Multiplexed, Single-Cell Profiling of Chromatin States with Expansion Microscopy

Marcus A. Woodworth

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Joshua C. Vaughan, Chair

Hao Yuan Kueh

Jesse Zalatan

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Multiplexed, Single-Cell Profiling of Chromatin States with Expansion Microscopy

Marcus A. Woodworth

Chair of the Supervisory Committee:

Joshua C. Vaughan

Department of Chemistry

Proper regulation of genome architecture and activity is essential for the development and function of multicellular organisms. Histone modifications can act in combination to specify these activity states at individual genomic loci. However, the methods used to study these modifications often require either a large number of cells or are limited to targeting one modification at a time. I developed a new method called Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE) that uses Expansion Microscopy (ExM) to visualize and quantify multiple histone modifications at in situ labeled genomic regions in single cells. Using SCEPTRE, I distinguished multiple histone modifications at different alleles of the same gene, quantified histone modification levels at multiple developmentally-regulated genes and identified a relationship between histone H3K4me3 and the loading of paused RNA polymerase II at individual loci. I observed extensive variability in epigenetic states between individual gene loci hidden from current population-averaged measurements, which emphasizes the need for single cell methods when defining the chromatin state of genes. Lastly, I further developed SCEPTRE for use with live-cell tracking to recognize how cell cycle and potential state inheritance impact H3K27me3 mark heterogeneity at a developmental gene. I therefore have demonstrated the potential SCEPTRE may have for uncovering the contribution epigenetic states have to cell fate decision making.

TABLE OF CONTENTS:

LIST OF FIGURES:

LIST OF TABLES:

ACKNOWLEDGEMENTS

As someone who has traveled much of his life, living in places like the coffee region of Colombia or the province of Buenos Aires in Argentina, graduate school was the longest continuous time I had ever been in a single place. I therefore did not have the expectation that I would possibly form meaningful friendships with such wonderful people, having learned to appreciate them more as time went on. Graduate school had its challenges, from learning to managing a long commute and balancing the life of a growing family with the crazy twists that research can throw at you. But each challenge was surpassed thanks to the contributions, strength, and willingness to help of everyone around me. I therefore have many people to thank for achieving my PhD.

I first would like to thank my co-mentors, Joshua Vaughan and Hao Yuan Kueh. Josh simply astounded me in how much he cared for not only the science in the lab, including my own project, but also how much he cared for the success of his students in life and beyond graduate school. I thank him for his mentorship, which helped me grow as a scientist and as a better person. His knowledge about any topic, and his continuous invitation to think of each question in diverse ways has helped me appreciate what one can do in both a long and brief period. I find myself quite fortunate that he allowed me to join his group! Kueh acted the same way, seeking opportunities to encourage people in the lab and invite them to think deeply of what they were doing and how they fit in the lab and greater scientific community. I thank him for how much I learned through him about the core purpose of my project. I hope that I may implement the same spirit he brought to the lab to whichever place I go next, encouraging others in their scientific journey. I thank him for his patience and understanding as I tried juggling my personal life and the schedules of two distinct groups during graduate school. Both my mentors are amazing people that were great to work under while we tackled the difficulties of each project.

I would also like to thank the rest of my thesis committee, Jesse Zalatan and David Shechner. My first rotation at UW in Jesse's lab is where I was inspired to study chromatin. I thank him for introducing me to such a fascinating topic, and for his support and encouragement as I progressed through graduate school and applied for fellowships and postdoc positions. His questions greatly improved the quality of my research while showing me how I can think critically of my own work and that of others. I thank David for each conversation we have had, since I seem to learn something new in every chat with him. His passion for trying new things and piecing concepts together is truly contagious!

I must also thank the people who I encountered during my time in the Vaughan group, such as Marco Howard, Lauren Gagnon, Tyler Chozinski, Chenyi Mao, Min Yen Lee, and Aaron Halpern. Tyler, Marco, and Lauren were fun to chat with and hear about their projects and ideas. I thank them for their graduate school perspective which helped me in choosing a lab I found comfortable and inspirational to work in. I thank Chenyi and Min Yen for allowing me to contribute to their exciting project, and for sharing meals and coffee in the office that made the hard days in the lab much more bearable. I was also happy to discuss frivolous topics such as how to perfect the best espresso in a 5 mL beaker with Min Yen and Aaron. I thank Aaron for showing me how a scientist of a higher caliber acts and thinks, while still being considerate and welcoming towards the people around them. I hope to act the same way in my postdoc!

I also had the pleasure of working with former members of the Kueh group, including Kenneth Ng, Nicholas Pease, and Sam Nguyen. Kenny really helped me out with the tough beginnings of the project, and I thank him for being a cheerful and dedicated lab partner as we tackled the tedious troubleshooting involved in making new methods. I thank Nick and Sam for not only being helpful in expanding my own knowledge and toolsets, but for also being a lot of fun to hang out with during long incubations or lunch times.

Of course, there are a lot of great current members of both labs that I must thank as well, given how much they contributed to my scientific progress. From the Vaughan group, there is Madeline Wong, Chris Kim, Hannah Perry, Adil Ali, Ziyu Guo, and Chetan Poudel. Even though I had the fantastic opportunity of mentoring Maddie, her big heart and ability to tackle big tasks without a breeze has been a constant inspiration to me. I thank her for her persistence in research, and her willingness to go beyond to get things done. I thank Chris and Hannah for their hard work as well, pursuing new feats of science I am excited to hear more about in the future. However, I probably will remember them most for their mad musical skills during my defense party (you truly rock!). I thank Adil and Ziyu for the great conversations we had in the lab. Their growth as scientists challenges me to try to improve in the same way. I thank Che for the wonderful friendship we built in the lab, though I do hope to beat him at ping pong some day and prove that my one victory against him was not a fluke.

From those of the Kueh group who were a huge help during my scientific journey at UW, there is Kathleen Abadie, Matt Wither, William White, Elisa Clark, Paul Leanza, and Lihua Chen. I thank Kathleen and Matt for their insight across the years. Talking with them and listening to their presentations helped broaden my knowledge, and seeing how they tackled their own projects helped me keep the right perspective on what to think about in a project. I was glad to have hung out with Elisa, who did an excellent job bringing the lab together for different events and who was happy to hear and share about different experiences inside and outside the lab. I thank Will for all the times we shared great ideas for image processing and data analysis, and I hope I was able to help him out as much as he helped me. I thank Paul for his hard work in the lab, not only helping my project out immensely but for any others to come. I thank Lihua for somehow making the Kueh lab function. Her commitment to getting things done is something I greatly admire.

I must thank my family, both my parents, David and Margot Woodworth, and my brothers, Davis and Lucas. My parents were key role models as I grew up, nurturing my creativity and curiosity, even

though it meant I took apart many of the electronics, doors, and toys in the house beyond repair. I thank them for their encouragement, which led me to where I am today, and I hope to be just like them with my own kids. My brothers fulfilled, effortlessly, their primary role of driving me crazy as I grew up. That being said, they were a blast to be with and to get into trouble with (though I am quite sure I was culprit number one in most cases). I thank them for sticking around, being encouraging and supportive in every step of my life and for being people I can rely on no matter what.

I must of course reserve a huge thanks to my wonderful and patient wife, Cara. I absolutely would not have been able to make it if it were not for her persistence in holding down the fort, with a toddler and a newborn in her arms might I add, while I frantically worked on experiments and drafting my thesis in the end. She has been my primary support since the beginning of graduate school and I owe her many times over for listening to my thoughts and stresses, for going on ridiculous adventures outside of work, for constantly bringing me joy and laughter, and for many other little and massive things she did that kept me moving. Thank you again and again!

There are of course others I want to thank, who throughout my life provided me with great lessons and support that placed me where I am now. I thank Victor Manuel, who in middle school imparted to me a passion for chemistry, biology, and the teaching of these subjects. I thank Jorge Klee, Jezlias Pava and Franky Duran, who were my roommates in Argentina. They helped me grow as a person and showed me how to be a great friend. I thank Majid Kashi, whose Calculus class was a neat wake up into being a better organized and dedicated student. I thank Luis Mota-Bravo, the first person to let me work on a project in a lab, showing me lots of the ropes on how to work as a scientist. I thank Lili Mesak, for helping me get started in research and think critically of an experiment. I also thank Marlene de la Cruz for her passion and support for advancing students in their scientific career. I thank Gregory Weiss for helping advance my knowledge and skill set in research and for supporting me throughout my scientific career path. I thank Luz Meneghini, who took the brunt of my lack of knowledge in being a

good scientist and helped me leave my undergraduate institution as a much better person and researcher. I thank Erin Kirschner for not only coordinating a phenomenal BPSD program but showing me how to do one's work with excellence, caring for those people who rely on you.

Lastly, I want to thank the person at the core of my faith, Jesus Christ. As a Christian, I believe that the intricate details that I observe in life and in my project are thanks to Him, and I thank Him for the great mysteries we find in every corner of nature. I believe that above anyone else, He has supported me and given me the strength to complete my PhD, providing hope above what I could possibly do by myself in life. I hope and pray that those around me may find comfort and support in the faith of something beyond themselves as well.

Dedication

To Cara Woodworth,

My wife and companion in all of life's crazy adventures.

CHAPTER 1: INTRODUCTION

1.1 DNA AS THE CODE OF LIFE

Every cell within our body is unique and plays a key role in allowing us to move, breathe and live. Despite their differences, each cell shares the same genetic code, having originated from a single cell that after trillions of divisions and differentiation steps, formed the body as we know it. Even though these divisions have been observed in countless scenarios, there is still great mystery behind what processes coordinate to allow for one code to account for such a diversity of cell types, from the cells that line our stomach to the neurons that stretch from one corner of the brain to the other. Although this code distinguishes us from other species, only 1% of it is transcribed into the blueprint RNAs that translate into the cell tools that give them their form and function.[1] The key is in how these genes are regulated within each cell.

Cells orchestrate, through a set of self-perpetuating cascades of biochemical reactions, the activity of thousands of genes to change their function and identity during development, but once committed to a unique fate, remove such possibility of change without disrupting their genetic code. This requires a set of mechanisms that cause semi-permanent gene states, that is, defined levels of expression (or lack thereof) for each gene that can be maintained over time and even across multiple divisions of a cell. The study of these mechanisms is known as epigenetics,[2,3] and disruptions to these mechanisms are broadly implicated in human health and disease, inspiring much exploration and innovation in how best to understand what makes each cell unique.[4]

What we know so far about the interplay of different epigenetic mechanisms and their close relationship with genomic structural features has been discovered due to technological developments in visualizing genome organization and associated regulatory elements. Yet, there remains unanswered questions about these mechanisms due to the technical limitations existing methodology. Therefore, there is a compelling need to develop novel techniques that address these limitations so as to further

reveal the principles of gene regulation that impact cell development and disease.[5] To better understand what such a method would look like, I will discuss how the genome has been studied until now, what are some of the current methods used in this study, what do we know so far about gene regulation vs. genome organization, and what further technical developments can arise to address the unanswered questions of chromatin biology today.

1.2 A BRIEF HISTORY OF GENOME STUDIES

The way our genetic code is structured and carried from cell to cell during division was a big mystery until the 19th century. In 1880, Walther Flemming encountered a "stainable material," which he termed "chromatin," as he worked on staining and fixing cells during division.[6,7] Using a light microscope, Flemming could tell chromatin was expanded, then partitioned into each daughter cell, a process important for cell replication. Soon after, peers of Flemming biochemically identified that chromatin was composed of nucleic acids, histone and other proteins.[8] However, as he lacked tools to delve deeper into chromatin organization, it took almost another century for the greater principles of genome function and structure to be uncovered.

There were, however, seminal discoveries that helped lay out the groundwork of chromatin studies. Among these are: the characterization of the double-helix structure of DNA and the base pair interactions of paired nucleotides (i.e., adenine, thymine, guanine and cytosine),[9] determination of histone fractional composition and observation of interactions with DNA within chromatin,[10] along with the surprising result that acetylation or methylation of chromatin proteins lead to changes in gene activity.[11] Researchers also discovered that genomic DNA was not completely covered by interacting protein, but about half of the DNA was considered "open," that is, accessible to degrading nucleases or stains.[12,13] Results like these helped provide an intriguing perspective as to what factors influenced genome structure and function.

In 1974, the subunits of the nucleosome, the fundamental unit of chromatin, were first characterized, with its 4 canonical histone pairs (H2A, H2B, H3 and H4) wrapped around ~147 base pairs of DNA.[14] This structural unit was repeatedly observed throughout chromatin filaments in electron microscopy images,[15,16] forming the revolutionary "beads on a string" model where nucleosomes are interspaced throughout free regions of genomic DNA. With this crucial structural feature elucidated, researchers began to observe how histones cluster together depending on which histone modifications were present.[17] These histone modifications, whose diversity and distribution across the nucleus was observed by antibody stains via fluorescence microscopy,[18] were increasingly linked to changes in genome structure and gene activity.[19] Scientists therefore began referring to covalent histone modifications as a "histone code," under the hypothesis that the combinatorial presence of histone modifications determines the regulatory state of genes present in their vicinity.[20]

It is important to note that, up to this point, the actual sequence of the genomic code was mostly unknown. Most of the focus had been on characterizing the composition of chromatin and its structural features. It was the advancement of high throughput genome sequencing, where short strands of genomic DNA are simultaneously amplified by PCR and sequenced in parallel via fluorescence imaging, that granted researchers access to whole genomes of organisms at a much more affordable price compared to the Sanger sequencing technology reliant on capillary electrophoresis for base pair distinction.[21] Genome after genome, from bacteria and yeast to worms and mice, were published, with the peak of this work being the human genome project in the early 2000s.[1] These tremendous feats not only provided profound revelations of each organisms genetic code, but it in turn opened up the possibility of performing genome wide studies of elements involved in gene regulation.

1.3 METHODS USED TO STUDY GENOME ORGANIZATION

The earliest methods used for looking at genome structure, beginning with Walther Flemming, was to stain chromatin and observe its broad shape and form on a microscope.[22] Although these original

staining methods were non-specific, they still provided a valuable perspective on how chromatin condensed into chromosomes before being partitioned into each cell during division. It was in the 1960s that researchers started using the complimentary aspect of DNA to employ probes that labeled unique regions of the genome.[23] Beginning first with radioactive labels, but soon after with fluorescence dyes, labeled bacterial-artificial chromosomes (BAC) complementary to a genomic region of interest were introduced into cells after intense denaturation of the genomic DNA, allowing them to bind to the genome by Watson-Crick base pairing, providing a detailed and unique perspective of how whole chromosomes (or sections thereof) were organized throughout the nucleus.[23] This technique, known as Fluorescence *in situ* Hybridization (FISH), soon became expanded and recognized as a gold standard in genome structure studies within each cell.[24]

One big improvement made for FISH labeling was the design and use of short single-stranded (ss) DNA libraries, termed oligopaint FISH, instead of bulky BAC probes.[25] Brought on by advancements in computational analysis of sequence complementarity and affordable DNA synthesis, many ssDNAs are designed to tile a genetic sequence with high stringency such that each probe will likely bind to only one location of a cell's genome (e.g., a gene of interest) while using suitable hybridization conditions. Using arrays of smaller probes, ~120 nucleotides each, allows for labeling of smaller regions of the genome compared to the Mb scale that BACs often covered (Figure 1.1, left). These smaller probe sets can therefore label less than 10 kilobases of genomic DNA with >90% labeling efficiency and haplotype distinction of gene alleles.[26] This approach allows for multiplex labeling schemes of genes as well, by introducing barcode sequences to each set of probes designed per gene. This use of complex barcode schemes with DNA FISH has generated a variety of genome structure targeting methods such as MERFISH,[27] seqFISH+,[28] ORCA[29] that observe thousands of genomic regions within each cell, revealing just how unique each cell organizes their DNA.

Alongside FISH, researchers have also used indirect methods to investigate the presence of histone modifications at genomic sites. Antibodies are well known for targeting protein structures with high specificity, which, if targeting a protein associated to a genome sequence, can allow for the precipitation out of such a sequence. This method, known as Chromatin ImmunoPrecipitation (ChIP), determines the enrichment of histone marks or other associated proteins present at genomic sites for a population of cells. It was originally performed on micro-arrays that compared modifications across the genomes of two different cell types or transformations,[30,31] but with the advent of affordable high-throughput sequencing, this method has expanded in both capability and diversity.

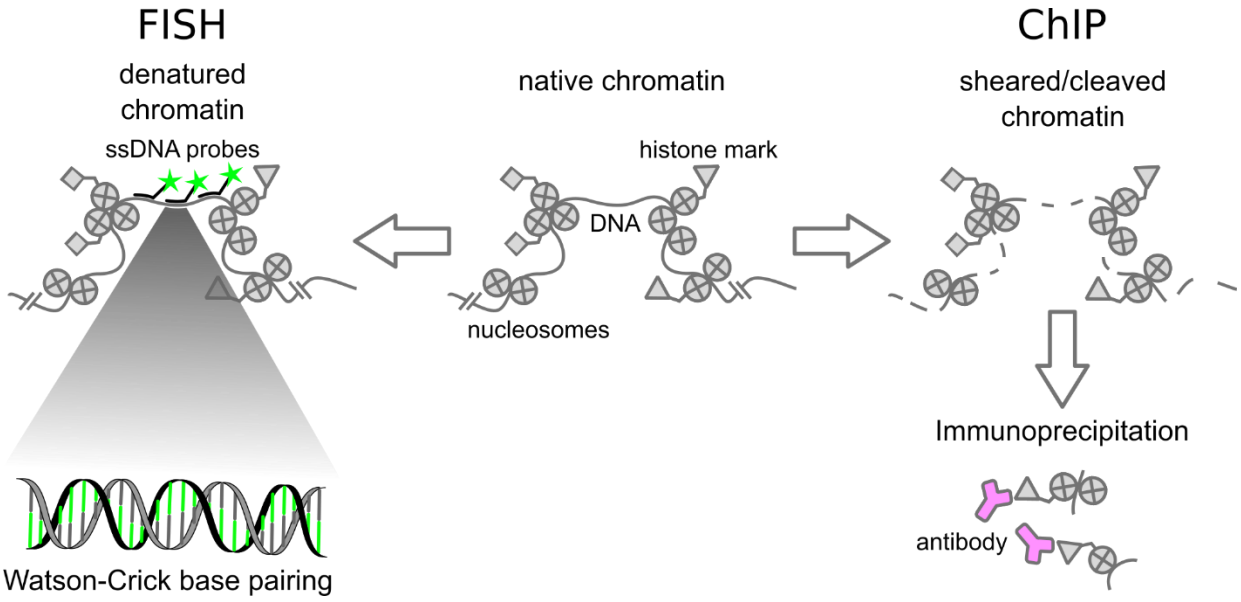


Figure 1.1 **Methods for studying genome structure and associated proteins**.
(Left) In Fluorescence in situ hybridization (FISH), single stranded DNA probes perform Watson-Crick base pairing to one of the two strands of genomic DNA after denaturation of chromatin, enabling visualization and localization of a previously determined genomic region within single cells. (Right) In chromatin immunoprecipitation (ChIP), genomic DNA is sheared or cleaved, then DNA strands with associated histone marks are precipitated out with an antibody for later sequencing and alignment to the known genome.

