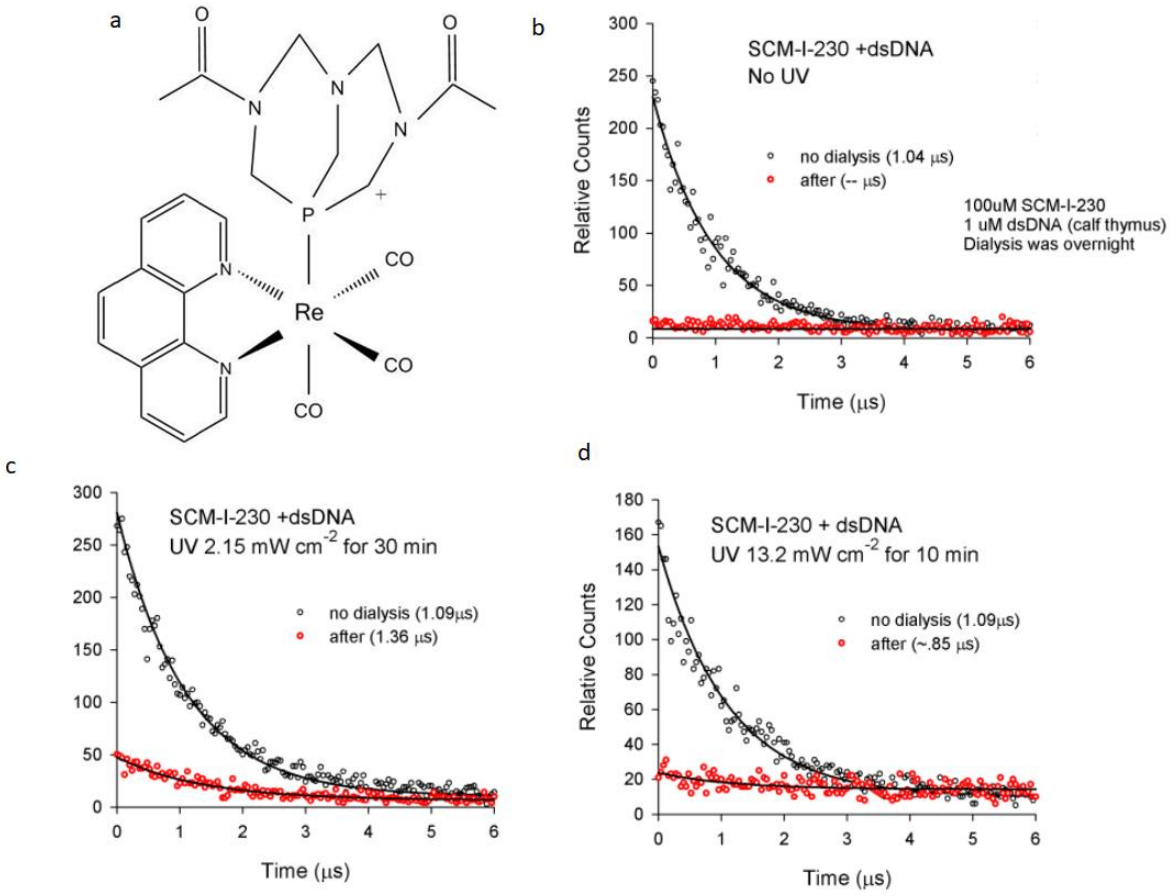


Figure 2.5. Photoactivity of a Rhenium-based Compound.

(a) Diagram of [Re(CO)2(Phen)(DAPTA)(Cl)] (SCM-I-230). The CO functional groups serve as binding sites after irradiation. The phosphine gives the compound its photoactivity while the DAPTA makes the compound soluble. (b) Phosphorescence lifetime of SCM-I-230 and DNA without UV irradiation. (c) Phosphorescence lifetime of SCM-1-230 and DNA with low power UV irradiation for extended time. (d) Phosphorescence lifetime of SCM-1-230 and DNA with high power UV irradiation for a short time.

in solution. In the absence of UV irradiation, we see no phosphorescence after dialysis, suggesting that the compound only binds DNA upon UV irradiation. This was confirmed by looking at phosphorescence lifetime curves generated after UV irradiation and dialysis, in which we see a slight residual phosphorescent signal after both UV irradiation doses, confirming residual SCM-I-230 is bound to DNA. Furthermore, we notice a difference in the phosphorescent signal when the compound is exposed to a lower UV dose over a longer period of time compared with a higher dose over a shorter period of time. In order to be useful to study genomic contacts we would require the addition of an affinity tag, but otherwise SCM-I-230 appears as a novel compound with the potential to be used to study genomic contacts. It appears to be sensitive to UV wavelengths and doses that would not form DNA breaks. Since most uses for DNA crosslinking compounds would require that the crosslink be reversed for downstream analysis, the fact that this compound forms mono-adducts is a benefit. Reversal treatments for mono-adducts may be less harsh than reversal treatments required to remove di-adducts.

2.3 Examination of Psoralen Based Compounds:

Besides examining novel photoactivatable DNA binding compounds, we also examined commercially available compounds. Psoralen-PEG3-Biotin (EZ-link) is a commercially available photoactivatable DNA crosslinker with an affinity tag making it an ideal compound for studying genomic contacts. The photoactivatable portion, psoralen, has seen extensive use in the past for crosslinking DNA [39]. These bonds are formed individually as a UV photon is required for each bond. Psoralen based photochemistry occurs at wavelengths between 340nm and 380nm [39]. These wavelengths are drastically longer than 240nm where you commonly see DNA damage [40].

While EZ-Link is commercially available, we had access to another compound 4'-AMT-PEG3-Biotion (AP3B) [41]. This compound is similar to EZ-link but swaps the psoralen group for a 4'-Aminomethyl trioxsalen. This compound has a higher binding affinity for DNA compared to the more common methoxsalen forms of Psoralen [39]. We sought to characterize various properties of both AP3B and EZ-link in order to examine their efficacy for examining genomic contacts.

2.3.1 Fluorescence and Absorption Spectra of 4'-AMT:

We measured fluorescence and absorption spectra for 4'-AMT. Fluorescence curves were taken using excitation at 330nm ±irradiation from a 360nm LED (Figure 2.6a). We performed irradiations at a variety of powers and times. Irradiation parameters were chosen to ensure many different doses were tested at both high and lower power (and thus long and short times with equivalent total dose). Several interesting characteristics were noticed. As doses increase the fluorescence peak shifts to a lower wavelength. At lesser doses in the case of lower power we do not yet see this shift occur. We attribute this shift in fluorescence to breakdown of the 4'-AMT molecules in solution upon irradiation. We measured absorbance spectra of 4'-AMT in the presence and absence of DNA both before and after irradiation using a 360nm LED (Figure 2.6b). Several interesting results are observed. We noticed that 4'-AMT shows activity in the 340-380nm range, on

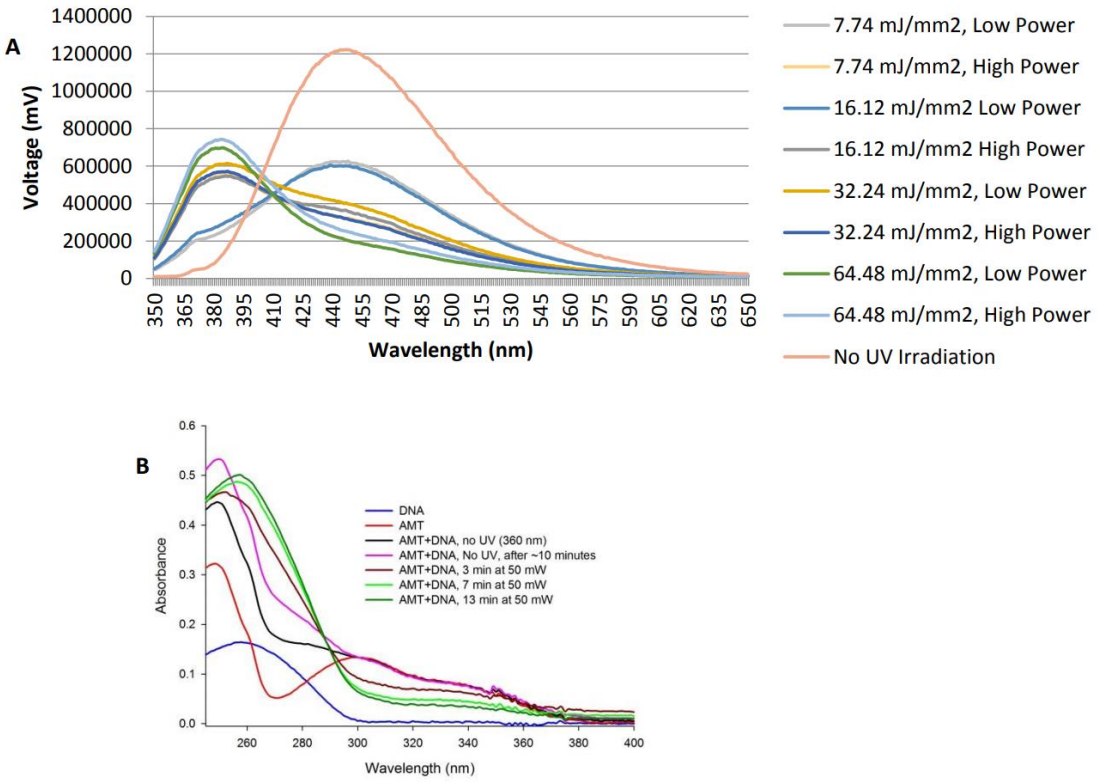


Figure 2.6. Absorbance and Fluorescent Spectra for 4'-AMT.

(a) Fluorescent spectra for 4'-AMT after irradiation of a variety of powers. 4'-AMT was irradiated using a 360nm LED for the doses listed. High power refers to irradiations at 13mW for a shorter time and low power at 2mW for a longer time for equivalent total dose. Spectra were taken after excitation at 330nm. We notice a distinct shift in fluorescence as irradiation occurs. (b) Absorbance spectra of 4'-AMT in the presence and absence of DNA after irradiation. We notice the activation absorbance of 4'-AMT is well away from the 260nm peak of DNA. We notice a longer shift in absorbance as 4'-AMT binds DNA.

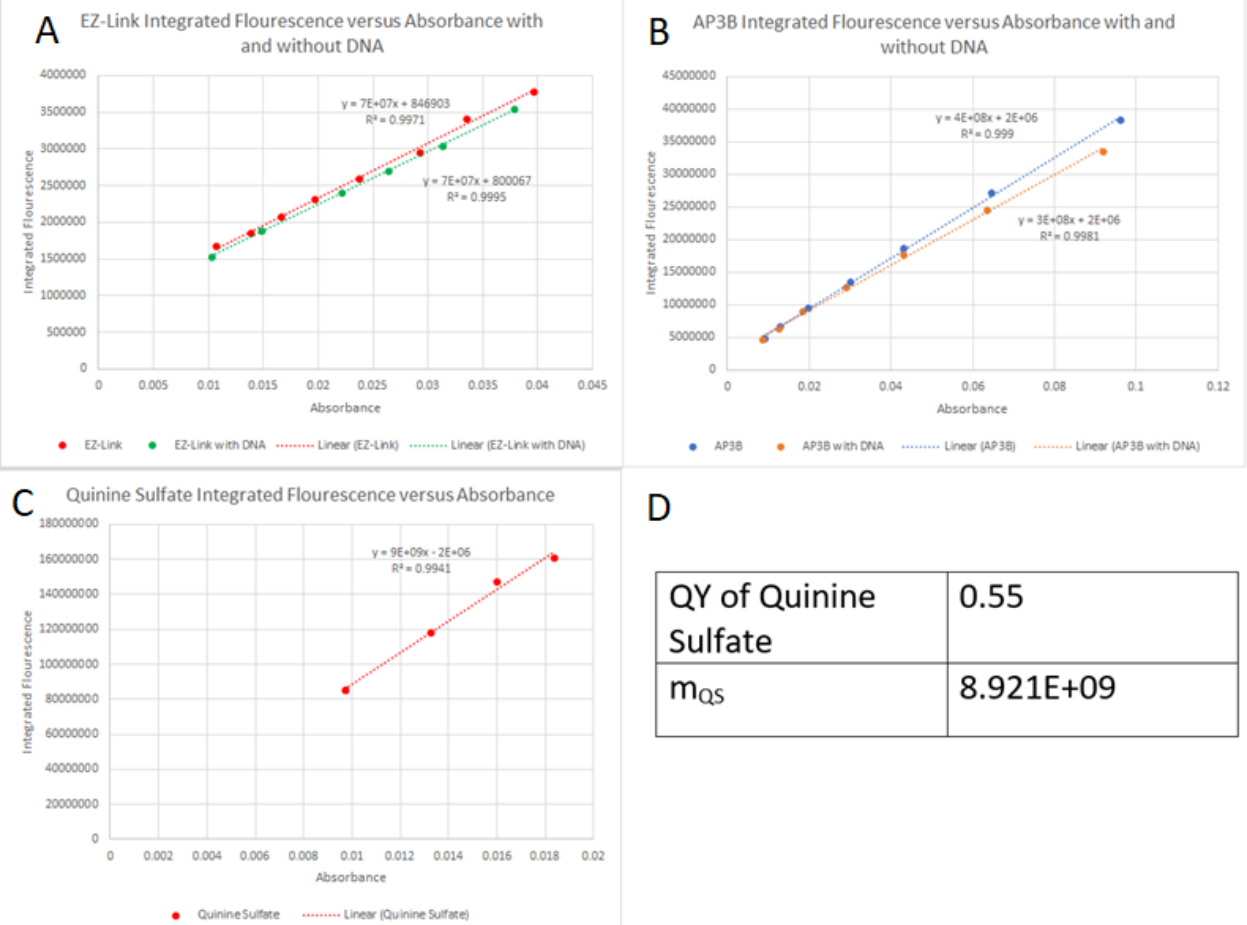
the longer wavelength shoulder of the spectrum. This is well outside the peak absorbance of DNA suggesting that we are able to activate 4'-AMT without causing significant DNA damage. Upon intercalation, but not crosslinking, we notice that the spectra appear fairly similar to a combination of the 4'-AMT and DNA spectra. After irradiation though, we notice that this changes. We notice a shift to a longer wavelength absorbance at lower UV. Additionally, we see a much more diminished shoulder feature from 300nm to 400nm. This shoulder feature continues to diminish with increasingly stronger irradiation doses.

2.3.2 Fluorescence Quantum Yield of AP3B and EZ-Link:

All Psoralen's are weakly fluorescent. Our ultimate goal is to measure the two-photon DNA cross-linking quantum yield of our AP3B probe, but to do this we first need to measure the absolute two-photon absorption cross-section, δ. The most commonly reported values are for the action cross section for fluorescence since this is easiest to measure. The action cross-section is the product of the absolution two-photon absorption cross section times the fluorescence quantum yield (ϕ_F). The fluorescence quantum yield of a fluorophore is the ratio of emitted photons to absorbed photons.

$$\Phi = \frac{\#\ emitted\ photons}{\#\ absorbed\ photons} \qquad (1)$$

Since fluorescence is a decay process of excited fluorophores the quantum yield can be written as



QY of Quinine Sulfate	0.55
m_{QS}	8.921E+09

Figure 2.7. Florescence Quantum Yield of AP3B and EZ-Link.

Plots of absorbance versus integrated fluorescence for (a) EZ-Link with and without DNA, (b) AP3B with and without DNA, and (c) Quinine Sulfate our quantum yield standard. (d) Literature values of the quantum yield of Quinine Sulfate [44] and the calculated m_{QS} value. Using this method and the described equation, the fluorescence quantum yield of AP3B and EZ-link can be calculated.

$$\Phi = \frac{k_f}{k_f + \sum k_n} \qquad (2)$$

Where k_f is the rate constant for fluorescence radiation and k_n is the rate constant for all other decay processes. Measuring the rate constants of all these processes is practically impossible. Instead, the fluorescence yield of a fluorophore with a known and well-defined quantum yield is used as a standard.

$$\Phi_{unknown} = \Phi_f \left(\frac{m_{unknown}}{m_f} \right) \qquad (3)$$

Where Φ_f is the quantum yield of a known fluorophore and m is the slope of integrated fluorescence emission versus absorbance for the unknown compound and for the known fluorophore. Plots of integrated fluorescence versus absorbance were collected for EZ-Link and AP3B both with and without DNA and Quinine Sulfide as a standard. Fluorescence quantum yields of 0.022 and 0.02 were found for AP3B without and with DNA respectively. Fluorescence quantum yields of 0.004 were found for EZ-link with and without DNA respectively.

2.3.3 Two-Photon Cross Section of AP3B and EZ-Link:

Of particular interest to our pursuits were two-photon characteristics of both Psoralen and 4'-AMT as these have not been quantified previously. Particularly we sought to characterize the two-photon absorption cross section. Since two-photon excitation is a second-order process, the number of two-photon absorptions per second is proportional to the two-photon absorption cross section δ times the square of the irradiance, I_0. The

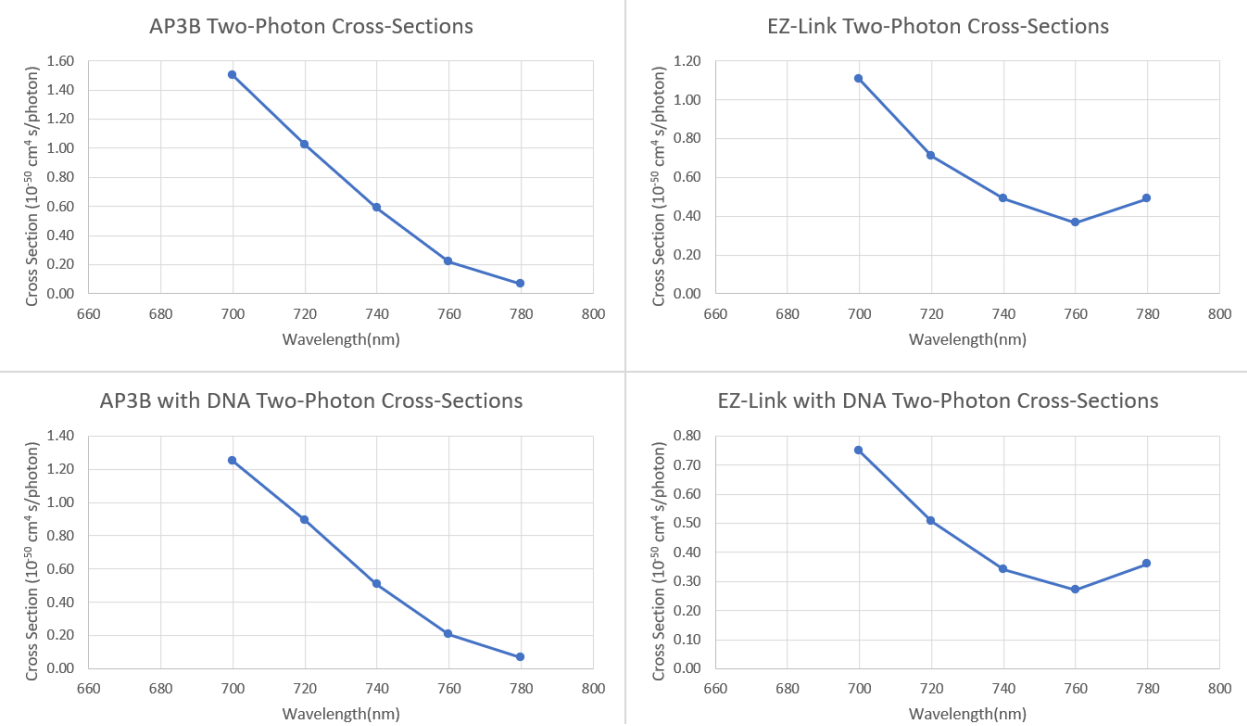


Figure 2.9. Two-Photon Cross-Sections for crosslinking compounds.

Cross-sections graphs for two photoactivatable DNA crosslinking compounds with and without DNA at wavelengths from 700nm to 780nm.

two-photon absorption rate can be written as

$$N_{abs}(t) = C\delta I_0^2(t) \int_V dVS^2(r) \qquad (4)$$

Where C is the concentration and S(r) is the spatial distribution function of the incident light (i.e. the point spread function). In the case of no stimulated emissions or self-quenching events, the number of fluorescently emitted photons as a function of time can be written as

$$F(t) = \frac{1}{2}\Phi_2 \eta N_{abs} \qquad (5)$$

Where Φ_2 is the fluorescence quantum yield and η is the fluorescence collection efficiency of the instrumental set up. The factor of ½ comes from the fact that two excitation photons are required for each fluorescently emitted photon. In practice, we measure the time-averaged fluorescence photon flux, ⟨F(t)⟩.

$$\langle F(t)\rangle = \frac{1}{2}\Phi\eta_2\, C\delta \frac{\langle I_0^2(t)\rangle}{R\tau} \int_V dVS^2(r) \qquad (6)$$

The time-averaged fluorescence flux is proportional to $\langle I_0^2(t)\rangle$, but most detectors give a signal proportional to $\langle I_0\ (t)\rangle$. Because of this we rewrite, substituting $\langle I_0^2(t)\rangle$ with $g\langle I_0\ (t)\rangle$.

$$\langle F(t)\rangle = \frac{1}{2}g\Phi\eta_2\, C\delta \frac{\langle I_0\ (t)\rangle^2}{R\tau} \int_V dVS^2(r) \qquad (7)$$

Here g is $\frac{\langle I_0^2(t)\rangle}{\langle I_0\ (t)\rangle^2}$, the second order temporal coherence of the excitation source. Additionally, since we are looking at the time-averaged fluorescence flux we need to introduce a factor for the duty cycle, R_T. In order to determine the two-photon cross

section, numerous factors need to be determined. Notably, the spatial distribution of incident light, the second order temporal coherence, the fluorescence collection efficiency, the concentration, and the quantum yield. These factors stem from a variety of sources. The concentration (C), the quantum yield (Φ), and the two-photon cross section (δ) are properties of the fluorophore solution itself. The remaining factors, the spatial distribution (S(r)), the duty cycle (R_T), the second order temporal coherence (g), and the fluorescent collection efficiency (η_2) are factors of the instrument setup. Determining the two-photon cross section of a compound would require either determining these values or utilizing a standard with a known two-photon cross section. First curves of fluorescent emission versus excitation intensity are generated for the compound of interest and a known standard. These are then fit to $y_0 + mI^2$ where the slope m contains all of the above-described factors. The ratio of fluorescent intensities can be written as

$$\frac{m_{cal}}{m_{new}} = \frac{\Phi_{cal}\eta_{cal}\, C_{cal}\delta_{cal}}{\Phi_{new}\eta_{new}\, C_{new}\delta_{new}} \qquad (8)$$

Notably, numerous factors drop out in this ratio, including the second order temporal coherence and the spatial distribution of the incident light. The two-photon cross section of interest can be determined by the concentration of the two compounds, the two-photon cross section of the standard, and the quantum yields of the two compounds correcting for the collection efficiencies of the instrument.

We utilized fluorescein as a standard to determine two-photon cross sections for AP3B and EZ-link both with and without DNA. Fluorescence intensity curves were generated for a variety of wavelengths. Using the known two-photon cross section of

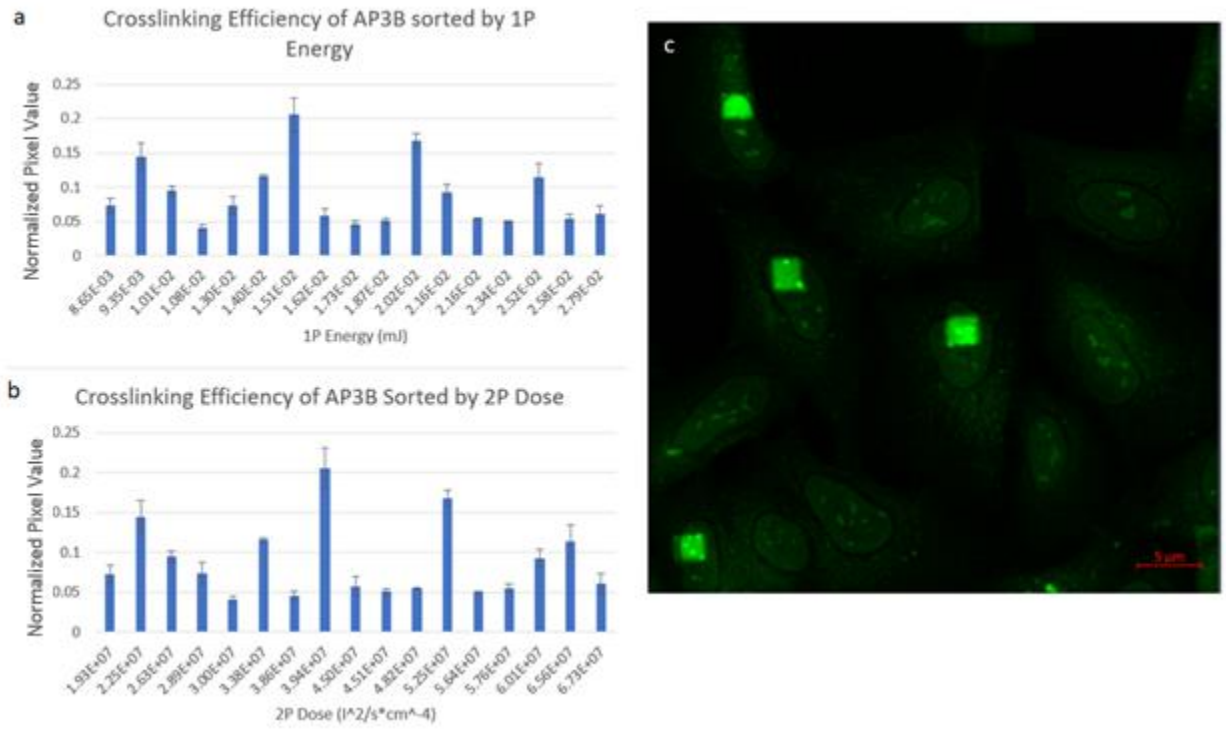


Figure 2.10. Determination of optimal crosslinking conditions using dye staining.

Cells were incubated with AP3B then small volumes were irradiated. After washing out unbound crosslinker these volumes were then stained with Streptavidin-DyLight488. After washing out unbound dye the pixel values of irradiated regions were determined and normalized to the maximum possible value. (a) Normalized Pixel Value (determined by looking at pixel values of crosslinked versus non-crosslinked areas) plotted against 1P Energy. (b) Similar to (a) but instead signal to noise is plotted against the 2-photon dose. (c) Example image following staining. Clear regions of Streptavidin-DyLight488 are visualized indicating AP3B crosslinking. The pixel values of these regions were determined for various irradiation settings.

fluorescein we were able to use the above equations to determine the two-photon cross section of both AP3B and EZ-link (Figure 2.9).

2.3.4 Optimization of the Two-Photon Excitation Conditions for Crosslinking:

While these previously calculated values are of interest, our primary goal with these compounds was to utilize them with two-photon excitation microscopy in order to selectively label the DNA in a small volume inside a nucleus. In order to examine their viability for this purpose, we created an assay to examine the crosslinking efficiency of numerous crosslinking conditions simultaneously. After plating cells on a gridded dish and adding our photoactivatable crosslinkers, we selectively irradiated small regions of nuclei at varying powers and times. We then stained with a streptavidin dye to allow visualization of the crosslinked regions (Figure 2.10a,b). By comparing the streptavidin dye signal of regions that had different illumination settings we were able to test a variety of irradiation conditions at once.

We utilized this assay to examine a variety of dose settings (Figure 2.10c). By examining the pixel value of irradiated regions, we determined the optimal dose consisted of 0.015 mJ 700nm light (corresponding to a ($3.94*10^7$ $J^2/(s*cm^{-4})$) dose of two-photon excitation). Generally, we noticed that at a certain power threshold we would see sufficient crosslinking signal with this signal increasing for longer duration irradiations. At extremely high powers we would notice fairly significant cell damage suggesting that these powers would not be suitable for a crosslinking experiment. In the previous experiments while we saw a very strong signal from streptavidin binding indicating good crosslinking, we noticed a fairly substantial background signal as well. We attributed this to nonspecific

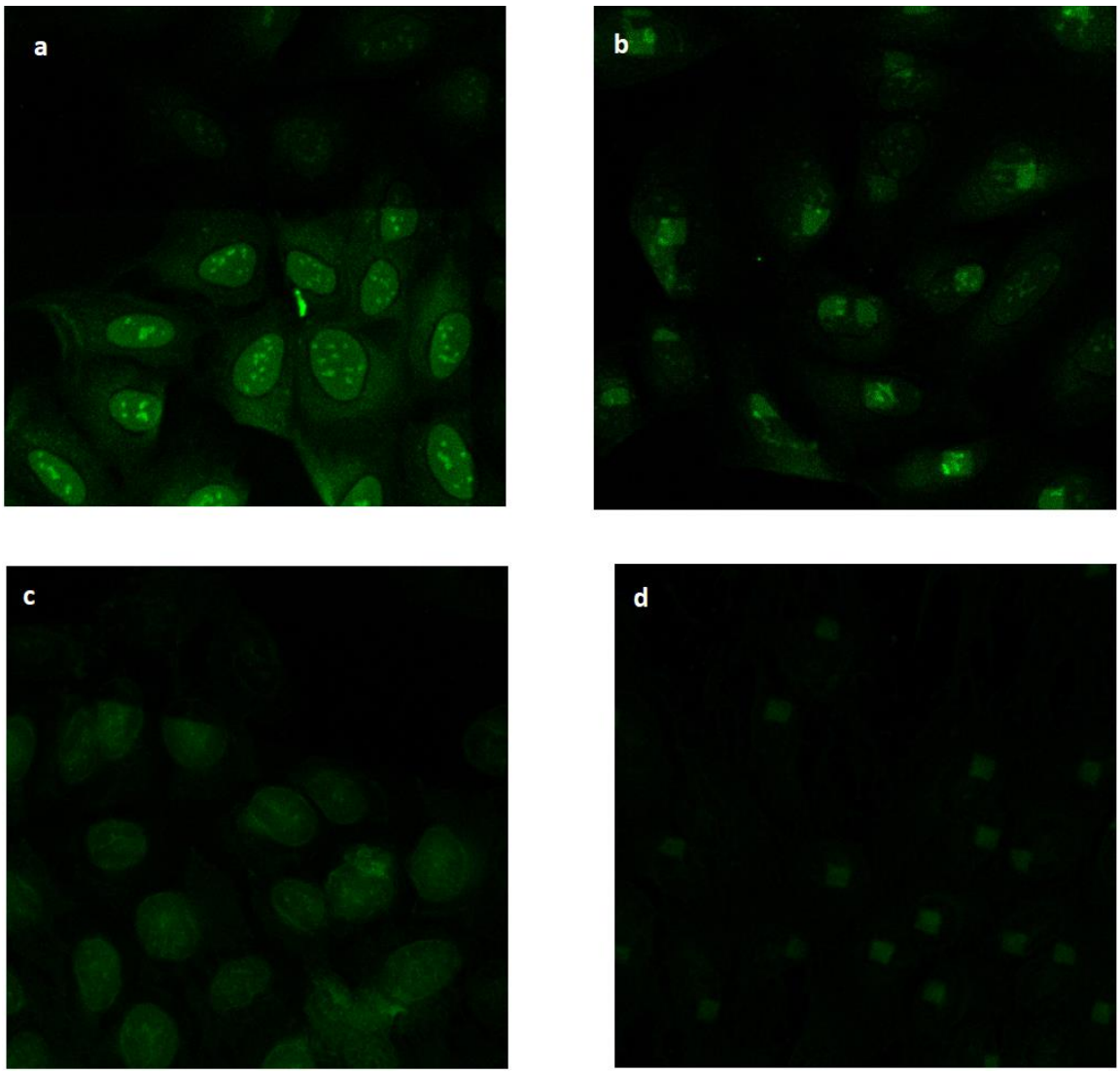


Figure 2.11. Use of 1,6-Hexanediol to remove background non-specific crosslinking

(a) Streptavidin staining showing both specific and non-specific binding of the crosslinking compound. The bottom half of the FOV has been excited using two-photon excitation. (b) Similar to (a) but with smaller regions crosslinked. (c) After excitation cells underwent a wash with 50mM 1,6-Hexanediol. The bottom half of the FOV was excited. (d) Similar to (c) but with small regions illuminated.