With ChIP-seq (chromatin immunoprecipitation followed by sequencing), antibody-targeted sequences associated with a protein of interest are separated after shearing of the genomic DNA, sequenced, then mapped back to the genome of the organism (Figure 1.2, right).[31] This process is done

over millions of cells to capture the average enrichment of histone modifications at particular regions, providing an ensemble-result for each cell type of interest. ChIP-seq has been crucial for understanding the relevance of histone marks in gene regulation, such as providing evidence of the abundance of H3K27me3 marks at repressed developmental genes,[32] and the abundance of H3K4me3 and acetylation marks at actively transcribed genes.[33] Researchers have applied ChIP-seq, along with other sequencing-based approaches, to study how histone marks denote chromatin states that are commonly recognized today, from active promoter to bivalent states.[34,35] The sequencing-based approach has been applied to other genomic features, including genomic-contact frequencies[36] and genome accessibility,[37,38] revealing facts about the genome's 3D organization in space. These new insights, complementary to the ones observed with FISH or other labeling techniques, have provided us with an interesting perspective of the architecture of the genome and how it relates to regulation.

1.4 MULTISCALE REGULATION SCHEMES OF THE GENOME

As seen by both microscopy-based and sequencing-based methods, regulation of the genome begins with the nucleosome. This structure involves a histone core wrapped by about 147 bases of DNA, which is the first step in compacting the 3 billion bases of our genome.[14] The size of the nucleosome is around ~10nm and prevents the coiled DNA from being accessed by other proteins.[39,40] These nucleosomes cluster together into higher ordered structures, which have been previously observed as 30 nm fibers by EM[41] or Cryo-EM,[42] but while using super-resolution microscopy in native cells, have been seen as clusters heterogeneous in size between 50-100 nm.[43] Such clustering and its impact on accessibility, a key factor relevant in gene expression,[44,38,45] is facilitated by histone modifications and other protein complexes,[17,46] with each histone modification associated with a different cluster size.[47] Depending on the epigenetic state, these clusters will span genomic regions ranging from 10 to 100 Kb of DNA.[48]

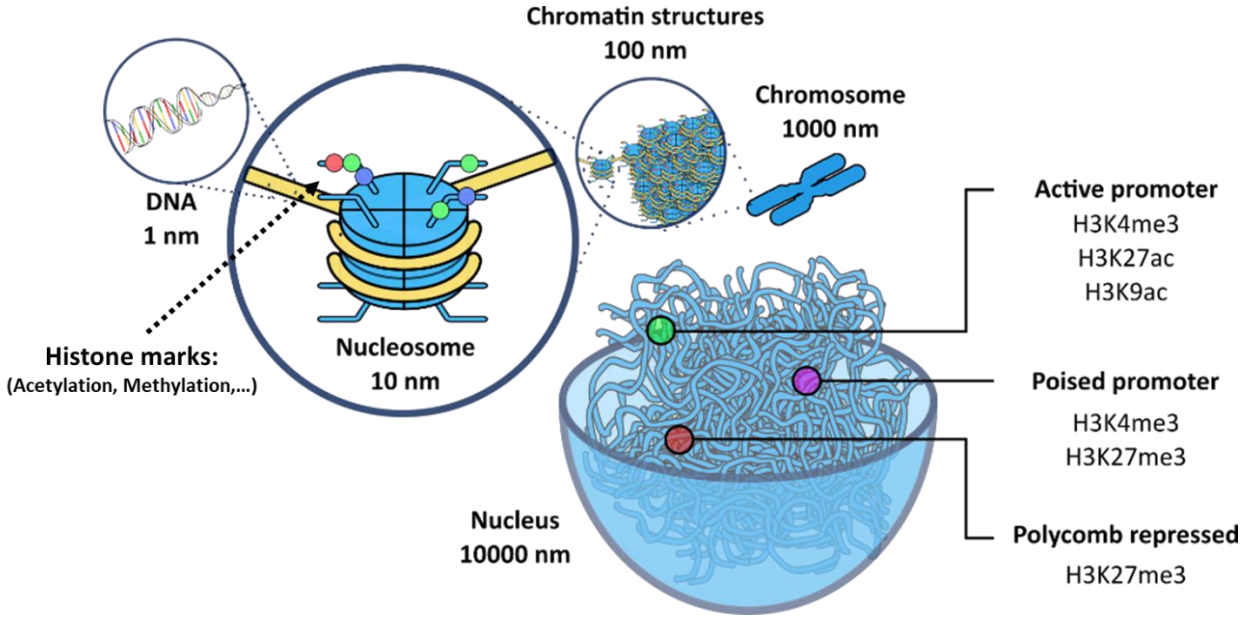


Figure 1.2 **The structured organization and epigenetic states of the genome.**
Eukaryotic cells compact DNA by first wrapping it around a complex of histones to form the nucleosome which, with the assistance of histone modifications (or histone marks), can compact further into higher order chromatin structures. In its most compact form, chromatin is seen as chromosomes, but during interphase, it is spread throughout the nucleus of each cell.

These clusters have been indirectly observed through contact-frequency maps as well, at scales like those seen by microscopy methods. With contact frequency mapping methods, researchers have found that genomes are partitioned into conserved regions of frequent contacts known as topologically associated domains (TADs), which share similar epigenetic states within each domain for a given cell type, but whose borders will change during development.[49,50] The sub-domains of TADs, referred to as sub-TAD, cover ~200 Kb of genomic sequence space,[51] with spatial sizes of ~160 nm.[52] TADs form part of broader compartments with common active vs. inactive gene regulation states across greater than 1 Mb scales,[36] and at µm length scales.[53,54] Lastly, whole chromosomes are known to maintain their own territories with distinct boundaries between each other observed in both microscopy and sequencing-based methods.[55] Thus, each cell is able to have a unique identity thanks in part to its ability to spatially organize their genome into various chromatin structures, from nucleosomes to chromosomes (Figure 1.2).

1.5 THE HISTONE CODE HYPOTHESIS.

The role of nucleosomes in gene regulation extends further than preventing accessibility of genes by other proteins for repressed states. Post-translational modification on the histones of these nucleosomes, such as methylation and acetylation, are thought to establish each gene's epigenetic state. This is known as the "histone code hypothesis," where the presence of covalent modifications to histone proteins stably associated with DNA loci alters the physical and regulatory state of each gene.[56,20] First of all, these modifications (referred to as histone marks) change the charge of histone tails, disrupting how nucleosomes interact with genomic DNA or promoting further interactions with other nucleosomes.[17,19] For example, acetylation of a histone lysine neutralizes its potential positive charge, decreasing the histone's ability to interact with negatively charged DNA, while methylation of this lysine facilitates interactions with other histones and proteins. Histone marks also recruit a variety of transcription factors and protein complexes involved in gene activation or repression.[57,58] There is, therefore, an interplay of proteins at each histone where enzymes known as "writers" will add or remove each mark so as to change how regulatory proteins, known as "readers," interact at each gene. This has been observed at the genome wide scale and is currently considered one of the main mechanisms by which cells maintain the epigenetic state of genes.[2,3]

Although there are hundreds of different histone marks,[59] the most studied are the acetylation or methylation of Lysines on each histone tail, given their presence in developmental contexts.[60] For this reason, broad regulatory states have been attributed to each mark. For example, acetylation marks (e.g., H3K27ac) deposited by histone acetyl-transferases are present at genes or enhancers that are active or interact with active genes.[61,62] For methylation marks deposited by methyl-transferases, their function varies from demarcating weak enhancers (H3K4me1),[63] promoters of genes (H3K4me3),[33] facultative heterochromatin that is repressed (H3K27me3),[64,65] or pericentromeric heterochromatin regions of the genome (H3K9me3).[66,67] The picture of each epigenetic state is much more complicated, given that

when viewing across the whole genome over multiple cell types, there are combinations of marks that denote a spectrum of epigenetic states.[35] Given that there is still much to learn about the impact these marks have on cell development, it is imperative to detect various histone marks at a time for each gene.

1.6 THE UNKNOWN IMPACT OF HISTONE MARKS ON CELL FATE BIAS

In Hematopoiesis, a fundamental model for studying differentiation, hematopoietic stem cells (HSCs) develop into the cells that make up the blood of an organism.[68] Ample research has been dedicated to studying the change in histone marks and other factors that drive this developmental process,[69] as HSCs relinquish pluripotency to commit to unique fates by asynchronously expressing various transcription factors.[70] Mounting evidence in this field has indicated that underlying epigenetic states at developmentally relevant genomic sites hold a latent "bias" in HSC fates, where each developing cell will consistently reconstitute blood cells in proportions that differ between cells.[71–73] It is possible that differing levels of histone marks at key fate determining genes allow some cells to more readily differentiate towards one fate over others. For example, HSCs that have higher repressive marks at a T-cell fate committing gene will more likely differentiate into alternative fates, given that to become a T-cell the high repressive marks would first have to be removed (Figure 1.3).

The potential bias of cell development by histone marks has continued to be observed as sequencing-based methods have improved to target low to single-cell levels of histone marks,[74,75] making it imperative to study this principle at the single-cell level. However, unanswered questions remain: How different are the levels of each histone mark at key developmental genes? What is the interplay between marks and other chromatin features in maintaining gene repression? How well are these marks inherited across different cell divisions and differentiation steps? Unfortunately, the limitations of current methods are what have left these questions mostly unanswered.

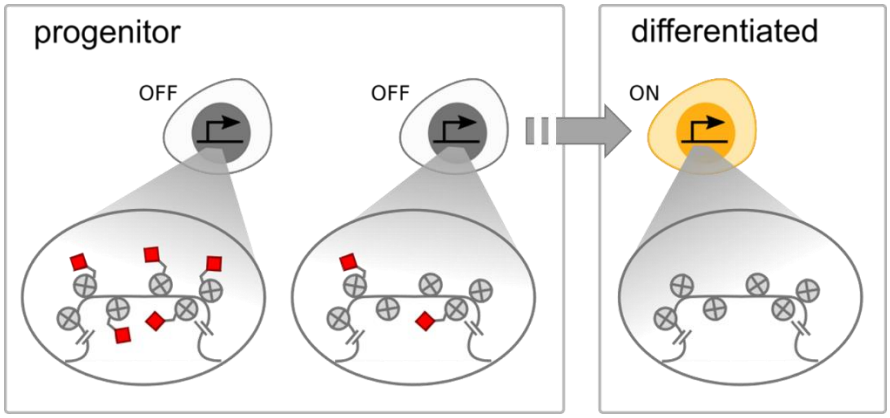


Figure 1.3 **The potential impact of histone marks on cell fate decision making**.
In a population of progenitor cells (left) with varying levels of a repressive histone mark (in red) at a fate committing gene, cells with a lower density of this mark at the gene may have a higher propensity to differentiate towards the committed state defined by the gene (right).

1.7 LIMITATIONS OF SEQUENCING-BASED METHODS FOR STUDYING EPIGENETIC STATES

There are major limitations that exist in sequencing-based methods, including their ability to target single-cells and the number of histone marks they can detect at a time in each cell. Most genome-wide analysis methods produce an average result from an ensemble of cells. This ensemble result masks heterogeneity of states potentially present in each population, particularly those that may lead to rare developmental trajectories compared to most cells. This limitation has been addressed in part by the development of new sets of single-cell sequencing techniques that, with the right microfluidics system or cell indexing procedure, separately profile histone marks for individual cells.[75,76] However, due to technical constraints and sample handling, the number of detected sequences associated with a given mark per cell remains low, with many cells having no marks detected. This limitation is particularly concerning for researchers studying cell populations of low abundance in an organism.

In addition to the challenge of profiling single-cells, sequencing-based techniques are often limited to targeting one histone mark at a time. Because ChIP methods rely on an antibody to precipitate sequences associated to each mark, targeting multiple histone marks provides a mixed and

often indistinguishable signal between each mark. Given that multi-histone mark profiling of individual cells has been a previously highlighted next step in epigenetic state profiling,[5] some have successfully profiled multiple histone marks using complex tagmentation methods which insert distinct barcodes per antibody at the location of each mark.[77,78] These methods unfortunately suffer from high technical difficulty and, as with single-cell ChIP-seq, very low detection efficiency.

Although sequencing-based methods may evolve to better detect multiple histone marks in individual cells, there are inherent limitations in the methodology that leave room for alternative approaches. Because protein associated DNA fragments are removed from each cell for sequencing and analysis, it is highly challenging to link the detected histone mark profiles to the physiological context of cells within tissue or disease. Additionally, single-cell sequencing provides a "present vs. not present" result per sequence, while ignoring other genomic features a gene is subject to within the nucleus, such as compaction state, proximity to other nuclear structures and complex interaction between more than one genetic element. Thus, there is still a compelling need for a novel method capable of profiling the epigenetic state of key genes at the single cell level, providing the density of various histone marks within the context of a cell's development stage or tissue position.

1.8 THE BARRIERS OF EPIGENETIC PROFILING WITH FLUORESCENCE *IN SITU* HYBRIDIZATION

Since histone modifications have been visualized within cells through immunofluorescence, and genetic elements with FISH, a microscopy-based method has the potential for profiling various histone marks at several genes with high detection efficiency. Thus, by combining these microscopy labeling techniques, one could colocalize histone mark signals to genomic regions and generate epigenetic profiles for each cell. There are, however, challenges for incorporating these labeling schemes together. First off, the harsh solvents and temperatures used in FISH techniques can be disruptive to antibody stains.[79,80] This challenge has inspired alternative labeling schemes such as: The digestion of one of the two genomic

DNA strands,[81,82] the nicking of DNA for local denaturation with a super-helicase,[83] or the design of DNA probes that hybridize to double helix DNA without the need for denaturation.[84] For the first case, non-canonical bases must be incorporated into the genomic strand by the targeted cell before digestion, limiting this application to cell culture work with few divisions. In the second case, the method requires additional libraries for nicking DNA and a whole set of enzymatic tools that hinder the scalability of the assay. In the third case, the alternative hybridization scheme limits the regions that can be targeted within the genome to mostly repetitive regions, such as ALU elements.

Two additional methods have helped facilitate the detection of both proteins and genomic DNA in the same cell, by using either inactivated Cas9 binding of DNA or crosslinking of DNA-barcoded antibodies before DNA FISH. The inactivated Cas9 used in genomic labeling allows the visualization of genomic loci in live cell and requires no denaturation steps.[85,86] However, targeting multiple loci per cell requires the use of fluorescently labeled guide RNAs, different Cas-derivatives, or other schemes (all which struggle with proper detection of signal over noise) that constrain the number of distinguishable loci per cell. Crosslinking of antibodies with DNA barcodes to the sample is another strategy that has enabled researchers to visualize more than a dozen histone marks at genomic loci within each nuclei.[87] However, even when preserving antibody labeling signal for colocalization of marks at a gene of interest, there is still the difficulty of resolving the correct epigenetic state of this gene thanks to the diffraction limit of light.

1.9 CIRCUMVENTING THE DIFFRACTION LIMIT OF LIGHT

The ability to distinguish two objects, such as fluorophores, is limited by the diffraction of the light used.[88,89] In the case of a conventional fluorescence microscope, this limit is approximately 250 nm, depending on the laser light and fluorophore used, which is much larger than the size of the 50-100 nm higher-order chromatin structures discussed above. Therefore, genomic regions that may be adjacent to each other in physical space, but maintain completely different epigenetic states, will have an overlap of

fluorescent signal that would cause the misidentification of each region's states. Thus, to study epigenetic states of genes under a microscope, one would need a method capable of surpassing the diffraction limit of light to resolve each epigenetic state.

Super-resolution microscopy techniques have allowed researchers to better visualize chromatin structures within nuclei. One of these methods is single molecule localization microscopy, such as Stochastic Optical Reconstruction microscopy (STORM), which uses sporadic "blinking" of photo-switchable dyes to better localize each dye and, after capturing enough blinks, reconstruct an image of the structure at a much improved resolution. [90,91] STORM has therefore been used in combination with DNA FISH to study the size of genes with differing regulatory states.[48] STORM was also used to determine the size of higher-order chromatin structures by resolving immunolabeled nucleosome clusters.[43,47] However, STORM and other super-resolution methods (e.g., Stimulated-Emission-Depletion microscopy)[92,93] require complicated microscope setups that are technically challenging to implement. These methods also require fluorophores with unique properties, limiting the number of structures that can be detected at a time within a sample, an aspect important for profiling multiple histone mark levels at various genes.

To summarize to this point, a microscopy-based method for profiling epigenetic states would complement current histone mark targeting sequencing-based methods, given that both genes and histone marks can be visualized at the single cell levels in a multiplexed manner, by using FISH to target genes and immunofluorescence to target histone marks. However, combining these two methods has been challenging for researchers due to the stringent conditions of FISH disrupting antibody labels and the need for studying these chromatin structures beyond the diffraction limit of light of conventional microscopes. Therefore, a new microscopy-based method should: Preserve histone mark labels for co-visualization with *in situ* labeled genes; allow for labeling of multiple structures at the same time; have better resolution capabilities than conventional microscopy.

1.10 EXPANSION MICROSCOPY: KILLING TWO BIRDS WITH ONE STONE

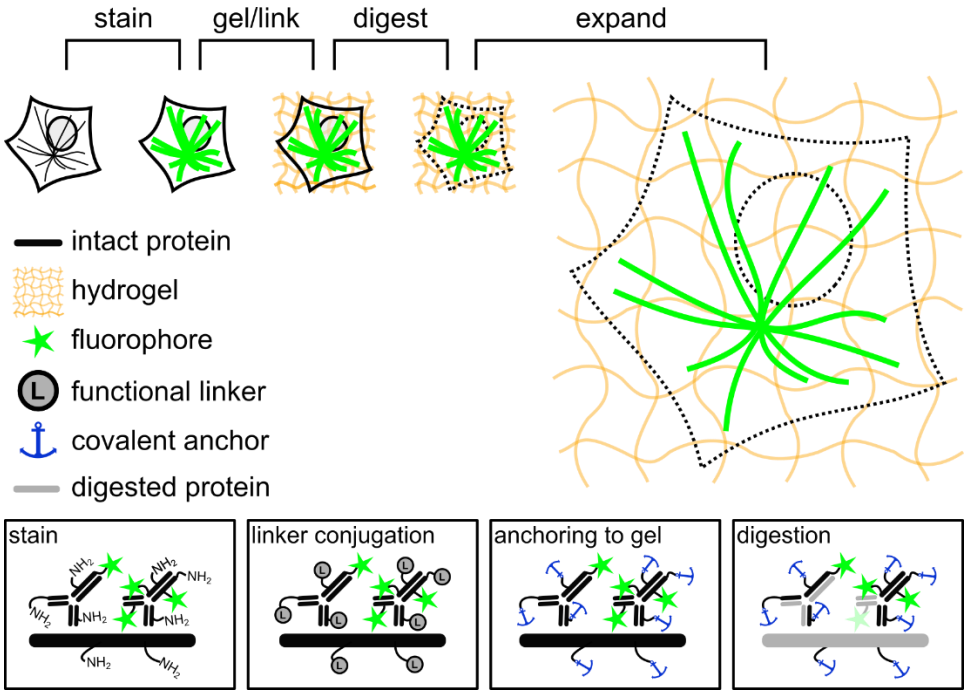


Figure 1.4 **Expansion Microscopy**
The structure of a fixed specimen is stained, then linked to a polymerized hydrogel grown within the sample. Protein structures are then digested or denatured to allow for isotropic expansion of the sample and stain by swelling of the gel in water. After expansion, the preserved stain signal is imaged with an improved resolution on a conventional fluorescent microscope.

In 2015, a group at MIT produced a radically innovative approach to surpass the diffraction limit of light. Instead of employing more complex microscopes with specialized fluorophores to circumvent the diffraction problem, what if the sample could be made bigger? If the fluorophores that label a given structure were to be isotropically pulled apart the features of the structure would be better resolved by simple physical magnification. With this concept, Expansion Microscopy (ExM) was developed.[94] In ExM, a fixed specimen is immunolabeled, anchored to a swellable gel and, after digestion or denaturation of rigid structures within the sample, placed in water so that by removal of salt, the increase in osmotic pressure will cause the sample to swell and isotropically expand. Since then, ExM has been adapted to work with conventional antibodies and fluorescent stains,[94,95] allowing for great ease of use and

adaptation, bringing super-resolution capabilities to what would be a diffraction-limited fluorescence microscope by simply varying how the sample is prepared.

ExM has been broadly applied to not only resolving various protein structures, but that of DNA and RNA components as well.[96] For example, ExM allowed for better counting of RNA transcripts in individual cells.[97] Given that immunostains are linked to the polymer gel to enable expansion, researchers were able to detect with DNA FISH the abundance of the HER2 gene in expanded breast cancer samples, while still preserving the signal from labeled proteins.[98] For this reason, we believed that expansion microscopy would enable the combination of histone mark immunostaining with genomic DNA FISH labeling for profiling of the epigenetic states at various genes. Thus, I developed Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE).[99]

In chapter 2 of my thesis, I will first describe the development of SCEPTRE, the experiments I performed to validate the use of expansion microscopy for chromatin profiling, how well SCEPTRE can profile the states of various developmentally-regulated genes, and some of the interesting insights determined by the method. In chapter 3, I will mention key results on how expansion microscopy enables use of FISH-based techniques in other labeling schemes, useful for answering biological questions beyond developmental contexts. In chapter 4, I will discuss a proposed application of SCEPTRE combined with live-cell tracking for studying the impact of cell cycle, epigenetic state inheritance, and the presence of alternative marks on histone mark heterogeneity at repressed developmental genes.

CHAPTER 2: MULTIPLEXED SINGLE-CELL PROFILING OF CHROMATIN STATES AT GENOMIC LOCI BY EXPANSION MICROSCOPY

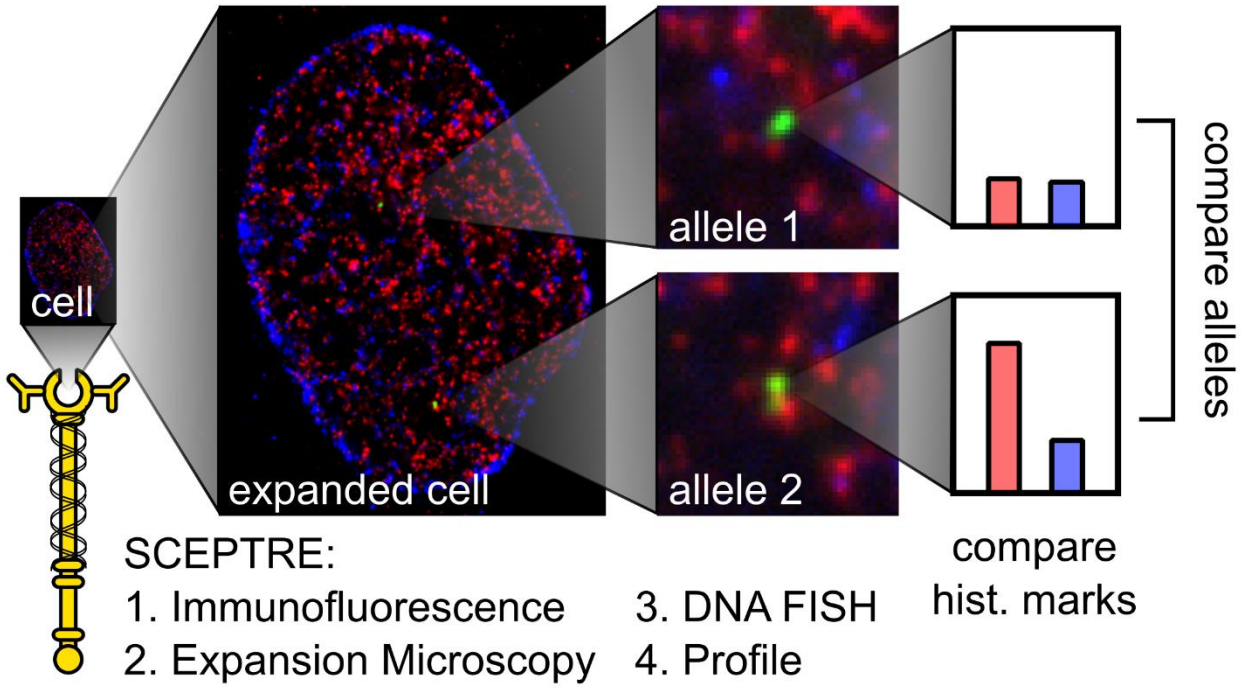
2.1 PREFACE:

At the time that I was joining Prof. Joshua Vaughan's group, they were collaborating with Prof. Hao Yuan Kueh's group to design a new microscopy-based method for single-cell epigenetic state profiling of genes. The Kueh group had observed in myeloid progenitors a delay between the expression of the two different alleles of *bcl11b*, a key developmental gene whose expression commits cells to the immune T-cell fate.[70] This delay meant an epigenetic mechanism was controlling the activation of this gene.[100] They were interested in probing further what mechanisms were behind the delay. Unfortunately, the traditional epigenetic state profiling methods at the time did not provide the single cell and locus resolution they would need to best dissect the mechanisms behind the delay. To help address this need, I developed the method Single Cell Evaluation of Pos-TRanslational Epigenetic Encoding (SCEPTRE), the subject of this chapter.[99] A more detailed protocol of SCEPTRE is also available on protocols.io.[101]

2.2 ABSTRACT:

Proper regulation of genome architecture and activity is essential for the development and function of multicellular organisms. Histone modifications, acting in combination, specify these activity states at individual genomic loci. However, the methods used to study these modifications often require either a

large number of cells or are limited to targeting one histone mark at a time. Here, we developed a new method called Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE) that uses Expansion Microscopy (ExM) to visualize and quantify multiple histone modifications at non-repetitive genomic regions in single cells at a spatial resolution of ~75 nm. Using SCEPTRE, we distinguished multiple histone modifications at a single housekeeping gene, quantified histone modification levels at multiple developmentally-regulated genes in individual cells, and evaluated the relationship between histone modifications and RNA polymerase II loading at individual loci. We find extensive variability in epigenetic states between individual gene loci hidden from current population-averaged measurements. These findings establish SCEPTRE as a new technique for multiplexed detection of combinatorial chromatin states at single genomic loci in single cells.



Abstract Figure 2.1 **SCEPTRE**
Single cell evaluation of post-translational epigenetic encoding (SCEPTRE) uses expansion microscopy to combine DNA in situ labeling with immunofluorescence and quantify histone mark levels at individual loci within cells.

2.3 INTRODUCTION:

Proper regulation of genome activity and architecture is critical for development, growth, and function of a multicellular organism.[2,3] Regulation occurs in large part at the nucleosome, where ~147 bp of DNA wrap around an octamer of 4 different histone pairs: H2A, H2B, H3 and H4.[14] Various residues found at the N and C-terminal tails of these histones can acquire post-translational modifications, such as acetylation and methylation, which grant nucleosomes the ability to either participate in organized compaction of chromatin or to recruit transcriptionally relevant protein complexes.[57,58] Researchers have therefore suggested that these modifications, also known as histone marks, act as a code for the epigenetic state of genomic regions.[56,102] Although several sequencing-based methods are available for studying distinct histone modifications (i.e. ChIP-seq),[31,75] chromatin accessibility,[103,104] genomic contact frequencies,[105,106] and genomic nuclear locations,[107] these methods are either unable to resolve cell-to-cell variations or are limited to studying one histone modification at a time. Therefore, the role these marks play in controlling chromatin structure and gene expression at the single cell and single locus level remains poorly understood and vigorously debated.

To tackle this problem, super-resolution fluorescence microscopy techniques have been used to observe more closely how histone marks impact chromatin organization within a cell's nucleus. Using Stochastic Optical Reconstruction Microscopy (STORM),[90,91] researchers saw that nucleosomes form clusters that vary in size and nuclear distribution depending on a cell's developmental stage or what histone marks they present.[43,47] Others have combined STORM with DNA Fluorescence *in situ* hybridization (FISH) to map spatial aspects of genomic loci with a spatial resolution comparable to the observed sizes of these nucleosomal clusters.[48] Collectively, these studies suggest that concurrent visualization of DNA and histone modifications with super-resolution microscopy could enable profiling chromatin states at the level of single loci. However, most studies to date have viewed histone marks and genes separately, because combining immunofluorescence and DNA FISH can be challenging due to

the harsh solvents and/or high temperatures used in FISH protocols.[79,80,108,109] Although researchers have visualized immunolabeled histone marks across whole chromosomes,[80,108] or at repetitive and highly abundant ALU elements regions labeled with an alternative hybridization strategy,[84] there are still no methods available to study multiple histone marks at individual non-repetitive genomic loci at the level of individual nucleosomal clusters. A better understanding of histone mark heterogeneity at individual loci would require a new method capable of further decoupling immunofluorescence and FISH labeling.

We therefore developed a new method, called Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE), which uses expansion microscopy (ExM)[94,95] to combine DNA FISH with immunofluorescence and quantify histone mark fluorescence signals at individual loci within the nucleus. ExM preserves the signal of antibody labels on protein structures by covalently linking antibodies and proteins to a swellable hydrogel that is grown within the sample.[94,95] This signal preservation enables subsequent use of relatively harsh conditions, such as high temperatures and organic solvents, for labeling of genomic DNA by FISH without loss of the antibody signal. At the same time, ExM enables the isotropic expansion of specimens with low distortion so that these specimens may be examined with a high spatial resolution (here ~75 nm) in the expanded state even when using conventional microscopes with a diffraction-limited resolution of ~250 nm. We demonstrate the capabilities of SCEPTRE for a variety of systems: (i) we compared signals of multiple histone marks at a housekeeping gene locus; (ii) we distinguished histone mark signals between developmentally-regulated genes in a single cell; (iii) we demonstrate a correlation between histone marks and paused RNA polymerase II in a single region. Together, these experiments establish SCEPTRE as a powerful tool to study the role histone marks have at individual genes within the nuclei of single cells.

2.4 MATERIALS AND METHODS:

Reagents: The following primary antibodies were purchased and used for immunofluorescence: Human anti-centromeres (Antibodies Incorporated, 15-235), Mouse anti-H3K4me3 (EMD Millipore, 05-1339-S),

Mouse anti-RNA polymerase II CTD repeat YSPTSPS phosphorylated at Serine 5 (Abcam, ab5408). The following primary antibodies, which were ENCODE-validated (www.encodeproject.org),[34] were purchased and used for immunofluorescence and/or CUT&RUN[110,111] followed by sequencing: Rabbit anti-H3K4me3 (Active motif, 39159), Rabbit anti-H3K27me3 (Active Motif, 39155), Rabbit anti-H3K27ac (Active Motif, 39133), Mouse anti-H3K27me3 (Active Motif, 61017). The following unconjugated secondary antibodies were purchased from Jackson ImmunoResearch: Donkey anti-rabbit (711-005-152) and Donkey anti-human (709-005-149). The following conjugated secondary antibodies were purchased from Jackson ImmunoResearch: Donkey anti-rabbit conjugated with Alexa Fluor 488 (711-545-152) and Donkey anti-mouse conjugated with Alexa Fluor 488 (715-545-150). The following enzymes were purchased: proteinase K (Thermo Fisher Scientific, EO0491), RNase A (Thermo Fisher Scientific, EN0531), alcohol oxidase (SigmaAldrich, A2404-1KU), catalase (Sigma-Aldrich, C100), Phusion Hot-start master mix (New England Biolabs, M0536L), DNase I (New England Biolabs, M0303A) and Maxima H Minus RT Transcriptase (Thermo Fisher Scientific, EP0752).

The following chemical reagents were purchased: 10× phosphate-buffered saline (PBS, Fisher Bioreagents, BP399-1), 32% paraformaldehyde aqueous solution (PFA, Electron Microscopy Sciences, RT15714), 4-(1,1,3,3-tetramethylbutyl)phenyl-polyethylene glycol (Triton X-100, Sigma-Aldrich, X100), Bovine serum albumin (BSA, Rockland Immunochemicals Inc., BSA-50), ATTO 488 NHS-ester (ATTO-TEC GmbH, AD 488-35), Alexa Fluor 568 NHS-ester (Thermo Fisher Scientific, A-20003), methacrylic acid NHS-ester (MA-NHS, Sigma-Aldrich, 730300), 40% acrylamide aqueous solution (Bio-Rad Laboratories, 1610140), 2% bis-acrylamide aqueous solution (Bio-Rad Laboratories, 1610142), 97% sodium acrylate powder (Sigma-Aldrich, 408220), ammonium persulfate (APS, Thermo Fisher Scientific, 17874), tetramethylethylenediamine (TEMED, Thermo Fisher Scientific, 17919), 10× tris-acetate-EDTA (TAE, Fisher Bioreagents, BP2434-4), guanidine hydrochloride powder (Sigma-Aldrich, G3272), sodium azide (Sigma-Aldrich, S2002), poly-L-lysine (Sigma-Aldrich, P8920), sodium bicarbonate (VWR, 470302),

formamide (Fisher Chemical, F84-1), 20× saline sodium citrate (SSC, Sigma-Aldrich, S6639), 50% OmniPur Dextran Sulfate (EMD Millipore, 3730), Tween 20 (Sigma-Aldrich, P9416), Hoechst 33258 (Sigma-Aldrich, B2883-25MG), Tris Base (Fisher scientific, BP152-500), methyl viologen dichloride hydrate (SigmaAldrich 856177), L-ascorbic acid (Fisher scientific, A61- 25), digitonin (EMD Millipore, 300410), glycogen (VWR, 97063-256), sodium chloride (NaCl, Thermo Fisher Scientific, S271500), Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA, Sigma-Aldrich, E6635), Ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid (EGTA, Sigma-Aldrich, E4378) and calcium chloride dihydrate (VWR, 0556).

Alpha-satellite, *GAPDH* set, adapter and conjugated reporter oligonucleotide probe sets were obtained from Integrated DNA Technologies (IDT). A Precise Synthetic Oligo Pool (SC1966-12) containing probes covering the *MYL6*, *HOXC* and *LINC-PINT* regions was obtained from GenScript (for a list of sequences, see supplementary spreadsheet 1).