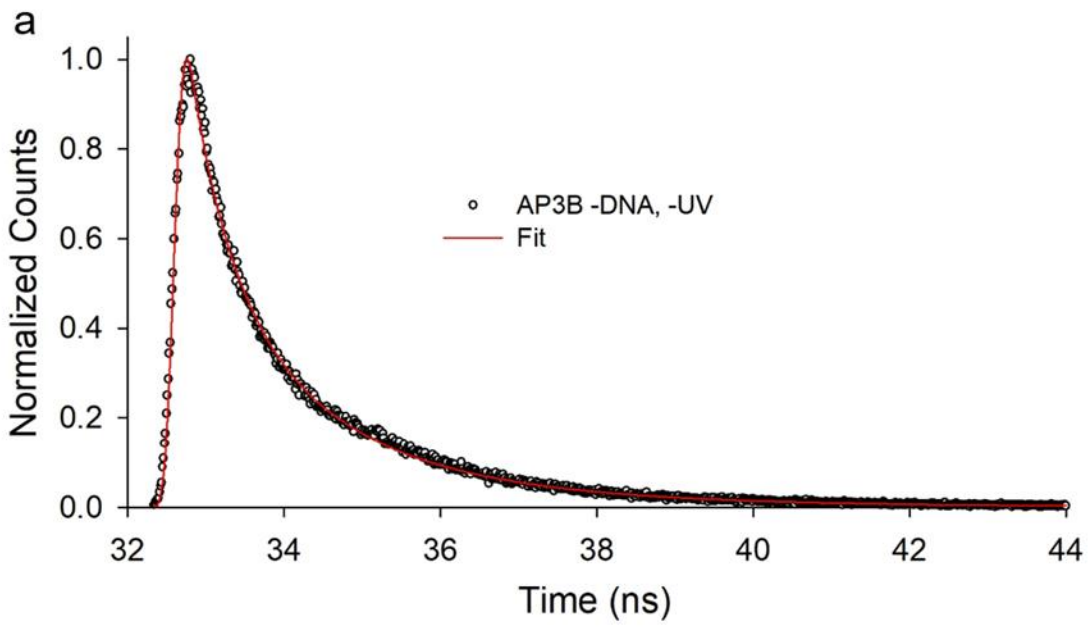
intercalation of the crosslinker molecule that was not getting properly washed out before staining with streptavidin dye. We had noticed this phenomenon in extracted gDNA that was enriched for the crosslinker as well as described in the next chapter. In order to correct for this, we introduced an additional washing step prior to staining with streptavidin dye. To ensure the unbound crosslinker was washed out we incubated the cells in 50mM 1,6-Hexanediol for 1 hour then washed that out prior to staining. 1,6-Hexanediol is known to break hydrogen bonds and other electrostatic interactions [45]. We did notice a lower signal that we expected from the irradiated regions, but found a significant drop in the background labeling suggesting that treatment with 1,6-Hexanediol helps remove intercalated, but non-covalently bound 4'-AMT. The addition of this step resulted in an increase in signal to background.

2.3.5 Fluorescence Lifetime Measurements of AP3B:

Another parameter of interest is the fluorescence lifetime of AP3B ±DNA. We measured the fluorescence lifetime both in the presence and absence of DNA utilizing the time-correlated single photon counting (TCSPC) [42]. The observed fluorescence decays required a two-exponential decay model to fit the data with acceptable $\chi 2$ values (<1.2):

$$F(t) = a_1 e^{\frac{-t}{\tau_1}} + a_2 e^{\frac{-t}{\tau_2}} \qquad (9)$$

Where a_1 and a_2 are the relative amplitudes of the two exponential decay components and τ_1 and τ_2 are the fluorescence decay times. The need for two exponential components to fit AP3B even in the absence of DNA is most likely an indication that there are a number radiative loss pathways of AP3B as the flexible structure



b. Summary of AP3B fluorescence lifetime measurements.

Sample	A_1	τ_1 (ns)	A_2	τ_2 (ns)	weighted τ	X^2
AP3B -DNA, -UV	0.663 ±0.004	0.697 ±0.027	0.337 ±0.004	2.22 ±0.033	1.21 ns	1.06 ±0.04
AP3b +DNA, -UV	0.863 ±0.003	0.871 ±0.012	0.137 ±0.003	3.75 ±0.041	1.26 ns	1.11 ±0.06
AP3B +DNA, + UV	0.861 ±0.006	0.770 ±0.017	0.139 ±0.005	3.63 ±0.063	1.16 ns	1.07 ±0.05

Figure 2.12. Fluorescence lifetime measurements of AP3B +/- DNA+UV.

(a) Typical TCSPC lifetime curve fit to a two exponential decay model. Sample was AP3B without added DNA. Fit parameters: a_1 = 0.66, τ_1 = 674 ps, a_2 = 0.34 and τ_2 = 2231 ps. ($\chi 2$ = 1.10) (b) Table showing the results for AP3B ±10mg/ml Salmon Sperm DNA and ± UV (2.15 mW of 360 nm illumination for 30 minutes). Mean and standard deviations shown for n = 5 traces for each condition.

takes on different configurations in solution. When intercalated into DNA, the shorter component shows a slight increase in both amplitude and decay time, while the longer lifetime component decreases in amplitude and increase to decay time. Overall, the weighted lifetime ($a_1 \times \tau_1 + a_2 \times \tau_2$) shows little change. Because the average lifetime was found to be ~1 ns, nonlinear saturation calculations can be carried out on a per pulse basis since our laser has a 80 MHz repetition rate (12 ns between pulses). Combined with two-photon absorption cross-section value of ~ 1GM at 700 nm we can calculate that at approximately 200 mW an AP3B molecule at the center of the focus would be two-photon excited approximately 150 times per second.

2.3.6 Examination of DNA damage from two-photon crosslinking:

Of some concern is the potential for high energy two-photon excitation to cause DNA damage. Particularly with low NA objectives, a relatively high power is required to achieve suitable crosslinking. While we don't expect traditional UV based DNA damage to occur, the high power focused through such a small volume could result in other forms of DNA damage. In order to examine the presence or absence of such damage we used cells that stably express fluorescently labeled Proliferating cell nuclear antigen (PCNA). PCNA is a DNA clamp that serves a variety of purposes including DNA replication and DNA repair [43]. We irradiated these cells using similar conditions as we tested earlier and tracked the fluorescently labeled PCNA to identify areas where DNA damage might be occurring. At optimal dose settings for crosslinking we don't observe any increase in PCNA around the irradiation site (Figure 2.12a,b). Upon increasing the power, however, we do observe a significant increase in PCNA activity around the irradiation site. These

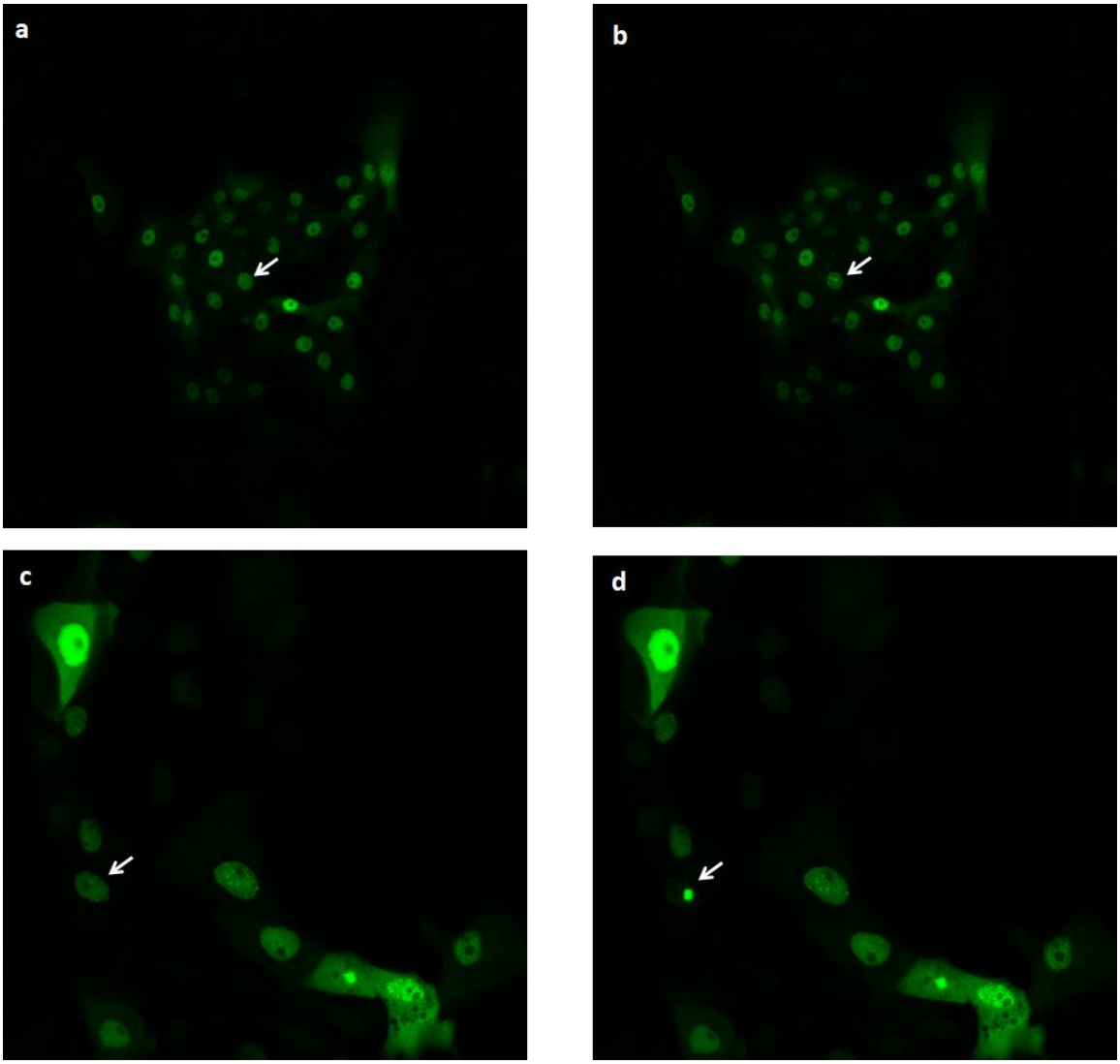


Figure 2.12. Examination of DNA damage by two-photon excitation conditions necessary for crosslinking.

(a) LLC-PK1 cells expressing eGFP-PCNA. The arrow indicates where an irradiation occurred. (b) After irradiation at 245mW for 41.2µsec no detectable eGFP-PCNA localization at the target site was observed. (c) and (d) Before and after irradiation at 315mW for 41.2µsec. Because the power is significantly higher, we did detect recruitment of PCNA to the target indicating double stranded DNA breaks are present.

results suggest that we are able to crosslink AP3B using powers which will not cause significant DNA damage.

2.4 Methods:

Rhenium Lifetime Measurements:

100μM SCM-I-230 was incubated with 1μM calf-thymus dsDNA for 1 hour. Phosphorescent lifetime curves were measured on our in-house set up consisting of a pulsed 405 nm laser with the time-resolved phosphorescence measured using a Stanford Research model SR-430 multi-channel scaler. To measure the effect of UV irradiation compound in the presence of DNA the solution was then placed in a small plastic dish and irradiated using a 360nm LED for the powers and times listed. After irradiation, the solution was dialyzed overnight (Slide-A-Lyzer Dialysis Cassette 10K cut off, ThermoFisher). The following day phosphorescent lifetime curves were taken again. Curves were fit to a single exponential decay model to obtain the phosphorescence lifetime of the solution.

4'-AMT Spectra Measurements:

To prepare for fluorescence spectra measurements, a solution of 200μM 4'-AMT in solution was prepared. UV irradiated samples were irradiated with a 360nm LED. The doses listed correspond to two levels of power (13mW, high) and (2mW, low). Changing the power along with the time was used to alter the doses. Fluorescent spectra were taken on a PTI QuantaMaster 400 Fluorimeter with an excitation wavelength of 330nm.

To prepare for absorbance spectra measurements, a solution with 3mM 4'-AMT and 25μM calf-thymus dsDNA was incubated for 30 minutes. Aliquots of this solution were

incubated for the times and powers listed using a 360nm LED. Absorbance spectra were then taken on a Cary300 UV-Vis Spectrometer.

Quantum Yield Measurements:

Solutions of 20μM AP3B and EZ-link were created with and without 505nM Salmon Sperm DNA. Integrated Fluorescence versus Absorbance spectra were generated using a PTI QuantaMaster 400 Fluorimeter. Slit widths were 8nm FWHM for the excitation path and 16nm FWHM for the detection. The excitation wavelength used was centered at 340nm. As a standard, Quinine Sulfide diluted to 0.4 absorbance was used.

Two-Photon Cross Sections:

Solutions of 200μM AP3B, 200μM EZ-Link and 0.5μM Fluorescein were created. Solutions with 200μM AP3B and 200μM EZ-Link with 10 mg/ml Salmon Sperm DNA were created as well and left to incubate for an hour. A tunable Spectra Physics MaiTai laser was used to excite the samples and the resulting fluorescence was collected using a photomultiplier tube. This was completed for wavelengths ranging from 680nm to 780nm for the 5 samples described earlier. The two-photon cross section was calculated from the measured fluorescence and the previously determined concentration, collection efficiency, quantum yields and fluorescein two-photon cross section.

Streptavidin Dye Staining Experiments:

U2OS 2-6-3 LacI-YFP #14 pTetOn C24 cells were grown in 10% FBS DMEM with 5mM IPTG present to prevent binding of the LacI-YFP to the transgene and promote normal cell growth. Before use, cells were passed and plated at 90% confluency on a gridded dish. 6 hours prior to an experiment IPTG was removed to allow the visualization of the

gene locus. An hour before an experiment cells were permeabilized with 0.2% Triton-X-100 in PBS. Cells were washed three times with 1x PBS before being incubated with 200µM AP3B for 1 hour. Cells were imaged on a LSM880 Zeiss combined confocal/multiphoton microscope. A 514nm laser line was used to locate the transgene locus which was then irradiated with a tunable Insight Laser. Individual irradiation doses were applied to specific grid squares. After suitable coverage of doses, cells were washed 3 times with 1mL PBS to ensure removal of unbound AP3B. Optionally, cells were incubated with 50mM 1,6-Hexanediol for 1 hour. 1,6-Hexanediol was washed out three times with 1mL PBS. Streptavidin-Dyelight488 (10µg/ml solution) was then added and allowed to incubate for 30 minutes. Unbound dye was washed out 3 times with 1mL PBS. Cells were reimaged looking for locations where binding of the dye was present.

Fluorescence Lifetime Measurements:

To measure fluorescence lifetime of AP3B we created a solution of 20µM AP3B with and without 10 mg/ml Salmon SpermDNA. Solutions were left for one hour, after which fluorescence lifetime curves were generated using a Zeiss LSM880 Confocal Multiphoton Microscope (u880). The tunable Spectra Physics Insight OPO laser tuned to 700nm provided ~100 fs pulsed excitation. Two-photon generated fluorescence was separated from the excitation using a 670 nm long pass dichroic filter (ZT670RDC, Chroma Technology Corp, VT), which directed the emission to a GaAsP photomultiplier tube after passing through a band-pass filter (ET440/30mm-2P, Chrome Technology Corp, VT). The laster power was attenuated using a near infrared (NIR) Acousto Optic Modulator (AOM) to keep the photon detection rate to less than 0.2% of the repetition rate to avoid photon pile-up. Time correlated photon counts were acquired using a TCSPC card (SPC-830,

Becker & Hickl GmbH). TCSPC data was fit a bi-exponential decay curve using the SPCImage software package (Becker & Hickl GmbH). The built-in Becker & Hickl instrument response function (IRF) was used for fitting.

2.5 References:

[1] Rana A, Dolmetsch RE. Using light to control signaling cascades in live neurons. Curr Opin Neurobiol 2010;20(5):617-22.

[2] Kaplan JH, Forbush B, Hoffman JF. Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts. Biochemistry 1978;17(10):1929-35.

[3] Emoto Y, Horiuti K, Tawada K, Yamada K. Tension relaxation induced by pulse photolysis of caged ATP in partially crosslinked fibers from rabbit psoas muscle. Proc Natl Acad Sci U S A 1995;92(5):1461-4.

[4] De Weer P, Salzberg B. (1986) Optical Methods in Cell Physiology (Society of General Physiologists Series). John Wiley & Sons Inc.

[5] Tsien RY, Zucker RS. Control of cytoplasmic calcium with photolabile tetracarboxylate 2-nitrobenzhydrol chelators. Biophys J 1986;50(5):843-53.

[6] Kaplan JH, Ellis-Davies GC. Photolabile chelators for the rapid photorelease of divalent cations. Proc Natl Acad Sci U S A 1988;85(17):6571-5.

[7] Dolphin AC, Wootton JF, Scott RH, Trentham DR. Photoactivation of intracellular guanosine triphosphate analogues reduces the amplitude and slows the kinetics of voltage-activated calcium channel currents in sensory neurons. Pflugers Arch 1988;411(6):628-36.

[8] Kaplan JH, Hollis R. External Na dependence of ouabain-sensitive ATP: ADP exchange initiated by photolysis of intracellular caged-ATP in human red cell ghosts. Nature 1980;288, 587–589

[9] Kaplan JH, Somlyo AP. Flash photolysis of caged compounds: New tools for cellular physiology. Trends in Neuroscience 1989;12(2):54-59.

[10] Lippincott-Schwartz J, Patterson GH. Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. Trends Cell Biol 2009;19(11):555-65.

[11] Patterson GH, Lippincott-Schwartz J. A photoactivatable GFP for selective photolabeling of proteins and cells. Science 2002;297(5588):1873-7.

[12] Verkhusha VV, Sorkin A. Conversion of the monomeric red fluorescent protein into a photoactivatable probe. Chem Biol 2005;12(3):279-85.

[13] Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, Salih A, Spindler KD, Nienhaus GU. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proc Natl Acad Sci U S A 2004;101(45):15905-10.

[14] Chudakov DM, Lukyanov S, Lukyanov KA. Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. Nat Protoc 2007;2(8):2024-32.

[15] Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. Proc Natl Acad Sci U S A 2002;99(20):12651-6.

[16] Chudakov DM, Belousov VV, Zaraisky AG, Novoselov VV, Staroverov DB, Zorov DB, Lukyanov S, Lukyanov KA. Kindling fluorescent proteins for precise in vivo photolabeling. Nat Biotechnol 2003;21(2):191-4.

[17] Kim PK, Mullen RT, Schumann U, Lippincott-Schwartz J. The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER. J Cell Biol 2006;173(4):521-32.

[18] Ferenz NP, Wadsworth P. Prophase microtubule arrays undergo flux-like behavior in mammalian cells. Mol Biol Cell 2007;18(10):3993-4002.

[19] Wiedenmann J, Nienhaus GU. Live-cell imaging with EosFP and other photoactivatable marker proteins of the GFP family. Expert Rev Proteomics 2006;3(3):361-74.

[20] Wacker SA, Oswald F, Wiedenmann J, Knöchel W. A green to red photoconvertible protein as an analyzing tool for early vertebrate development. Dev Dyn 2007;236(2):473-80.

[21] Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging intracellular fluorescent proteins at nanometer resolution. Science 2006;313(5793):1642-5.

[22] Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods 2006;3(10):793-5.

[23] Lukyanov KA, Chudakov DM, Lukyanov S, Verkhusha VV. Innovation: Photoactivatable fluorescent proteins. Nat Rev Mol Cell Biol 2005;6(11):885-91.

[24] Dolmans DE, Fukumura D, Jain RK. Photodynamic therapy for cancer. Nat Rev Cancer 2003;3(5):380-7.

[25] Dougherty TJ, Lawrence G, Kaufman JH, Boyle D, Weishaupt KR, Goldfarb A. Photoradiation in the treatment of recurrent breast carcinoma. J Natl Cancer Inst 1979;62(2):231-7.

[26] Bown SG, Rogowska AZ, Whitelaw DE, Lees WR, Lovat LB, Ripley P, Jones L, Wyld P, Gillams A, Hatfield AW. Photodynamic therapy for cancer of the pancreas. Gut 2002;50(4):549-57.

[27] Barr H, Krasner N, Boulos PB, Chatlani P, Bown SG. Photodynamic therapy for colorectal cancer: a quantitative pilot study. Br J Surg 1990;77(1):93-6.

[28] Zhang P, Wu MX. A clinical review of phototherapy for psoriasis. Lasers Med Sci 2018;33(1):173-80.

[29] Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. Nat Rev Drug Discov 2005;4(4):307-20.

[30] Dai Z, Wang Z. Photoactivatable Platinum-Based Anticancer Drugs: Mode of Photoactivation and Mechanism of Action. Molecules 2020;25(21):E5167.

[31] Oun R, Moussa YE, Wheate NJ. The side effects of platinum-based chemotherapy drugs: a review for chemists. Dalton Trans 2018;47(19):6645-53.

[32] Wang X, Wang X, Jin S, Muhammad N, Guo Z. Stimuli-Responsive Therapeutic Metallodrugs. Chem Rev 2019;119(2):1138-92.

[33] Deans AJ, West SC. DNA interstrand crosslink repair and cancer. Nat Rev Cancer 2011;11(7):467-80.

[34] Hearst JE. Psoralen photochemistry and nucleic acid structure. J Invest Dermatol 1981;77(1):39-44.

[35] Marker SC, MacMillan SN, Zipfel WR, Li Z, Ford PC, Wilson JJ. Photoactivated in Vitro Anticancer Activity of Rhenium(I) Tricarbonyl Complexes Bearing Water-Soluble Phosphines. Inorg Chem 2018;57(3):1311-31.

[36] Koike K, Okoshi N, Hori H, Takeuchi K, Ishitani O, Tsubaki H, Clark IP, George MW, Johnson FP, Turner JJ. Mechanism of the photochemical ligand substitution reactions of fac-[Re(bpy)(CO)(3)(PR(3))](+) complexes and the properties of their triplet ligand-field excited states. J Am Chem Soc 2002;124(38):11448-55.

[37] Darensbourg DJ, Ortiz CG, Kamplain JW. A New Water-Soluble Phosphine Derived from 1,3,5-Triaza-7-phosphaadamantane (PTA), 3,7-Diacetyl-1,3,7-triaza-5-phosphabicyclo[3.3.1]nonane. Structural, Bonding, and Solubility Properties. Organometallics 2004;23(8)1747-1754.

[38] Koike K, Tanabe J, Toyama S, Tsubaki H, Sakamoto K, Westwell JR, Johnson FP, Hori H, Saitoh H, Ishitani O. New synthetic routes to biscarbonylbipyridinerhenium(I) complexes cis,trans-[Re(X2bpy)(CO)2(PR3)(Y)n+ (X2bpy = 4,4'-X2-2,2'-bipyridine) via photochemical ligand substitution reactions, and their photophysical and electrochemical properties. Inorg Chem 2000;39(13):2777-83.

[39] Hearst JE. Photochemistry of the psoralens. Chem Res Toxicol 1989;2(2):69-75.

[40] Besaratinia A, Yoon JI, Schroeder C, Bradforth SE, Cockburn M, Pfeifer GP. Wavelength dependence of ultraviolet radiation-induced DNA damage as determined by laser irradiation suggests that cyclobutane pyrimidine dimers are the principal DNA lesions produced by terrestrial sunlight. FASEB J 2011;25(9):3079-91.

[41] Wielenberg, K., et al., An Improved 4’-Aminomethyltroxsalen-Based DNA Crosslinker for Biotinylation of DNA. bioRxiv, 2020;2020.02.29.971317.

[42] Berezin MY, Achilefu S. Fluorescence lifetime measurements and biological imaging. Chem Rev 2010;110(5):2641-84.

[43] Shivji KK, Kenny MK, Wood RD. Proliferating cell nuclear antigen is required for DNA excision repair. Cell 1992;69(2):367-74.

[44] Eaton DF. Reference Materials for Fluorescence Measurements. Pure and Applied Chemistry. 1988;60:1107

[45] Nagy PI. Competing intramolecular vs. intermolecular hydrogen bonds in solution. Int J Mol Sci 2014;15:19562-633.

Chapter 3

Femto-Seq: A Novel Method to Examine Genomic Contacts of a Genomic Locus

3.1 Overview:

Previously I have discussed both the importance of understanding gene regulation and numerous mechanisms through which this regulation occurs. In particular, I have focused on how spatial organization of the genome can affect gene expression regulation. I have described several techniques that have been used to examine these types of interactions including both imaging based and biochemical techniques. These techniques have provided a wealth of valuable data looking at a variety of different systems; however, research is still being done to both improve these techniques and create novel techniques that overcome some of the shortcomings of these older methods. In this chapter, I will discuss several of these newer methods. I will then present work aimed at the creation of a novel technique called Femto-Seq. This technique combines two-photon excitation microscopy with the previously discussed photoactivatable crosslinking compounds to examine genomic contacts in a small volume of the nucleus. We utilize the small excitation volume provided by two-photon excitation microscopy to activate the crosslinking compound in a small volume in the nucleus (on the order of a femtoliter). Enriching for the crosslinking compound itself allows us to enrich for all the genomic DNA in that small volume. The DNA can then be analyzed by sequencing to provide a snapshot of the sequences in that small volume. While I have previously described characterization of photoactivable crosslinkers, in this chapter I will describe trial runs of the Femto-Seq method using a model system and potential applications of such a method.

3.2. Improvements in examining genomic architecture:

Previously I have discussed a few common imaging techniques that can be used to examine genomic contacts. Fluorescence in situ hybridization (FISH) uses short fluorescently labeled DNA oligos. This technique requires both permeabilization of the cell using methanol to allow probe strands to enter the nucleus and denaturation of the strands using heat and formamide (Figure 3.1c). In order to examine genomic contacts, individual loci must be labeled. Genomic contacts are determined semi-arbitrarily by looking at the distance between fluorescently labeled loci. Colocalization usually corresponds to a genomic contact, but the spatial distance of the signals upon colocalization can vary from 50nm to 1µm depending on the probe and detection system set up.

3.2.1 Improvements made to Fluorescence in Situ Hybridization:

Several important considerations must be made in FISH (Figure 3.1b) experiments to examine genomic contacts. Typically, DNA probes are designed to cover a large swath of the genome (up to an entire chromosome). While this allows researchers to look at large scale organizational motifs, looking at smaller more subtle genomic contacts may not be able to be detected without special consideration. Likewise, the microscope being used must be able to distinguish fluorescent spots well enough to ensure that genomic contacts are actual events instead of just random events. Finally, the harsh treatment in order to insert the probes into the nucleus as well

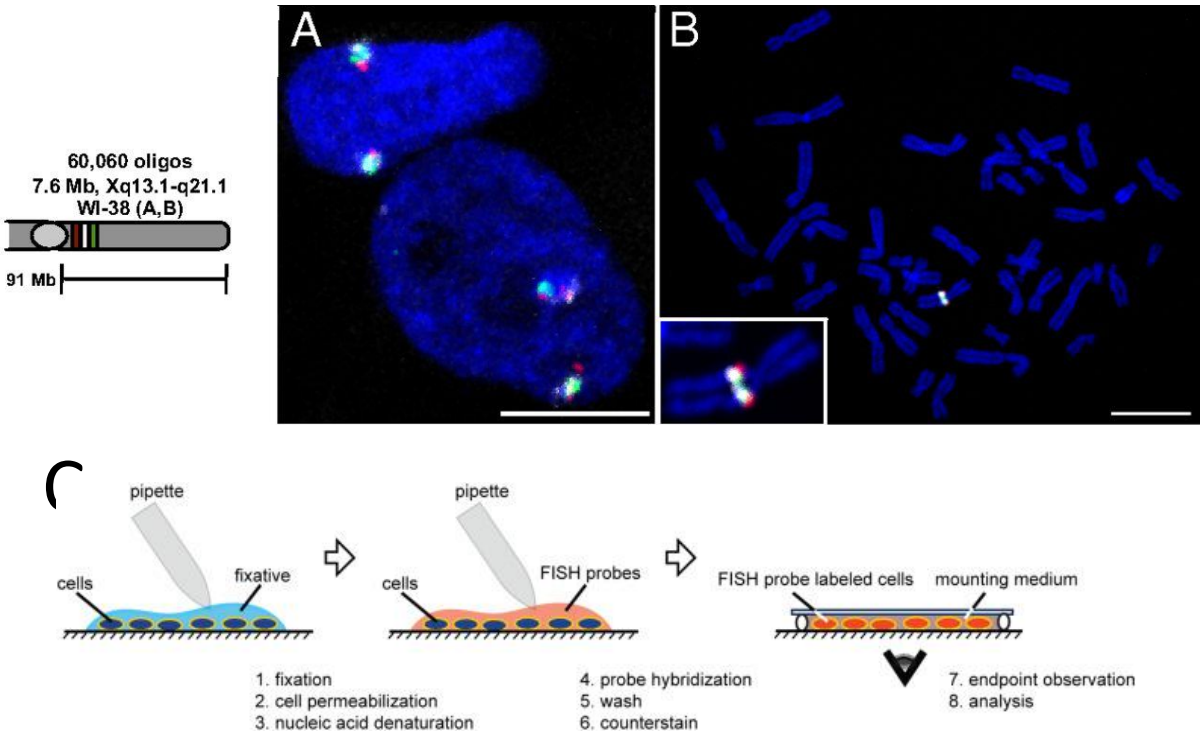


Figure 3.1: Example of FISH Imaging.

(a) and (b) Multicolor FISH using Oligopaints. Human chromosome regions Xq13.1, Xq13.2, Xq.13.3-q21.1 were targeted to perform FISH. Cells were in interphase (a) or metaphase (b). (c) Illustration showing traditional FISH. Cells are first fixed, then permeabilized. Nucleic acids are denatured to allow probes to hybridize. This is followed by washing steps and imaging. Adapted from [29] and [30].

as the denaturation procedure may change the organization of the genome. In order to overcome some of these issues several improvements have been made to FISH. Cyro-FISH utilizes fixed cryosections to maintain nuclear architecture while allowing high resolution electron microscopy imaging [1]. Cells are fixed for electron microscopy, placed in a cryoprotectant and froze in liquid nitrogen. Cells are then sectioned and after probe introduction are imaged. CryoFISH has been used to examine chromosome territories and the positioning of transcription factories relative to these territories [2]. It has also been used to examine long range chromatin interactions [3]. Other improvements include improvements made to the selection and generation of oligo probes. COMBO-FISH utilizes genome database searching to select several oligos that are chosen in such a way to colocalize at a given target [4]. Oligopaints (Figure 3.1a), utilizes an oligo library as opposed to segments of purified genomic DNA to create short fluorophore tagged oligos for use in FISH experiments [5]. These easier to handle oligos have allowed researchers to utilize super resolution microscopy to visualize single copy genomic regions in Drosophila [6].

3.2.2 Improvements made to Fluorescently Labeled DNA Binding Proteins:

Live cell imaging requires additional considerations and usually utilizes fluorescently labeled DNA binding proteins. A key consideration when utilizing fluorescently labeled DNA binding proteins is signal to noise ratio. Since fluorescently labeled DNA binding proteins will be expressed in bulk, a sufficient number of them

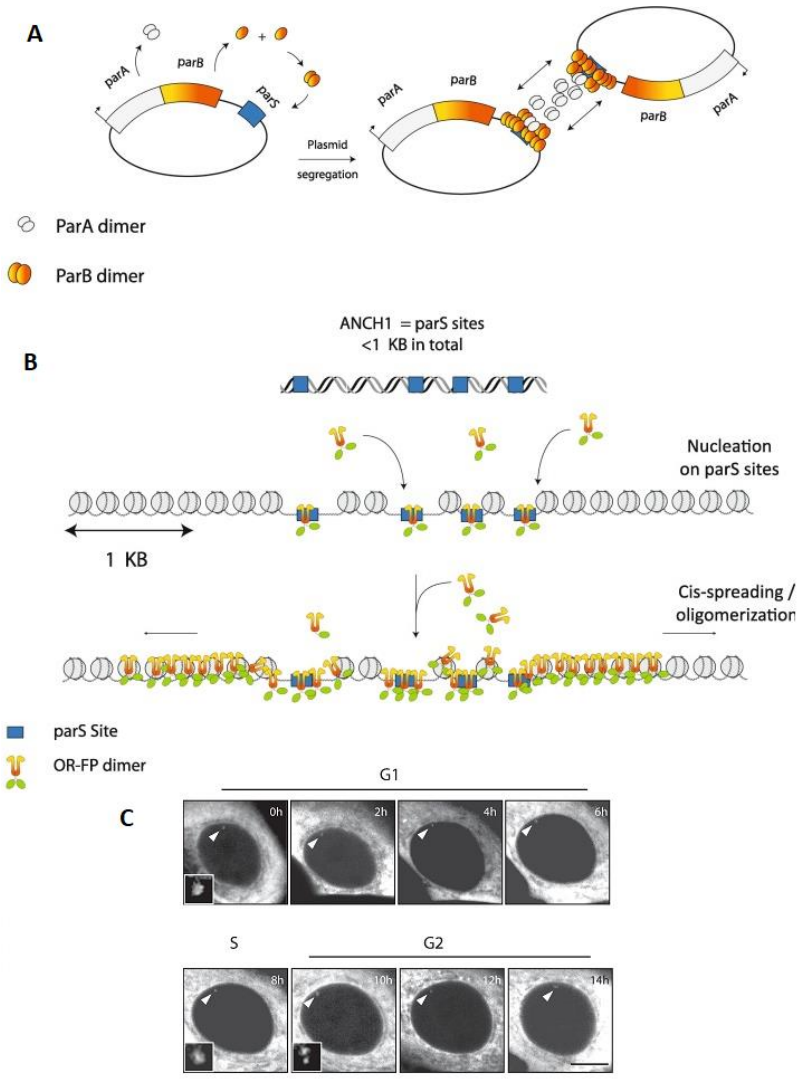


Figure 3.2: ANCHOR binding system to examine genomic loci.

(a) Overview of the plasmid partitioning system. Upon replication, ParB binds to ParS and spreads. (b) The ANCHOR labeling system. ANCH1 is inserted into the genome at a locus of interest. ANCH1 consists of several ParS sites spread over a 1kb DNA sequence. The sites recruit a ParB-FP to fluorescently label the site. The ParB-FP conjugate spreads over nearby chromatin to the ANCH1 site. (c) MCF7 cells have a CCND1 transgene labeled using the ANCHOR system. Adapted from [9].

must be localized at the gene locus of interest to allow one to visualize the locus against the background of unbound fluorescently labeled proteins. Numerous approaches can solve this problem. In the past, this required inserting unique sequences into the genome at the gene locus of interest to allow a sufficient number of fluorescently labeled DNA binding proteins to bind the targeted site. By inserting Lac operon along with a transgene, researchers were able to visualize the transgene locus using a fluorescently labeled Lac repressor [7]. A similar approach has been taken using the Tet operator-repressor system [8]. Other methods utilize split or recruitment-based approaches to reach sufficient signal to background. The ANCHOR system (Figure 3.2) utilizes the ParABS partitioning system from bacterial chromosomes [9]. ParS sites are inserted over a 1kb region of the genome. Fluorescently labeled ParB is recruited to these sites. After recruitment, additional fluorescently labeled ParB is recruited to the site through protein-protein interactions. These further recruited ParB proteins nonspecifically bind DNA near the original recruitment site. This spreading effect allows enough fluorescently labeled ParB to be localized at the locus of interest to provide visualization [9]. While these approaches have proven to be fruitful, the insertion of DNA sequences into the genome can affect some small-scale interactions that can occur.

While these previous approaches have utilized fluorescently labeled DNA binding proteins that bind a specific DNA sequence, the use of engineerable DNA binding proteins is of particular interest. There are several proteins that have been shown to be particularly useful for this purpose. Transcription Activator Like Effectors (TALE)

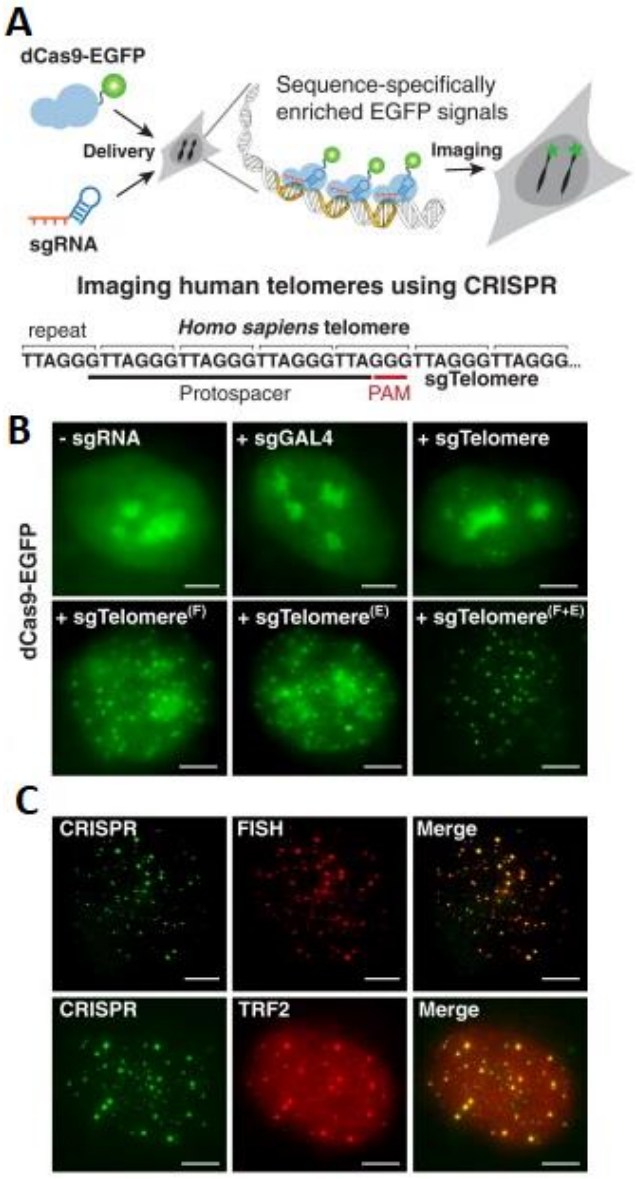


Figure 3.3: Use of Fluorescently labeled dCas9 to visualize genomic loci.

(a) Overview of the dCas9 system for use as a tool to examine genomic loci. dCas9 is fluorescently labeled and bound to a genomic locus of interest using a specific sgRNA. Multiple dCas9 are required to be bound for sufficient signal to noise. (b) CRISPR imaging of human telomeres. Two different sgRNA designs were utilized with sgGAL4 as a negative control. (c) Co-labeling of telomeres using dCas9-eGFP, PNA-FISH and TRF2. Adapted from [10]

proteins are DNA binding proteins with a programmable central repeat domain. These proteins can be engineered to bind a particular DNA sequence. A more thorough discussion of TALE proteins can be found in the next chapter.

The clustered regularly interspaced short palindromic repeats (CRISPR) Cas9 system is one of the most interesting recent developments in the gene editing and genomic labeling field. This system is used in prokaryotes to provide resistance to foreign viruses and bacteria. The prokaryotes capture small snippets of DNA which are stored as a CRISPR array. RNA segments are produced from these repeats allowing the Cas9 protein to target the foreign body's DNA. In order to use this system for imaging purposes, a deactivated Cas9 protein (dCas9) is used to prevent cleaving of the targeted DNA. A guide RNA (gRNA) is used to specifically target a gene locus of interest. The gRNA consists of a scaffold sequence necessary to bind the Cas9 protein and the 20bp spacer that defines the target binding location. The target location must also contain a Protospacer Adjacent Motif adjacent to the binding site.

Fluorescently labeling the dCas9 complex has two approaches. The dCas9 protein itself can be fluorescently labeled or, alternatively, the gRNA can be fluorescently labeled. Fluorescently labeling the dCas9 protein has been used to visualize repetitive DNA sequences. In living human cells eGFP-dCas9 has been used to visualize telomeres (Figure 3.3) and repeats in the Muc4 gene using optimized sgRNAs [10]. Likewise, researchers have looked at the optimal number of sgRNA required to provide sufficient signal to noise ratio to visualize non- repetitive sequences

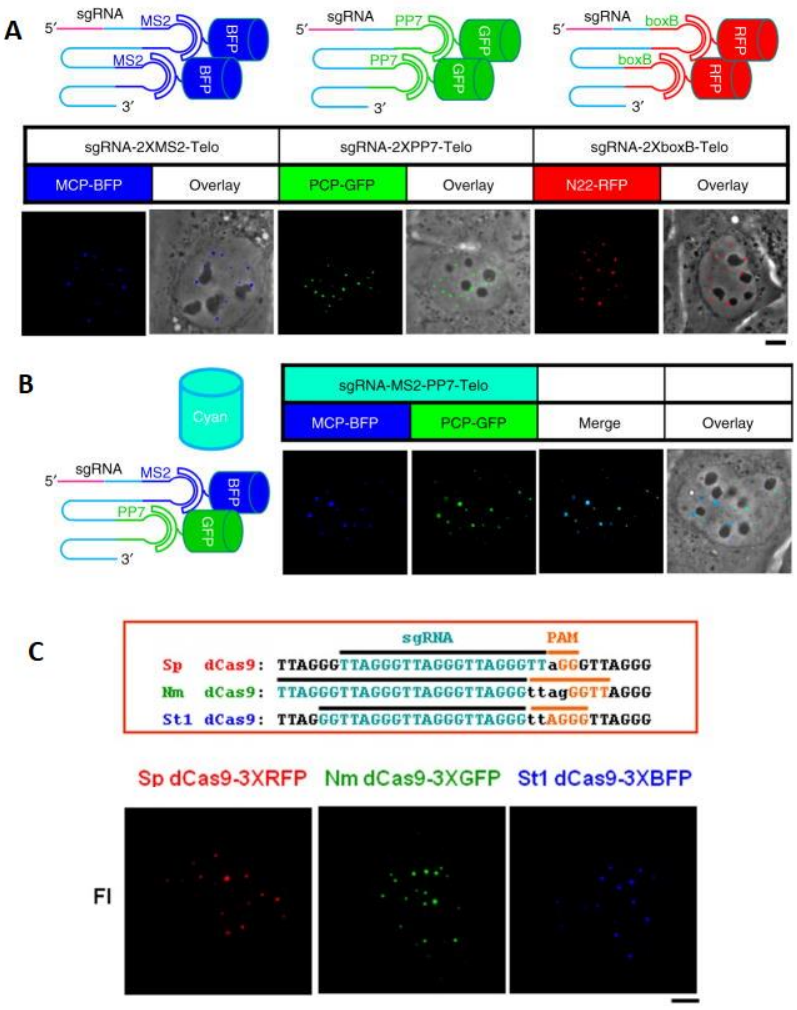


Figure 3.4: Variations of dCas9 labeling of genomic loci.

(a) Labeling of sgRNA rather than the dCas9 protein. The sgRNA includes hairpin domains such as MS2. These domains recruit a fluorescently labeled RNA binding protein such as MCP-BFP. Different hairpin and binding proteins can be used to provide a dCas9 system with multicolor capability. (b) Example of CRISPRainbow. A similar set up to (a) is used in which sgRNA is fused with hairpins. By utilizing different combinations of hairpins, several additional color combinations can be utilized to allow additional options of colors. (c) Fluorescently labeled dCas9 orthologs give researchers the capability to label different genomic loci simultaneously in different colors. Adapted from [13] and [14].

in the Muc4 gene [10]. The benefits of using live cell imaging are shown by tracking genomic loci through the cell cycle [10]. Further improvements to this technique include the optimization of this technique for use in mouse embryonic stem cells [12]. The major shortcoming of fluorescently labeling dCas9 is that only a single locus is able to be imaged at one time. However, several techniques are able to circumvent this issue. Researchers have utilized Cas9 orthologs to label distinct loci (Figure 3.4C) [11, 13]. Cas9 orthologs were labeled different colors and their different PAM specificities were used to target different genomic loci. Another approach to circumvent the limited throughput of labeling the dCas9 protein involves labeling the sgRNA. By fusing hairpins MS2, PP7, and boxB to stem loops of different sgRNA molecules, fluorescently labeled proteins bound to RNA hairpin binding domains were recruited to the sgRNA labeling the targeted locus (Figure 3.4a,b) [14]. By fusing different combinations of hairpins, researchers were able to not only utilize primary colors, but also secondary colors to increase the labeling throughput of this method. Finally, improvements have been made in the labeling of non-repetitive genomic loci. The chimeric array of gRNA oligonucleotides (CARGO) technique utilizes a molecular assembly strategy and chimeric array of gRNA to insert 12 different gRNA into a cell. This technique allowed labeling of 2kb of non-repetitive DNA [15].

3.2.3 Improvements to Chromosome Capture Techniques:

As described previously, techniques utilizing imaging are only one approach used to look at nuclear organization. While the previously described chromosome capture techniques have existed for decades numerous improvements continue to be made to them.

Several improvements have been made to the Hi-C technique itself. DNase Hi-C utilizes DNase I instead of restriction enzymes to fragment the cross-linked chromatin (Figure 3.5a). This leads to mixed fragment ends rather than the consistent ends generated by using restriction enzymes. Researchers used this technique on hESCs and K562 cells to validate that data generated from DNase Hi-C was consistent with previous Hi-C studies. While these two methods proved consistent, DNase Hi-C showed better genome coverage when compared to traditional Hi-C. DNase Hi-C also shows potential for overcoming the resolution limit present while using restriction enzymes [16].

In-Situ Hi-C is similar to traditional Hi-C but performs the proximity ligation in intact nuclei (Figure 3.5b). Following crosslinking by formaldehyde, cells are permeabilized leaving the nuclei intact. DNA cutters are introduced fragmenting the chromatin. Overhangs are filled in (including a biotinylated nucleotide) and blunt ends ligated together. The DNA is sheared and pulled down using streptavidin beads then analyzed with paired end sequencing. Performing this in an intact nucleus provides a variety of benefits. In-Situ Hi-C reduces illegitimate contacts that are usually obtained by doing ligations in a dilute solution. The In-Situ Hi-C protocol is also faster than traditional Hi-C and enables higher resolution by allowing the use of 4-cutters rather than 6-cutters to fragment DNA [17].

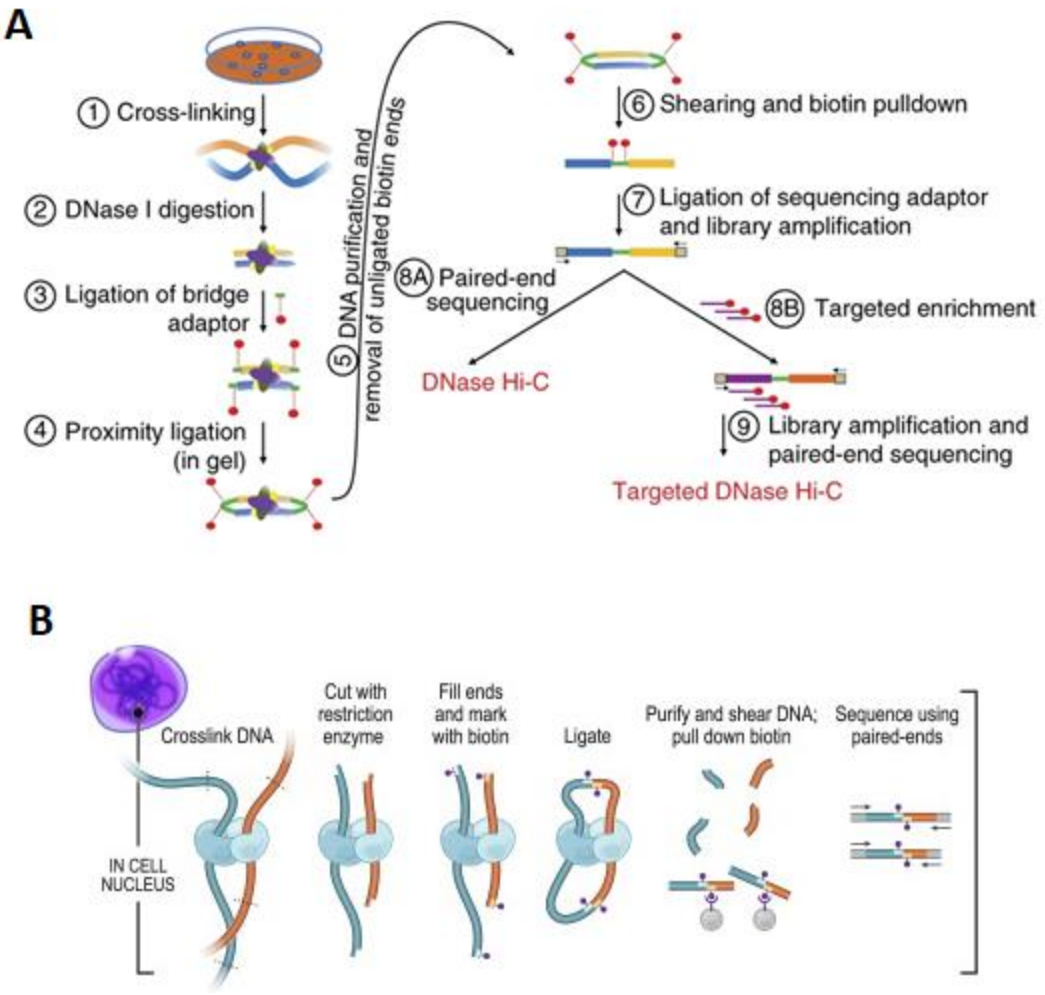
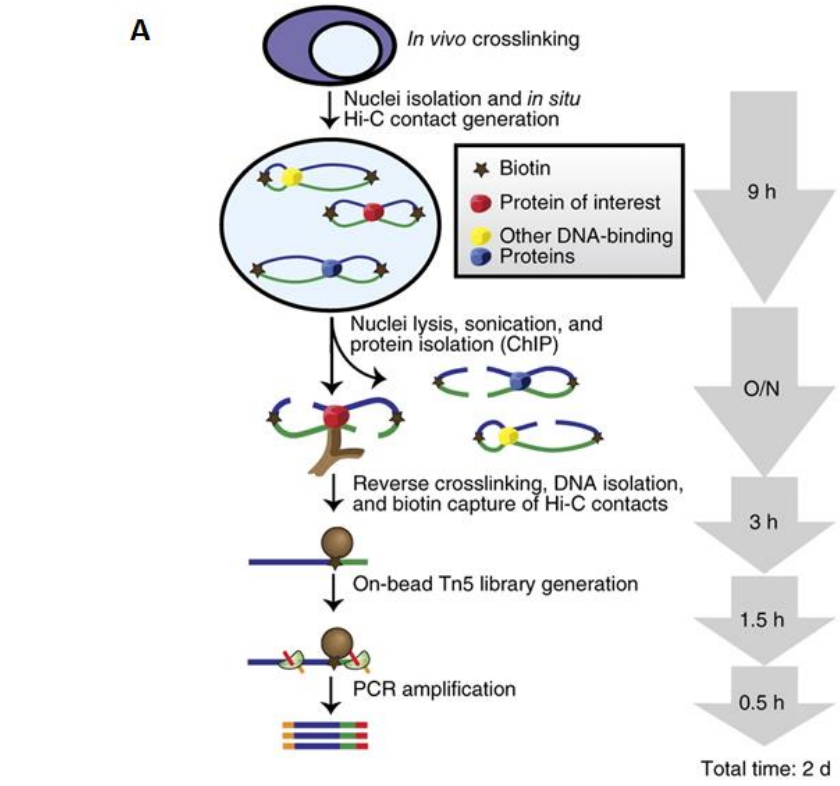


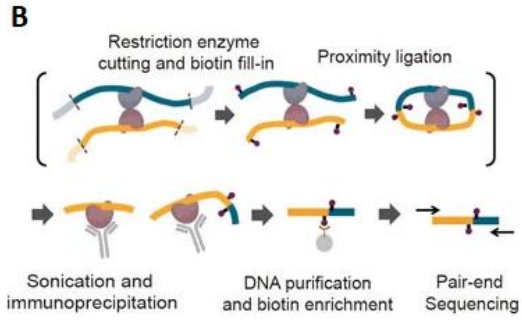
Figure 3.5: Improvements of Hi-C.

Example of two improved methods over traditional Hi-C. (a) Overview of DNase Hi-C. Traditional Hi-C utilizes a restriction enzyme to fragment cross-linked chromatin. DNase Hi-C replaces this step with a DNase I digestion. This allows for better genome coverage and avoids the resolution limit placed by restriction enzymes. (b) In-Situ Hi-C is very similar to traditional Hi-C but takes place entirely in the nucleus. This is done to avoid extraneous crosslinks that can form in solution ligations. Adapted from [16] and [17].

Other improvements to Hi-C experiments involve limiting the number of required cells. Since the Hi-C protocol traditionally studies a population of cells, examination of interesting contacts in individual or a subpopulation of cells proves difficult. Strategies to isolate individual or small populations of cells have been developed to solve this issue. Single cell Hi-C uses In-Situ Hi-C followed by separating nuclei into individual tubes where the nuclei undergo cross link reversal, DNA shearing, streptavidin bead pulldown and library preparation. While this technique showed some success, notably genome coverage was limited with 2.5% coverage at best [18]. As an alternative to separating out individual nuclei, single-cell combinatorial indexed Hi-C utilizes sequential barcoding prior to ligation to ensure each nucleus is tagged by a unique barcode combination. After sequencing, both population and single cell data can be analyzed [19].

Finally, improvements to chromosome capture techniques have come in the form of including chromatin immunoprecipitation techniques. Chromatin Immunoprecipitation (ChIP) when combined with chromosome capture technologies can be used to look at how genomic contacts may be regulated by transcription factors or chromatin modifications. The development of techniques such as ChIP-Seq [20] and Chromatin Immunoprecipitation with paired-end ditag (ChIA-Pet) [21] have led researchers to explore combining Hi-C with chromatin immunoprecipitation techniques [21]. Proximity Ligation-Assisted Chromatin Immunoprecipitation Sequencing (PLAC-Seq) is similar to ChIA-PET, but ligates prior to shearing (Figure 3.6b) [23]. Hi-ChIP utilizes in-situ Hi-C to ensure long range DNA contacts are captured in the nucleus (Figure 3.6a). Then, ChIP is performed further enriching for contacts mediated by a protein of interest [22].





Statistic	Pol II PLAC-seq	Pol II ChIA-PET[14]
Cell number	5M	100M
Total pairs	175M	16M
PCR duplication	30%	44%
cis pairs (%)	89%	52%
long-range *cis* pairs	67% (>10kb)	8.6% (>8kb)
Processing time	3-4 Days	6 Days

Figure 3.6: ChIP based techniques to study chromatin contacts.

(a) Overview of Hi-ChIP. Similar to Hi-C, chromatin is crosslinked in the nucleus. Following in-situ ligation, ChIP for a protein of interest is performed. After crosslink reversal, cleanup and sequencing protein mediated contacts are able to be determined. (b) PLAC-Seq overview. By reversing the ligation and shearing steps of ChIA-PET, researchers were able to gain much higher accuracy and efficiency. Adapted from [22] and [23].

3.2.4 Non-Ligation Based Techniques to Examine Genomic Contacts:

The aforementioned methods all utilize crosslinking and ligation to examine genomic contacts, but a variety of methods have been created to look at contacts without utilizing ligation. Genomic Architecture Mapping (GAM) utilizes cryo-sectioning to create random thin slices of the nucleus (Figure 3.7a). These slices are then isolated and undergo a DNA extraction. The DNA is amplified using whole genome amplification and ligated to indexed sequencing adapters. Following this, sequencing is then performed. The presence of genomic contacts is determined by examining the frequency that two genomic sequences are found in the same nuclear slice. GAM as a method offers a number of advantages over ligation-based methods to examine genomic contacts. GAM requires a small number of cells relative to the amount needed for ligation-based methods. GAM is able to examine higher multiplicity contacts rather than the pairwise interactions that ligation-based methods have been limited to in the past [24].

Tyramide Signal Amplification Sequencing (TSA-seq) is another ligation free method to examine genomic contacts. TSA-seq utilizes an antibody coupled Horseradish Peroxidase (HRP) to generate biotin-tyramide free-radicals in a nuclear compartment of interest (Figure 3.7b). These radicals form covalent bonds with nearby molecules, effectively biotinylating sequences nearby the HRP molecule. These sequences can then be enriched and analyzed. The exponential decay of radicals is used as a “ruler” to determine relative distances from discretely labeled nuclear

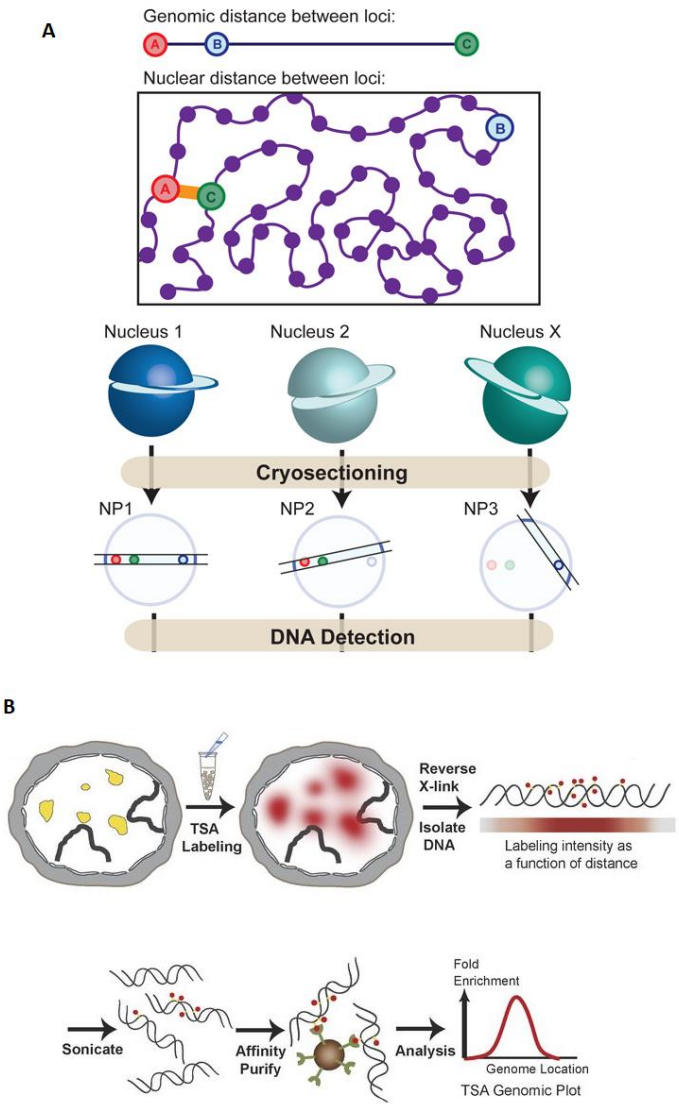


Figure 3.7: Non-ligation techniques to look at chromosome contacts.