Cell culture: h-TERT RPE1 cells were cultured and grown to ∼80% confluency using Dulbecco's modified eagle medium (Gibco, 11995065) supplemented with 100 units/ml of penicillin and streptomycin (Gibco, 15140122), 1% nonessential amino acids (Gibco, 11140050), and 10% fetal bovine serum (Gibco, 26140079). Cells were then trypsinized with 0.25% trypsin–EDTA (Gibco, 25200056) and seeded at ∼75,000 cells per well on top of round coverslips (no. 1.5, ∼12 mm diameter) placed within 24-well culture plates. After growing overnight (∼18 h), the cells were briefly rinsed with 1× PBS then fixed with either 4–10% PFA in 1× PBS for 10 min at room temperature (∼22°C), or in cold EtOH:MeOH (1:1) for 6 min at –20°C. Fixed cells were washed three times with 1× PBS, then stored in 1× PBS azide (1× PBS with 3 mM sodium azide) at 4°C before use (see Supplementary table 2.1 for more details). Secondary antibody fluorophore conjugation was performed by mixing 40 µl of a secondary antibody solution at a concentration of 1.3 mg/ml with 5 µl of a 1 M sodium bicarbonate solution, then adding 2–5 µg of an NHS ester functionalized fluorophore. The mixture was left to react for 30 min protected from ambient light and the crude reaction mixture was passed through a NAP5 column (GE Healthcare Life Sciences,

17085301) for collection and purification of the fluorophore-conjugated secondary antibody. Further characterization of the secondary antibody was done by ultraviolet/visible absorption spectroscopy.

Immunostaining procedure: The immunostain procedure was adapted from previous protocols,[43,47] and goes as follows: fixed RPE1 cells were incubated first in permeabilization solution (1× PBS with 0.1% (v/v) Triton X-100) for 10 min, then washed three times with 1× PBS. After permeabilization, cells were incubated in block solution (1× PBS with 10% (w/v) BSA and 3 mM sodium azide) for 1 h at room temperature, followed by incubation in primary solution (2–5 μg/ml of primary antibodies diluted in block) overnight at 4°C. The sample was washed with block three times (10 min each time), then incubated in secondary solution (2–3 μg/ml of secondary fluorophore-conjugated antibodies in block) for 1–2 h at room temperature. The sample was washed once for 10 min with block, then three times with 1× PBS azide. Samples which had been originally fixed in EtOH:MeOH were postfixed in 4% PFA in 1× PBS for 10 min, then washed three times with 1× PBS azide. Immunostained samples were either immediately gelled or stored in 1× PBS azide at 4°C for up to ∼1 week for later use (see Supplementary Table 2.1 for more details).

Cell gelation, digestion and expansion: Expansion microscopy was adapted from a previous protocol,[95] and goes as follows: immunolabeled cells were treated with freshly prepared 5 mM MA-NHS in 1× PBS for 10 min, then washed three times with 1× PBS. Cells were incubated in monomer solution (1× PBS with 2 M NaCl, 2.5% (w/w) acrylamide, 0.15% (w/w) N,N'- methylenebisacrylamide and 8.625% (w/w) sodium acrylate) for 10 min before gelation with 0.15–0.2% (w/v) APS and 0.2% TEMED (w/w) at room temperature for at least 30 min in a sealed container backfilled with nitrogen gas. After polymerization, the cell-embedded hydrogel was gently removed from the 12 mm coverslip, then incubated in digestion solution (1× TAE with 0.5% (v/v) Triton X-100, 0.8 M guanidine HCl and 8 units/ml proteinase K) overnight at 37°C. The digested sample was both washed and expanded by placing the sample in

deionized water, which was replaced every 15–20 min for at least three times. Hydrogels were stored in 2× SSC at 4°C, typically up to ∼1 month.

DNA fluorescence in situ hybridization: The general DNA FISH procedure for non-repetitive genomic regions (*GAPDH*, *MYL6*, *HOXC* and *LINC-PINT*) was adapted from previous protocols,[25,112] and goes as follows: Briefly, a small (∼3.5 mm × 3 mm × 2 mm) piece of gel from each expanded cell sample was first incubated in hybridization buffer (2× SSC with 50% (v/v) formamide and 0.1% (v/v) Tween 20) for 10 min at room temperature. Samples were incubated in pre-heated hybridization buffer for 30 min at 60°C. A hybridization mixture (2× SSC with 50% formamide (v/v), 10% dextran sulfate (w/v), 0.1% (v/v) Tween 20, 3 mM sodium azide, ∼10-20 nM oligo probe library per kb of targeted genomic region, and 1– 1.5× concentration of oligo reporters and adapters to oligo probe library) specific to each sample was preheated to 90°C for 5–10 min and then added to each sample at an approximate 2:1 volume ratio. Samples were denatured at 90– 92.5°C for 2.5–10 min and hybridized at 37–42°C overnight. Samples were washed three times, 15 min each time: first with preheated 2× SSCT (2× SSC with 0.1% (v/v) Tween 20) at 60°C, then with preheated 2× SSCT at 37°C, and lastly with 2× SSCT at room temperature. Samples were stored at 4°C in 0.2× SSCT (0.2× SSC with 0.01% (v/v) Tween 20) until needed (within a week). Samples were fully expanded to ∼4× the original size with deionized water at 4°C, replacing the water twice every 10 min (see Supplementary Table 2.1 for more details).

The DNA FISH procedure for the repetitive alpha-satellite region was done as follows: expanded RPE1 cells were incubated for 1 h at room temperature in 1× PBS. The sample was then incubated in 1× PBS supplemented with 100 µg/ml of RNase A for 1 h at 37°C. After RNA digestion, the sample was incubated in 2× SSCT for 30 min at room temperature. The samples were then incubated in hybridization buffer for 30 min at room temperature. The gel was transferred to a hybridization buffer containing 200 nM of alpha-satellite oligonucleotide probe. The sample was denatured for 15 min at 95°C. Gels were washed once in 20× SSC for 15 min at 37°C, then in 2× SSC for 1 h at 37°C. The samples were incubated

in 2× SSC with 200 nM alpha-satellite adapter probe and 600 nM of reporter probe A for 30 min at 37°C. The sample was washed with 20× SSC for 20 min at 37°C and lastly with 2× SSC for 20 min at room temperature. After this, the alpha-satellite sample was expanded to ∼3× the original size by incubating the sample in 0.2× SSC, then a second time in 0.2× SSC with 1 µg/ml of Hoechst 33258 (see Supplementary Table 2.1 for more details).

Sample mounting and imaging: For expanded samples using Alexa Fluor 750 fluorophore conjugated reporters, samples were incubated in imaging buffer (10 mM Tris buffer (pH 8) with 1 mM Methyl viologen, 1 mM ascorbic acid and 2% (v/v) MeOH) for 10 min. Before imaging, samples were first adhered to a poly-L-lysine–coated rectangular no. 1.5 coverslip, then they were supplemented with ∼30 units/ml alcohol oxidase and 0.2% (w/v) catalase. Samples that did not have Alexa Fluor 750 were adhered to a poly-L-lysine–coated rectangular no. 1.5 coverslip. All samples were imaged with either a Leica SP5 inverted confocal point scanning microscope at the University of Washington Biology Imaging Facility using a Plan Apo CS 63× 1.2 numerical aperture (NA) water-immersion objective, a homebuilt spinning disk confocal microscope using a Nikon CFI60 Plan Apochromat 60× 1.27 NA water-immersion objective, or a conventional wide-field epifluorescence inverted Nikon Ti-S microscope using a Nikon CFI Plan Apo VC 60× 1.2 NA water-immersion objective (see Supplementary Table 2.1 for more details).

CUT&RUN: H3K4me3, H3K27me3 and H3K27ac profiling CUT&RUN was performed as previously described,[111] with the following adaptations: 250 000 trypsinized RPE1 cells were used per antibody condition. Cells were bound to Concanavalin A coated magnetic beads (Bangs Laboratories, BP531), permeabilized with 0.025% (w/v) digitonin, then incubated overnight with 5 µg of either antiH3K4me3 (Active Motif, 39159), anti-H3K27me3 (Active Motif, 39155) or anti-H3K27ac (Active Motif, 39133) in 500 µl of solution at 4°C. Cells were washed then incubated with protein A-MNase fusion protein (a gift from S. Henikoff, FHCRC) for 15 min at room temperature. After another wash, cells were incubated with 2 mM calcium chloride for 30 min at 0°C to induce MNase cleavage activity. The reaction was

stopped by adding 100 µl of 2× STOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 50 µg/ml glycogen and 2 pg/ml of yeast spike-in DNA) to each sample. Cleaved histone–DNA complexes were isolated by centrifugation and DNA was extracted with a NucleoSpin PCR Clean-up kit (MachereyNagel, 740609).

Library preparation for each CUT&RUN antibody condition was done with a KAPA Hyper Prep Kit (VWR, 89125-040) with the PCR amplification settings adjusted to have simultaneous annealing and extension steps at 60°C for 10 seconds. Library products between 200–300 base pairs were selected using Agencourt AMPure XP beads (Beckman Coulter, A63880) then sequenced with an Illumina MiSeq system at the University of Washington Northwest Genomics Center with paired-end 25 bp sequencing read length and TruSeq primer standard for ~6 million reads per condition.

Paired-end sequencing reads were aligned separately to human (GRCh38/hg38) and yeast genomes using Bowtie2[113] with the previously suggested specifications for mapping CUT&RUN sequencing data:[111] –local –verysensitive-local –no-unal –no-mixed –no-discordant -I 10 - X 700. Alignment results were converted to BAM files with SAMtools[114] and then to BED files with BEDTools.[115] Reads were sorted and filtered to remove random chromosomes, then, with BEDTools genomecov, histograms were generated for the mapped reads using spiked-in yeast reads and the number of cells for each condition as scaling factors. The results were visualized using the WashU Epigenome Browser (https://epigenomegateway.wustl.edu/).[116]

Oligonucleotide probe design and amplification: DNA FISH probes were designed using OligoMiner,[117] with standard buffer, length and melting temperature conditions, with the exception of the target *MYL6*, which had the following adaptations: base length between 28–42 nucleotides and melting temperature between 38 and 46°C. Orthogonal DNA sequences, which were previously screened for DNA FISH purposes,[48,118] were appended to each probe as adapter/reporter hybridizing regions specific to each gene, along with a primer set for amplification. Designed probes were purchased as part of an oligo pool

from GenScript, and the probes were amplified using a T7/Reverse-Transcriptase amplification protocol previously published,[112] in an RNase-free environment with the following adaptations: After PCR amplification with a Phusion Hot-start master mix and purification with a DNA Clean & Concentrator-5 kit (Zymo Research, D4013), probes were T7 amplified with a HiScribe T7 Quick High Yield RNA Synthesis Kit (New England BioLabs, E2050S) supplemented with 1.3 units/l RNaseOUT (ThermoFisher Scientific, 10777019) for 16 h at 37°C. DNA was digested with DNase I for 1 h at 37°C. RNA was purified from the sample by first adding LiCl solution from the HiScribe Kit at a 1:7 ratio to the RNA solution, incubating the solution at −20°C for 30 min and pelleting the precipitated RNA by centrifugation (∼17 000 g) for 15 min at 4°C. The supernatant was removed from the tube and the pellet was washed with 70% EtOH. The RNA pellet was centrifuged (∼17 000 g) for 5 min at 4°C and, after carefully removing the supernatant, the pellet was left to dry for 3 min. The RNA was dissolved in water and ∼50 µg of RNA was added to Maxima H Minus RT buffer with 2.86 units/l Maxima H Minus RT Transcriptase, 2.3 units/l RNaseOUT, 1 mM dNTP and 14 µM Forward project primer. The solution was incubated at 50°C for 2 h, then samples were digested with 100 µg/ml RNase A for 1 h at 37°C. After RNase digestion, oligonucleotide probes were purified using a DNA Clean & Concentrator-25 kit (Zymo Research, D4033) with Oligo binding buffer (Zymo Research, D4060-1-10). The final product was assumed to have full yield (for a list of sequences, see supplementary spreadsheet).

Image processing and analysis for SCEPTRE profiling: Image processing and analysis was performed using MATLAB. First, raw images obtained from immunofluorescence channels were smoothed with a gaussian filter using 1–2 standard deviations within a 3 × 3 × 3 matrix. Smoothed images were contrast adjusted, where background pixel levels were clipped at an adaptively determined threshold for each image set at 2–9 third quartiles away from the median of each image stack histogram. The contrast adjusted images were binarized, either by an Otsu method or a Laplace filter with alpha = 0.2 followed by selection of all negative values. A nuclear mask (generated as described below) was applied to the

binarized immunofluorescence channel and, after small components (volume < 20 voxels) were removed, a watershed transformation was applied to the segmented clusters. Features, including mean fluorescence intensity of every immunofluorescence channel and overlap with clusters of other segmented immunofluorescence channels, were identified for each segmented cluster (see Supplementary Table 2.2 for more details). For the first image stack, each step was visually inspected to confirm proper threshold levels. The nuclear mask was generated by applying the same segmentation process from above to either a Hoechst stain channel or the same immunolabeled channel with a contrast adjustment done with 1–3 third quartiles clipping. The segmented channel was subject to morphological dilation with a sphere of 3 pixel radius, to fuse clusters within the nucleus. The convex hull was computed for the largest component (i.e., the nucleus) after all other clusters were removed. The segmented nucleus was morphologically eroded with the same sphere that was used to morphologically dilate the channel (see Supplementary Table 2.2 for more details). For the first image stack, each step was visually inspected to confirm proper threshold levels.

After segmentation of the nuclear channel and immunofluorescence channels, the FISH raw channel was segmented in the same manner with the following exceptions: (i) the nuclear mask was applied after smoothing and before contrast adjustment; (ii) clipping during contrast adjustment was performed with a threshold of 10–15 third quartiles away from the median; (iii) no watershed transformation was applied to FISH segmented regions; (iv) clusters intersecting the periphery of the nuclear mask (e.g. highly fluorescent contaminant in FISH channel next to nucleus) were removed. Features, including mean fluorescence intensity of every immunofluorescence channel and overlap with each immunofluorescence segmented cluster, were identified for each segmented FISH cluster. Since the segmented FISH channel can contain small and dim clusters that, by visual inspection, do not correspond to the FISH labeled genomic loci, small clusters (volume < 20–80 voxels) were filtered out before analyses. After segmentation of the FISH channel, randomly selected cubic regions were

generated throughout the nuclear region of each image stack with a volume approximately equal to the mean volume of the selected FISH clusters. Mean fluorescence intensities of each immunofluorescence channel were determined for these random clusters (see Supplementary Table 2.2 for more details). For the first image stack, each step was visually inspected to confirm proper threshold levels.

Data obtained from the segmented clusters were inspected using contour and scatter plots with MATLAB built-in functions, or violin plots, using the MATLAB script violinplot (https://github.com/bastibe/Violinplot-Matlab). Contours were smoothed with a gaussian filter using 1 standard deviation within a 5 × 5 matrix. Pearson correlation coefficients were determined using the MATLAB function corrcoef.

Statistical analyses: Each figure, along with their related supplementary figures, represents an individual experiment where all cells were labeled, expanded, imaged and processed under the same conditions. Cell numbers for each experiment were: 1 (Figure 2.3), 50 (Figure 2.4, Supplementary Figure 2.12 and 2.14), 52 (Figure 2.5A, Supplementary Figures 2.15A and 2.16A), 38 (Figure 2.5C, Supplementary Figures 2.15B and 2.16B), 54 (Figure 2.6, Supplementary Figure 2.19−2.20), 1 (Supplementary Figure 2.9), 36 (Supplementary Figure 2.10A), 10 (Supplementary Figure 2.10B), 48 (Supplementary Figure 2.11), 10 (Supplementary Figure 2.12), 10 (Supplementary Figure 2.13), 40 (Supplementary Figures 2.21 and 2.22), 20 (Supplementary Figures 2.23 and 2.24).

Fluorescence signal, defined as the mean fluorescence intensity for a given immunolabeled channel within a cluster of the same experiment, was used as the main measurement for comparing histone mark or paused RNA polymerase II levels within the segmented clusters of immunolabeled, FISH-labeled or randomly selected regions, or between the distribution of fluorescence signals for each set of clusters. Background fluorescence signal is defined as the mean fluorescence intensity for a given immunolabeled channel within the randomly-selected groups of background voxels for H3K4me3 marks in Figure 3.4, H3K27me3 marks in Figure 2.5C, or for both immunolabeled channels in Supplementary

Figure 2.13 and 2.17. The groups of background voxels were equal in size to the median size of segmented clusters for each channel.

An arbitrary 'on' threshold (equal to the 5th percentile of the fluorescent signal found within a respective immunolabeled cluster set) is represented in all graphs excluding violin plots, as a qualitative determinant of high or low fluorescence signal within each set of clusters. Correlation coefficients were determined for each comparison between fluorescence signals within a set of clusters. A right-tailed Wilcoxon rank sum test was used to determine if for a given experiment the median fluorescence signal of a cluster set was significantly higher than the median signal in randomly selected regions or a separate set. All numbers corresponding to fraction of overlap and distance are represented as mean ± standard deviation.

To better understand the detection limitations for immunostaining a particular histone mark, cells were concurrently labeled with an antibody of interest and an alternative antibody, both targeting the same mark. The detection efficiency based on colocalization is defined as the percent of voxels from the clusters of the alternative antibody that intersect with the clusters of the antibody of interest. This colocalization is interpreted as a lower bound measurement of detection efficiency due to the challenges of targeting the same epitope with two competing antibodies. A fluorescence-based analysis was used instead to evaluate the detection efficiency of the antibody of interest. In this analysis, we determine the percent of clusters from the alternative antibody distribution that are above the background (i.e., >99.9% of the background fluorescence signal distribution) for the fluorescence signal of the antibody of interest.