(a) Overview of Genomic Architecture Mapping (GAM). This process cryosections nuclei into very thin slices. The chromatin in each slice is then sequenced. Since genomic loci that are located spatially near each other will be more likely to show up in the same nuclear section, this technique can be used to map 3D contacts. (b) Overview of TSA-Seq. Biotin-Tyramide free radicals are generated in a nuclear volume of interest, labeling DNA located there. Enriching for the Biotinylated DNA provides a snapshot of sequences from the nuclear volume. Adapted from [24] and [25].

compartments [25]. TSA-Seq has been used to look at genomic contacts around splicing speckles. DamID is a similar method to TSA-seq. It utilizes a DNA adenine methyltransferase (DAM) attached to a DNA binding protein of interest. DAM specifically methylates adenines in GATC sequences near the binding site of the DNA binding protein of interest. Genomic DNA is extracted and the methylated GATC sites are enriched. This provides a snapshot of sequences spatially near the DNA binding protein of interest. DamID has been used to study Lamina Associated Domains located near the nuclear periphery [26].

Split-pool recognition of interactions by tag extension (SPRITE) is a proximity ligation free method to study chromatin interactions (Figure 3.8a). Similar to chromatin conformation capture techniques, chromatin is first crosslinked. Nuclei are isolated and the chromatin are fragmented. The crosslinked chromatin fragments are split across a 96-well plate where a specific tag sequence is ligated on. This process is repeated several times. Because chromatin interactions are crosslinked together, they are sorted into the same pools every cycle. After crosslink reversal and sequencing the unique barcode tag previously ligated on can be used to determine which sequences were crosslinked together. The proximity ligation free nature of this technique allows it to examine higher order interactions rather than the traditional pairwise interactions examined with Hi-C. Another unique facet of this technique is its ability to study RNA interactions [27].

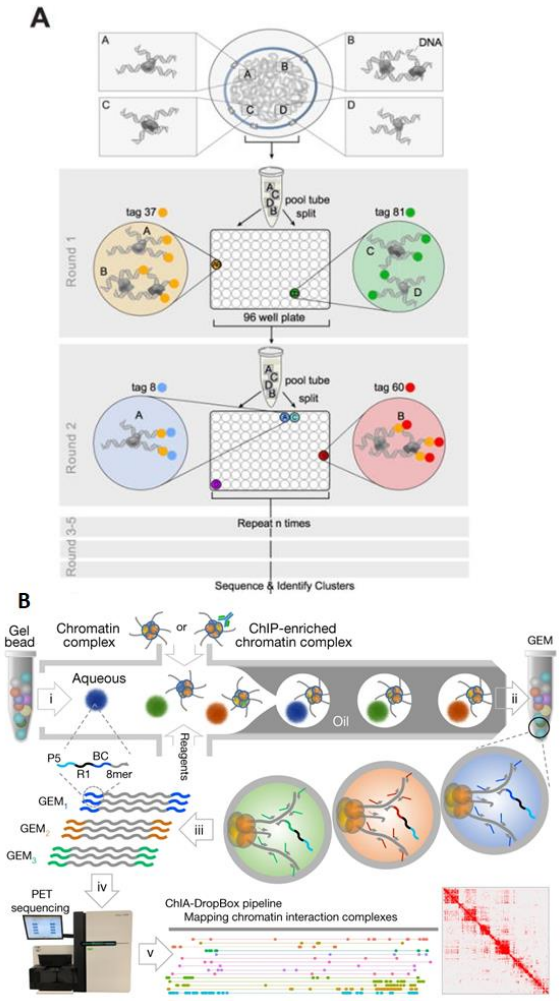


Figure 3.8: Pooling Approaches to examining Chromatin Contacts.

(a) Overview of Split-pool recognition of interactions by tag extension (SPRITE). After crosslinking, chromatin fragments are split into several pools where a specific tag sequence is ligated on. This process is repeated several times before sequencing. The ligated tag sequences are used to determine what sequences were originally crosslinked together. (b) Overview of Chromatin-interaction analysis via droplet-based and barcode linked sequencing (ChIA-Drop). After crosslinking and fragmenting, microfluidics is used to load chromatin complexes into droplets. The droplets are used to barcode the chromatin complexes. After sequencing, the barcodes from the droplets are used to identify crosslinked sequences. Adapted from [27] and [28].

Lastly, Chromatin-interaction analysis via droplet-based and barcode linked sequencing (ChIA-Drop) utilizes microfluidics to examine chromatin interactions (Figure 3.8b). Chromatin is crosslinked and fragmented then loaded into a microfluidic device. This device partitions chromatin complexes in a gel-bead-in-emulsion droplet. This droplet contains a unique DNA barcode along with ligation and amplification reagents. Following barcoding and amplification, chromatin is pooled together and analyzed using high throughput sequencing. The barcode is used to identify what sequences were located in the same droplet and thus what chromatin fragments were originally ligated together. Since this is not a ligation-based approach like Hi-C, this technique can be used to detect non-pairwise interactions. [28]

With the benefits and limitations of these methods in mind we sought to develop a novel method to examine genomic contacts of a genomic locus. Many of these techniques are ensemble techniques. Specific sub population information may be lost without prior screening. While examination of pairwise interactions has been accomplished since the creation of 3C, examination of higher order interactions is not accomplished in traditional Hi-C. Many of these techniques rely on formaldehyde crosslinking or other chemistry to link together nearby chromatin fragments. While this does provide a statistically determined map of chromatin contacts, it does not measure real distances between them. Rather it presents likelihoods that contacts would be found linked together. Finally, many of these techniques lack the ability to analyze specific nuclear volumes, but rather examine sequences located around a genomic locus or sequences associated with specific proteins.

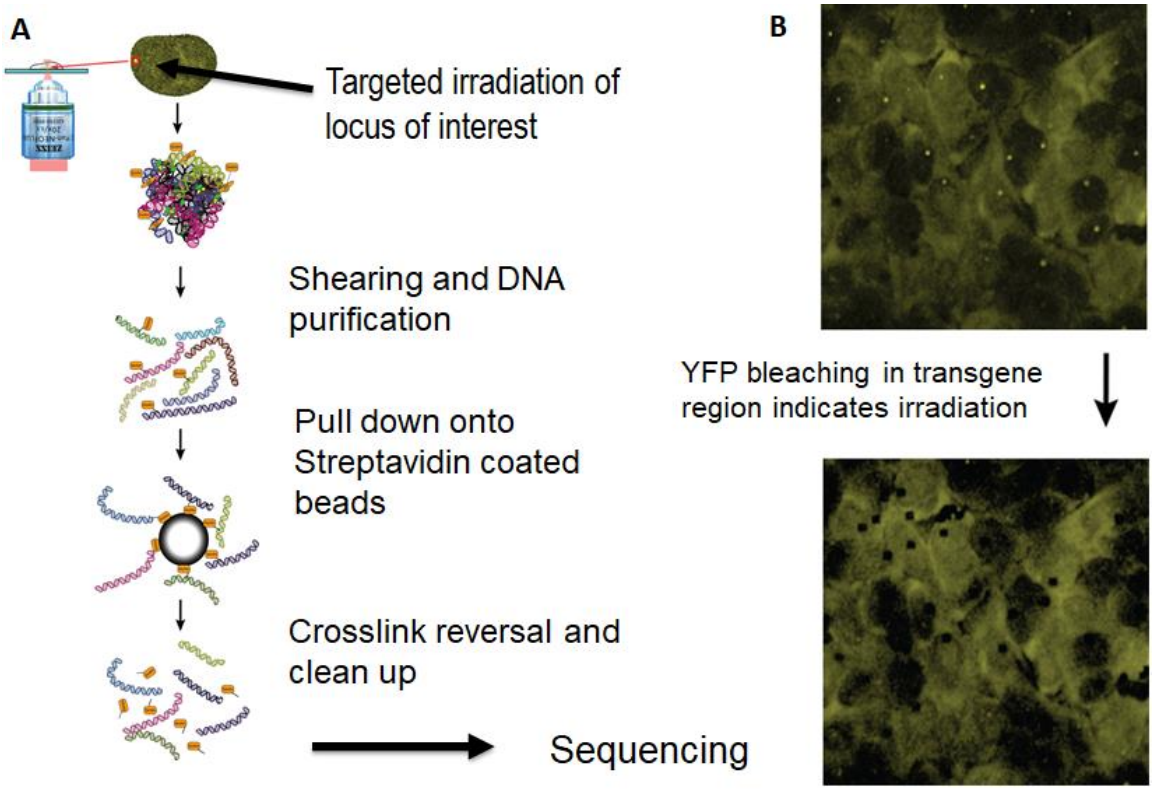


Figure 3.9: Overview of Femto-Seq.

(a) Overview of the Femto-Seq technique. A genomic volume of interest is fluorescently labeled. A photoactivatable crosslinker with a biotin tag is then added. 3D localized two-photon excitation is used to excite the crosslinker around the volume of interest. DNA is purified and the crosslinked DNA is enriched via streptavidin bead pulldown. After crosslink reversal and clean up, sequencing is performed. Since we enrich for the chromatin from the volume of interest, a snapshot of sequences from the volume of interest is obtained. (b) Example images before and after targeted irradiation. A transgene is marked by punctate YFP spots and after targeted two-photon excitation clear bleaching patterns at the loci are seen.

3.3 Development of a Novel Method to Examine Genomic Contacts:

Our method, which we call Femto-Seq, is a novel method to look at genomic sequences in a specific volume of the nucleus. Similar to 4C, this method can look at genomic contacts around a specific genomic locus; however, it is not limited to that. Femto-Seq is able to look at any nuclear volume that can be fluorescently labeled. Femto-Seq does not require formaldehyde crosslinking of chromatin fragments. Instead, it utilizes photoactivatable compounds which, as discussed in the previous chapter, can provide spatial resolution. Femto-seq is performed using a two-photon microscope which allows examination of specific sub-populations that may be missed in ensemble-based measurements.

The experimental process of Femto-Seq is as follows (Figure 3.9a). Prior to the experiment the nuclear volume of interest must be fluorescently labeled (Figure 3.9b). Many methods can accomplish this and have been discussed previously. After labeling a nuclear volume of interest, nuclei are incubated with a psoralen based crosslinker with an affinity tag. The previously discussed AP3B and EZ-link are two examples of compounds suitable for this purpose, as they contain a psoralen like crosslinker and an affinity tag. After allowing this compound to intercalate, nuclei are imaged on a two-photon microscope. The photoactivatable crosslinker is selectively activated in the region of interest with two-photon excitation using the fluorescent label as a target. This process covalently bonds the photoactivatable crosslinker with an affinity tag to only the DNA located in the irradiated volume. Chromatin is then extracted and sheared. Chromatin from the nuclear volume of interest is enriched by performing a bead

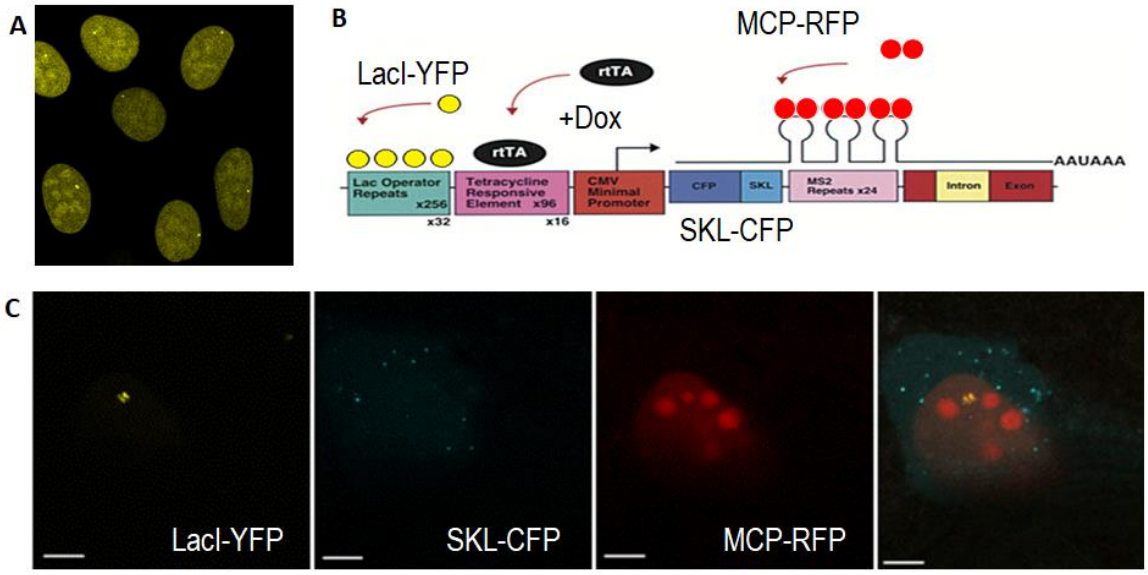


Figure 3.10: Overview of U2OS 2-6-3 LacI-YFP pTetOn Cell Line

(a) Punctate spots are clearly seen; these spots mark the location of the transgene. (b) Illustration of the transgene. Upstream of the transgene several Lac Operator repeats are present. The cells stably express LacI-YFP which binds to these repeats. The cell line is Doxycycline inducible via the Tetracycline Responsive Element. The transgene encodes a CFP-SKL protein. The fluorescent protein CFP is fused to a peroxisome targeting signal to allow visualization of the active transgene. The mRNA produced by the transgene contains numerous MS2 repeats which allows visualization of transcription by utilizing a MCP fluorescent protein. (c) Transgene was activated. LacI-YFP clearly labels the location of the transgene. Expression of CFP-SKL is seen indicating transgene activity. MCP-RFP is used to visualize active transcription. The presence of this MCP-RFP signal colocalizes clearly with the LacI-YFP signal.

pulldown using the affinity tag attached to the crosslinker. The crosslinking compound is removed from the chromatin and the chromatin is then sequenced, effectively providing a snapshot of the sequences located in the nuclear volume of interest.

We received a U2OS cell line from the Spector Lab at Coldspring Harbor This cell line contains 200 copies of a transgene located on Chromosome 1 (Figure 3.10b). The transgene is tetracycline inducible and encodes a Cerulean Florescent Protein that localizes to the peroxisome (CFP-SKL), allowing confirmation that the transgene is active. The transgene also includes 24 MS2 repeats which can be used with MS2 Coat Protein (MCP) to visualize transcription. Upstream of the transgene are 256 Lac Operon repeats. Upon binding of fluorescently labeled LacI imaging of the transgene locus is possible.

The cell line was modified to stably express LacI-YFP allowing consistent imaging of the transgene (Figure 3.10a). Upon tetracycline activation co-localization of the transgene (labeled with LacI-YFP) and the mRNA transcript (visualized with MCP labeling the MS2 repeats) is seen. The expression of CFP-SKL can also been seen in the peroxisome (Figure 3.10c).

The easy visualization of the transgene makes this cell line a suitable option for a pilot experiment using Femto-Seq. Additionally, the activatable nature of the transgene provides two distinct states to perform a comparison on genomic contacts. While determining the presence of genomic contacts is the end goal of Femto-Seq, as a pilot experiment we hoped to see enrichment of the transgene itself over untargeted background sequences.

Several considerations needed to be taken when designing this experiment. Perhaps most importantly is the number of irradiated nuclei. Since a small amount of DNA per nuclei is being irradiated and pulled down, a large number of nuclei are required to ensure sufficient enrichment. Previous experiments examining mixed cell populations have shown that approximately 30000 nuclei would prove sufficient for a Femto-Seq experiment. Enrichment of the target sequence over background sequences depends on a variety of factors. The irradiation volume compared to the nuclear volume is critical. At one extreme, the irradiation volume could be as small as the focal volume of a high NA objective (on the order of <0.1 femtoliter for a 1.4NA objective). The downsides of using such an objective and limiting the excitation region to the focal volume are (1) given that a typical nucleus is ~250 fL in volume, the minimum volume possible would correspond to $<5 \times 10^{-3}$ of the nucleus's total DNA, making isolation of the biotinylated chromatin from the background more difficult, and (2) high NA implies higher magnification which reduces the number of nuclei that can be irradiated in a given field of view. A lower NA objective allows crosslinking of numerous nuclei in a single field of view greatly speeding up the time an experiment can take at the slight loss of theoretical signal to background. Additionally, in order to make it clear irradiation is occurring a small box is irradiated rather than using a focal volume. This box allows visualization of the bleaching of the fluorescently labeled nuclear volume. Likewise, size of the fluorescently labeled target may require a larger volume of irradiation to ensure the entire target is irradiated. Lastly and perhaps most difficult to consider is the non-uniformity of crosslinking. In an ideal case, crosslinking would occur equally at all

genomic loci; however, psoralen based crosslinking compounds tend to favor open chromatin over closed chromatin.

We performed a pilot Femto-Seq experiment using 30000 nuclei. The transgene in our U2OS cell line was activated in order to open the chromatin and provide optimal conditions for crosslinking. As a positive control we used nuclei with an activated transgene incubated with AP3B and exposed to UV excitation using a UV oven. As a negative control we used isolated nuclei with an activated transgene. Following genomic DNA extraction and pulldown a library was prepared (Figure 3.11). As expected, the positive control showed a large amount of enriched DNA while the negative control showed little enrichment. We see slight enrichment of the two-photon irradiated sample. We quantified the amount of enriched transgene by qPCR of the min-CMV promoter. To have a comparison we also quantified the amount of rDNA by qPCR. rDNA was chosen as it has shown strong enrichment upon crosslinking in previous experiments. In order to make a true comparison the difference in both copy number and heterogeneity in crosslinking potential must be considered. Using the positive control to determine the difference in likelihood of crosslinking of both these sequences we compared the amount of enriched untargeted rDNA to the amount of enriched min-CMV promoter. We were able to find a ~5-fold enrichment of the targeted transgene relative to the untargeted background rDNA sequences compared to the controls (whole cell UV illumination).

We sought to repeat the previously described pilot Femto-Seq experiment with several improvements. First, we shrunk the irradiation volume by 75% (to 14μm^3).

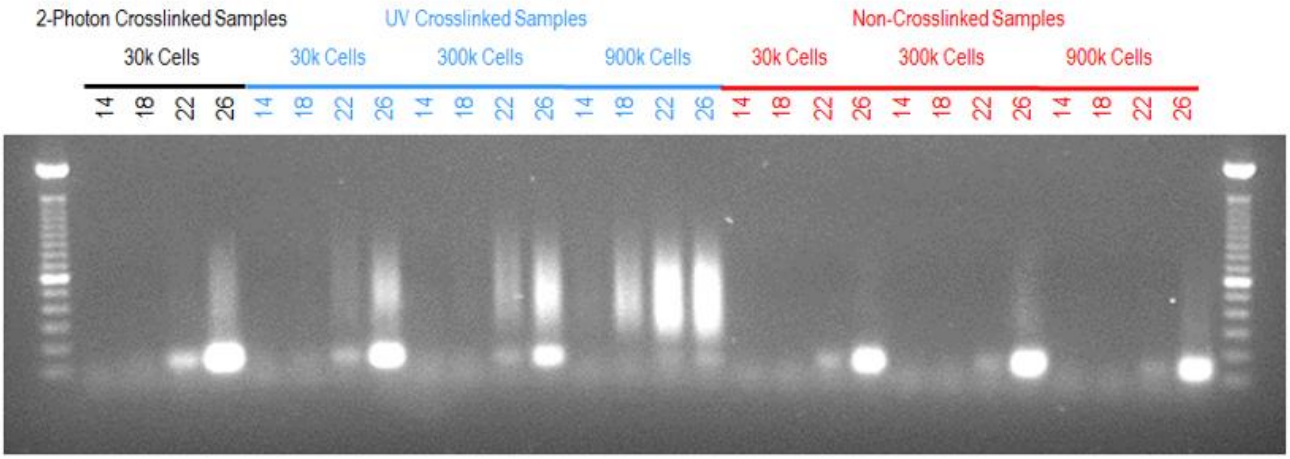


Figure 3.11: Pilot PCR Gel of First Femto-Seq Pilot Experiment.

Thirty Thousand U2OS 2-6-3 LacI-YFP pTetOn nuclei had their transgene activated, were incubated with 200µM AP3B and then specifically irradiated targeting the activated transgene. Following gDNA extraction, streptavidin coated bead pulldown and crosslink reversal a library was prepared. A gel showing the results from the pilot PCR of the library is shown. Besides the two-photon crosslinked samples, whole nuclei UV crosslinked samples at a variety of cell numbers and completely uncross linked samples are shown. At higher PCR cycle number DNA is found in the two-photon crosslinked samples. UV crosslinked nuclei show a large amount of enriched DNA. Non-crosslinked nuclei show very little DNA after streptavidin bead pull down.

Ideally, this would drastically decrease the number of off-target sequences being photo-biotinylated by the two- photon excitation. Two additional sources of background signal were concerning. Intensity control of the two-photon excitation beam was achieved using an AOM. We found that even at the minimum setting of the AOM we had ~1mW of 700nm light on the sample and were worried that this was causing low level crosslinking across entire nuclei. The other potential source of background is residual AP3B that is not covalently bound and fails to get washed out. We have seen this occur in the streptavidin staining experiments described in the previous chapter. While more intense washing conditions have been described in the previous chapter, we sought to implement them in a Femto-Seq experiment. We designed six conditions to test (Figure 3.12a). The first condition consisted of nuclei with AP3B scanned over with only two-photon excitation that leaked through the AOM. Our second sample consisted of isolated nuclei with no AP3B or irradiation. The third sample consisted of nuclei incubated with AP3B irradiated using a UV oven to perform full nucleus crosslinking. This sample is necessary as a positive control and for comparison purposes. Our fourth sample was identical to our second, but was incubated with AP3B. This sample was created to determine the impact of intercalated but non-covalently bound AP3B on background signal. The last two samples were both full Femto-Seq experiments. Two-photon excitation was used to excite the crosslinker in a small volume around the fluorescently labeled transgene. One of these samples was washed in 1,6-Hexanediol, which earlier experiments had shown could reduce background signal from uncrosslinked regions.

A

	Irradiation	Crosslinker	Washing
Sample 1	two-photon at 0 Power	AP3B	PBS
Sample 2	None	None	PBS
Sample 3	UV	AP3B	PBS
Sample 4	None	AP3B	PBS
Sample 5	two-photon	AP3B	PBS
Sample 6	two-photon	AP3B	1,6 Hexanediol

B

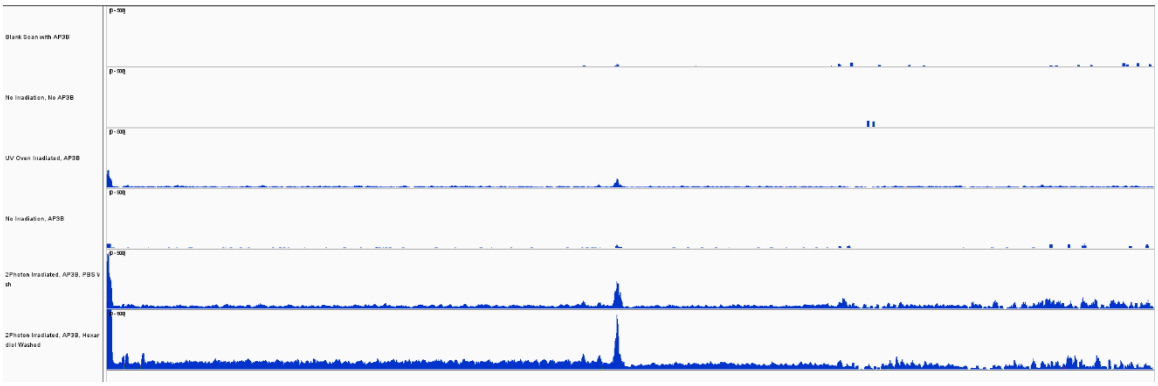


Figure 3.12: Femto-Seq with additional Washing Steps.

(a) Samples for the second pilot Femto-Seq experiment. Samples 1, 2, and 4 were control samples to examine background caused by leakage of the AOM (Sample 1), non-specific sticking of chromatin (Sample 2) and non-specific intercalation of AP3B (Sample 4). Sample 3 was our positive control, nuclei fully irradiated in a UV oven. Our last 2 samples were two-photon irradiated specifically targeting the activated transgene with (Sample 6) and without (Sample 5) a 1,6-Hexanediol washing step. (b) Mapping of sequencing results to the transgene. two-photon irradiated samples targeting the transgene show higher enrichment over untargeted irradiation patterns. Mapping across the transgene does not appear to be uniform.

All samples underwent the previously described extraction, pulldown and clean up procedures described previously followed by high throughput sequencing. This pilot experiment proved more fruitful than the first. Very little background was seen from leakage of two-photon excitation through the AOM, but the presence of AP3B alone presented quite a bit of background. A similar result of approximately 3-fold enrichment of the transgene was seen in the normal Femto-Seq sample. However, we did see a 15-fold enrichment of the transgene over untargeted sequences when the sample undergoes a 1,6-Hexanediol wash. This level of enrichment is about what would be expected based on the irradiation volume compared to the nuclear volume (Figure 3.12b).

Several important features of the data collected from this experiment were noticed. Reads across the transgene were not consistent. Particularly, reads at the LacI sites and min-CMV promoter were more commonly found than other regions of the transgene. Chromosome 1 was found enriched compared to other chromosomes. As the transgene was located on chromosome 1, we expected this. We noticed increased detection in Chromosome 18 reads as well. It is unclear why Chromosome 18 appears enriched compared to other chromosomes, but it could indicate the presence of a chromatin contact or chromosome compartment.

Numerous improvements could be made to the Femto-Seq protocol. While I will discuss some of them in more detail in a future chapter, many of them will be discussed here. First, the irradiation process for Femto-Seq is partially automated (Figure 3.13a) on a Zeiss i880 system. While the spot finding and irradiation region determination is automated, stage manipulation is not. Furthermore, the Zeiss i880 bleaching function,

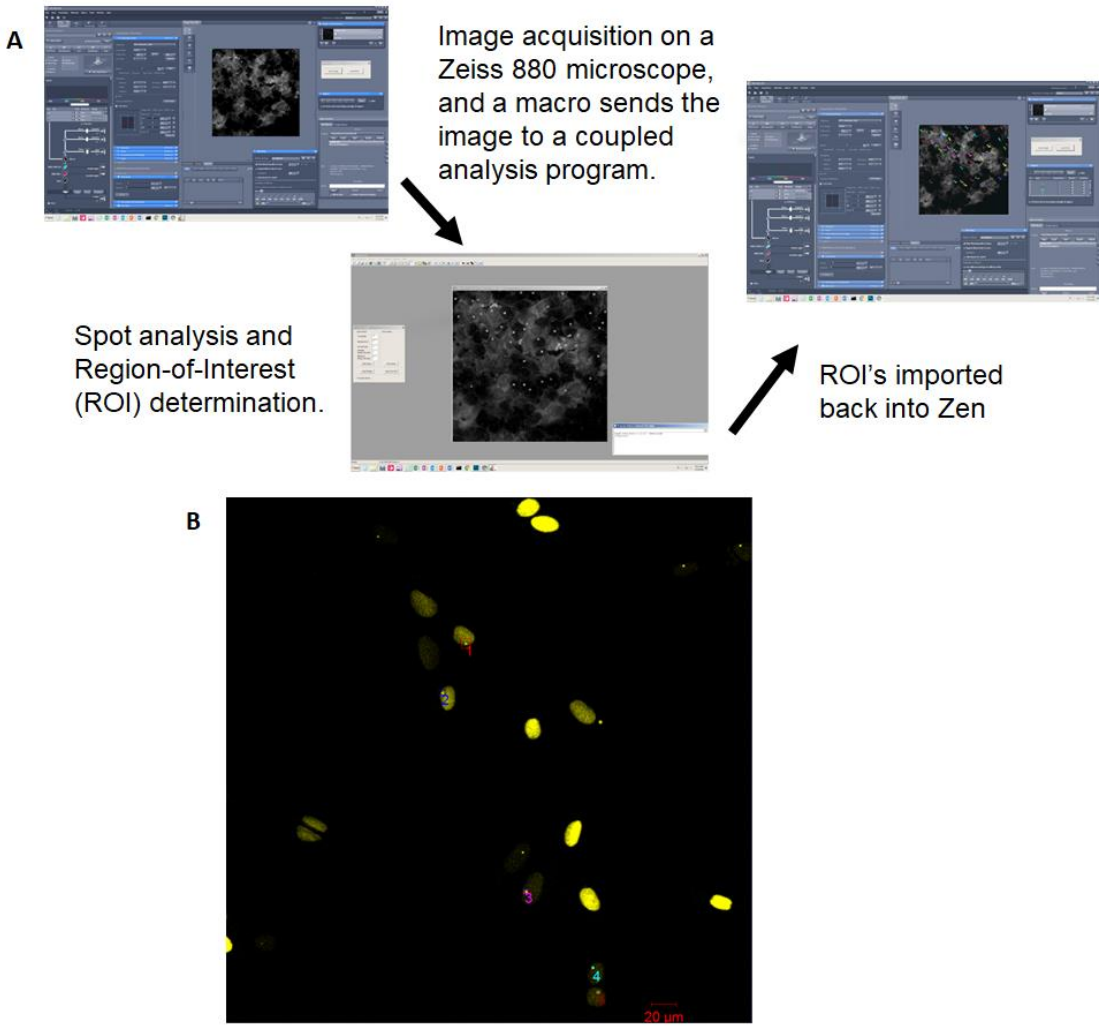


Figure 3.13: Improvements to Femto-Seq.