2.5 RESULTS:

2.3.1 SCEPTRE USES EXM TO CO-LOCALIZE IMMUNOLABELED PROTEINS AT DNA FISH LABELED GENOMIC REGIONS

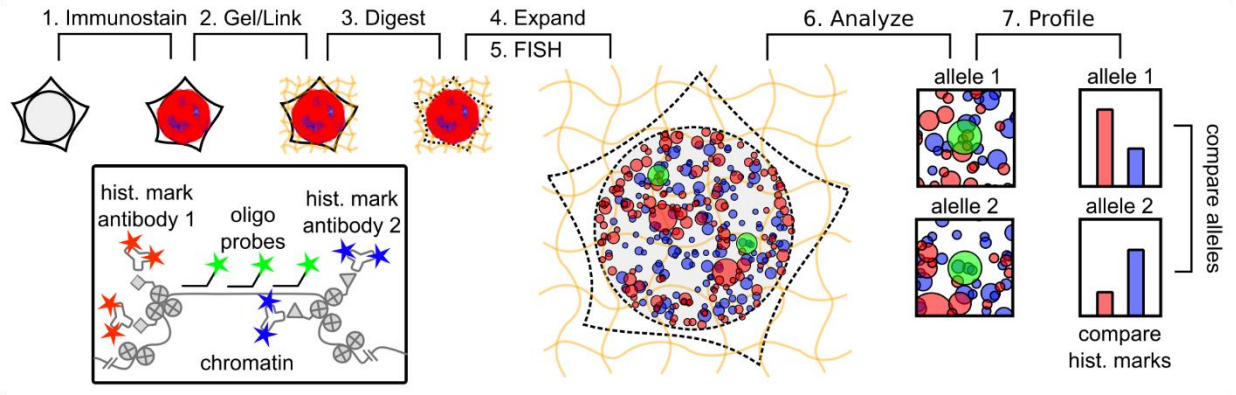


Figure 2.2 **Workflow of SCEPTRE.**
(1) Histone marks or other protein structures are antibody-labeled in fixed cells. (2) The sample and antibodies are linked to a swellable hydrogel grown within the sample. (3) The sample is digested by proteinase K. (4) The hydrogel is expanded in water. (5) DNA loci, alleles from the same or different genes, are labeled by FISH. (6) The sample is imaged, and relevant features are extracted for analysis. (7) An epigenetic profile is constructed for each cell, comparing histone mark levels between alleles or genes.

The labeling of individual genomic loci by DNA fluorescence in situ hybridization (FISH) has provided a powerful tool for visualizing chromatin structure in single cells.[23–26] While DNA FISH could be combined with immunofluorescence labeling to enable concurrent visualization of chromatin modification states and associated proteins, integration of these two techniques has been challenging, because the harsh conditions required to melt double stranded genomic DNA during labeling (e.g., treatment with hot formamide) may remove antibody labels applied before FISH or may compromise the antigenicity of relevant epitopes for post-FISH immunolabeling (Supplementary Figure 2.7).[79,80,108,109] To overcome this challenge, we employed expansion microscopy (ExM) as a means to preserve the signal of immunolabeled protein structures during DNA FISH labeling. In ExM, immunolabeled structures are covalently linked to a swellable hydrogel polymer scaffold that is isotropically expanded in deionized water in order to reveal features closer than the ~250 nm diffraction limit of light in the expanded

state.[94,95] ExM not only provides a high spatial resolution (~75 nm or better when using a standard confocal microscope with ~4× expanding gels), but also enables antibody labels to be covalently tethered to the hydrogel scaffold, such that DNA FISH can subsequently be performed without loss of antibody fluorescence. ExM has previously been combined with DNA FISH to either visualize the HER2 gene in tissue,[98] or to visualize repetitive centromere regions in plants.[119] However, this combination has not yet been used to determine the density of a protein structure, such as histone mark clusters, at specific genomic regions. We refer to this new methodology as Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE), as a tool to quantify the fluorescence signal of immunolabeled histone marks or proteins structures at individual FISH-labeled genomic loci within individual cells (Figure 2.2).

To test the ability of SCEPTRE to report on DNA-protein associations within the nucleus, we immunostained centromere-associated proteins while using DNA FISH to co-stain the repetitive alpha-satellite DNA of centromeres (Figure 2.3A). ExM images revealed discrete regions corresponding to alpha-centromeres and centromere-associated proteins, as well as significant overlap between these two regions. To quantify this degree of overlap, we created an automated image analysis software routine in MATLAB to segment individual regions corresponding to centromeres and centromere-associated proteins, and then quantified their degree of co-localization (Supplementary Figure 2.8). From this analysis, we found that DNA-labeled regions had almost complete fractional overlap with centromere-associated proteins (0.97 ± 0.06). Furthermore, the distance between the nearest-neighbor centroids of the protein and DNA labeled regions was small relative to the average radius for either region (77 ± 85 nm versus 234 ± 68 nm (anti-cen.) and 224 ± 65 nm (cen.), respectively). We then quantified the fluorescence signal of labeled centromere-associated proteins at individual centromeric DNA clusters, along with randomly selected regions of comparable size to these FISH clusters (Figure 2.3B). While immunofluorescence and FISH-labeled regions maintained similar anti-centromere

fluorescence signals, these regions showed much higher signals compared to randomly selected regions. Therefore, due to the high overlap between the FISH-labeled and immunolabeled regions, the high anti-centromere signal in the FISH-labeled regions, and the fact that the anti-centromere labeled structures did not shift much between pre- and post-expansion (Supplementary Figure 2.9), we concluded that ExM can co-localize the signal of protein and DNA components of a genomic region within a nucleus.

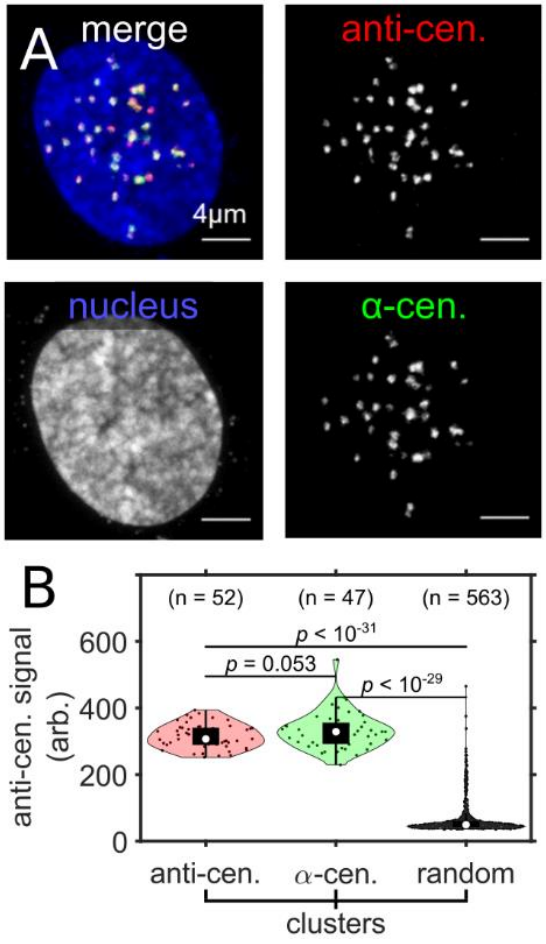


Figure 2.3 **ExM reveals colocalization between centromere-associated proteins with repetitive centromeric DNA.**
(A) Maximum intensity projection image of an entire expanded RPE1 cell nucleus with immunolabeled centromere associated proteins (anti-cen., red), FISH labeled alpha-satellite DNA of centromeres (α-cen., green) and Hoechst-stained nucleus (blue). (B) The distribution of anti-centromere fluorescence signal (arb. = arbitrary units) in anti-centromere, α-centromere and in randomly selected region (random) clusters within the nucleus of the cell in (A). Significance determined by a right-tailed Wilcoxon rank-sum test for anti-centromere against random, α-centromere against random, and α-centromere against anti-centromere cluster distributions. All scale bars are in pre-expansion units.

2.3.2 SCEPTRE RESOLVES MULTIPLE HISTONE MODIFICATIONS AT SINGLE GENE LOCI IN SINGLE CELLS

To determine whether SCEPTRE can distinguish between multiple histone marks at a single, non-repetitive genomic region, we concurrently visualized two histone marks, H3K4me3 and H3K27me3, at the house-keeping gene *GAPDH* (Figure 2.4). *GAPDH* encodes for glyceraldehyde3-phosphate dehydrogenase, which is highly expressed in many cell types[120] due to its essential role in metabolism;[121] therefore, histone H3K4me3, commonly found at active gene promoters,[33] is expected to be present at *GAPDH*, whereas histone H3K27me3, which is associated with repressed regions,[64] is expected to be absent. Using SCEPTRE, we measured the fluorescence signals of immunolabeled H3K4me3 and H3K27me3 marks at the FISH-labeled *GAPDH* locus, along with the fraction of overlap between *GAPDH* and H3K4me3 or H3K27me3 clusters (Figure 2.4).

From this analysis, we observed that *GAPDH* had much higher H3K4me3 fluorescence signal compared to H3K27me3 signal (Figure 2.4E and F). To our surprise, the H3K4me3 signal found at *GAPDH* varied greatly between loci, with some loci having high signals while others a more baseline level (Figure 3G). These results were the same when only one of the two histone marks was labeled and imaged with *GAPDH* (Supplementary Figure 2.10), or when a different set of antibodies was used to label H3K4me3 and H3K27me3 in RPE1 cells (Supplementary Figure 2.11).

Interestingly, histone mark signals were uncorrelated between *GAPDH* alleles in the same cell (Supplementary Figure 2.12A-B), suggesting histone mark levels at alleles from the same gene are independently regulated. When comparing these fluorescence signals to those obtained from randomly selected regions across the nucleus, *GAPDH* showed lower H3K27me3 signals and much higher H3K4me3 signals than those found at random regions. Similar to the fluorescence signal results, the mean fraction of overlap of *GAPDH* with H3K4me3 clusters was higher than with H3K27me3 clusters (0.21 ± 0.21 versus 0.045 ± 0.11, respectively). To corroborate these results, we measured the density of H3K4me3 and H3K27me3 marks across the RPE1 genome for an ensemble of cells using CUT&RUN

followed by sequencing.[110,111] Analysis of the CUT&RUN sequencing results revealed that a substantial presence of H3K4me3 marks was found at the targeted *GAPDH* region and only background levels of H3K27me3 marks were found for this same region (Figure 2.4H), with the closest repressed region observed ∼500 kb away. These results demonstrate that SCEPTRE can distinguish between the abundance of two histone modifications at individual non-repetitive genomic regions within a nucleus.

To determine whether the heterogeneity of H3K4me3 marks at *GAPDH* results from detection limitations of the antibody used above (Rb × H3K4me3), we concurrently labeled H3K4me3 with Rb × H3K4me3 and an alternative antibody targeting the same mark but stained using a different fluorophore (Ms × H3K4me3, Supplementary Figure 2.13). As an initial estimate of detection efficiency, we calculated the percent of all voxels segmented for the alternative antibody that also segmented for the antibody of interest. This analysis yielded a nominal detection efficiency of 26% for Rb × H3K4me3; however, this value likely under-estimates the true detection efficiency, as the two antibodies likely compete for the same epitope (i.e., the methylated lysine moiety), reducing their ability to co-segment within the same voxel, particularly at high spatial resolution. Therefore, as an alternative measurement of detection efficiency, we directly quantified the cumulative fluorescence signal distribution for Rb × H3K4me3 within clusters segmented with the alternate antibody (Ms × H3K4me3), and we calculated the fraction of this signal above background. This analysis yielded a higher detection efficiency of 88% (Supplementary Figure 2.13). In contrast, the probability of detecting an antibody signal above background for *GAPDH* loci was considerably lower, at 67% (Supplementary Figure 2.13E); thus a substantial degree of heterogeneity in H3K4me3 levels at this locus likely reflects biological variability rather than limitations of antibody detection.

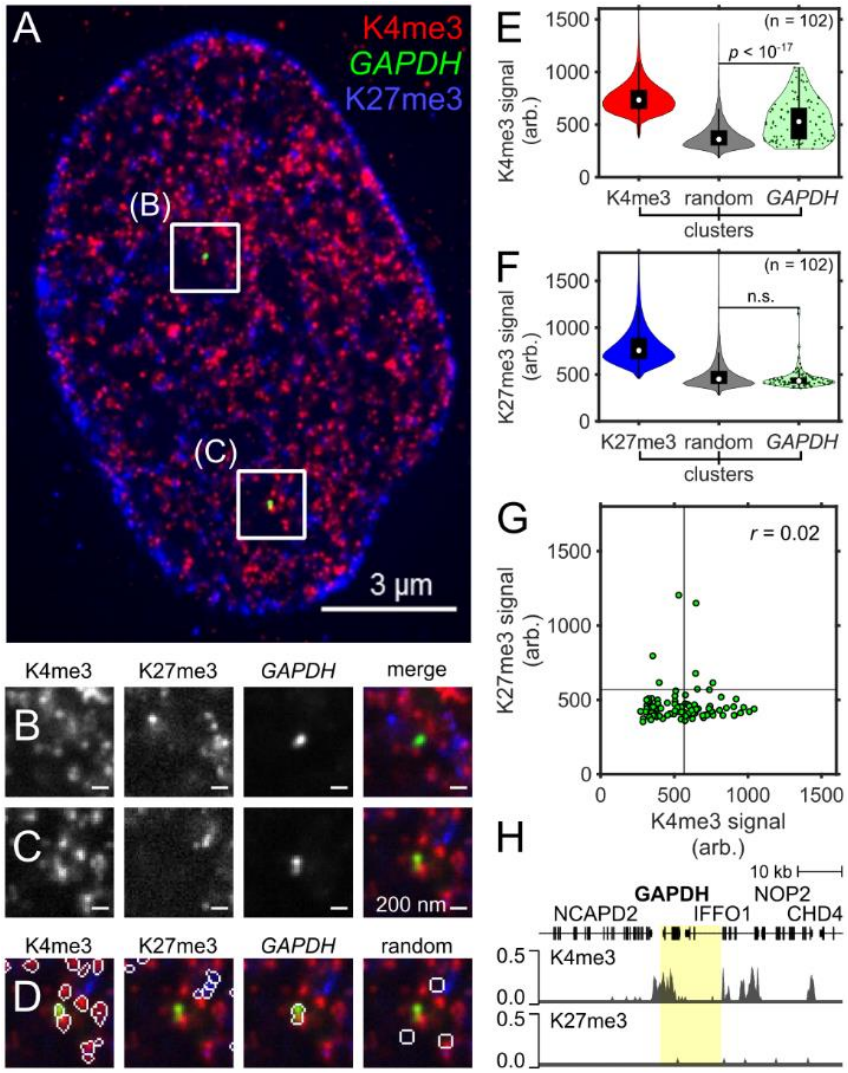


Figure 2.4 **SCEPTRE distinguishes two histone marks at one genomic region.**
(A) An expanded RPE1 cell with immunolabeled H3K4me3 marks (K4me3, red) and H3K27me3 marks (K27me3, blue), and FISH-labeled *GAPDH* (green). (B, C) Zoomed in views of the approximate center plane of an image stack for each *GAPDH* allele in the cell seen in (A). (D) Outline of the segmented regions for H3K4me3, *GAPDH*, H3K27me3 and randomly selected region clusters for the image plane seen in (C). (E) Distribution of H3K4me3 fluorescence signal (arb. = arbitrary units) within H3K4me3, randomly selected regions (random) and *GAPDH* clusters. (F) Distribution of H3K27me3 fluorescence signal within H3K27me3, randomly selected regions and *GAPDH* clusters. (G) H3K27me3 and H3K4me3 fluorescence signals within *GAPDH* clusters (green). Black lines represent the threshold 'on' level for each fluorescence signal. The correlation coefficient (r) between fluorescence signals within *GAPDH* is shown in the top-right corner of the plot. (H) CUT&RUN normalized counts for H3K4me3 (top) and H3K27me3 (bottom) marks in RPE1 cells for the FISH-targeted *GAPDH* region (highlighted). Cluster numbers for (E). and (F). are K4me3 = 343334, K27me3 = 478825, random = 8322, *GAPDH* = 102. Significance determined by a right-tailed Wilcoxon rank-sum test of histone mark fluorescence signals in *GAPDH* against random cluster distributions. All scale bars are in pre-expansion units.

H3K4me3 and H3K27me3 are generally thought to mark distinct chromatin states, though they have been reported to colocalize to form 'bivalent domains' on genes primed for transcription.[32,122] We therefore investigated the relationship between H3K4me3 and H3K27me3 clusters across the nuclei of the RPE1 cells. As previously observed,[47] H3K27me3 clusters preferentially inhabited the periphery of the nucleus, whereas H3K4me3 clusters were more evenly distributed (Figure 3A). There was a low fraction of spatial overlap between H3K4me3 clusters and H3K27me3 clusters (0.079 ± 0.14 H3K4me3 with H3K27me3, 0.12 ± 0.16 H3K27me3 with H3K4me3). The H3K4me3 fluorescence signal in H3K27me3 clusters, as well as the H3K27me3 signal in H3K4me3 clusters, was substantially low, albeit higher than the distribution of random regions (Supplementary Figure 2.12C-D). We therefore plotted the frequency of H3K4me3 and H3K27me3 signals within each of the other's histone mark's clusters, along with these signals found in randomly selected regions (Supplementary Figure 2.14). These plots show that H3K4me3 and H3K27me3 form largely non-overlapping clusters, though there exists a small fraction of clusters having high signal from both histone marks. These results suggest that H3K4me3 and H3K27me3 mostly form disjoint clusters, though a very small fraction may colocalize, similar to what has been observed in other differentiated cell types.[123]

2.3.3 SCEPTRE QUANTIFIES HISTONE MODIFICATION LEVELS AT MULTIPLE GENOMIC LOCI IN SINGLE CELLS

To test whether SCEPTRE can quantify histone mark signals at multiple genomic loci within the same cell, we designed a library of FISH probes to simultaneously label three different genomic regions in RPE1 cells. The first region contains *MYL6*, a gene on Chr12 encoding myosin light chain-6 that is actively expressed in the eye;[124] the second contains the *HOXC* gene cluster, which is normally active in progenitors but repressed upon differentiation;[64,65] the third covers an internal region of long intergenic non-coding P53 induced transcript (*LINC-PINT*) variant, a non-coding transcript that is found on Chr7, which is broadly expressed across multiple tissues (Figure 2.5).[125] Previous RNA-seq results in RPE1 cells[126] indicated that *MYL6* is expressed at high levels, that all genes within the *HOXC* cluster are silent,

and that *LINC-PINT* is transcribed at low levels (Supplementary Figure 2.9). As expected, bulk analysis of histone modifications at these loci using CUT&RUN revealed H3K4me3 peaks at *MYL6* and *LINC-PINT*, but not within the *HOXC* cluster. H3K27me3 marks, on the other hand, covered a large region encompassing the *HOXC* cluster, but were largely absent from *MYL6* and *LINC-PINT* (Figure 2.5E).

In agreement with population-level results, H3K4me3 fluorescence signals measured using SCEPTRE were significantly higher at *MYL6* and *LINC-PINT* than at randomly chosen clusters (Figure 2.5B, $P < 10^{-5}$, *MYL6*; $P<10^{-6}$, *LINC-PINT*), indicating enrichment of this histone modification at these two loci. Interestingly, both genes showed high H3K4me3 variability between individual loci, similar to what was observed at *GAPDH* loci. In contrast, H3K4me3 signals at the *HOXC* cluster were not significantly higher than those found at random regions. Consistently, H3K4me3 clusters showed greater overlap with the *MYL6* and *LINC-PINT* loci than at the *HOXC* cluster, where it appeared to be visibly excluded (0.23 ± 0.26, *MYL6* and 0.20 ± 0.24, *LINC-PINT* versus 0.068 ± 0.16, *HOXC*). Therefore, SCEPTRE detects differences in H3K4me3 levels between multiple genomic regions seen with population-level analysis, agreeing with the results obtained by CUT&RUN. We note that H3K4me3 mark signals were largely uncorrelated between two alleles of each gene in each cell (Supplementary Figure 2.15A), as well as between the loci of different genes in the same cell (Supplementary Figure 2.16A), indicating that the levels of this histone mark are largely independent across loci within individual cells.

Also in agreement with bulk analysis, The *HOXC* cluster showed higher H3K27me3 fluorescence signal compared to random clusters, and higher H3K27me3 signal compared to *MYL6* or *LINC-PINT* (Figure 2.5D). Strikingly, H3K27me3 levels varied substantially between different *HOXC* clusters, with a substantial fraction of *HOXC* clusters with either low or even baseline levels of H3K27me3, even though all genes in this cluster were silent in their expression (Supplementary Table 2.3). To determine whether this heterogeneity represents limits in H3K27me3 detection versus bona fide biological heterogeneity, we calculated H3K27me3 detection efficiency for this antibody (Rb × H3K27me3, Figure 4C) by co-

staining with an alternative antibody, as above (Ms × H3K27me3, Supplementary Figure 2.17). This analysis yielded a H3K27me3 detection efficiency of 25%, as measured by co-segmentation of the two antibodies, and 72%, determined by quantifying the fraction of Rb × H3K27me3 signal above background for clusters segmented with the alternative antibody (Supplementary Figure 2.17). In contrast, only 11% of antibody signals at *HOXC* loci were detected above background (Supplementary Figure 2.17E), suggesting that the majority of H3K27me3 variability at *HOXC* reflects innate biological heterogeneity as opposed to technical limitations of antibody detection.

Although *MYL6* did not have significantly higher H3K27me3 fluorescence levels compared to random regions, *LINC-PINT* did, despite an absence of H3K27me3 marking seen in CUT&RUN data. The presence of H3K27me3 at some *LINC-PINT* loci may reflect the looping of this locus to a different genomic region where H3K27me3 is present; to investigate this possibility, we consulted a previously published Hi-C data set for the RPE1 cell line;[127] indeed *LINC-PINT* maintained high frequency contacts with an adjacent H3K27me3 domain (Supplementary Figure 2.18), which may contribute to its lower transcriptional levels compared to *MYL6* (Supplementary Table 2.3). These results suggest that, at the current spatial resolution, adjacent genomic regions can influence each other's histone mark levels detected by SCEPTRE. That being said, the SCEPTRE results broadly agree with the results obtained by CUT&RUN and can distinguish between the chromatin modification states of multiple genes in the same cell (e.g., *MYL6* and *HOXC*). As seen with the H3K4me3 marks at these genomic regions, we observed no relationship between the H3K27me3 levels for two alleles of the same gene (Supplementary Figure 2.15B) or for alleles from different genes (Supplementary Figure 2.16B) in the same cell.

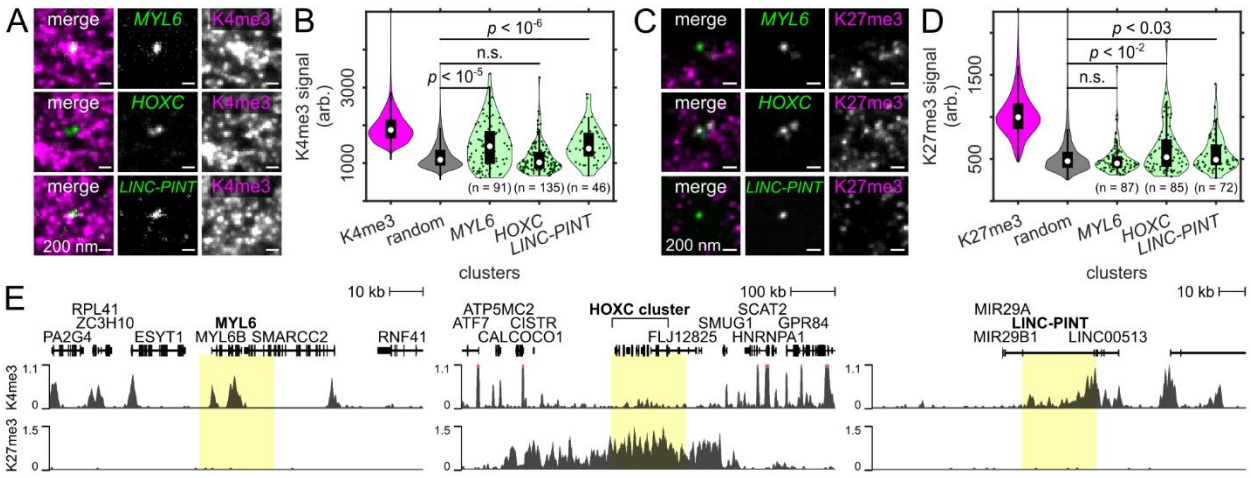


Figure 2.5 **SCEPTRE quantifies one of two histone marks at three genomic regions.**
(A) Example images of the approximate center plane for each image stack of simultaneously FISH-labeled *MYL6*, *HOXC* or *LINC-PINT* loci (green) from the same expanded RPE1 cell immunolabeled for H3K4me3 marks (K4me3, magenta). (B) Distribution of H3K4me3 fluorescence signals (arb. = arbitrary units) within H3K4me3, randomly selected regions (random), *MYL6*, *HOXC* and *LINC-PINT* clusters (cluster numbers are K4me3 = 390331, random = 7421, *MYL6* = 91, *HOXC* = 135, *LINC-PINT* = 46). (C) Example images of the approximate center plane for each image stack of simultaneously FISH-labeled *MYL6*, *HOXC* or *LINC-PINT* loci (green) from the same expanded RPE1 cell immunolabeled for H3K27me3 marks (K27me3, magenta). (D) Distribution of H3K27me3 fluorescence signals within H3K27me3, randomly selected regions, *MYL6*, *HOXC* and *LINC-PINT* clusters (cluster numbers are K27me3 = 196 798, random = 6041, *MYL6* = 87, *HOXC* = 85, *LINC-PINT* = 72). (E) CUT&RUN normalized counts for H3K4me3 (top) or H3K27me3 (bottom) at the FISH-labeled *MYL6*, *HOXC* and *LINC-PINT* regions (highlighted). Significance determined by a right-tailed Wilcoxon rank-sum test of fluorescence signals in each FISH-labeled set against the random cluster distribution. All scale bars are in pre-expansion units.

2.3.4 H3K4ME3 MODIFICATIONS COINCIDE WITH PAUSED RNA POLYMERASE II AT A TRANSCRIPTIONALLY ACTIVE LOCUS

H3K4me3 levels have been reported to correlate with active transcription[34] and a model has been proposed where H3K4me3 facilitates the loading of RNA polymerase II, which remains paused proximally to the gene's promoter until a subsequent release step.[128] However, this model was based on separate population-level measurements of H3K4me3 and RNA polymerase II, and did not distinguish whether both components coincide directly at the same time at single loci in cells. To test whether both H3K4me3 and paused RNA polymerase II were present simultaneously at *GAPDH*, we performed SCEPTRE with H3K4me3 and the post-translationally modified form of paused RNA polymerase II during

transcription initiation, where the Serine 5 of the repeat C-terminal domain of RNA polymerase II is phosphorylated (Figure 2.6).[129–131]

We detected a large coincidence between H3K4me3 and paused RNA polymerase II, both at the *GAPDH* locus and also more broadly in the nucleus. At individual *GAPDH* loci, there were high signals from both H3K4me3 and paused RNA polymerase II (Figure 2.6B-E), such that there was also a strong correlation between these signals (Figure 2.6F, r = 0.70). Consistently, *GAPDH* overlapped with both H3K4me3 and paused RNA polymerase II clusters (0.23 ± 0.21 and 0.21 ± 0.19, respectively). Similarly to H3K4me3, paused RNA polymerase II signals were uncorrelated between *GAPDH* loci in the same cell (Supplementary Figure 2.19). In the nucleus more broadly, there was also substantial colocalization between H3K4me3 clusters and paused RNA polymerase II clusters (Figure 2.6B, fraction of overlap 0.19 ± 0.21), as well as a strong correlation between these two signals in randomly selected region clusters (Supplementary Figure 2.20C, r = 0.68).

In contrast, no correlation was seen at *GAPDH* between H3K4me3 and H3K27me3 signals (Figure 2.4G, r = 0.02), or between H3K27me3 and paused RNA polymerase II (Supplementary Figure 2.21F, r = 0.04). On a broader level, there was also little to no correlation in random regions between H3K4me3 and H3K27me3 (Supplementary Figure 2.14C, r = 0.18), or between H3K27me3 and paused RNA polymerase II (Supplementary Figure 2.22C, r = 0.17). Interestingly, when H3K27ac, another active histone mark,[62] was concurrently visualized with paused RNA polymerase II, some correlation was seen between these two signals, with r = 0.43 at *GAPDH* (Supplementary Figure 2.23F), and r = 0.59 at random regions (Supplementary Figure 2.24C). However, the fraction of *GAPDH* loci with high H3K27ac signals was smaller compared to that with high paused RNA polymerase II signals, suggesting that H3K27ac and the phosphorylation indicative of paused RNA polymerase II play distinct roles in the transcriptional cycle. Together, these results are consistent with a close regulatory relationship between

H3K4me3 modifications and the loading of paused RNA polymerase II, both at *GAPDH* and more broadly across the nucleus.

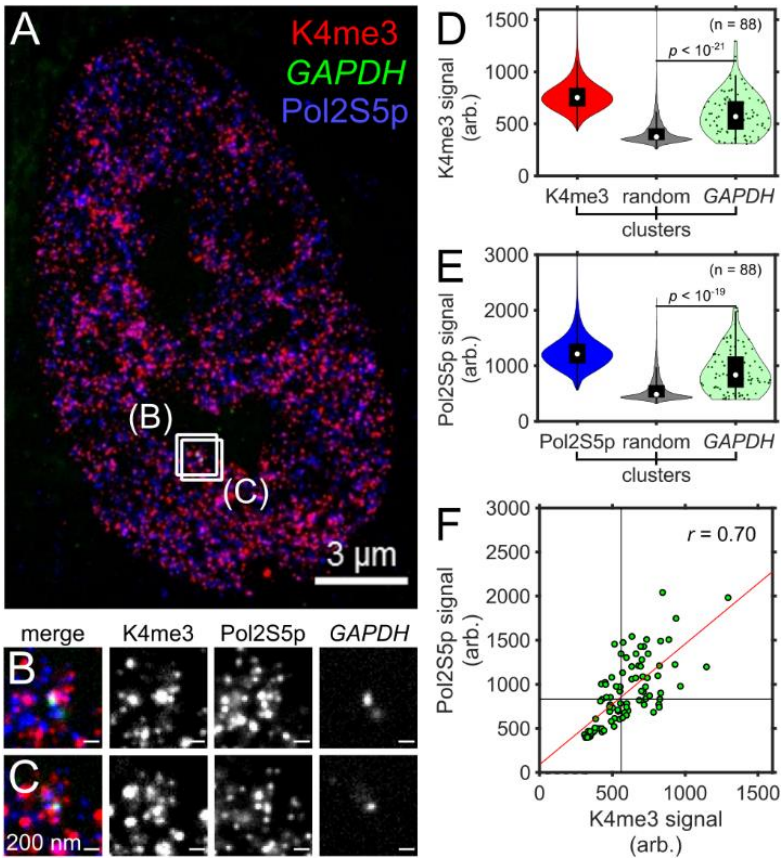


Figure 2.6 **SCEPTRE compares H3K4me3 and paused RNA polymerase II signals at a single genomic region.**

(A) An expanded RPE1 cell with immunolabeled H3K4me3 marks (K4me3, red) and paused RNA polymerase II (Pol2S5p, blue), and FISH-labeled *GAPDH* (green). (B, C) Zoomed in views of the approximate center plane of an image stack for each *GAPDH* allele in the cell seen in (A). (D) Distributions of H3K4me3 fluorescence signal (arb. = arbitrary units) within H3K4me3, randomly selected regions (random) and *GAPDH* clusters. (E) Distribution of paused RNA polymerase II fluorescence signal within paused RNA polymerase II, randomly selected regions and *GAPDH* clusters. (F) H3K4me3 and paused RNA polymerase II fluorescence signals within *GAPDH* clusters (green). Black lines represent the threshold 'on' level for each fluorescence signal, while the red line represents the linear regression. The correlation coefficient (r) between fluorescence signals within *GAPDH* is shown in the top-right corner of the plot. Cluster numbers for (D). and (E). are K4me3 = 440298, Pol2S5p = 542245, random = 8240, *GAPDH* = 88. Significance determined by a right-tailed Wilcoxon rank-sum test of histone mark fluorescence signals in *GAPDH* against random cluster distributions. All scale bars are in pre-expansion units.

2.6 DISCUSSION:

SCEPTRE is a new method capable of profiling chromatin states at multiple genomic loci within the 3D nuclear context of a cell by combining immunofluorescence with DNA in situ labeling by means of ExM. This combination enables efficient detection of histone mark fluorescence signals at a resolution of ~75 nm, sufficient to quantify histone mark abundance at individual genomic loci. In contrast to sequencing-based methods, SCEPTRE provides quantitative measurements of physical properties, such as overlap, density, and position within the nucleus for more than one histone mark at multiple genomic regions. Such measurements reveal a heterogeneity in chromatin states that has been previously masked in ensemble sequencing-based methods.

There are limitations to SCEPTRE compared to other histone mark profiling methods. Sequencing based methods can achieve nucleosome level resolution for histone mark mapping across an entire genome, such as in the case of CUT&RUN.[110] Since SCEPTRE relies on DNA FISH, detection of genes by in situ labeling often is limited to a minimum labeling size of over 10 kb, since smaller regions are detected with lower efficiency. However, genome organization is thought to occur at a larger scale than that of the single nucleosome. Nucleosomes are known to organize as clusters throughout the nucleus, with spatial sizes ranging around 50–100 nm,[43] a size that corresponds to ~10 kb of genomic DNA, depending on the region's activity state.[48] The scale increases further when observing Topologically Associating Domains (TADs), which are genomic regions of ~200 kb to 1 Mb in size that maintain similar epigenetic and regulatory landscapes,[49,50] or the smaller sub-TADs that are ~185 kb,[51] with spatial sizes of ~160 nm.[52] Even larger than 1 Mb are chromatin A and B compartments which are associated with broader open (active) and closed (repressed) states,[36] with spatial sizes on the µm scale.[53,54] Since SCEPTRE has allowed us to profile multiple genes at the lower scale of this genomic organization, there is potential to build upon this technique in order to target larger genomic regions by using multiplex FISH methods, such as MERFISH,[27] seqFISH,[28] ORCA[29] or OligoFISSEQ.[132] These methods

would allow SCEPTRE to interrogate the relationships between histone modifications and gene activity at a variety of developmentally-regulated genes, and at increasingly larger scales of genome organization.

SCEPTRE revealed heterogeneity in the levels of H3K4me3 at active genomic regions such as *GAPDH*, *MYL6* and *LINC-PINT* beyond what would be expected from technical limitations of SCEPTRE. This variability, which has been suggested in sequencing-based single-cell chromatin profiling studies,[75,76] suggests that active gene loci can adopt different states with different levels of H3K4me3 modification. Moreover, because SCEPTRE revealed a close correlation between H3K4me3 and paused RNA polymerase II levels at individual loci, this heterogeneity may reflect differences in the transcriptional stat of each gene. Given that genes are transcribed in bursts,[133] where polymerase recruitment happens intermittently,[134] it is plausible that H3K4me3 marks and polymerase II phosphorylation at Serine 5 are concurrently added during a transcriptional burst, but removed at a later stage in the transcriptional cycle. Moving forward, it would be useful to utilize SCEPTRE to further visualize H3K4me3 and other histone marks alongside different stages of transcription, to elucidate how histone marks participate in the regulation of gene transcription.