(a) Semi-automated workflow of the irradiation procedure. An image of the fluorescent volumes is obtained using Zeiss Microscope control software. This image is exported to a separate image analysis program to determine spot locations. This analysis program generates regions for irradiation which are then imported back into the Zeiss Microscope control software. (b) Illustration of various irradiation region sizes. Regions vary from 56µm^2 to 0.69µm^2. Previous Femto-Seq experiments used box sizes of 56µm^2 and 11µm^2, but smaller regions can easily be used if desired.

raster scans the entire field of view, activating the two-photon excitation only when a region marked for irradiation is reached. By having the beam move directly from region to region rather than scanning the entire field of view, irradiation times could be drastically lowered. Both the increased automation and improved scanning could be implemented on a custom-built system rather than a commercially available system. Decreasing the time it takes to irradiate nuclei has several advantages. Chromosome organization may change over the several hours it would take to irradiate a sufficient number of nuclei. Likewise, fluorescent labeling methods used in Femto-Seq may lose their efficacy with prolonged irradiation times.

Another major improvement that could be made to Femto-Seq is shrinking the irradiation volume. We utilized a larger irradiation volume to allow visualization of irradiation through bleaching of the YFP signal. Likewise, the size of the nuclear volume being imaged required a somewhat sizeable irradiation volume to ensure full coverage. If imaging of the bleached irradiation volume is not required, a significantly smaller irradiation volume can be used (Figure 3.13b). At best, the smallest volume that can be used is the focal volume of the objective. Here we utilized a 20x/0.5NA air objective to allow many nuclei to be captured in a single field of view. This increased the speed of our Femto-Seq experiments at the loss of potentially higher resolution.

3.4 Potential Applications of Femto-Seq:

As previously mentioned Femto-Seq can be used to examine chromatin located near any nuclear volume that can be fluorescently labeled. The pilot experiments described above examined the nuclear volume around a labeled transgene. The most obvious extension would be using Femto-Seq to examine a gene locus that has

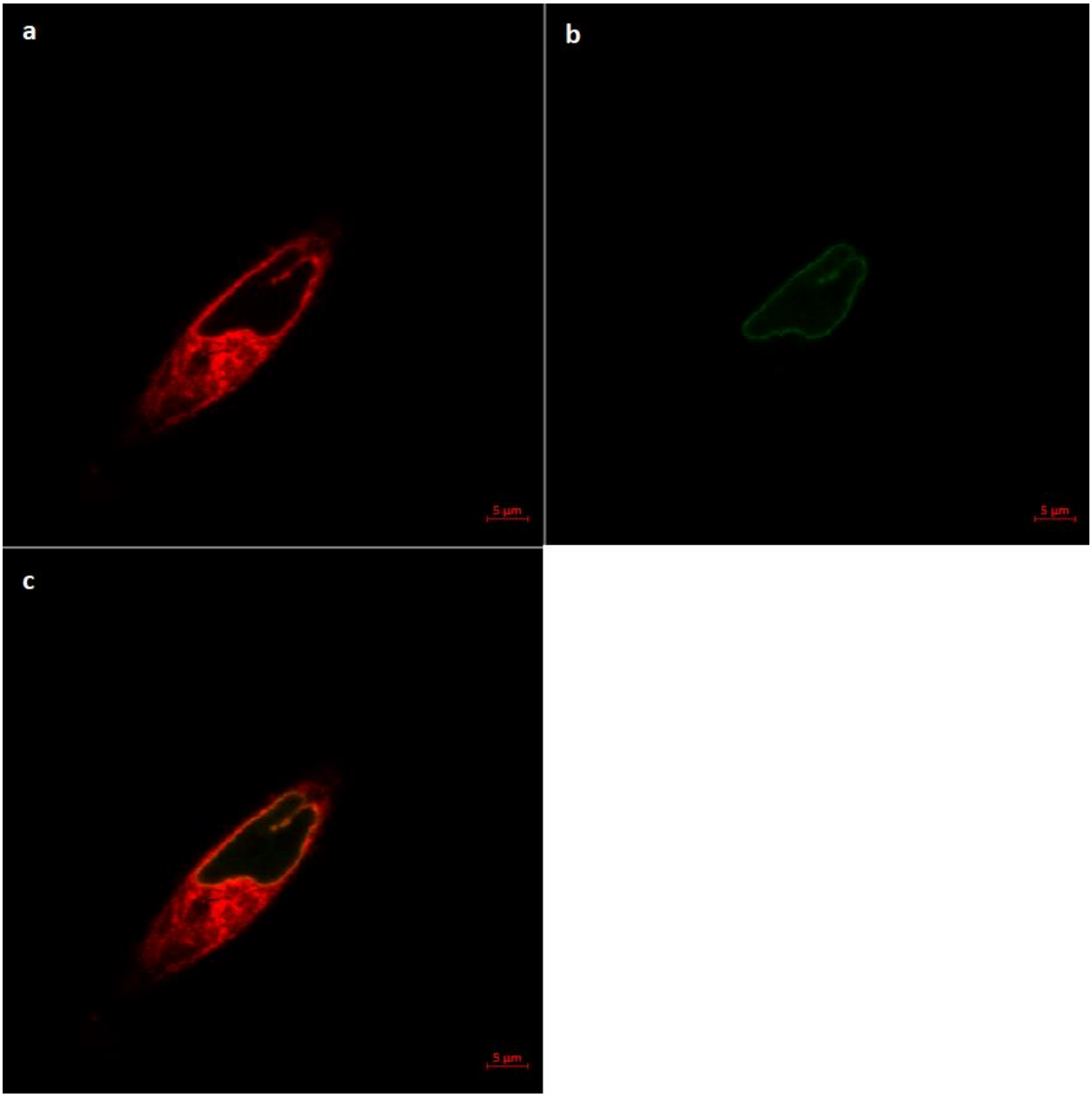


Figure 3.14: Nucleoplasmic Reticulum.

Example image of the nucleoplasmic reticulum. (a) ER stained in red show invagination into the nucleus. (b) LaminA-eGFP allows visualization of deformation of the nuclear lamina where the invagination forms. (c) Composite image showing colocalization of the ER-stain and lamina deformation.

illustrated long distance chromatin contacts like β-globin Outside of examining specific genes, other nuclear structures can be examined using Femto-Seq as well. Invaginations of the Endoplasmic Reticulum into the nucleus, called the Nucleoplasmic Reticulum has been an observed but not fully understood phenomena [31]. The Nucleoplasmic Reticulum has been observed in a variety of animal and plant cells. This behavior is observed in both healthy and abnormal cells and is not limited by tissue or developmental stage. Two types of invaginations exist. The first type are invaginations that only cause the inner nuclear membrane to protrude inward which does not form a cytoplasmic core. The second type are invaginations of the double wall nuclear membrane which forms a cytoplasmic core which often contains vesicles and other cellular bodies. The nucleoplasmic Reticulum offers the potential for deep nuclear calcium signaling, as high calcium concentration inside NR has been seen. Nuclear calcium signaling has been shown to be utilized in transcriptional regulation as seen with modulation of the transcription factor CRE-binding protein and its coactivator CREB-binding protein. The nucleoplasmic Reticulum proves to be an interesting target for a Femto-Seq experiment. The nucleoplasmic Reticulum can be easily visualized using an ER stain and a cytoskeleton stain (Figure 3.14). The number of individual invaginations of the Nucleoplasmic Reticulum usually varies from 0 to 10 allowing visualization and irradiation of distinct spots. Since little is known about the function of the nucleoplasmic reticulum, examination of the genes and sequences located consistently nearby them may provide interesting insights of their function.

3.5 Use of AP3B to Examine Genomic Landscapes:

As mentioned previously, AP3B and other psoralen based crosslinkers do not crosslink chromatin uniformly. It is not entirely clear what regions are favored by AP3B, but we suspect that euchromatin is favored based on the affinity of AP3B for actively transcribing regions. This behavior while requiring consideration in a Femto-Seq experiment could be utilized to look at other chromatin dynamics, similar to DNase-Seq.

In previous Femto-Seq experiments, a high concentration of AP3B was used to ensure as much crosslinking as possible. By using a lower concentration of AP3B, we hoped to allow visualization of chromatin that was not ideal for AP3B crosslinking. Similar to previously described experiments, nuclei were incubated with AP3B then stained with Streptavidin-Alexa647. STORM imaging was performed in order to more clearly visualize regions which show preferred binding for AP3B. As expected, and previously seen, nucleoli show very strong AP3B binding (Figure 3.15c). After activation of the transgene, areas of increased binding by AP3B are visualized (Figure 3.15ab). It is unclear if this is due to more open chromatin state due to the actively transcribing nature of the transgene.

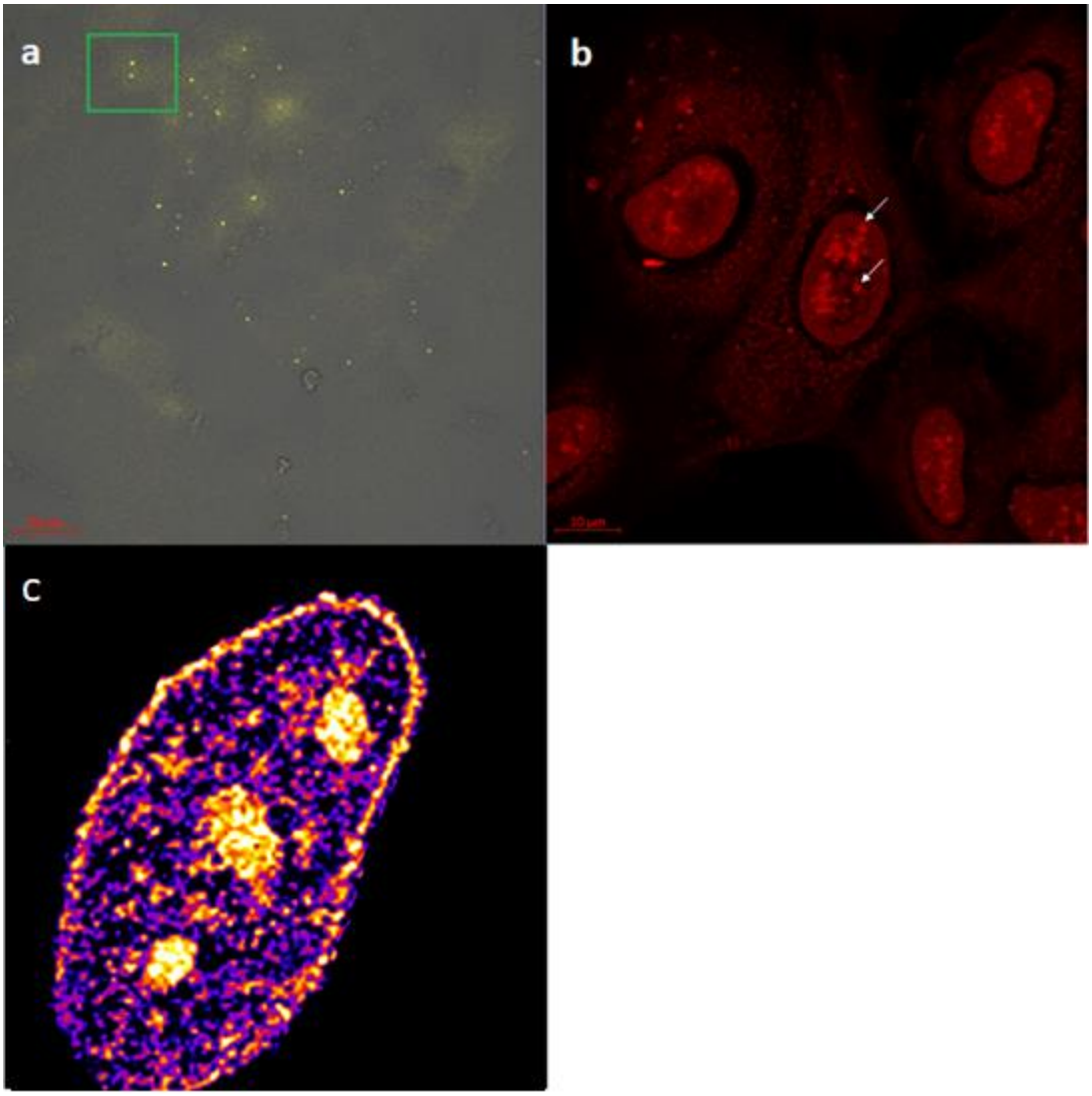


Figure 3.15: Low Concentration AP3B staining

(a) Active transgene of U2OS 2-6-3 LacI-YFP pTetOn visualized using confocal imaging. (b) After irradiation of the green highlighted region and staining with Streptavidin-Alexa647. Areas where the transgene was previously visualized indicated with white arrows. (c) STORM imaging of Streptavidin-Alexa647 after irradiation. Clear areas of nucleoli are seen, small additional spots appear but are less easily visualized.

3.6 Methods:

Cell Culture:

U2OS 2-6-3 cells were kept in a 37C incubator in 1x DMEM Media. The media contained 5µM of IPTG. The presence of IPTG serves to ensure that LacI-YFP is not bound to the transgene during cell growth cycles.

Cell Preparation:

10000 U2OS 2-6-3 cells stably expressing LacI-YFP and pTetOn were plated at the center of a 7mm Mattek dish in a 50 uL volume. After allowing the cells to settle on the surface of the plate, 2mL of DMEM containing 5mM IPTG and 5µg/mL doxycycline was added to the dish. Cells were then incubated overnight at 37C. The following day, 6 hours before irradiation, media was removed and 2mL of fresh DMEM media was added. One hour before irradiation cells were permeabilized by incubating them in 0.2% Triton-X-100 in Buffer S for 10 minutes. Cells were then washed thrice with 1mL PBS and left to incubate in 200µM AP3B in PBS for 1 hour.

Full plate Scanning:

Nuclei were imaged on a Zeiss i880 confocal microscope. The center of the plate was used to focus the two-photon excitation. Using a 20x/0.5NA air objective the two-photon excitation beam was raster scanned over the entire dish, covering the 7mm well and 6µm above and below the focus at the center of the plate. Nuclei were then washed with 1x PBS and left for an hour at 37C. Nuclei were then removed from the plate using 0.25% Trypsin-EDTA, collected in a 1.7mL non-stick Eppendorf tube.

Two-photon Excitation:

Nuclei were then imaged on a Zeiss i880 confocal microscope using a 514nm laser line. A 20x/0.5 NA Air objective was used to image the EYFP labeled genomic loci. The EYFP signal was used to identify regions for two-photon irradiation. After regions were identified a $32\ \frac{J}{mm^2}$ dose (unless otherwise described) of 700nm light (pulse width ~100fs, peak power of the pulse ~33kW), was applied to a 56 um^3 or 14 μm^3 volume around the EYFP signal for each Femto-Seq experiment respectively.

After irradiation of all cells on the plate, the buffer was removed and replaced with 1mL of either PBS or 50mM 1,6-Hexanediol. Cells were incubated in this solution for 1 hour at 37C. Cells were then removed from the plate using 0.25% Trypsin-EDTA and collected in a 1.7mL non-stick Eppendorf tube. Tubes were then frozen at -80C° until ready for future processing.

gDNA Extraction and Pulldown:

Genomic DNA was isolated is isolated from cells using a simple protocol entailing alkaline lysis, KAc precipitation of proteins, Phenol/Chloroform extraction, and ethanol precipitation. Isolated genomic DNA is spectroscopically verified to be high quality and yield for the number of cells used, sheared to 200-600bp fragments using Bioruptor (Diagenode) sonicator (30 sec ON and 30 sec OFF at high setting for 15 min). DNA was eluted into 50µL Elution Buffer, 10µL of which was saved while 40µL continued to sonication and bead pulldown.

Bead pulldown was performed using Dynabeads M-280 Streptavidin (Thermofisher). 10µL Dynabeads were washed 3 times with 1mL Tween Wash Buffer (5mM Tris, 0.5mM

EDTA, 1M NaCL, 0.05% Tween 20, pH 8). Beads were then suspended in 100µL 2x Binding Buffer (10mM Tris, 1mM EDTA, 2M NaCL, pH 8). Bead solution was then mixed with the previously extracted 40µL gDNA, 260µL H2O and 200µL 2X Binding Buffer. This solution was gently mixed for 60 minutes at room temperature. After allowing the DNA to bind, the beads were washed with 500µL Tween Wash Buffer three times. Illumina sequencing adapter were added to the bound crosslinked genomic DNA fragments either by ligation or by tagmentation with Tn5 transposase.

DNA was eluted off of the beads by incubating it in 56µL Elution Buffer (95% Formamide, 10mM EDTA) at 65C° for 5 minutes. The supernatant was transferred to a new non-stick tube. To reverse the crosslinks, 44µL of crosslink reversal buffer (3M Urea, 0.1M KOH, 1mM EDTA) was added to the elution. This mixture was incubated at 90C° for 10 minutes. An Ethanol precipitation was then performed.

Sequencing:

DNA was then PCR amplified using Illumina sequencing primers and size-selected/purified either by agarose gel electrophoresis or by SPRI beads (AmpureXP) in the range of 200-600bp. After purification, the amplified library is paired-end sequenced using an Illumina platform (2x37 bp at minimum).

Low Concentration AP3B Staining and Imaging:

U2OS 2-6-3 cells stably expressing LacI-YFP and pTetOn were plated on a 14mm Mattek Dish. `Cells were permeabilized using `0.2% Triton-X-100 in Buffer S for 10 minutes. Cells were then washed thrice with 1mL PBS and left to incubate in 20µM AP3B for 1 hour. Regions were then irradiated as described previously using a Zeiss

i880 confocal microscope with a 20x/0.75NA air objective. After irradiation, the buffer was removed and replaced with 1mL of 1x PBS. Nuclei were incubated in this solution for 1 hour at 37C. Buffer was then removed and 5µM Streptavidin-Alexa647 was added and left at room temperature for 1 hour. Optionally, 5µM of Syto-RNASelect stain was added to visualize potential regions of higher RNA concentration.

3.7 References:

[1] Xie SQ, Lavitas LM, Pombo A. CryoFISH: fluorescence in situ hybridization on ultrathin cryosections. Methods Mol Biol 659: (2010);219-30.

[2] Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 4: (2006);e138.

[3] Palstra RJ, Simonis M, Klous P, Brasset E, Eijkelkamp B, de Laat W. Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. PLoS One 3: (2008);e1661.

[4] Müller P, Schmitt E, Jacob A, Hoheisel J, Kaufmann R, Cremer C, Hausmann M. COMBO-FISH enables high precision localization microscopy as a prerequisite for nanostructure analysis of genome loci. Int J Mol Sci 11: (2010);4094-105.

[5] Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard JM, Wu CT. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proc Natl Acad Sci U S A 109: (2012);21301-6.

[6] Beliveau BJ, Boettiger AN, Avendaño MS, Jungmann R, McCole RB, Joyce EF, Kim-Kiselak C, Bantignies F, Fonseka CY, Erceg J, Hannan MA, Hoang HG, Colognori D, Lee JT, Shih WM, Yin P, Zhuang X, Wu CT. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. Nat Commun 6: (2015);7147.

[7] Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, Prasanth KV, Ried T, Shav-Tal Y, Bertrand E, Singer RH, Spector DL. From silencing to gene expression: real-time analysis in single cells. Cell 116: (2004);683-98.

[8] Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, Cramard J, Faure AJ, Ralser M, Blanco E, Morey L, Sansó M, Palayret MGS, Lehner B, Di Croce L, Wutz A, Hendrich B, Klenerman D, Laue ED. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature 544: (2017);59-64.

[9] Germier T, Audibert S, Kocanova S, Lane D, Bystricky K. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. Methods 142: (2018);16-23.

[10] Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155: (2013);1479-91.

[11] Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. Proc Natl Acad Sci U S A 112: (2015);3002-7.

[12] Anton T, Bultmann S, Leonhardt H, Markaki Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. Nucleus 5: (2014);163-72.

[13] Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. Proc Natl Acad Sci U S A 112: (2015);3002-7.

[14] Ma H, Tu LC, Naseri A, Huisman M, Zhang S, Grunwald D, Pederson T. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat Biotechnol 34: (2016);528-30.

[15] Gu B, Swigut T, Spencley A, Bauer MR, Chung M, Meyer T, Wysocka J. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science 359: (2018);1050-5.

[16] Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, Hesson J, Cavanaugh C, Ware CB, Krumm A, Shendure J, Blau CA, Disteche CM, Noble WS, Duan Z. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat Methods 12: (2015);71-8.

[17] Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: (2014);1665-80.

[18] Nagano T, Lubling Y, Yaffe E, Wingett SW, Dean W, Tanay A, Fraser P. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. Nat Protoc 10: (2015);1986-2003.

[19] Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J. Massively multiplex single-cell Hi-C. Nat Methods 14: (2017);263-6.

[20] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462: (2009);58-64.

[21] Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, Liu J, Zhao XD, Chew JL, Lee YL, Kuznetsov VA, Sung WK, Miller LD, Lim B, Liu ET, Yu Q, Ng HH, Ruan Y. A global map of p53 transcription-factor binding sites in the human genome. Cell 124: (2006);207-19.

[22] Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods 13: (2016);919-22.

[23] Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, Ren B. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Res 26: (2016);1345-8.

[24] Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, Fraser J, Dostie J, Game L, Dillon N,

Edwards PA, Nicodemi M, Pombo A. Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543: (2017);519-24.

[25] Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, van Steensel B, Ma J, Belmont AS. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol 217: (2018);4025-48.

[26] Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453: (2008);948-51.

[27] Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, Trinh V, Aznauryan E, Russell P, Cheng C, Jovanovic M, Chow A, Cai L, McDonel P, Garber M, Guttman M. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell 174: (2018);744-757.e24.

[28] Zheng M, Tian SZ, Capurso D, Kim M, Maurya R, Lee B, Piecuch E, Gong L, Zhu JJ, Li Z, Wong CH, Ngan CY, Wang P, Ruan X, Wei CL, Ruan Y. Multiplex chromatin interactions with single-molecule precision. Nature 566: (2019);558-62.

[29] Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard JM, Wu CT. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proc Natl Acad Sci U S A 109: (2012);21301-6.

[30] WIP

[31] Malhas A, Goulbourne C, Vaux DJ. The nucleoplasmic reticulum: form and function. Trends Cell Biol 21: (2011);362-73.

Chapter 4

Making Improvements for Femto-Seq: Examination of TALE proteins as a fluorescent labeling platform for Femto-Seq and Examination of the Use of Microfluidics to Increase Downstream Efficiency

4.1 Overview:

In the previous sections I have described efforts undertaken to design a novel method to examine genomic contacts, Femto-Seq. One of the requirements of a Femto-Seq experiment is the fluorescent labeling of the nuclear volume of interest. In many cases when one is interested in a particular gene, commonly used DNA loci labeling methods can be used; however, in other cases different fluorescent labeling techniques may be required. The two most popular methods for labeling a gene locus are Fluorescence in Situ Hybridization (FISH) [1, 2] and the utilization of a fluorescently labeled DNA binding proteins. FISH utilizes fluorescently labeled oligonucleotides to visualize genomic loci [1] while fluorescently labeled DNA binding proteins encompass a wide variety of proteins that can be used for this purpose. To use many of these DNA binding proteins, artificial sequences would need to be inserted near the genomic loci of interest [3]. Alternatively, specific DNA binding proteins would need to be selected to fit the genomic loci attempting to be studied [4,5]. Utilizing engineerable DNA binding proteins which don't require targeting very specific DNA sequences or inserting binding sites would be ideal in a Femto-Seq experiment. While the CRISPR-dCas9 system is one of the most well characterized DNA binding protein constructs that fulfill this purpose [6], it is certainly not the only one.

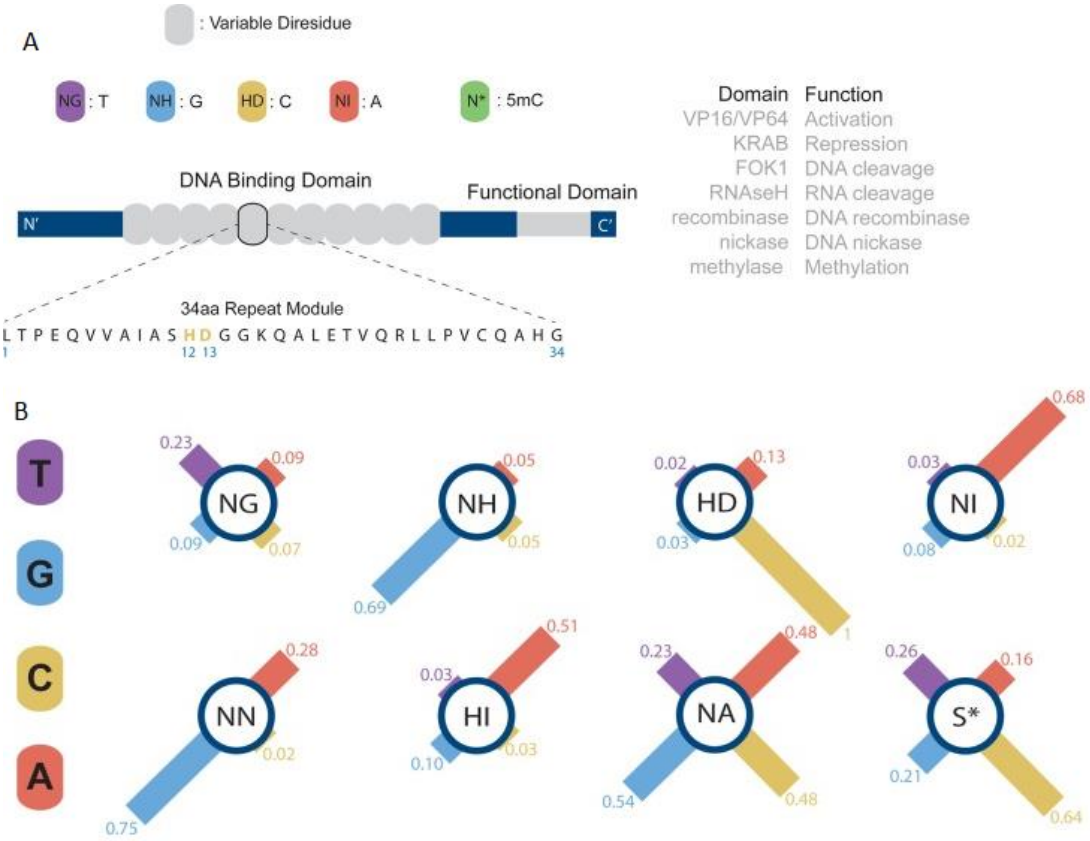


Figure 4.1. Transcription Activator Like Effectors

(a) Illustration showing the TALE protein. The N'-terminal domain which is necessary for translocation, the DNA binding domain with the Repeat Variable Diresidue (RVD) and the functionally important C'-terminal domain are shown. (b) Illustration of the RVD code. Specific RVD combinations correspond to a specific targeted nucleic acid. There is variability in the specificity of the targeted nucleic acid (adapted from [17]).

Transcription Activator-Like Effectors (TALEs) are another set of DNA binding proteins secreted by *Xanthomonas*, a plant pathogenic bacterium [7]. TALE proteins contain a DNA binding domain which can be engineered to bind a specific DNA sequence [8]. This domain contains a central repeat domain which consists of numerous repeats that are 34 residues in length (Figure 4.1a) [9]. A typical repeat sequence is LTPEQVVAIAS**HD**GGKQALETVQRLLPVLCQAHG. The bolded residues at the 12th and 13th positions are the Repeat Variable Diresidue (RVD) and determine what base pair the repeat targets [10]. The numerous repeats thus correspond to a specific DNA sequence the entire TALE protein will target [11]. The RVD code and the corresponding base pairs have been well characterized (Figure 2.1b) [12]. The C-terminal domain contains a nuclear localization signal but is not required for TALE proteins to bind DNA. The N-terminal domain; however, is necessary for translocation [13]. TALEs have been shown to be a useful tool not only for gene editing purposes [15], but for gene locus visualization by fluorescent labeling [16].

TALE protein binding kinetics and dynamics have been the focus of some research but it has been limited. Previously, the search mechanism has been determined by utilizing extended DNA oligos attached to a PEG surface [18]. TALEs search for a binding site by micron scale hopping which fully dissociates the protein from DNA and a much slower 1D diffusion along the DNA oligo [18]. The lack of target sites for the TALE protein in these experiments did not allow the researchers to characterize target binding kinetics. We sought to study binding dynamics of TALE proteins using a few methods: DNA curtains, Single Molecule PEG chambers, and FCS. We believed characterizing these proteins was an important step in determining their

utility in a Femto-Seq experiment. In this section I will also describe efforts we undertook to improve the bead pulldown step in Femto-Seq experiments.

4.2 Results:

4.2.1 DNA Curtains:

DNA curtains are a method that utilize a microfluidic flow chamber to stretch out DNA to be visualized on a TIRF microscope [19]. DNA is first tethered to a lipid bilayer. Hydrodynamic force is then applied by washing buffer through the microfluidic chamber. The tethered DNA molecules get caught on chrome barriers inside the microfluidic chamber. The constantly applied flow causes the DNA molecules to get stretched out. They can then be visualized using fluorescent compounds that bind DNA or fluorescently labeled DNA binding molecules. We sought to utilize this assay to look at the binding kinetics of TALE proteins.

We utilized photolithography rather than electron beam lithography [20] to create the microfluidic chamber used for DNA curtains (Figure 4.2). Photolithography is cheaper and faster than electron beam lithography and is suitable to create the microfluidic features we were hoping to create. In photolithography, a silicon wafer is coated in photoresist. This photoresist material becomes soluble after exposure to UV light. A photomask defining the features we wish to present is created in CAD software and printed onto a chrome and photoresist coated plate.

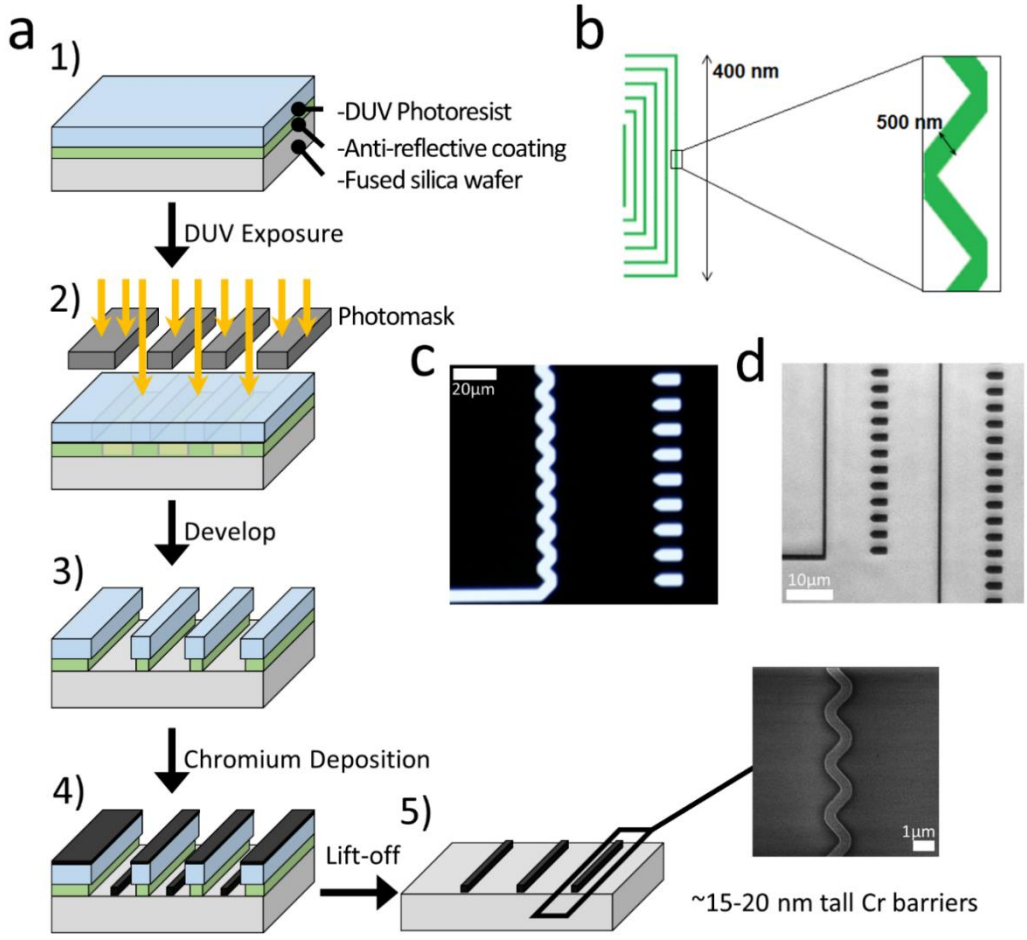


Figure 4.2. Use of photolithography to create a DNA curtain microfluidic device (a) Overview of the photolithography process. A wafer is coated with photoresist then undergoes UV exposure through a photomask to detail the curtain rods. Exposed Photoresist removed by developing allowing for chromium to be deposited onto the glass and form the curtain rods. (b) L-Edit image of the designed curtain rod pattern. (c) Darkfield image showing the photomask used in the exposure step. (d) Brightfield image showing the chrome patterns deposited onto the glass surface.

Once the photomask is exposed, it is used to pattern features on the previously coated silicon wafers. UV light is directed through the photomask by a lens to selectively expose select regions of photoresist on the wafer. These exposed regions of photoresist are then washed away using a solvent. An evaporation chamber is used to deposit chrome onto the bare glass regions (from where the previously exposed photoresist was washed away). Finally, all the remaining photoresist is washed away leaving a set of chrome barriers where they were deposited earlier.

DNA curtains provide a unique platform to study the binding dynamics of DNA binding proteins, including TALEs. We utilized a unique TALE protein, TAL1535, designed by Fabio Cupri Rinaldi of the Bogdanove laboratory at Cornell University, as a model protein in this system. TAL1535 has been shown to have a picomolar binding affinity (Figure 4.3c). This extremely tight binding makes it an ideal candidate for study using the DNA curtains assay. The TAL1535 binding site was inserted into Lambda DNA at 21kb into the Lambda-DNA sequence. TAL1535 was incubated with Lambda-DNA prior to injection into the DNA curtains flow chamber. Regardless of changing parameters such as TAL1535 concentration (30-1500 nM), significant sticking of the TAL1535 protein was noticed on the chrome curtain rods. TAL1535 that did manage to bind to the Lambda-DNA remain attached when flow was removed (Figure 4.3f). Due to our continued struggles with using DNA curtains to examine TAL proteins we sought to other methods to study their binding kinetics.

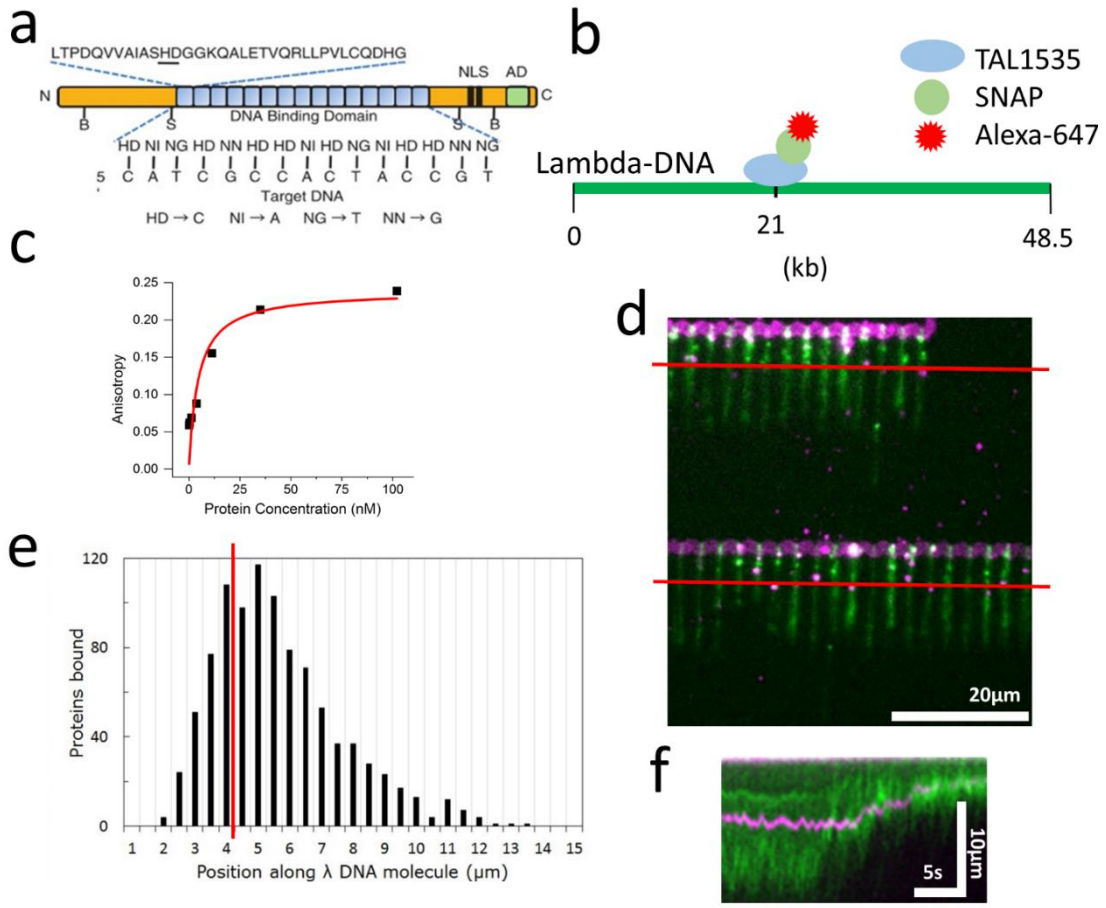


Figure 4.3. Using DNA curtains to examine the binding of TAL1535

(a) Example layout of a TALE protein. The underlined residues correspond to the RVD which can be engineered to make the TALE protein bind a specific sequence. (b) Illustration of the labeling scheme. Lambda-DNA contains the TAL1535 binding site around the 21kb mark. TAL1535 is labeled using a SNAP-CLIP labeling platform. (c) Fluorescence anisotropy of TAL1535. (d) DNA curtains experiment with Alexa647 labeled TAL1535 (red) binds YOYO-1 stained DNA (green). The red line corresponds to the expected location of the TAL1535 binding site. (e) Histogram of TAL1535 location along Lambda-DNA. Red line represents the expected binding location based on the target site. (f) Kymograph showing the location of TAL1535 as the DNA curtains are relaxed.

4.2.2 Single Molecule PEG Chambers:

Another commonly used method to look at protein interactions is the use of polyethylene glycol (PEG) coated chambers. These types of chambers have been utilized in the past for FRET [21], SiMPull [22], and a variety of other single molecule biophysical experiments [23]. Due to the nonspecific sticking issues found in the DNA curtains experiments we decided a PEG coated slide may provide a better pacified surface to analyze TALE proteins. PEG coated chambers are simply glass microfluidic chambers with a PEG coated surface inside. The PEG surface contains a small amount of biotin labeled PEG. The PEG itself is present to prevent nonspecific sticking of the molecules of interest to the glass surface while the biotinylated PEG serves as a platform for molecules to bind. To study binding kinetics of TAL1535 we utilized the previously discussed TAL1535 labeled with Alexa647 and a 50bp DNA oligo containing the binding site for TAL1535 labeled with Cy3 (Figure 4.4a). We planned to examine the colocalization of these signals to determine characteristics of the binding kinetics of TAL1535. While many different statistical analysis methods could be used to examine colocalization, we utilized the Pearson Correlation Coefficient (PCC).

$$r_{xy} = \frac{\sum_{i=1}^{n}(B_i - B_{avg}) * (R_i - R_{avg})}{\sqrt{\sum_{i=1}^{n}(B_i - B_{avg})^2} * \sqrt{\sum_{i=1}^{n}(R_i - R_{avg})^2}} \qquad (1)$$

This statistic can be used to interpret how well colocalized two fluorescent signals are, with a result of +1 meaning complete colocalization and -1 meaning anti-colocalization [24]. Tetraspeck fluorescent beads were used to calibrate the Zeiss Elyra Microscope

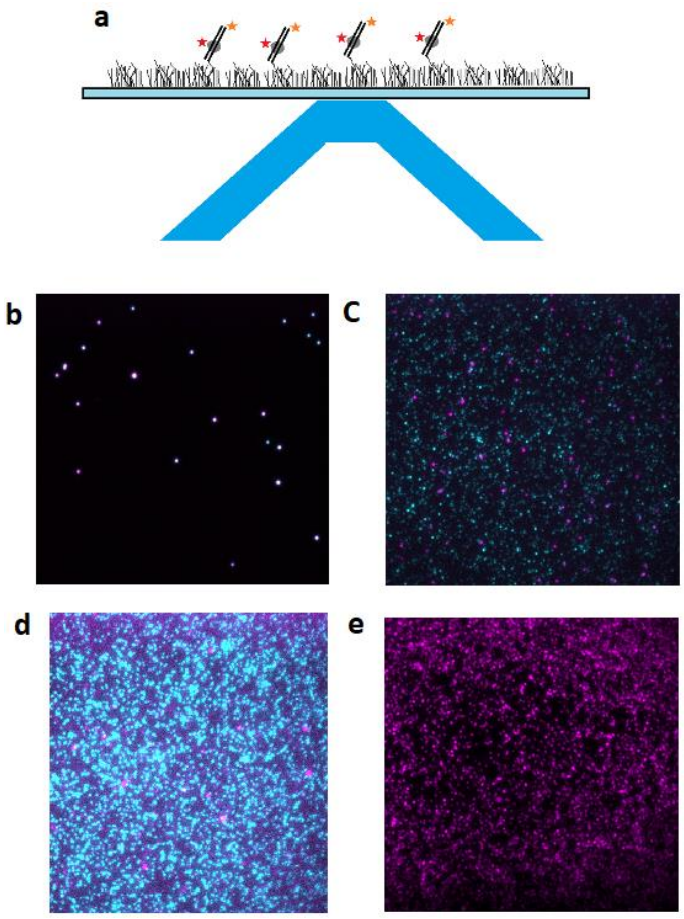


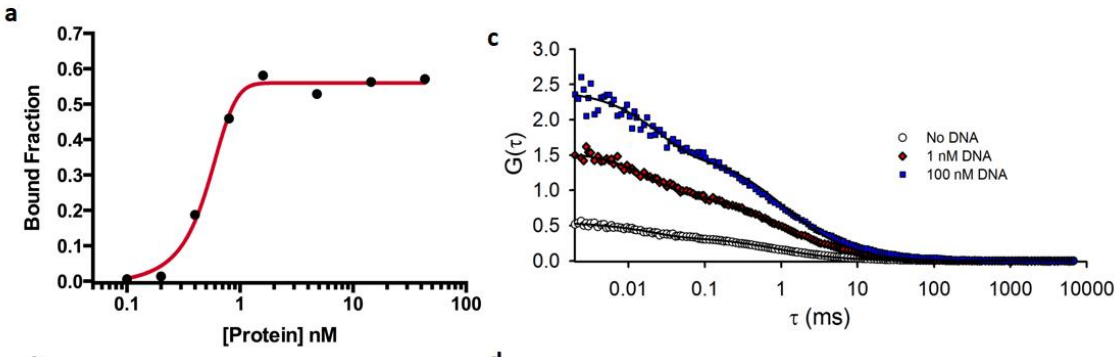
Figure 4.4 Examination of TAL1535 on PEG and Lipid Surface

(a) Illustration of DNA bound to a PEG surface. The 50bp DNA oligo is labeled with Cy3 while the TAL1535 protein (grey) is labeled with Alexa647. The surface is imaged using TIRF microscopy. (b) Tetraspeck beads imaged in multiple channels to ensure proper microscope alignment and correct calculation of Pearson's Correlation Coefficient (PCC=0.85). (c) 50pM Alexa647 labeled TALE protein (magenta) and 20pM 50bp Cy3 labeled DNA oligo (cyan) on a PEG coated coverslip. Very little colocalization is observed and a PCC of 0.02 was calculated. (d) 50pM Alexa647 labeled TALE protein (magenta) and 20pM Alexa488 labeled DNA oligo (cyan) on a lipid bilayer. However, less non-specific binding is observed compared with (c). A PCC of -0.77 was calculated. (e) 20pM of Alexa647 labeled TALE on a BSA blocked glass surface. Indicating the efficiency of blocking methods.

to ensure that alignment or correction was not necessary (Figure 4.4b). We calculated a PCC of 0.85 for Tetraspeck beads suggesting the system was appropriately aligned. We calculated a PCC of 0.02 for TAL1535 and the short DNA oligo suggesting very little colocalization. We attributed this to nonspecific sticking of TAL1535 to the surface amongst other factors discussed later. By using a lipid bilayer, we are able to eliminate the majority of non-specific binding (Figure 4.4d); however, this results in a PCC of -0.77 suggesting anti-colocalization.

4.2.3 Fluorescence Correlation Spectroscopy Experiments:

The high binding affinity of TAL1535 made the anti-colocalization results from the PEG chambers unexpected. In order to explore why we were seeing this, we decided to utilize Fluorescence Correlation Spectroscopy (FCS) Microscopy. FCS analyzes fluctuation of fluorescence intensity through a volume. Using a confocal microscope, the beam is focused and the fluorescence intensity fluctuations due to diffusion are autocorrelated. From this, numerous properties of the fluorescently labeled molecule can be determined including the diffusion coefficient and concentration. We conducted an FCS experiment using 3nM of our Alexa647 labeled TAL1535 both in the presence and absence of a 400bp DNA oligo containing the binding site and Cy3 (1nM and 100nM). Curves were generated using the average of thirty 10 second correlations and fit to a single diffusion model with an added triplet component (Figure 4.5c). Upon analysis several interesting phenomena were observed (Figure 4.5d). As DNA concentration increases the calculated concentration of TAL1535 decreases. We



d

	3 nM TALE protein	+ 1 nM DNA	+ 100 nM DNA
N	2.94 ± 0.25	1.04 ± 0.09	0.65 ± 0.03
τ_D	0.909 ± 0.08 ms	1.040 ± 0.13 ms	1.118 ± 0.21 ms
τ_{trip}	0.020 ± 0.004 ms	0.021 ± 0.002 ms	0.024 ± 0.008 ms
Triplet Fraction	0.375 ± 0.03	0.386	0.369

Figure 4.5. FCS and EMSA Analysis of TAL1535.

(a) EMSA data showing increased binding as TAL1535 concentration increases. A K_d of 0.6nM was calculated. (b) Gel showing increased shift upon binding of TAL1535 to DNA. (c) Correlation curves of Alexa647 labeled TAL1535 with and without a 400bp DNA oligo containing the binding site and Cy3. The curves are an average of thirty 10 second correlations fit to a single diffusion model with added triplet component. (d) Table of calculated concentration, residence times, and triplet fraction.

attribute this to quenching of the Alexa647upon binding. We do see increased residence times upon addition of DNA suggesting that binding is occurring for the TAL1535 that can be measured. We believe both FCS and the use of single molecule fluorescence techniques can be used to study the binding kinetics of TALE proteins. Unfortunately, these experiments had issues with sticking and quenching, but provided these issues could be solved single molecule analysis could be utilized to examine TALE protein binding kinetics.

4.2.4 Microfluidic Platform for downstream cleanup:

While the previous sections described efforts to characterize binding characteristics of TALE proteins for use in Femto-Seq experiments, we also worked towards improving the efficiency of the DNA pulldown step of the Femto-Seq process. While there are many different methods that could be implemented into improving DNA pulldown, we focused on utilizing a technique called MOWChIP [25]. This technique utilizes a microfluidic chamber to perform Chromatin Immunoprecipitation with as few as 100 cells. While we planned on using more cells than this in a Femto-Seq experiment, we were targeting a small fraction of the DNA per cell and thus sought to use this method to prevent potential losses of crosslinked DNA. MOWChIP uses a PDMS flow cell formed by imprinting PDMS on a silicone mold. This fluidic layer combined with a very thin PDMS layer used for pneumatic control and a valve layer forms the microfluidic chamber used to perform the pulldown. Beads are flowed into the chamber and forced into a packed bed using a magnet. Genomic DNA fragments are then flowed in and allowed to bind to the magnetic beads. Oscillatory washes are then

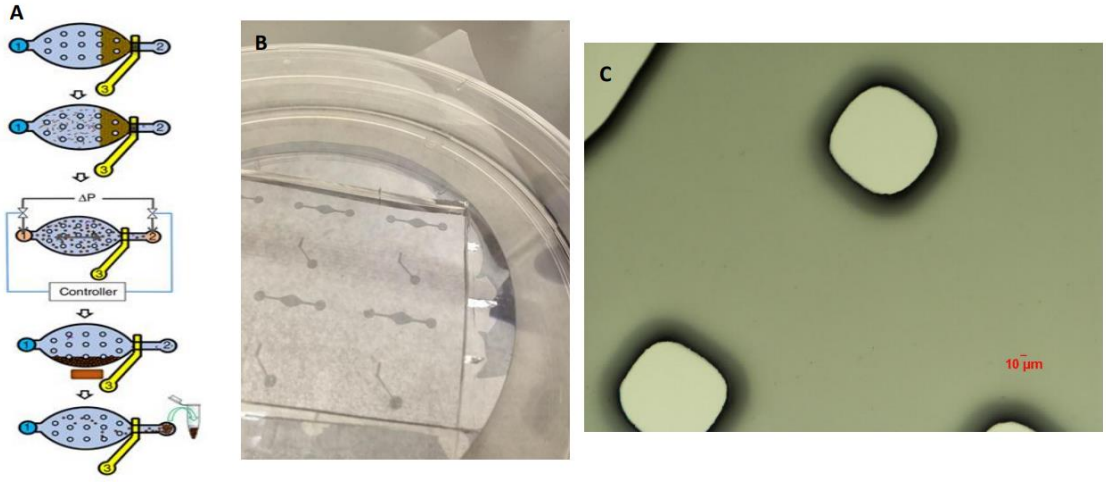


Figure 4.6. Overview of MOWChIP and Creation of Master Mold.

(a) Overview of the MOWChIP process. A packed bed of beads is made by closing the valve layer, allowing fluid to pass but not the beads. Extracted chromatin is then flowed through the bed. The beads then undergo an oscillatory washing process. A magnet is used to pull down the beads and buffer is replaced. The beads are then extracted. (Adapted from [25]). (b) A photograph of the mold for the device. PDMS is poured on the mold in order to create numerous channels. Shown here are two of the three layers, the valve layer and the microfluidic layer. Sandwiched between these layers is an extremely thin control layer (spun on a blank wafer) used to shut the valve. (c) Microscopy image of the pillar features on the device mold.

performed to wash away unbound chromatin. Finally, the DNA is eluted off the beads (Figure 4.6a). In order to utilize a MOWChIP set-up in a Femto-Seq experiment we needed to create the master molds for forming the PDMS chambers. We utilized photolithography at the Cornell Nanofabrication Facility to create master molds for forming both the valve layer and fluidic layer (Figure 4.6b, Figure4.6c). After creation of the master molds, we used PDMS to form a fluidic chamber. Unfortunately, these chambers were subject to leaking due to improper sealing. Despite this, we believe this or other microfluid platforms could be viable alternatives to the traditional magnetic bead pulldown.

4.3 Conclusion:

Here we showed that despite issues with non-specific surface binding, TALE proteins remain an attractive option for labeling a specific gene locus of interest. They can be engineered to bind non-repetitive DNA and show very robust binding properties. We showed that a variety of single molecule analysis methods such as colocalization methods and DNA curtains have the potential to be useful for studying TALE binding kinetics.

We also demonstrated that further downstream improvements to Femto-Seq can be made. The magnetic bead pulldown process can lead to losses in material that can be critical when working with a small amount of crosslinked DNA. Microfluidic platforms such as MOWChIP have the potential to be used to improve the overall efficiency of Femto-Seq.

4.4 Methods:

The DNA curtains work was performed in collaboration with Alex Van Slyke at Cornell University.

Protein and DNA preparation:

Lambda DNA containing the TAL1535 binding site was prepared by Alex Van Slyke. TAL1535-SNAP was cloned and expressed by Fabio Cupri Rinaldi of the Bogdanove Labrotory at Cornell University. TAL1535-Alexa647 was created by incubating TAL1535-SNAP with SNAP-Alexa647. Labeled DNA oligos containing the TAL1535 binding site were created by PCR. A 400bp g-block containing the TAL1535 binding site (CTGGATCATTCCCGAGCGCT) was ordered from IDT. A DNA primer labeled with Cy3 and several primers labeled with Biotin were ordered as well. The Biotin labeled primers were picked to allow for customization of oligo length after PCR. PCR was performed to ensure all the products contained both a Cy3 and Biotin label.

DNA curtains:

Before construction of the DNA curtain chamber, a photomask was designed using L-Edit and written onto a chromium mask plate then exposed using the CNF DWL 2000 High Resolution Pattern Generator (Heidelberg Instruments). A 100mm diameter 1mm thick silica wafer was coated with 40nm of anti-reflective coating (DS-K101-4, Brewer Science) and 300nm of photoresist (UV210, Dow Chemical Co.) using the CNF Gamma Cluster Tool (SUSS MicroTec). The photomask was used to expose (248nm) the curtain pattern on the coated wafer using the CNF AASML PAS 5500/300C DUV Stepper. Photoresist was removed using AZ 726 MIF photoresist developer. A chromium layer

20nm thick was deposited on the wafer using the CNF SC4500 Thermal/Electron-beam evaporator. The wafer was then cut into 60 x 25 mm rectangles to fit standard microscope stages. The remaining photoresist was then washed away using Microposit Remover 1165.

The glass slides had 1.5mM holes drilled into them using a Dremel Drill. Glass chambers were washed with Hellmanex, 1M NaOH, EtOH, and MilliQ water. Slides were then dried with Nitrogen and plasma cleaned for 5 minutes. Slides were attached to coverslips using double sided tape. Slides were then heated in a 75C° oven for 30 minutes to ensure proper sticking of the coverslip to the slide. Nanoports were attached to the previously drilled holes using hot glue.

A mixture of 95% DOPC, 4.4% mPEG2000-DOPE, 0.6% Biotin-DOPE and 0.1% Lissamine Rhodamine-DOPE (if imaging the bilayer) was created and dehydrated using nitrogen. The dehydrated lipid mixture is then stored overnight in a vacuum chamber. The lipid mixture was rehydrated with lipid buffer (100mM NaCL, 50mM Tris pH 8.0). The lipid mixture was stored on ice for 10 minutes, sonicated for 1 minute halfway through. The lipid mixture is then sonicated using a probe sonicator in a 4C fridge. Lipids were then spun down at 16000 rcm for 5 minutes.

DNA curtain chambers were flushed with lipid buffer, then 1mL of the previously described lipisomes. This was repeated 3 times with 5 minutes in between injections. The chamber was then washed with lipid buffer to wash away any unbound lipisomes. The chamber was left at room temperature to allow the lipisomes to settle. The chamber was then washed with BSA buffer (40mM Tris pH 7.8, 1mM MgCl2, 1mM DTT, 0.2mg/mL BSA) to block nonspecific binding. Chambers sat at room temperature for 20

minutes to allow blocking to occur. 30nM Tal1535-Alexa647 was incubated with 10pM Lambda DNA containing the TAL1535 binding site in a volume of 100uL for 25 minutes. This solution was diluted to 1mL then injected into the flow chamber. Finally, the chamber was washed with imaging buffer (BSA buffer with 20uL GLOX and 200uL 40% Glucose). Imaging took place on a custom microscope using an Olympus UPlanApo 60x/1.2 Water Immersion lens.

Construction of PEG Chambers:

Using a Dremel Drill a 1mM hole is drilled into quartz slides. Slides are then scrubbed with MeOH and distilled water and placed in a glass coplin jar. Slides are rinsed with MilliQ water and sonicated in a bath sonicator in 10% alconox solution for 20 minutes. Slides are again rinsed with MilliQ water then sonicated in MilliQ water for 5 minutes then rinsed with MilliQ water. Slides are rinsed with acetone then sonicated in acetone for 15 minutes. Slides are rinsed and sonicated in 1M KOH for 20 minutes. Slides are rinsed then dried using nitrogen. Coverslips are washed with water then sonicated in 1M KOH. Coverslips are then rinsed and dried using nitrogen. Slides and coverslips are coated in a Silanization buffer (100mL MeOH, 5mL Acetic Acid, 1mL Aminopropylsilane) and left for 10 minutes at room temperature. Slides and coverslips are rinsed with MeOH and Water then dried using Nitrogen. 320uL of PEGylation buffer (100mM Sodium Bicarbonate) is combined with 2mg Biotin-PEG-SVA and 80mg mPEG-SVA. Solution is mixed then applied in between the coverslips and slides. Coverslips and slides are left for 3 hours in a humid environment. Tubes are attached to the previously drilled holes and affixed with epoxy. Excess tubing is cut away and the coverslip is

affixed using double sided tape. PEG chambers are stored in vacuum sealed 50mL tubes at -20C° until use.

Single Molecule PEG Experiments:

PEG chamber was first flushed using T50 buffer then with 0.2mg/mL Streptavidin solution in T50 buffer. 20pM of a 50bp DNA oligo containing the TAL1535 binding site and labeled with Cy3 and Biotin was mixed with 50pM TAL1535-Alexa647 in T50 buffer with 0.1 mg/ml BSA. This solution was then flowed into the chamber and allowed to bind for 15 minutes at room temperature. The flow chamber was then washed twice with T50 with BSA. Flow chambers were imaged on a Zeiss Elyra TIRF microscope using 60x/1.4 NA objective. Tetraspeck beads images were taken with a Tetraspeck slide.

Fluorescence Correlation Spectroscopy:

Solutions were created with 3nM TAL1535 labeled with Alexa647 with no DNA, 1nM 400bp DNA oligo TAL1535 binding site labeled with Cy3 and Biotin, and 100nM of the previously described DNA oligo. These solutions were left at room temperature for 25 minutes before FCS experiments. FCS curves were taken on a Zeiss i880 Inverted Confocal Microscope using a 633nm laser line.

MOWCHIP Mold Procedure:

The photomask detailing the fluidic wafer was designed in L-Edit and written onto a chromium mask plate. A clean silicone wafer had 3g of Su-8 Photoresist spun onto it (500 rpm for 10s, 2500 rpm for 30s). The photoresist was baked onto the wafer at 95C° for 8 minutes then the wafer underwent UV exposure (160mJ/cm2 for 17 seconds) through the photomask using the CNF AASML PAS 5500/300C DUV Stepper. The

wafer was baked again then developed using SU-8 developer. The wafer is washed with isopropanol, dried, then baked again at 150C° for 15 minutes. To make the PDMS microfluidic chamber, the mold was placed in a large petri dish and covered with a 5:1 PDMS mixture. To form the control layer, a 20:1 PDMS mixture was dessicated then spun onto a clean slilicone wafer (500 rpm for 10s, 1100 rpm for 30s). The PDMS was heated in an 80C° oven to partially set. The fluidic layer was peeled off, and attached to the control layer. The chamber was then heated at 80C° again to set the full chamber.

4.5 References:

[1] Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science 1990;247(4938):64-9.

[2] Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci USA. 1982;79(14):4381-4385.

[3] Robinett CC, Straight A, Li G, Willhelm C, Sudlow G, Murray A, Belmont AS. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol 1996;135(6 Pt 2):1685-700.

[4] Hellwig D, Münch S, Orthaus S, Hoischen C, Hemmerich P, Diekmann S. Live-cell imaging reveals sustained centromere binding of CENP-T via CENP-A and CENP-B. J Biophotonics 2008;1(3):245-54.

[5] Wang, X., Kam, Z., Carlton, P.M. et al. Rapid telomere motions in live human cells analyzed by highly time-resolved microscopy. Epigenetics & Chromatin 1, 4 (2008). https://doi.org/10.1186/1756-8935-1-4

[6] Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 2013;155(7):1479-91.

[7] Römer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science 2007;318(5850):645-8.