Similar to H3K4me3, H3K27me3 also appeared to show substantial heterogeneity in its levels at individual *HOXC* loci, with a considerable fraction of loci having low or baseline levels of this modification. As the *HOXC* maintains a transcriptionally silent state in these cells (Supplementary Table 2.3), our results suggest that the *HOXC* cluster is able to maintain a silent chromatin state even with low or baseline levels of H3K27me3 modification. In line with these findings, our recent study has pointed to a necessity for H3K27me3-independent mechanisms in the maintenance of compacted, polymerase-inaccessible state at repressed developmental gene loci.[135] Gene repression at the Hox gene cluster requires PRC1, a protein complex that mediates chromatin compaction and gene silencing.[136] PRC1 binds to H3K27me3, an interaction that explains the co-localization of these two factors in the genome;

however, PRC1 can also bind genomic loci independently of H3K27me3, and thus could conceivably maintain a repressed state at the *HOXC* locus in the absence of H3K27me3.[137] To further investigate these ideas, it will be helpful to use SCEPTRE to interrogate polycomb domains at other genomic loci.

Lastly, there are certain factors that influence the way SCEPTRE profiles the epigenetic state of genes. As demonstrated in the example of the *LINC-PINT* region (Figure 2.5 and Supplementary Figure 2.15), the 3D context of a region can influence its epigenetic profile. This 'crosstalk' from neighboring regions is most-likely due to the fact that genes with different epigenetic states may be found closer than the resolution of ExM provides at a 4× expansion factor (~75 nm). If so, methods that achieve better resolution, such as iterative ExM[138] or the combination of ExM with structured illumination microscopy,[139] would provide a greater distinction between epigenetic states of neighboring genes while using SCEPTRE. Another factor that may play a role in profiling are the genetic elements within a targeted region. The labeled *LINC-PINT* region was an internal sequence of a gene, which showed a different distribution of H3K4me3 signals compared to the *MYL6* region, whose promoter was found at the center of the labeled region. Therefore, considering the 3D context of chromatin within cells (seen by Hi-C from a previous study)[127] and the epigenetic landscape of a genomic sequence (seen by CUT&RUN in this study) can help with either selecting each targeted region for SCEPTRE, or in determining the epigenetic state for each region within a cell. Further improvements in multiplexing and resolution would allow SCEPTRE to systematically profile chromatin states in the genome, providing new insights into our understanding of genome structure and function.

2.7 DATA AVAILABILITY:

CUT&RUN sequencing data for H3K4me3, H3K27me3 and H3K27ac were submitted to the NCBI gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo/) and are available under the accession number GSE160784. These results, which were obtained using ENCODE-validated antibodies, can be viewed using the following UCSC genome browser session: https://genome.ucsc. edu/s/marcwood/SCEPTRE

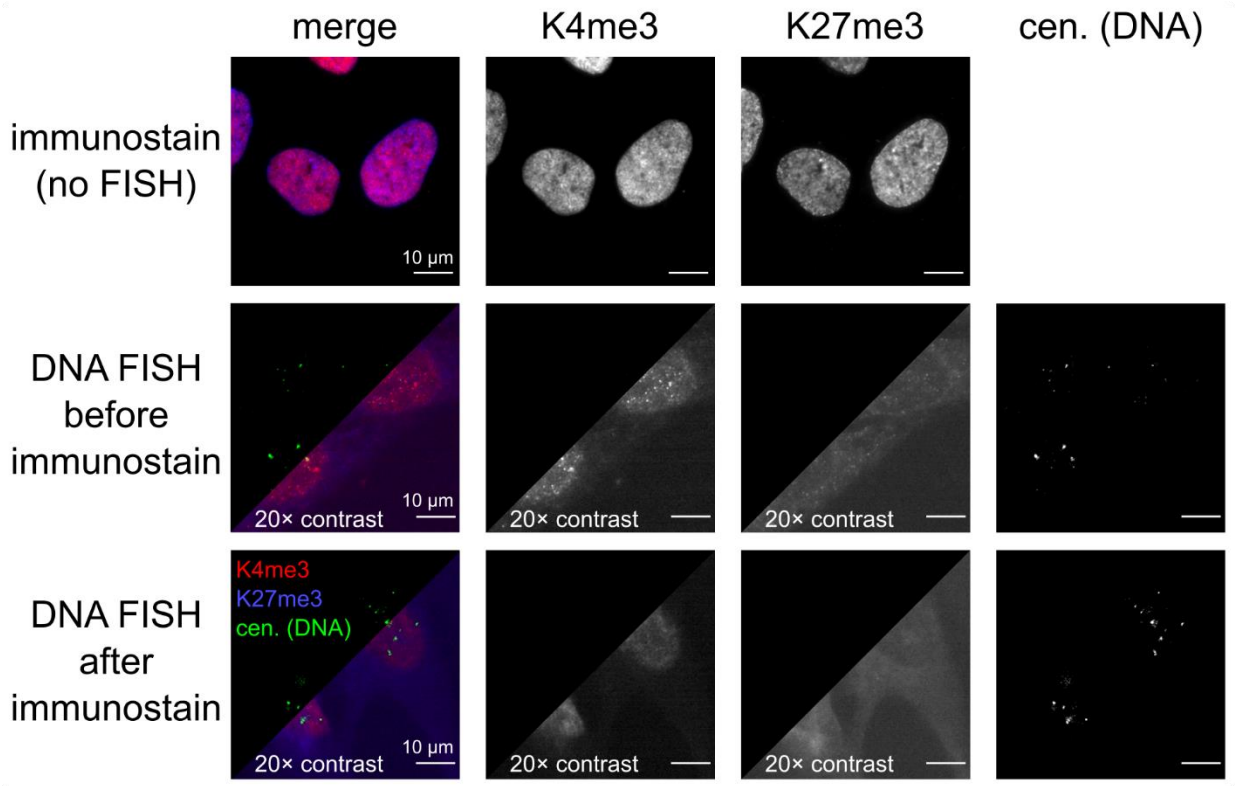
RPE1 CnR hg38. For FISH-labeled regions, CUT&RUN results were confirmed by comparison to previously published ChIP-seq results.[140,141]

MATLAB scripts for image processing and analysis are available at GitHub (https://github.com/marcwood13/ SCEPTRE pipeline). Additional data related to this paper will be made available by the corresponding authors upon reasonable request.

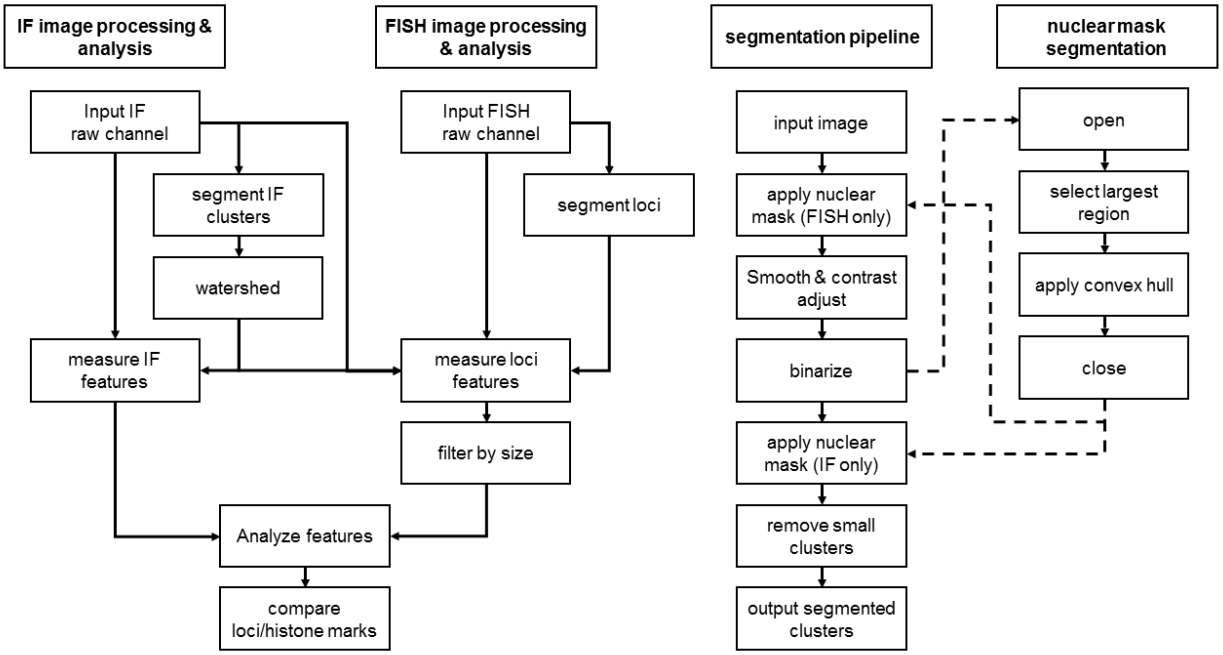
2.8 NOTE ADDED IN PROOF:

A recently published paper combined immunofluorescence and DNA FISH to visualize multiple histone modifications across many genomic regions using sequential hybridization,[87] though imaging was performed using conventional light microscopy that limited their analysis to chromosomal scale domains. This sequential hybridization approach developed can potentially be combined with SCEPTRE to allow for finer-scale profiling of chromatin states across a large number of gene loci.

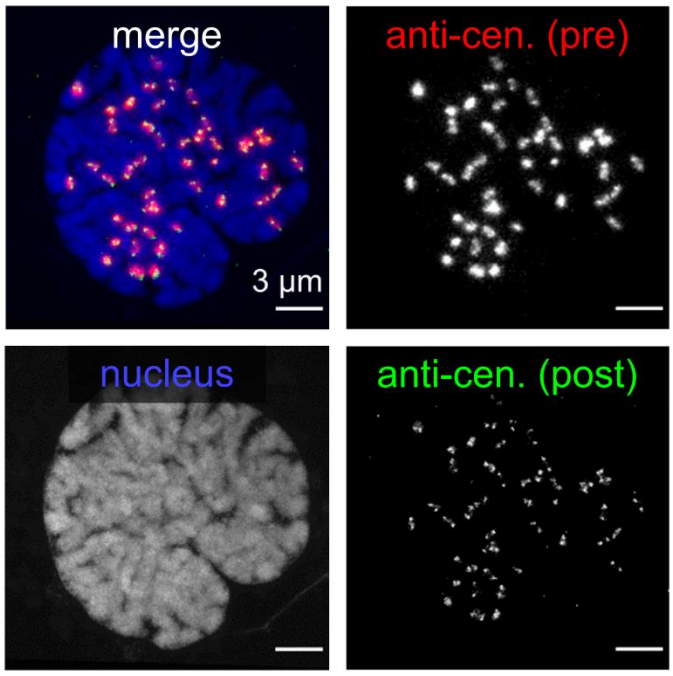
2.9 SUPPLEMENTARY DATA:



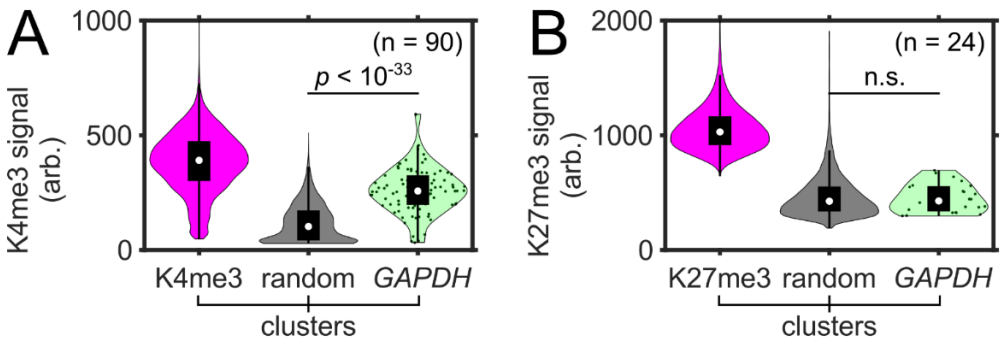
Supplementary Figure 2.7 **DNA FISH can disrupt the immunolabeling of nuclear structures.**
Widefield images of DNA FISH labeled centromeric DNA (green) using hot formamide at 90 °C before (middle row) or after (bottom row) immunolabeling of H3K4me3 (K4me3, red) and H3K27me3 (K27me3, blue). DNA FISH causes a dramatic loss of fluorescent signal for both histone marks compared to immunolabeled cells with no DNA FISH treatment (top row). For the middle and bottom rows, the lower right corner of the histone mark channel images are 20× contrast adjusted compared to the upper left corner of each image and the corresponding top row histone mark image.



Supplementary Figure 2.8 **Image processing schematic for SCEPTRE.**
Raw images obtained from the immunofluorescence (IF) of protein structures are segmented with the following steps: smooth with a gaussian filter, then contrast adjust with an adaptively determined threshold per cell; binarize either by an Otsu method, or by a Laplace filter followed by selection of all negative values; apply a nuclear mask; apply a watershed transformation. After the segmentation of the nuclear channel and the immunofluorescence channels, the FISH raw channel is then segmented in the same manner with the following exceptions: a nuclear mask is applied after smoothing and before contrast adjustment, and no watershed transformation is applied. Features, including mean fluorescence intensity and fraction of overlap with segmented clusters from each immunofluorescence channel are identified for all segmented clusters within a channel. FISH clusters are further filtered by size. The nuclear mask is generated with the following additional segmentation steps: smooth and contrast adjust either a Hoechst stain channel or one of the present immunofluorescence channels; dilate image to fuse clusters within the nucleus; select largest region encompassing the nucleus; determine the convex hull; erode segmented nucleus; apply to immunofluorescence and FISH channels (for more details, see Materials and Methods and **sup. table 2**).

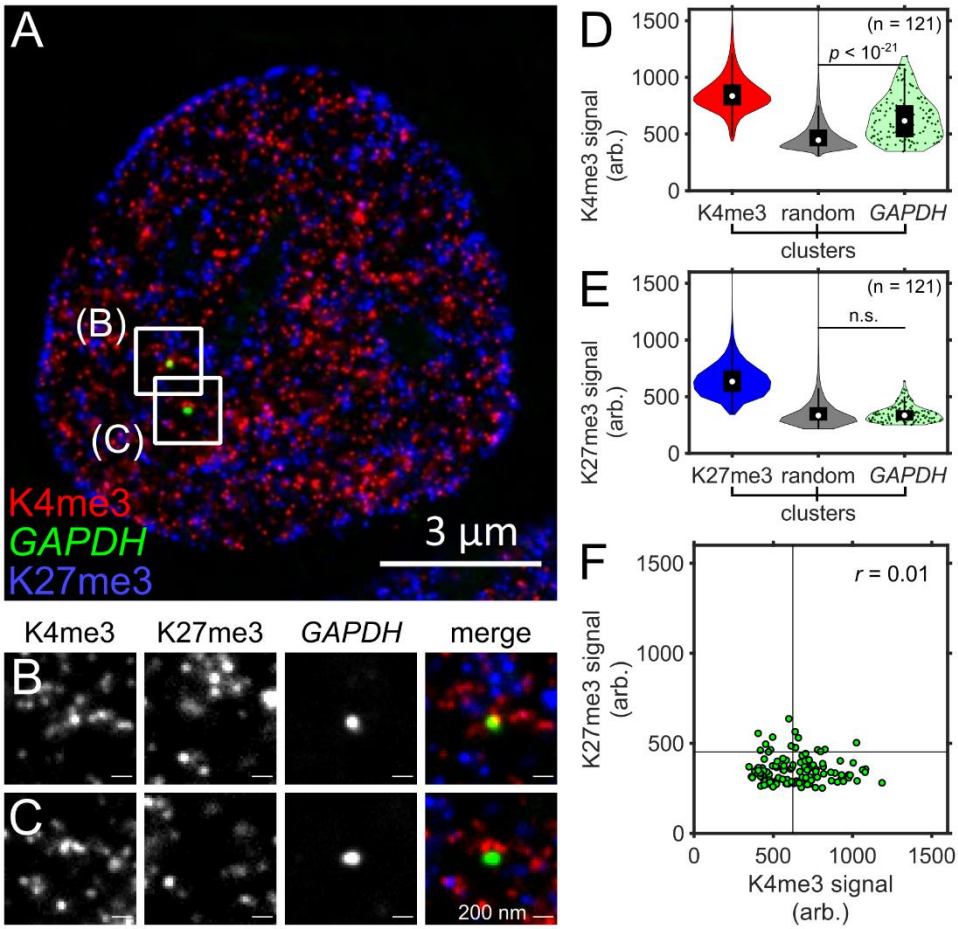


Supplementary Figure 2.9 **Correlative imaging of anti-centromere stain before and after expansion.** Anti-centromere imaged post-expansion (post, green), is aligned by similarity transform to the same stain imaged pre-expansion (pre, red) and visualized in the context of the post-expansion nucleus labeled by Hoechst (blue). All scale bars are in pre-expansion units.