[8] Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 2011;39(12):e82.

[9] Kay S, Hahn S, Marois E, Hause G, Bonas U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 2007;318(5850):648-51.

[10] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 2009;326(5959):1509-12.

[11] Garg A, Lohmueller JJ, Silver PA, Armel TZ. Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res 2012;40(15):7584-95.

[12] Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. Science 2011;333(6051):1843-6.

[13] Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res 2011;39(21):9283-93.

[14] Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J. Transcriptional activators of human genes with programmable DNA-specificity. PLoS One 2011;6(5):e19509.

[15] Politz MC, Copeland MF, Pfleger BF. Artificial repressors for controlling gene expression in bacteria. Chem Commun (Camb) 2013;49(39):4325-7.

[16] Miyanari Y, Ziegler-Birling C, Torres-Padilla ME. Live visualization of chromatin dynamics with fluorescent TALEs. Nat Struct Mol Biol 2013;20(11):1321-4.

[17] Moore R, Chandrahas A, Bleris L. Transcription activator-like effectors: a toolkit for synthetic biology. ACS Synth Biol 2014;3(10):708-16.

[18] Cuculis, L, Abil, Z, Zhao, H, Shroeder C. Direct observation of TALE protein dynamics reveals a two-state search mechanism. Nat Commun 6, 7277 (2015). https://doi.org/10.1038/ncomms8277

[19] Visnapuu ML, Fazio T, Wind S, Greene EC. Parallel arrays of geometric nanowells for assembling curtains of DNA with controlled lateral dispersion. Langmuir 2008;24(19):11293-9.

[20] Fazio T, Visnapuu ML, Wind S, Greene EC. DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging. Langmuir 2008;24(18):10524-31.

[21] Roy R, Hohng S, Ha T. A practical guide to single-molecule FRET. Nat Methods 2008;5(6):507-16.

[22] Jain A, Liu R, Xiang YK, Ha T. Single-molecule pull-down for studying protein interactions. Nat Protoc 2012;7(3):445-52.

[23] Ojala H, Ziedaite G, Wallin AE, Bamford DH, Hæggström E. Optical tweezers reveal force plateau and internal friction in PEG-induced DNA condensation. Eur Biophys J 2014;43(2-3):71-9.

[24] Adler J, Parmryd I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry A 2010;77(8):733-42.

[25] Zhu B, Hsieh YP, Murphy TW, Zhang Q, Naler LB, Lu C. MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. Nat Protoc 2019;14(12):3366-94.

Appendix A

Transcription and Transcriptional Regulation

A.1 Transcription:

Transcription is the process of turning the information stored in DNA into RNA. The transcription process consists of 3 stages: initiation, elongation and termination (Figure A.1a). The transcription process is performed by RNA Polymerases. In eukaryotes there are 3 different types of RNA polymerase. RNA Polymerase I is located in the nucleolus. It is responsible for producing ribosomal RNAs (28S, 18S, 5.8S) subunits [1]. RNA Polymerase III is responsible for producing small RNAs such as transfer RNA, 5S Ribosomal RNA, and signal recognition particle RNA [2]. Perhaps the most interesting is RNA Polymerase II. This RNA polymerase is responsible for producing messenger RNA which is then translated into proteins, small nuclear RNAs, small interfering RNAs and microRNAs [3]. RNA Polymerase II contains 12 subunits, 4 of which are shared with the other RNA Polymerases [4] (Figure A.1b). While each subunit plays a unique role, RPB1 is of particular importance. The C-terminal domain of RPB1 contains numerous repeats of YSPTSPS [5]. These amino acids can undergo modifications such as phosphorylation during the transcription cycle in order to control transcription [6]. Transcription begins by formation of the preinitiation complex. RNA Polymerase II and numerous transcription factors described later bind to the gene. Once formed, promoter melting occurs and the transcription bubble is formed. Before fully transcribing the gene, several 2-15 nucleotide RNA chains are generated. These transcripts are aborted before elongation begins [7]. Once the transcript reaches 10

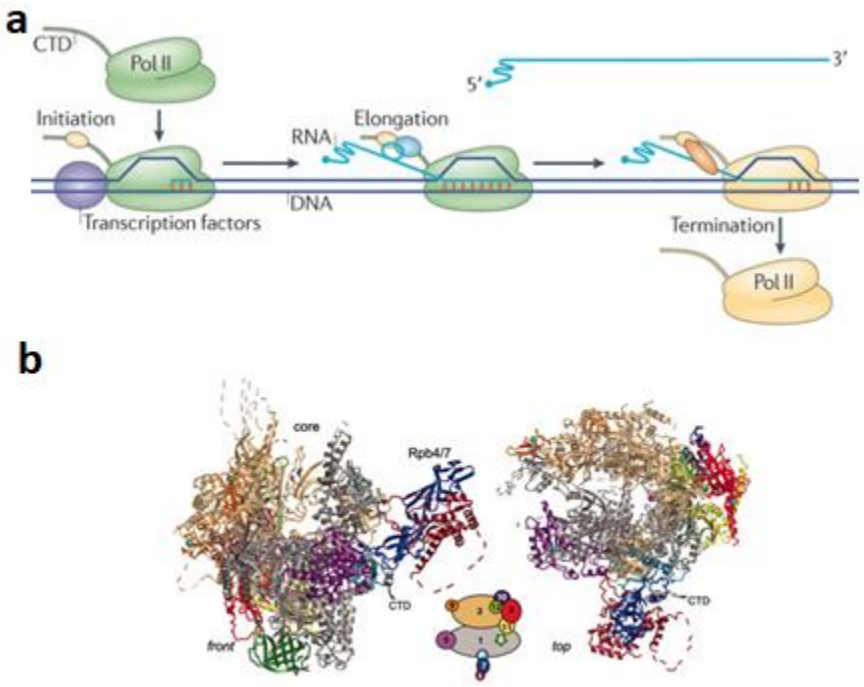


Figure A.1: Overview of transcription and RNA Polymerase II.

(a) RNA Polymerase II is recruited to a gene promoter by transcription factors. The template strand is exposed and an 8-9 nucleotide long DNA-RNA hybrid is created. RNA Polymerase II then extends the transcript. For many genes, approximately 20-60 bp into transcription, RNA Polymerase II undergoes a process called promoter proximal pausing. Following release from this paused state, transcription continues until termination occurs. (b) Ribbon diagram of RNA Polymerase II. Front and top view. Subunits are colored according to the diagram at the bottom. Adapted from [97] and [4]

nucleotides promoter escape occurs. This process requires phosphorylation of the C-terminal domain of RNA Polymerase II [8]. After promoter escape elongation begins. Usually, the elongation process pauses within the first 100 nucleotides in a process called promoter proximal pausing. Two pause-inducing factors, DRB sensitivity inducing factor and negative elongation factor, have been shown to be critical to the pausing process. The process of releasing RNA Polymerase II from its paused state requires P-TEFb. P-TEFb phosphorylates NELF, DSIF, and the C-terminal Domain of RNA Polymerase II. The phosphorylation of NELF causes it to dissociate from the paused transcription complex. The phosphorylation of DSIF turns it into a form that helps the elongation process. The phosphorylation of the C-terminal Domain of RNA Polymerase II allows it to bind additional factors that assist in the elongation process [9].

A.2 Gene Expression Regulation:

The majority of gene expression regulation occurs at the transcription level. This regulation occurs via two modes, trans-regulatory and cis-regulatory [10]. Trans-regulatory elements are molecules separate from DNA that enhance or suppress transcription. These proteins [11] can target and bind to specific DNA sequences. MicroRNAs [12] can provide post-transcriptional control. Conversely cis-regulatory elements consist of non-coding DNA sequences which regulate gene expression [13]. These sequences include promoters and enhancers. Cis-regulatory elements consist of targets for trans-regulatory elements to bind. Together in tandem, they work to regulate gene expression [10].

A.2.1 Trans-regulatory elements:

Perhaps the most important set of trans-regulatory elements are the general transcription factors that bind promoter regions to facilitate transcription. Additional trans-regulatory elements include sequence specific transcription factors. These are proteins which target specific genes or sets of genes unlike general transcription factors which are utilized in all gene transcription [14]. Other trans-regulatory elements include chromatin remodelers [15] which change the chromatin landscape in order to facilitate or repress transcription.

General Transcription Factors:

In bacteria there is a single general transcription factor, sigma factor [16]. This factor along with RNA polymerase are the only proteins needed to facilitate transcription in bacteria. In eukaryotes many more general transcriptions factors exist including: TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH (Figure A.2).

TFIID:

TFIID begins the formation of the pre-initiation complex. TFIID consists of several sub-units including the TATA-binding protein (TBP) and TBP-associated factors (TAFs). TBP binds the TATA-box motif (a cis-regulatory element upstream of gene promoters). In genes without a TATA-box, the TAFs will target Initiator Elements or the core promoter and force TBP to bind non-specifically. Since TFIID is the first general transcription factor to bind DNA, it effectively aligns RNA polymerase to the gene. It acts as a station for other transcription factors and RNA polymerase to bind and form the pre-initiation complex [17].

TFIIA:

TFIIA assists with the formation of the pre-initiation complex. It interacts with TBP to help it bind to the TATA-box. It helps to stabilize the interaction between TBP and the TATA-box [18].

TFIIB:

In eukaryotes, TFIIB serves a similar function as sigma factor in bacteria. TFIIB works by interacting with the TBP subunit of TFIID. It additionally interacts with the B Recognition Element (downstream of the TATA Box). TFIIB then recruits RNA Polymerase II. Following this, TFIIB assists RNA Polymerase II in unwinding DNA and choosing the transcription start site. Once transcription starts TFIIB is ejected from the transcription complex [19].

TFIIF:

TFIIF binds to RNA Polymerase II at first. It remains bound and stabilizes it while it is interacting with TFIIB and TBP. Additionally, it prevents RNA Polymerase II from binding non-specific DNA sequences [20].

TFIIE:

TFIIE binds upstream of the transcription start site. It is responsible for recruiting TFIIH and activating the kinase on RNA Polymerase II. It is thought to be involved in DNA melting at the promoter region [21].

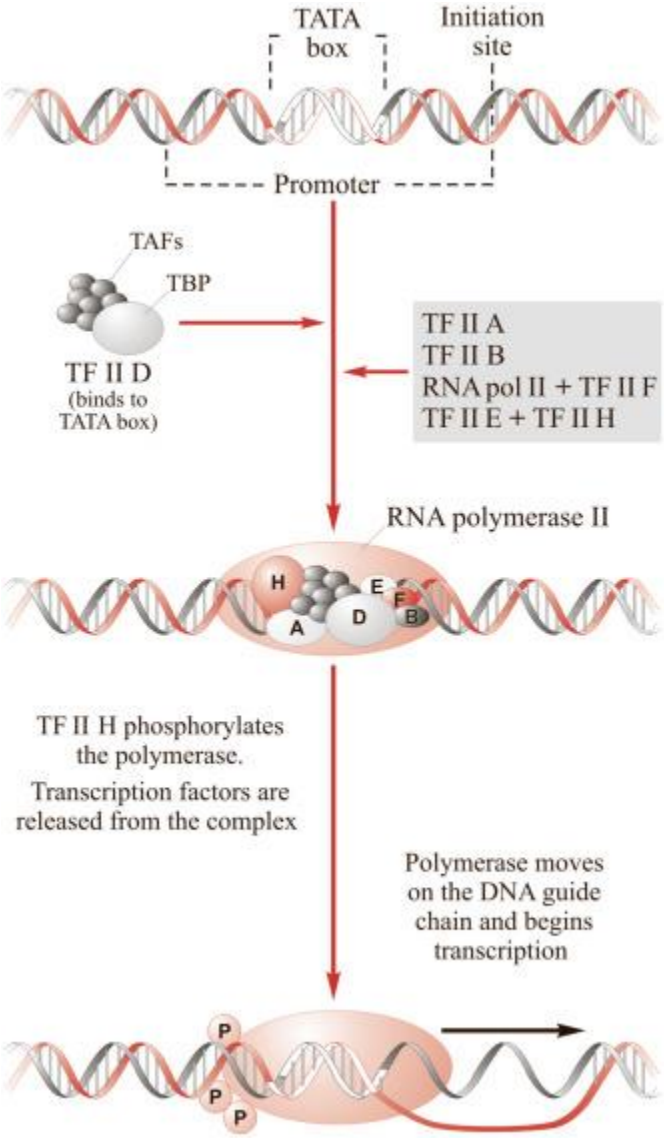


Figure A.2. Overview of general transcription factors.

(a) Several general transcription factors are responsible for formation of the pre-initiation complex. TFIID binds first targeting the TATA box. It then helps recruit many additional factors (TFIIA-TFIIH). After formation of the complex, TFIIH phosphorylates the polymerase to start transcription. Adapted from [98].

TFIIH:

TFIIH is the final general transcription factor to join the transcription complex. TFIIH contains a helicase which unwinds DNA to form the transcription bubble. Additionally, it has a kinase subunit which phosphorylates the C-terminal domain to transition transcription to elongation [22].

Sequence Specific Transcription Factors:

Sequence specific factors are what truly regulate transcription when compared with general transcription factors. There are estimated to be approximately 2600 proteins with DNA binding domains in the human genome and 270 in E. Coli [23]. Transcription factors work with other proteins in order to promote or repress transcription. Sequence specific transcription factors are able to respond to numerous stimuli including developmental signals (Hox) [24], environmental signals (HSF) [25], and cell cycle signals (Myc) [26]. Transcription factors are modular and contain domains with many different functions. They have a DNA binding domain which allows them to specifically target a DNA sequence to perform their regulatory duties [27]. There are many DNA binding domains that have been identified including basic helix-turn-helix [28], basic leucine zipper [29], and srf-like [30]. They also contain an activation domain. This domain binds other transcription coregulators (such as chromatin remodelers) in order to regulate transcription [31]. Transcription factors can also contain a signal-sensing domain in order to respond to external stimuli [32].

A.2.2 Chromatin Structure:

Generally, for transcription to occur the chromatin must be accessible by proteins. In eukaryotes, chromatin is tightly packed in units called nucleosomes. Nucleosomes consist of 147 base pairs of DNA wrapped around a histone octamer [33]. The purpose of this packaging is two-fold. First it ensures large DNA molecules are able to fit in the nucleus. Additionally, it acts as a mechanism to control transcription. Chromatin remodeler complexes are required to make chromatin accessible. Similarly, packaging of chromosomes into nucleosomes can act as a transcription suppression mechanic [34].

Covalent Histone Modifications:

Numerous protein complexes are known to covalently modify histones. The modifications that can occur to histones includes acetylation, methylation, phosphorylation and ubiquitination (Figure A.3). Methylation is one of the most commonly studied histone modifications. Methylation can occur to lysine and arginine residues and is performed by a family of enzymes called Histone Methyltransferases [35]. These proteins transfer up to 3 methyl groups to histones. Methylation of H3K9Me3 has been shown to repress transcription [36]. Methylation of residues H3K4 and H3K36 has been shown to indicate regions of open chromatin and gene expression [37]. Histone acetylation occurs at the lysine residues within the N-terminal tails. Histone acetylation is performed by Histone Acetyltransferases while deacetylation is performed by Histone Deacetylases [38]. These proteins don’t contain DNA binding domains and need to be recruited by other sequence specific transcription factors [39]. Histone acetyltransferases correspond to

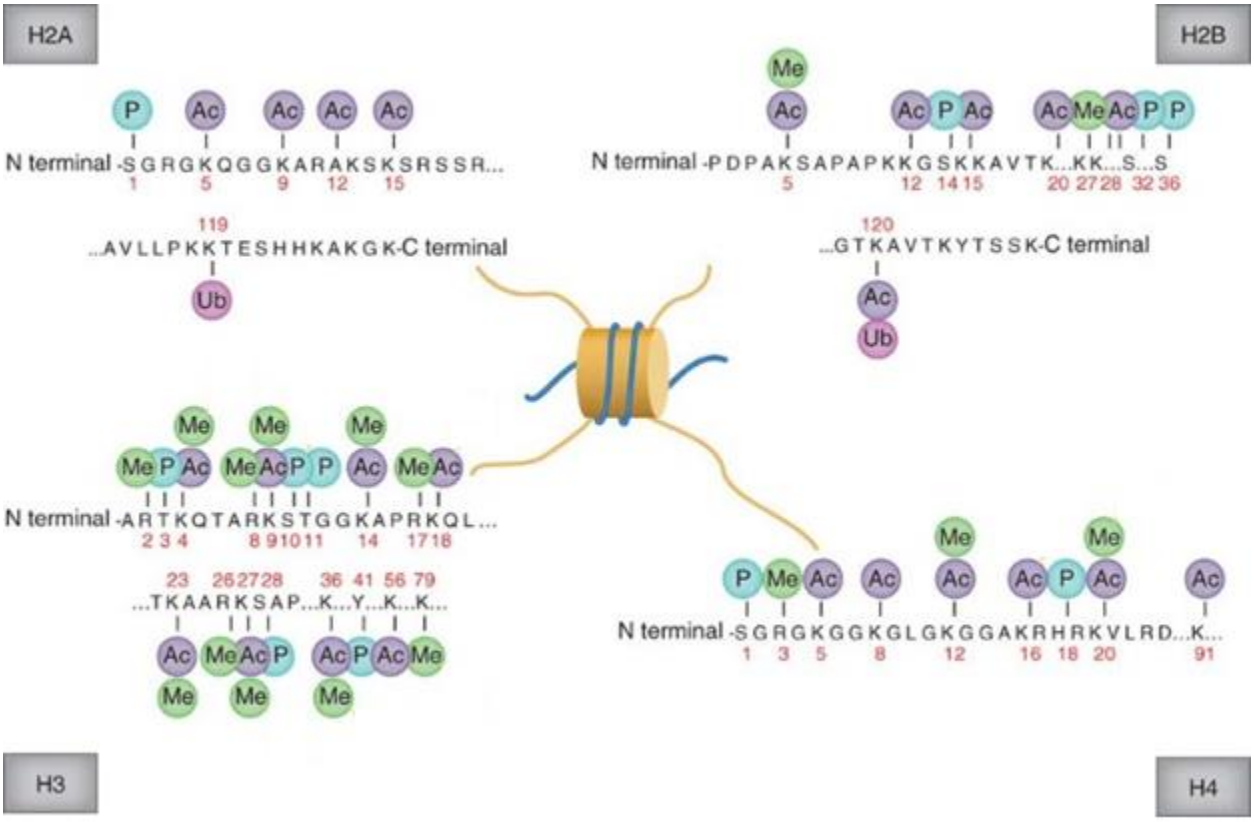


Figure A.3. Histone Modifications.

Overview of many possible histone modifications. Four types of modifications can occur, methylation (Me), Phosphorylation (P), Acetylation (Ac) and Ubiquitination (Ub). Histone modifications tend to occur on the N-terminal tail but can occasionally be found on the C-terminal tail. Many different amino acids can be modified. These modifications can lead to open and closed chromatin states which can be used to regulate transcription. Adapted from [99].

transcriptional activation while histone deacetylases correspond to transcription repression [40]. Many phosphorylated histone residues are directly related to gene expression regulation [41]. Phosphorylation of serines 10 and 28 of H3 has been shown to be associated with transcription regulation of epidermal growth factor [42]. Phosphorylation of H3S10, T11 and S28 have been associated with H3 acetylation which is related to transcription activation [43]. Perhaps the least is known about histone ubiquitination, but it does appear to play a role in transcription regulation. Ubiquitination is the process of adding a 76 amino acid protein, ubiquitin to histones [44]. H2A and H2B are the most frequently ubiquitinated proteins in the nucleus [45]. Ubiquitination of H2A is correlated with gene silencing [46] while ubiquitination of H2B is associated with transcriptional activity [47].

Chromatin Remodeling Complexes:

There are many characterized chromatin remodeling complex families such as SWI/SNF, ISWI, NURD/Mi2, INO80, and SWRI [48]. Although SWI/SNF [49] and ISWI [50] are the most well characterized chromatin remodeling complexes. Both these remodeling complexes are ATP dependent and utilize ATP-hydrolysis to evict or slide nucleosomes [51,52] (Figure A.4). These families are distinct due to the additional subunits surrounding the ATPase domain. SWI/SNF remodelers are large complexes composed of about 15 protein subunits. Several of these subunits including SMARCB1 and SMARCC1 are conserved between SWI/SNF complexes [53]. Activity by these remodelers results in the release or movement of nucleosomes which allows transcription factor binding and accompanying transcription.

A.2.3 DNA Methylation:

DNA methylation is a widely noticed phenomena in a variety of species [54]. It can occur on both cytosine and adenine residues [55]; however, cytosine methylation is much more well characterized and studied. The methylation of DNA is performed by DNA methyltransferases (DMNTs). DMNT1 serves a maintenance function and is responsible for copying DNA methylation following replication [56]. DMNT3A and DMNT3B are responsible for methylating previously unmethylated CpG sites [57]. CpG sites are found in either short regions in regulatory DNA or large regions of repetitive DNA [58]. Methylation in regulatory DNA has been shown to be linked to developmental [58] and differentiation phenomena [59]. It is believed that DNA methylation at promoters is linked to repression of transcription [54].

A.2.4 Cis-Regulatory Elements:

While trans-regulatory elements consist of separate molecules that regulate transcription, cis-regulatory elements are DNA sequences that operate to regulate transcription (Figure A.5). These elements can be located both proximal or distal to the gene they are regulating. Cis-regulatory elements include promoters, enhancers, repressors, and insulators.

Promoters:

The promoter region (Figure A.5a) for RNA Polymerase II consists of multiple different elements. The core promoter is located between -35 and +35 base pairs of the transcription start site. The core promoter contains the TATA box which consists

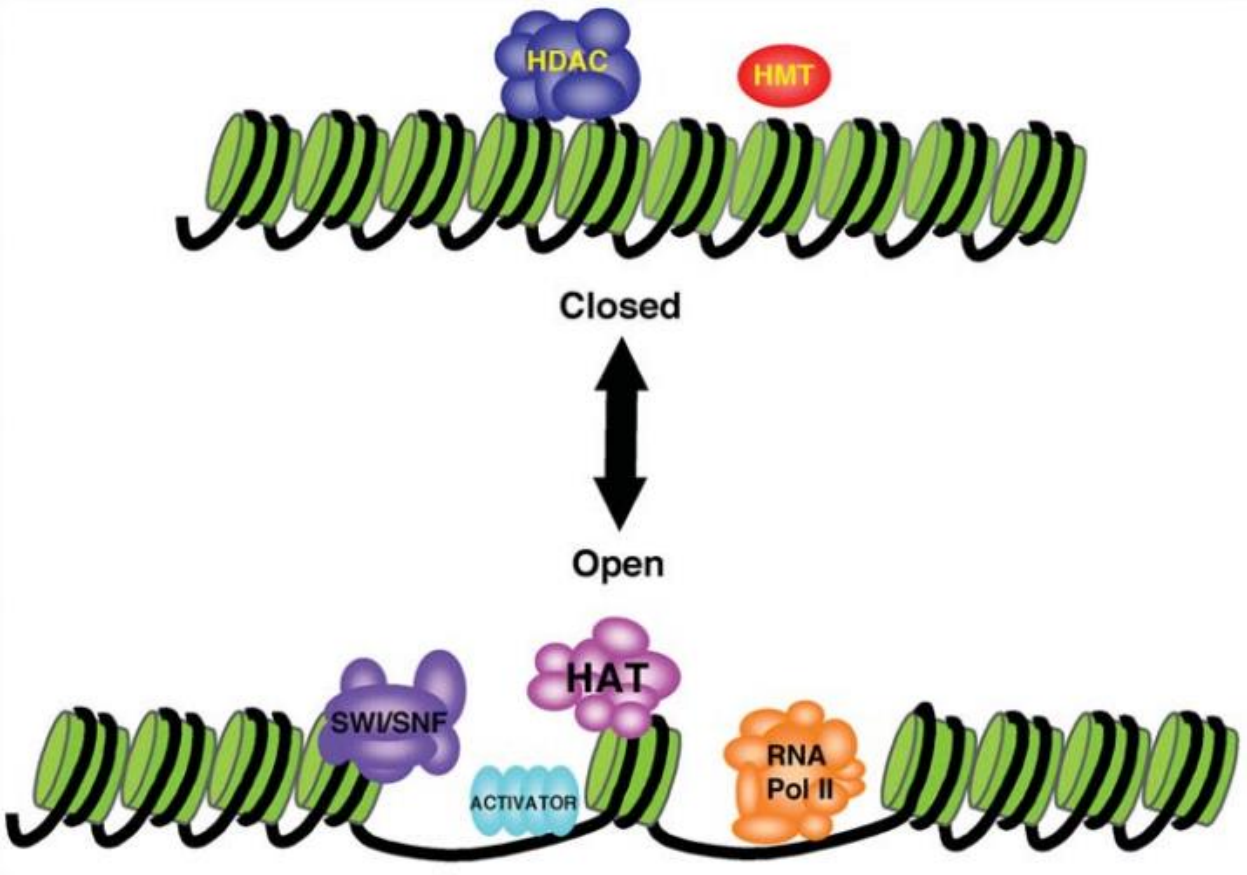


Figure A.4. Overview of Chromatin Remodeling Complexes.

Open chromatin is required for transcription to allow access for transcription machinery. Conformational changes in chromatin structure are regulated by a variety of mechanisms. Closed chromatin is maintained by Histone deacetylases (HDAC) and histone methyltransferases (HMT). Open chromatin is maintained by transcription factors (activators), Histone acetyltransferases (HAT) and ATP dependent remodelers (SWI/SNF). Transcription can be regulated by these proteins by creating open chromatin states which facilitate transcription and closed chromatin states which limit transcription. Adapted from [100].

of the sequence TATAAAA approximately 25 base pairs upstream of the transcription start site in eukaryotes. The TATA box is the target binding site of TFIID [60]. The initiator element is located at the transcription start site from -3 to +5 and has the sequence YYANWYY [61]. It is found in approximately 49% of human promoter sequences, significantly more than the TATA box which is found in approximately 22% of human promoter sequences [62]. Approximately 32 to 37 base pairs upstream of the transcription start site is the TFIIB Recognition Element. This element is composed of the sequence SSRCGCC. As the name suggests, this element is recognized by TFIIB [63]. Downstream of the transcription start site is the downstream promoter element. This element consists of the sequence RGWYV and is located approximately 28-32 base pairs downstream of the transcription start site [64]. Approximately 70% of human promoters contain a CpG Island through which transcriptional regulation can occur by DNA methylation [55].

Enhancers:

Transcriptional machinery only requires the sequences around the transcription start site known as the promoter region. While this is enough to start transcription, it is fairly weak without other regulatory DNA sequences. The presence of enhancer sequences which can be moderately distal from the transcription start site are required to promote strong transcription (Figure A.5b) [65]. Enhancer sequences are short DNA sequences that promote the binding of sequence specific transcription factors. As stated earlier, these transcription factors recruit additional proteins to help facilitate transcription. Enhancers have other features which distinguish them from other cis-regulatory elements, in that they are able to operate at a very long distance from the

transcription start site they are regulating [66]. There are many different suggested mechanisms for how this occurs. Perhaps the most widely accepted mechanism is the formation of chromatin loops. In this suggested model, chromatin forms a loop to bring a distal enhancer/promoter pair spatially close together and activate transcription [66]. Another feature which distinguishes enhancers is their ability to work regardless of what promoter they are near. Enhancers tend to continue to function if moved to a separate reporter gene [67]. Finally, enhancers behave in a modular method. Reporter assays with multiple different enhancers report a pattern consistent with their combined activity [67].

Silencers:

Silencers are functionally the opposite of enhancers (Figure A.5c). While enhancers recruit activator proteins to facilitate transcription, silencers recruit repressor proteins to decrease transcription [68,69]. They can be located in the promoter region, distal from the transcription start sire or downstream of the gene they regulate. Several types of silencers have bene reported. Classical silencers bind proteins to directly block transcription initiation [70]. Negative Regulatory Elements can be located further away from genes and prevent transcription factors from binding to their target or inhibit transcription elongation from proceeding [71]. Repressors can also recruit chromatin-remodeling enzymes to remove regions of open chromatin and prevent the binding of transcription factors.

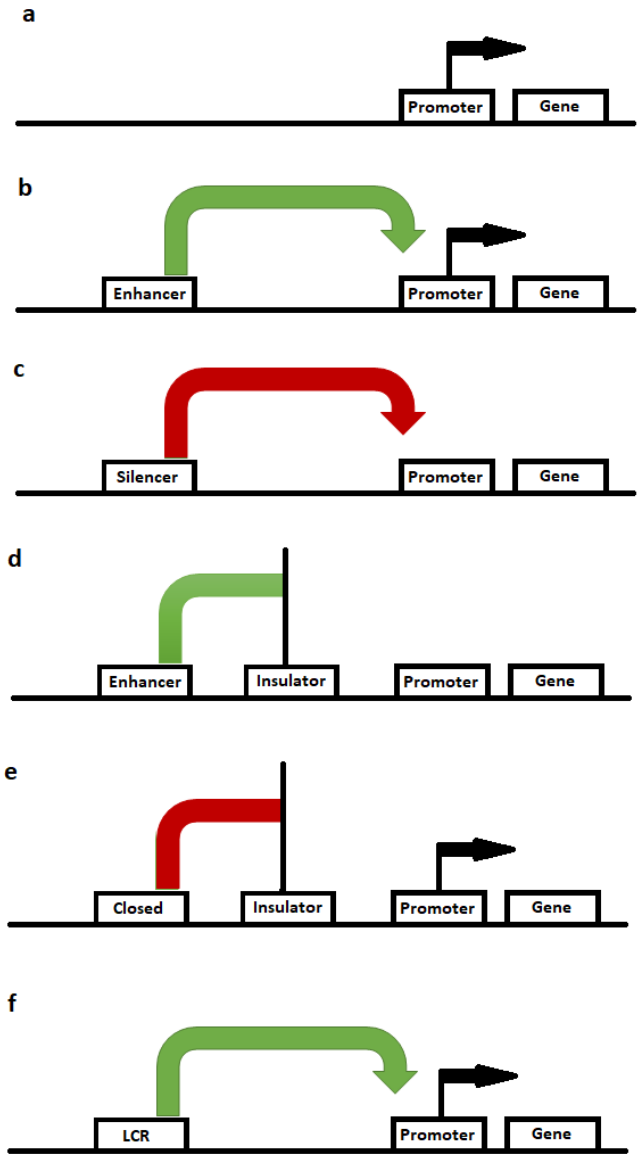


Figure A.5. Cis regulatory elements.

(a) Illustration of a simple gene. A promoter element upstream of the gene body helps form the pre-initiation complex. (b) Distal enhancer upstream contributes to transcriptional activity. (c) Silencer upstream of the gene complex limits transcription. (d) Insulator between an enhancer and gene prevents the enhancer from effecting transcriptional activity of the gene. (e) Insulator blocks heterochromatin from spreading to a gene, allowing the gene to be freely transcribed. (f) Locus Control Region regulating expression of a gene.

Insulators:

Insulators serve a unique function compared to enhancers and silencers. Insulators are regions of DNA 300 to 2000 base pairs in length that function as either a barrier or a blocker for enhancers [72]. They can prevent distal enhancers by acting on nearby genes (Figure A.5d). The other function they serve is preventing the spread of heterochromatin (Figure A.5e) [73]. Insulators are critical when genes that have different transcription patterns are near each other.

Locus Control Regions:

Locus Control Regions (LCRs) are cis-regulatory elements which functions to enhance gene expression at distal loci (Figure A.5f). The first LCR was identified as a regulator of the human β-globin gene locus [74]. Early studies of the β-globin gene showed that a 5kb gene segment did not contain the necessary regulatory elements to provide meaningful levels of transcription [74]. LCRs have several properties that make them unique. First, LCRs show extremely strong transcription enhancing activity. In the absence of the LCR, transcription of the human β-globin gene is less than 1% of the normal transcription level [75]. LCRs can be located several kilobases away (6 to 22kb upstream in the case of β -globin) [74]. LCR activity is tissue specific, for the β-globin LCR activity is limited to Erythroid cells [76]. LCRs tend to be colocalized with DNase I hypersensitive sites [77]. While enhancers share many of these features, LCRs differ in their ability to provide copy number-dependent expression [74]. Copy number dependent expression is something commonly associated with open chromatin states which is substantiated by the colocalization of DNase I hypersensitive sites [78].

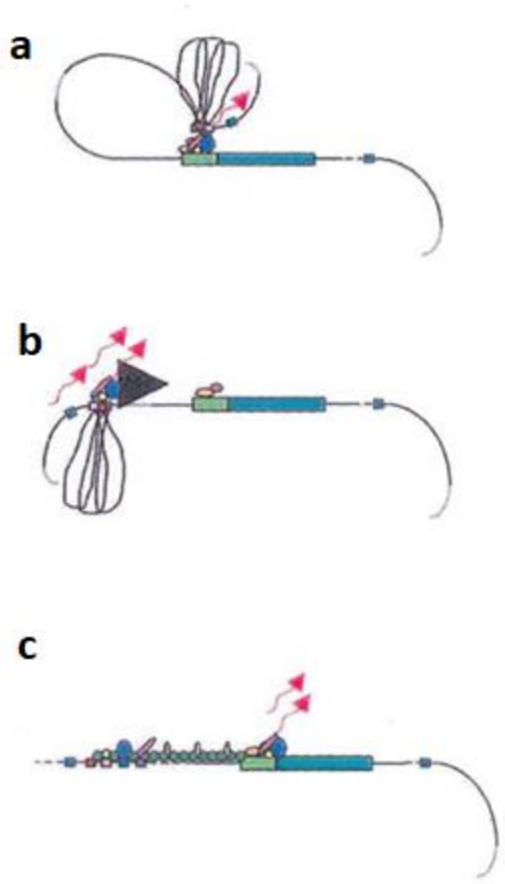


Figure A.6. Locus Control Region Model Mechanisms.

(a) Looping model. This model suggests that a chromatin loop is formed. This brings the distal LCR spatially close to the promoter of the gene it is regulating. (b) Tracking model. Transcription factors first bind the LCR forming a complex. This complex then moves down the chromatin until it reaches additional transcription factors at a promoter. Thus, the transcription factors originally bound to the LCR end up at the promoter to facilitate transcription. (c) Linking model. Transcription factors continually bind the LCR to create a link between the LCR and the gene it regulates. Adapted from [101].

However, the mechanism through which the β -globin LCR operates remains poorly understood. Several models have been hypothesized including looping, tracking, and linking (Figure A.6). One shared feature of these models is the ability of the LCR to create an open chromatin structure. The looping model suggests that the LCR forms a loop to bring the LCR to the regulated promoter [79]. The tracking model has transcription factors bind the LCR forming a complex which then moves along chromatin until it reaches a promoter [80]. The linking model suggests that facilitator proteins mediate the binding of transcription factors to regulate transcription [81]. Each model is still being thoroughly researched in order to determine the method through which LCRs operate. While the most well characterized LCR is that of the β-globin gene, several other similar elements have been identified in a variety of species [82].

A.3 Methods to examine regulatory elements:

Due to the importance of understanding transcription regulation, several techniques have been implemented to look at regulatory factors. These techniques include both experimental and computation methods. These methods allow researchers to examine where proteins that regulate transcription bind and what sequences serve a regulatory function.

A.3.1 DNase I Hypersensitivity Assays:

On the genome wide scale, understanding what regions of chromatin are bound and unbound by DNA binding proteins is important to understand. One way this is accomplished is by using DNase I hypersensitivity assays [83] (Figure A.7). In this

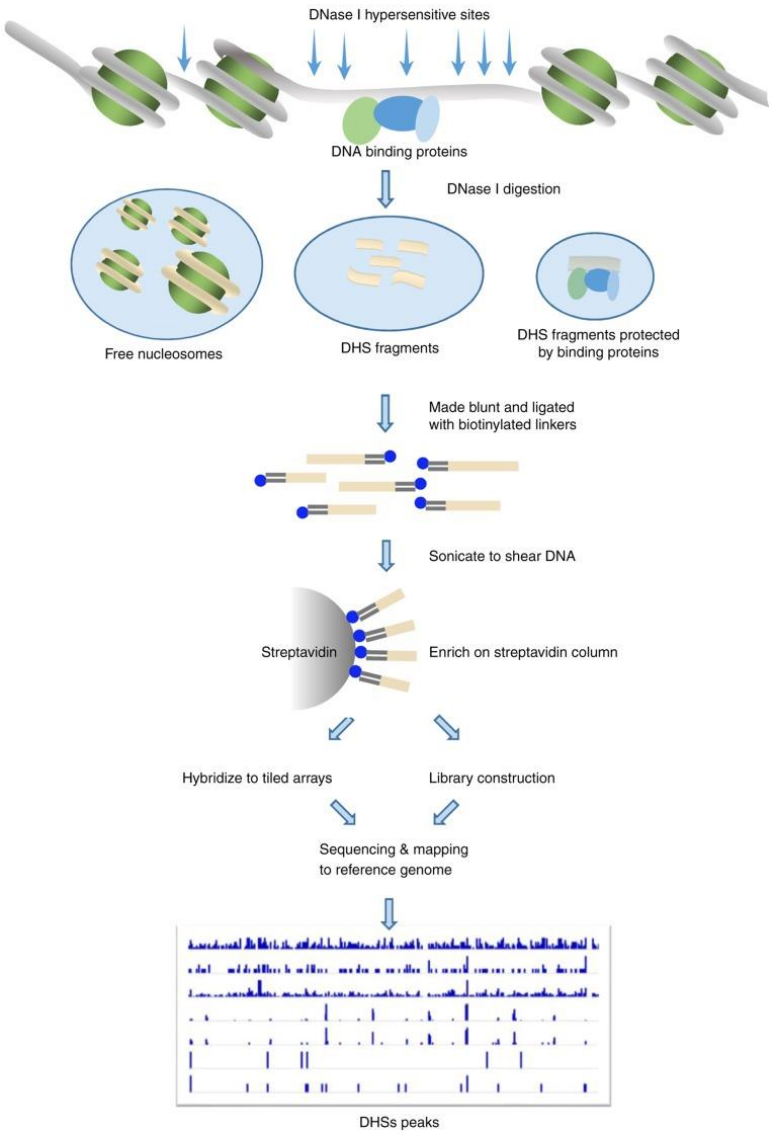


Figure A.7. DNase I Seq.

Overview of DNase I Seq. First nuclei are cleaved using DNase I endonuclease. This endonuclease is effective at cleaving open chromatin while leaving condensed heterochromatin intact. Cleaved fragments are made blunt then biotinylated linkers are ligated on. After further shearing the linker is used to enrich these fragments using a streptavidin pulldown. Analysis can then occur using a microarray or high throughput sequencing. The resulting mapped sequences were from open chromatin regions, generally including coding regions and non-coding functional cis regulatory elements. Adapted from [102].

technique nuclei are exposed to endonuclease DNase I. This endonuclease cleaves open chromatin, leaving behind small DNA fragments. These fragments from open chromatin regions can then be analyzed through sequencing [84], microarray analysis [85] or electrophoresis [86]. DNase I Hypersensitive Sites have several interesting features. They primarily reside around promoters and actively transcribing genes [87]. They can vary between cell type and gene expression profile [83]. They are usually correlated with histone modifications that correspond to open chromatin such as H3K27Me3 [88].

A.3.2 Reporter Gene Assays:

Identification of sequences bound by DNA binding proteins is useful information; however, it does not truly illustrate how those sequences function in a regulatory fashion. Reporter assays were developed in order to fill that niche. In these assays, suspected regulatory sequences are inserted into a plasmid containing a gene which encodes for a reporter protein (such as luciferase or green fluorescent protein). The plasmid is then transfected into cells to measure the transcription of the reporter gene [89] (Figure A.8). This method has been adapted as a genome wide approach to look for cis regulatory elements. In STARR-seq, for example, random genomic fragments are created then inserted into a reporter plasmid downstream of a minimal promoter allowing transcription of an mRNA transcript containing the genomic fragment sequence and a poly-A tail. The reporter plasmid is then transfected into cells and allowed to transcribe. All RNA is extracted and the poly-A tail is isolated. Reverse transcription is performed on the isolated RNA and a library is created and sequenced to provide a set of sequences with enhancer like qualities [90].

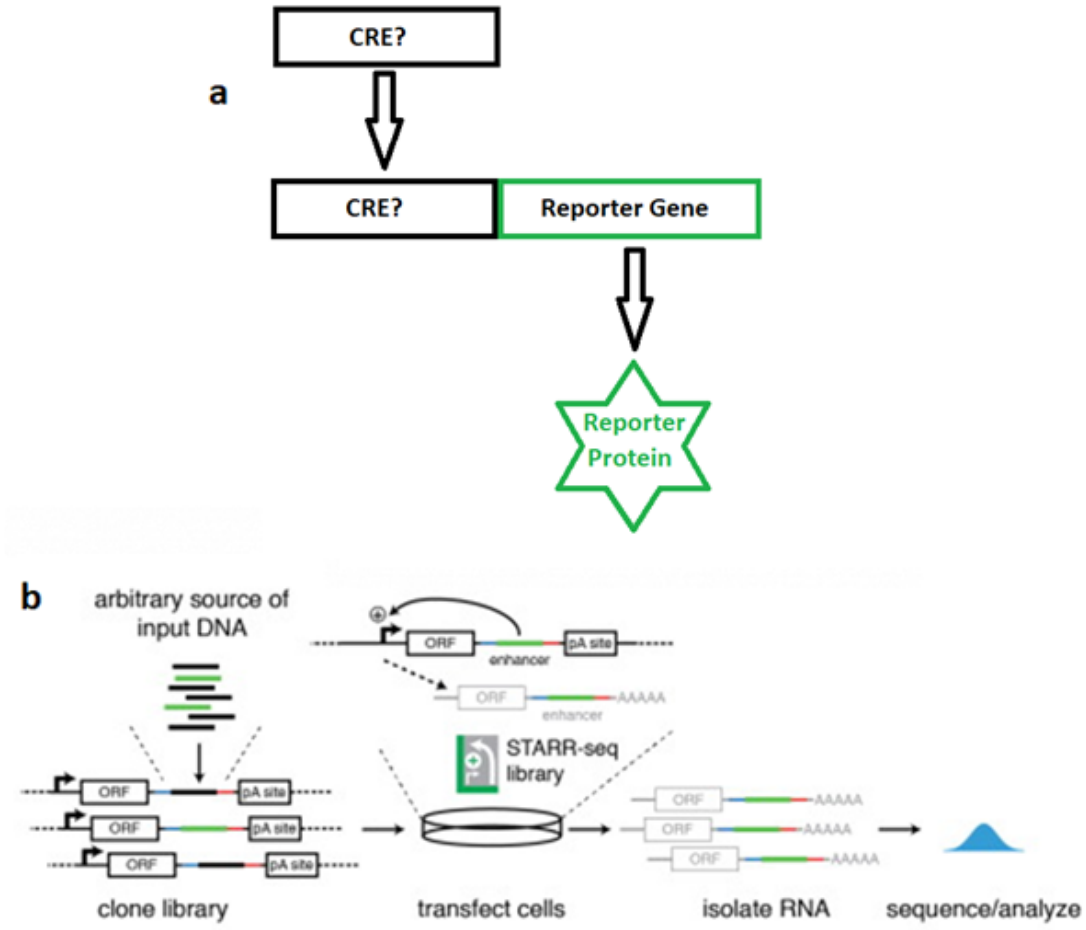


Figure A.8. Overview of Reporter Gene Assays.

(a) Simple illustration of a reporter gene assay. A promoter of interest (or other regulatory sequence) is inserted into a plasmid containing a gene which encodes for a reporter protein. Measurement of the reporter protein is used to determine the functionality of the candidate inserted regulatory sequence. (b) Overview of STARR-Seq. Candidate sequences are inserted downstream of a promoter. Sequences with enhancer functionality will cause activation of the nearby promoter, creating a transcript which contains the enhancer sequence. By extracting all the transcribed RNA and performing reverse transcription and sequencing. Mapped reads will show sequences with enhancer functionality. Adapted from [103].

A.3.3 Chromatin Immunoprecipitation:

Chromatin Immunoprecipitation (ChIP) is a very robust technique to look at protein DNA interactions [91]. In a ChIP experiment, DNA and DNA binding proteins are crosslinked using formaldehyde. After this, the DNA-protein complexes are sheared into smaller sizes. The DNA fragments are then enriched for a specific DNA binding protein using an antibody. The crosslinks are reversed and the enriched DNA is analyzed to provide a snapshot of the binding sites of specific DNA binding proteins [92]. A variety of analysis techniques have been used to map the extracted DNA to the genome including microarray [93] and sequencing [94] analysis. This technique has been used to look at the specific sites of activity for certain transcription factors [95]. Similarly, it has been used to map the location of various histone modifications [96]. Along with other described techniques, ChIP experiments have helped provide a more detailed protein-DNA interaction map for the genome.

A.4 References:

[1] Russell J, Zomerdijk JC. The RNA polymerase I transcription machinery. Biochem Soc Symp (2006);203-16.

[2] White R.J. (1998) *RNA Polymerase III Transcription*. Springer-Verlag, New York, NY.

[3] Kornberg RD. Eukaryotic transcriptional control. Trends Cell Biol 9: (1999);M46-9.

[4] Armache KJ, Mitterweger S, Meinhart A, Cramer P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem 280: (2005);7131-4.

[5] Ahearn JM, Bartolomei MS, West ML, Cisek LJ, Corden JL. Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J Biol Chem 262: (1987);10695-705.

[6] Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell 61: (2016);297-304.

[7] Goldman SR, Ebright RH, Nickels BE. Direct detection of abortive RNA transcripts in vivo. Science 324: (2009);927-8.

[8] Rodriguez CR, Cho EJ, Keogh MC, Moore CL, Greenleaf AL, Buratowski S. Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Mol Cell Biol 20: (2000);104-12.

[9] Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet 13: (2012);720-31.

[10] Wang Q, Jia Y, Wang Y, Jiang Z, Zhou X, Zhang Z, Nie C, Li J, Yang N, Qu L. Evolution of cis- and trans-regulatory divergence in the chicken genome between two contrasting breeds analyzed using three tissue types at one-day-old. BMC Genomics 20: (2019);933.

[11] Ghosh S, Baltimore D. Activation in vitro of NF-kappa B by phosphorylation of its inhibitor I kappa B. Nature 344: (1990);678-82.

[12] Bartel DP. Metazoan MicroRNAs. Cell 173: (2018);20-51.

[13] Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12: (2011);283-93.

[14] Matsui T, Segall J, Weil PA, Roeder RG. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem 255: (1980);11992-6.

[15] Workman JL, Kingston RE. Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science 258: (1992);1780-4.

[16] Kang JG, Hahn MY, Ishihama A, Roe JH. Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from Streptomyces coelicolor A3(2). Nucleic Acids Res 25: (1997);2566-73.

[17] Burley SK, Roeder RG. Biochemistry and structural biology of transcription factor IID (TFIID). Annu Rev Biochem 65: (1996);769-99.

[18] Liu Q, Gabriel SE, Roinick KL, Ward RD, Arndt KM. Analysis of TFIIA function In vivo: evidence for a role in TATA-binding protein recruitment and gene-specific activation. Mol Cell Biol 19: (1999);8673-85.

[19] Ha I, Lane WS, Reinberg D. Cloning of a human gene encoding the general transcription initiation factor IIB. Nature 352: (1991);689-95.

[20] Čabart P, Újvári A, Pal M, Luse DS. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. Proc Natl Acad Sci U S A 108: (2011);15786-91.

[21] Compe E, Genes CM, Braun C, Coin F, Egly JM. TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. Nat Commun 10: (2019);2084.

[22] Rimel JK, Taatjes DJ. The essential and multifunctional TFIIH complex. Protein Sci 27: (2018);1018-37.

[23] Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr Op in Structural Biology 14: (2004);283-91.

[24] Lewis EB. A gene complex controlling segmentation in Drosophila. Nature 276: (1978);565-70.

[25] Pelham HR. A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. Cell 30: (1982);517-28.

[26] Blackwood EM, Eisenman RN. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science 251: (1991);1211-7.

[27] Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245: (1989);371-8.

[28] Anderson WF, Ohlendorf DH, Takeda Y, Matthews BW. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature 290: (1981);754-8.

[29] Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240: (1988);1759-64.

[30] Minty A, Kedes L. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol Cell Biol 6: (1986);2125-36.

[31] Ma J, Ptashne M. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 48: (1987);847-53.

[32] Shingler V. Signal sensing by sigma 54-dependent regulators: derepression as a control mechanism. Mol Microbiol 19: (1996);409-16.

[33] Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: (1997);251-60.

[34] Polach KJ, Widom J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J Mol Biol 254: (1995);130-49.

[35] Wood A, Shilatifard A. Posttranslational modifications of histones by methylation. Adv Protein Chem 67: (2004);201-22.

[36] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell 129: (2007);823-37.

[37] Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell 128: (2007);707-19.

[38] Yang XJ. Lysine acetylation and the bromodomain: a new partnership for signaling. Bioessays 26: (2004);1076-87.

[39] Roth SY, Denu JM, Allis CD. Histone acetyltransferases. Annu Rev Biochem 70: (2001);81-120.

[40] Vidali G, Gershey EL, Allfrey VG. Chemical Studies of Histone Acetylation. The Distribution of ϵ-*N*-Acetyllysine in Calf Thymus Histones. J. Biol. Chem. 243: (1968); 6361– 6366.

[41] Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics 7: (2012);1098-108.

[42] Lau AT, Lee SY, Xu YM, Zheng D, Cho YY, Zhu F, Kim HG, Li SQ, Zhang Z, Bode AM, Dong Z. Phosphorylation of histone H2B serine 32 is linked to cell transformation. J Biol Chem 286: (2011);26628-37.

[43] Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, Marmorstein R, Berger SL. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell 5: (2000);917-26.

[44] Goldknopf IL, Taylor CW, Baum RM, Yeoman LC, Olson MO, Prestayko AW, Busch H. Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein. J Biol Chem 250: (1975);7182-7.

[45] West MH, Bonner WM. Histone 2B can be modified by the attachment of ubiquitin. Nucleic Acids Res 8: (1980);4671-80.

[46] Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science 296: (2002);1132-6.

[47] Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol 10: (2008);483-8.

[48] Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol 18: (2017);407-22.

[49] Côté J, Quinn J, Workman JL, Peterson CL. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science 265: (1994);53-60.

[50] Dang W, Bartholomew B. Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. Mol Cell Biol 27: (2007);8306-17.

[51] Tsukiyama T, Daniel C, Tamkun J, Wu C. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell 83: (1995);1021-6.

[52] Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer 11: (2011);481-92.

[53] Han Y, Reyes AA, Malik S, He Y. Cryo-EM structure of SWI/SNF complex bound to a nucleosome. Nature 579: (2020);452-5.

[54] Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: (2010);916-9.

[55] Ji P, Wang X, Xie N, Li Y. N6-Methyladenosine in RNA and DNA: An Epitranscriptomic and Epigenetic Player Implicated in Determination of Stem Cell Fate. Stem Cells Int 2018: (2018);3256524.

[56] Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 279: (2004);48350-9.

[57] Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19: (1998);219-20.

[58] Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 16: (2002);6-21.

[59] Ehrlich M, Lacey M. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. Epigenomics 5: (2013);553-68.

[60] Carninci P, et. al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38: (2006);626-35.

[61] Smale ST, Baltimore D. The "initiator" as a transcription control element. Cell 57: (1989);103-13.

[62] Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389: (2007);52-65.

[63] Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12: (1998);34-44.

[64] Burke TW, Kadonaga JT. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev 10: (1996);711-24.

[65] Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33: (1983);729-40.

[66] Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell 16: (2009);47-57.

[67] Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. Development 124: (1997);1851-64.

[68] Gilbert W, Müller-Hill B. Isolation of the lac repressor. Proc Natl Acad Sci U S A 56: (1966);1891-8.

[69] Ptashne M. Specific binding of the lambda phage repressor to lambda DNA. Nature 214: (1967);232-4.

[70] Kim MK, Lesoon-Wood LA, Weintraub BD, Chung JH. A soluble transcription factor, Oct-1, is also found in the insoluble nuclear matrix and possesses silencing activity in its alanine-rich domain. Mol Cell Biol 16: (1996);4366-77.

[71] Dong JM, Lim L. The human neuronal alpha 1-chimaerin gene contains a position-dependent negative regulatory element in the first exon. Neurochem Res 21: (1996);1023-30.

[72] Kellum R, Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. Cell 64: (1991);941-50.

[73] Pikaart MJ, Recillas-Targa F, Felsenfeld G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. Genes Dev 12: (1998);2852-62.

[74] Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51: (1987);975-85.

[75] Magram J, Chada K, Costantini F. Developmental regulation of a cloned adult beta-globin gene in transgenic mice. Nature 315: (1985);338-40.

[76] Blom van Assendelft G, Hanscombe O, Grosveld F, Greaves DR. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. Cell 56: (1989);969-77.

[77] Tuan D, Solomon W, Li Q, London IM. The "beta-like-globin" gene domain in human erythroid cells. Proc Natl Acad Sci U S A 82: (1985);6384-8.

[78] Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M. A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev 4: (1990);1637-49.

[79] Dunaway M, Dröge P. Transactivation of the Xenopus rRNA gene promoter by its enhancer. Nature 341: (1989);657-9.

[80] Tuan D, Kong S, Hu K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. Proc Natl Acad Sci U S A 89: (1992);11219-23.

[81] Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev 13: (1999);2465-77.

[82] Fu L, Mambrini M, Perrot E, Chourrout D. Stable and full rescue of the pigmentation in a medaka albino mutant by transfer of a 17 kb genomic clone containing the medaka tyrosinase gene. Gene 241: (2000);205-11.

[83] Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, Weintraub H. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell 20: (1980);451-60.

[84] Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, Zhou D, Luo S, Vasicek TJ, Daly MJ, Wolfsberg TG, Collins FS. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res 16: (2006);123-31.

[85] Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods 3: (2006);503-9.

[86] Wu C. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature 286: (1980);854-60.

[87] Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. Annu Rev Biochem 57: (1988);159-97.

[88] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell 132: (2008);311-22.

[89] Nguyen VT, Morange M, Bensaude O. Firefly luciferase luminescence assays using scintillation counters for quantitation in transfected mammalian cells. Anal Biochem 171: (1988);404-8.

[90] Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339: (2013);1074-7.

[91] Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci U S A 81: (1984);4275-9.

[92] O'Neill LP, Turner BM. Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. EMBO J 14: (1995);3946-57.

[93] Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, Farnham PJ. Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques 43: (2007);791-7.

[94] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: (2007);553-60.

[95] Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409: (2001);533-8.

[96] Hebbes TR, Thorne AW, Clayton AL, Crane-Robinson C. Histone acetylation and globin gene switching. Nucleic Acids Res 20: (1992);1017-22.

[97] Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol 12: (2011);283-94.

[98] Blanco A, Blanco G. Chapter 21 - The Genetic Information (I). Medical Biochem: (2017);465-492.

[99] Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nat Med 17: (2011);330-9.

[100] Davis PK, Brackmann RK. Chromatin remodeling and cancer. Cancer Biol Ther 2: (2003);22-9.

[101] Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. Blood 100: (2002);3077-86.

[102] Chen Y, Chen A. Unveiling the gene regulatory landscape in diseases through the identification of DNase I-hypersensitive sites. Biomed Rep 11: (2019);87-97.

[103] Muerdter F, Boryń ŁM, Arnold CD. STARR-seq - principles and applications. Genomics 106: (2015);145-50.

[108] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 295: (2002);1306-11.

[109] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10: (2002);1453-65.

[110] Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38: (2006);1341-7.

[111] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38: (2006);1348-54.

[112] Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16: (2006);1299-309.

[113] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: (2009);289-93.

[114] Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 43: (2011);1059-65.

[115] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502: (2013);59-64.

[116] Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A 79: (1982);4381-5.

[117] Amakawa G, Ikemoto K, Ito H, Furuya T, Sasaki K. Quantitative analysis of centromeric FISH spots during the cell cycle by image cytometry. J Histochem Cytochem 61: (2013);699-705.

[118] Meeker AK, Gage WR, Hicks JL, Simon I, Coffman JR, Platz EA, March GE, De Marzo AM. Telomere length assessment in human archival tissues: combined telomere fluorescence in situ hybridization and immunostaining. Am J Pathol 160: (2002);1259-68.

[119] Vitturi R, Libertini A, Armetta F, Sparacino L, Colomba MS. Chromosome analysis and FISH mapping of ribosomal DNA (rDNA), telomeric (TTAGGG)n and (GATA)n repeats in the leech Haemopis sanguisuga (L.) (Annelida: Hirudinea). Genetica 115: (2002);189-94.

[120] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2: (2001); 292–301.

[121] Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol 16: (2015);180.

[122] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: (2012);376-80.

[123] Kim YJ, Cecchini KR, Kim TH. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. Proc Natl Acad Sci U S A 108: (2011);7391-6.

[124] van Bemmel JG, Galupa R, Gard C, Servant N, Picard C, Davies J, Szempruch AJ, Zhan Y, Żylicz JJ, Nora EP, Lameiras S, de Wit E, Gentien D, Baulande S, Giorgetti L, Guttman M, Hughes JR, Higgs DR, Gribnau J, Heard E. The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. Nat Genet 51: (2019);1024-34.

[125] Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A, Gelfand MS, Shevelyov YY, Razin SV. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res 26: (2016);70-84.

[126] Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG, Flicek P, Dekker J, Merkenschlager M. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res 23: (2013);2066-77.

[127] Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A 111: (2014);996-1001.

[128] Kaushal A, Mohana G, Dorier J, Özdemir I, Omer A, Cousin P, Semenova A, Taschner M, Dergai O, Marzetta F, Iseli C, Eliaz Y, Weisz D, Shamim MS, Guex N, Lieberman Aiden E, Gambetta MC. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat Commun 12: (2021);1011.

[129] Wang Q, Sun Q, Czajkowsky DM, Shao Z. Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. Nat Commun 9: (2018);188.

[130] Liu C, Cheng YJ, Wang JW *et al.* Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. Nature Plants 3: (2017); 742–748.

[131] Parada LA, McQueen PG, Misteli T. Tissue-specific spatial organization of genomes. Genome Biol 5: (2004);R44.

[132] Tanabe H, Müller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci U S A 99: (2002);4424-9.

[133] Bridger JM. Chromobility: the rapid movement of chromosomes in interphase nuclei. Biochem Soc Trans 39: (2011);1747-51.

[134] Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 4: (2006);e138.

[135] Strambio-De-Castillia C, Niepel M, Rout MP. The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11: (2010);490-501.

[136] Houben F, Willems CH, Declercq IL, Hochstenbach K, Kamps MA, Snoeckx LH, Ramaekers FC, Broers JL. Disturbed nuclear orientation and cellular migration in A-type lamin deficient cells. Biochim Biophys Acta 1793: (2009);312-24.

[137] Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD. Nuclear lamins. Cold Spring Harb Perspect Biol 2: (2010);a000547.

[138] Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B. Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38: (2006);1005-14.

[139] Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38: (2010);603-13.

[140] Amendola M, van Steensel B. Mechanisms and dynamics of nuclear lamina-genome interactions. Curr Opin Cell Biol 28: (2014);61-8.

[141] Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L, van Steensel B. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res 23: (2013);270-80.

[142] Brickner JH, Walter P. Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol 2: (2004);e342.

[143] Casolari JM, Brown CR, Komili S, West J, Hieronymus H, Silver PA. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. Cell 117: (2004);427-39.

[144] Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol 12: (2011);283-94.

[145] Blanco A, Blanco G. Chapter 21 - The Genetic Information (I). Medical Biochem: (2017);465-492.

[146] Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nat Med 17: (2011);330-9.

[147] Davis PK, Brackmann RK. Chromatin remodeling and cancer. Cancer Biol Ther 2: (2003);22-9.

[148] Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. Blood 100: (2002);3077-86.

[149] Chen Y, Chen A. Unveiling the gene regulatory landscape in diseases through the identification of DNase I-hypersensitive sites. Biomed Rep 11: (2019);87-97.

[150] Muerdter F, Boryń ŁM, Arnold CD. STARR-seq - principles and applications. Genomics 106: (2015);145-50.

[151] de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev 26: (2012);11-24.

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