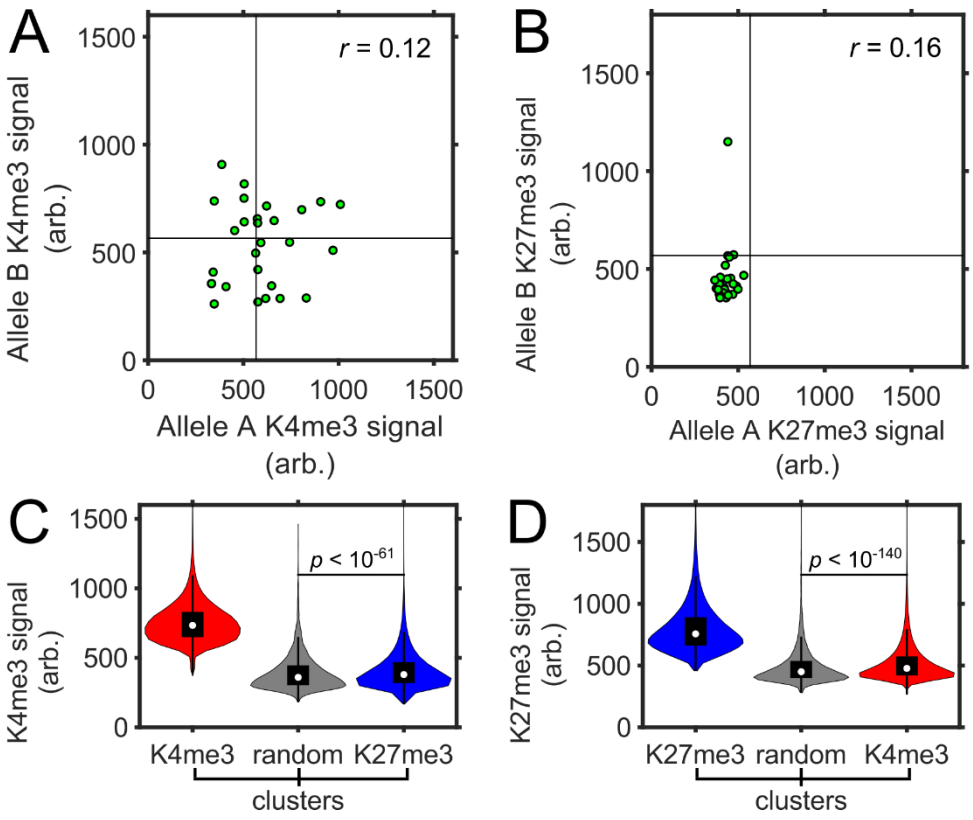


Supplementary Figure 2.10 **SCEPTRE measures signal of single-immunolabeled histone marks at *GAPDH* in RPE1 cells.**

(A) Distributions of H3K4me3 (K4me3) fluorescence signal (arb. = arbitrary units) within H3K4me3, randomly selected regions (random) and *GAPDH* clusters from single-immunolabeled expanded RPE1 cells. Cluster numbers are K4me3 = 196194, random = 5744, *GAPDH* = 90. **(B)** Distribution of H3K27me3 (K27me3) fluorescence signal within H3K27me3, randomly selected regions and *GAPDH* clusters from single-immunolabeled expanded RPE1 cells. Cluster numbers are K27me3 = 60235, random = 6504, *GAPDH* = 24. Significance determined by a right-tailed Wilcoxon rank-sum test of histone mark fluorescence signal in *GAPDH* against random cluster distributions.

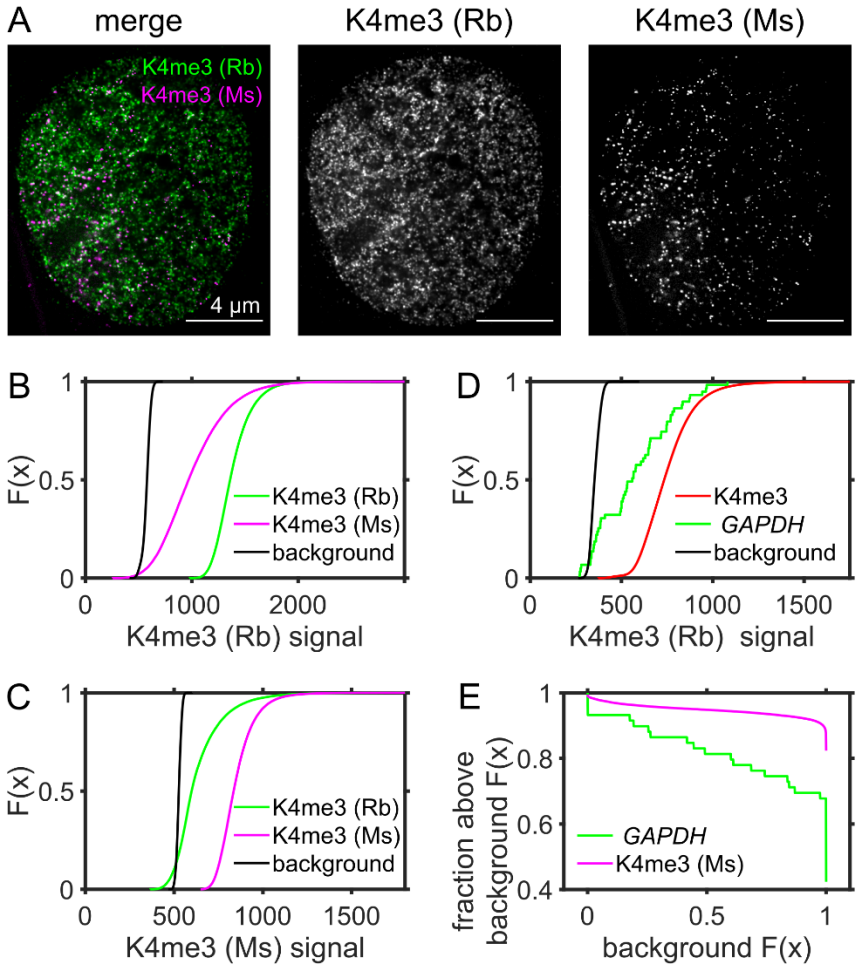


Supplementary Figure 2.11 **SCEPTRE shows reproducible results with a different set of antibodies.** **(A)** An expanded RPE1 cell with immunolabeled H3K4me3 marks (K4me3, red) and H3K27me3 marks (K27me3, blue), and FISH-labeled *GAPDH* (green), using an alternative set of antibodies to **figure 2.4**. **(B-C)** Zoomed in views of the approximate center plane of an image stack for each *GAPDH* allele in the cell seen in **A.** **(D)** Distributions of H3K4me3 fluorescence signal (arb. = arbitrary units) within H3K4me3, randomly selected regions (random) and *GAPDH* clusters. **(E)** Distribution of H3K27me3 fluorescence signal within H3K27me3, randomly selected regions and *GAPDH* clusters. **(F)** H3K27me3 and H3K4me3 fluorescence signals within *GAPDH* clusters (green). Black lines represent the threshold "on" level for each fluorescence signal. Cluster numbers for **D.** and **E.** are K4me3 = 250644, K27me3 = 262307, random = 7406, *GAPDH* = 121. The correlation coefficient (*r*) between fluorescence signals within *GAPDH* is shown in the top-right corner of the plot. Significance determined by a right-tailed Wilcoxon rank-sum test of histone mark fluorescence signals in *GAPDH* against random cluster distributions. All scale bars are in pre-expansion units. Although the antibodies used in **figure 2.4** and for targeting H3K27me3 in **A.** were validated by ENCODE, the H3K4me3 targeting antibody in **A.** was not.



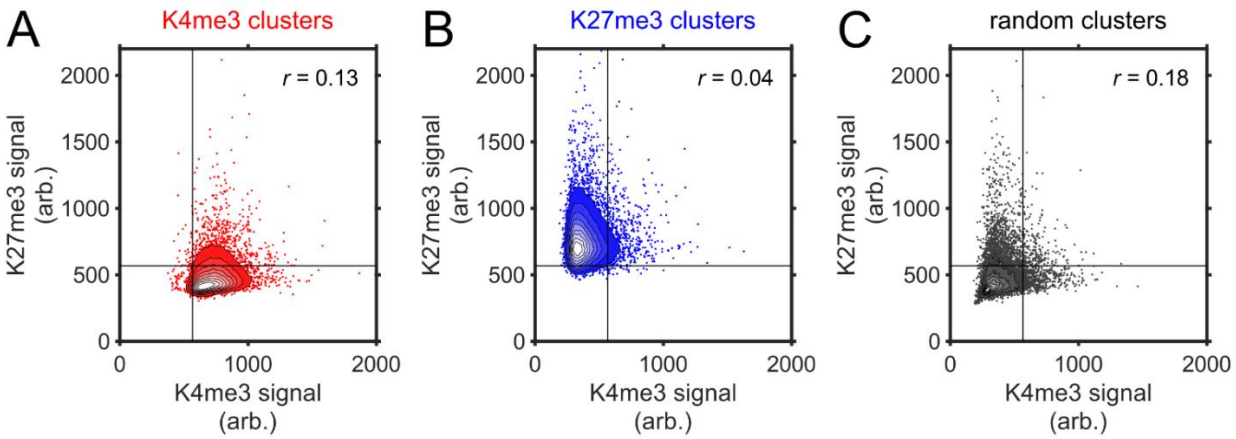
Supplementary Figure 2.12 **SCEPTRE compares H3K4me3 and H3K27me3 signals between different *GAPDH* alleles in the same cell, or between histone mark cluster distributions.**

(A-B) Fluorescence signal (arb. = arbitrary units) of either H3K4me3 (K4me3) in **A.**, or H3K27me3 (K27me3) in **B.**, in *GAPDH* alleles within the same cell from the data set in **figure 2.4** (one locus from each cell containing 2-4 loci is randomly assigned as allele A, and a second locus as allele B). Black lines represent the threshold "on" level for each histone mark fluorescence signal. The correlation coefficient (*r*) is shown on the top-right corner of each plot. **(C-D)** Fluorescence signal of either H3K4me3 in **C.**, or H3K27me3 in **D.**, for each distribution of H3K4me3 (red), H3K27me3 (blue) and randomly selected region (random, gray) clusters within the cells in **figure 2.4**. Significance determined by a right-tailed Wilcoxon rank-sum test of fluorescence signals in each histone mark cluster set against random cluster distributions.



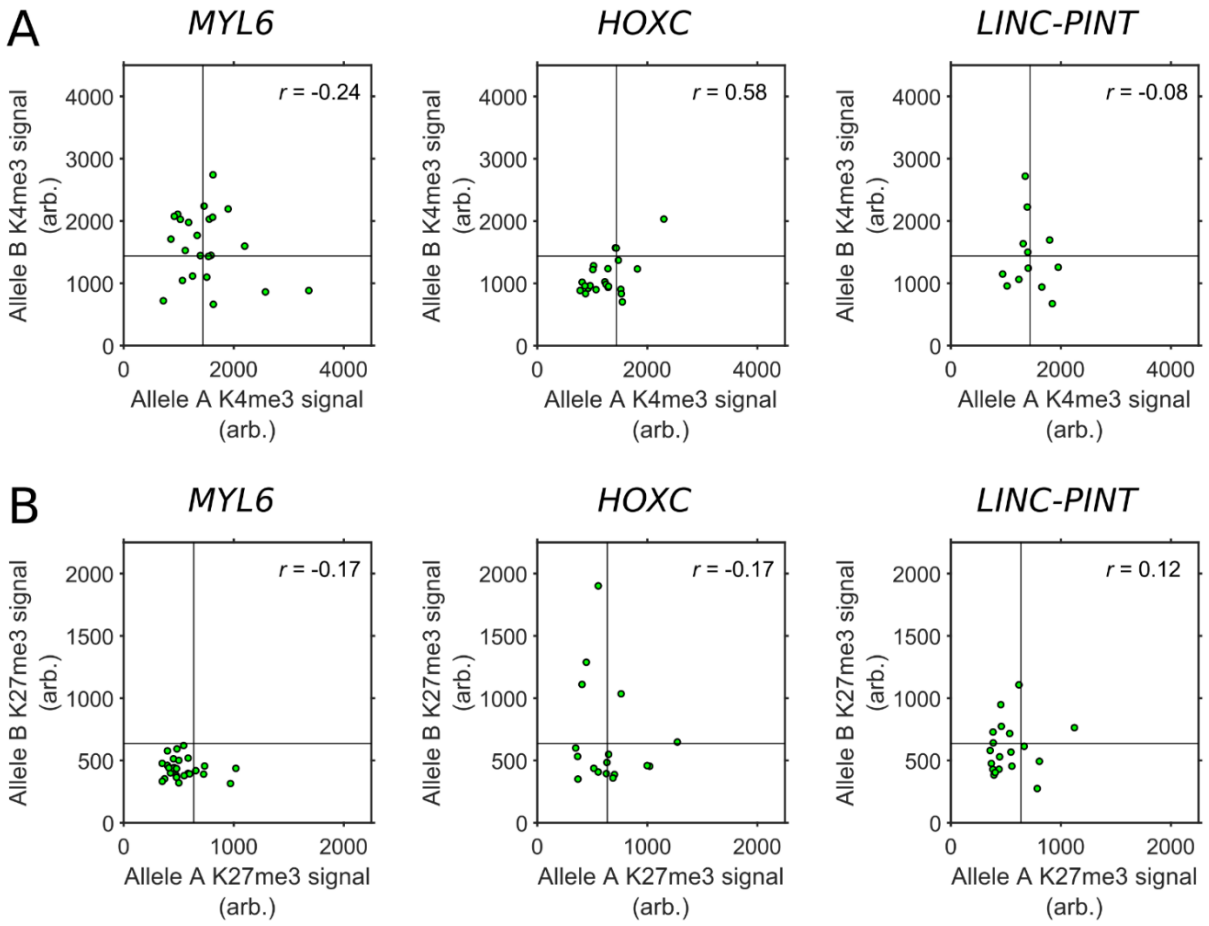
Supplementary Figure 2.13 **Dual labeling of H3K4me3 marks with two different antibodies provides information on the detection efficiency of each antibody.**

(A) An expanded RPE1 cell with H3K4me3 (K4me3) marks immunolabeled simultaneously by two different antibodies: Rb×H3K4me3 (Rb, green) and Ms×H3K4me3 (Ms, magenta). **(B-C)** Empirical cumulative distribution functions (F(x)) for the fluorescence signal of Rb×H3K4me3 in **B.**, Ms×H3K4me3 in **C.**, for Rb×H3K4me3 clusters (green), Ms×H3K4me3 clusters (magenta), or background (black) from the same group of cells as the one in **A. (D)** Empirical cumulative distribution functions (F(x)) for the fluorescence signal of Rb×H3K4me3 in *GAPDH* clusters (green), Rb×H3K4me3 clusters (red), or background (black) from the cells in **figure 2.4. (E)** The fraction of either *GAPDH* clusters (green) from **figure 2.4,** or the Ms×H3K4me3 clusters (magenta) from the same group of cells as the one in **A.**, that are above the Rb×H3K4me3 fluorescence signal of background F(x).



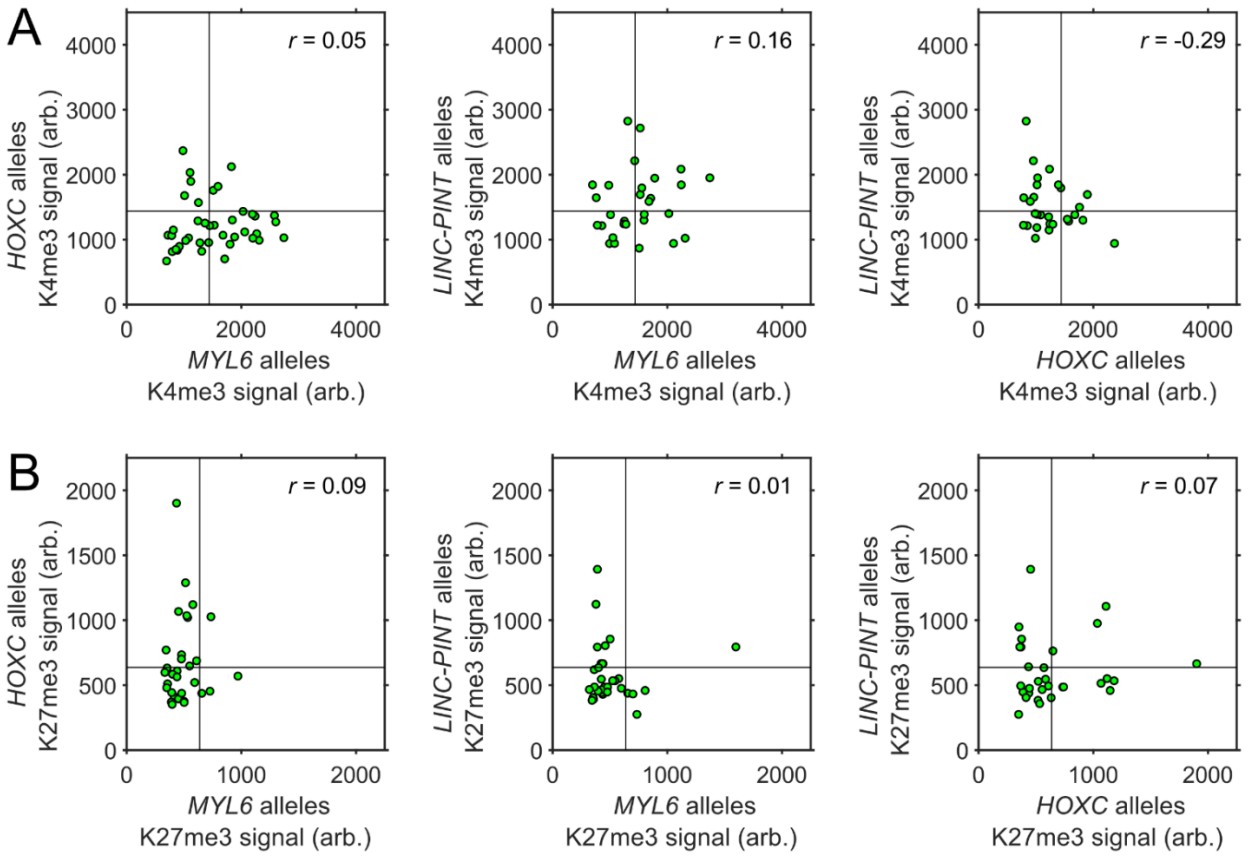
Supplementary Figure 2.14 **SCEPTRE compares H3K4me3 and H3K27me3 signals within segmented immunostained and random clusters.**

(A-C) Contours for the fluorescence signal (arb. = arbitrary units) frequency of H3K4me3 (K4me3) and H3K27me3 (K27me3) in the cluster sets of H3K4me3 (red) in **A.**, H3K27me3 (blue) in **B.**, and randomly selected regions (random, gray) in **C.**. Straight black lines represent the threshold "on" level for each fluorescence signal. Contours have uniformly spaced steps ranging from 0.1 to 0.9 frequency and represent all clusters obtained for cells in **figure 2.4**. The remaining scatter in **A.** and **B.** is a 100-fold downsample of the original data by random selection for plot representation purposes. Correlation coefficients (*r*) for each data set, which are calculated before downsampling, are shown in the top-right corner of each plot.



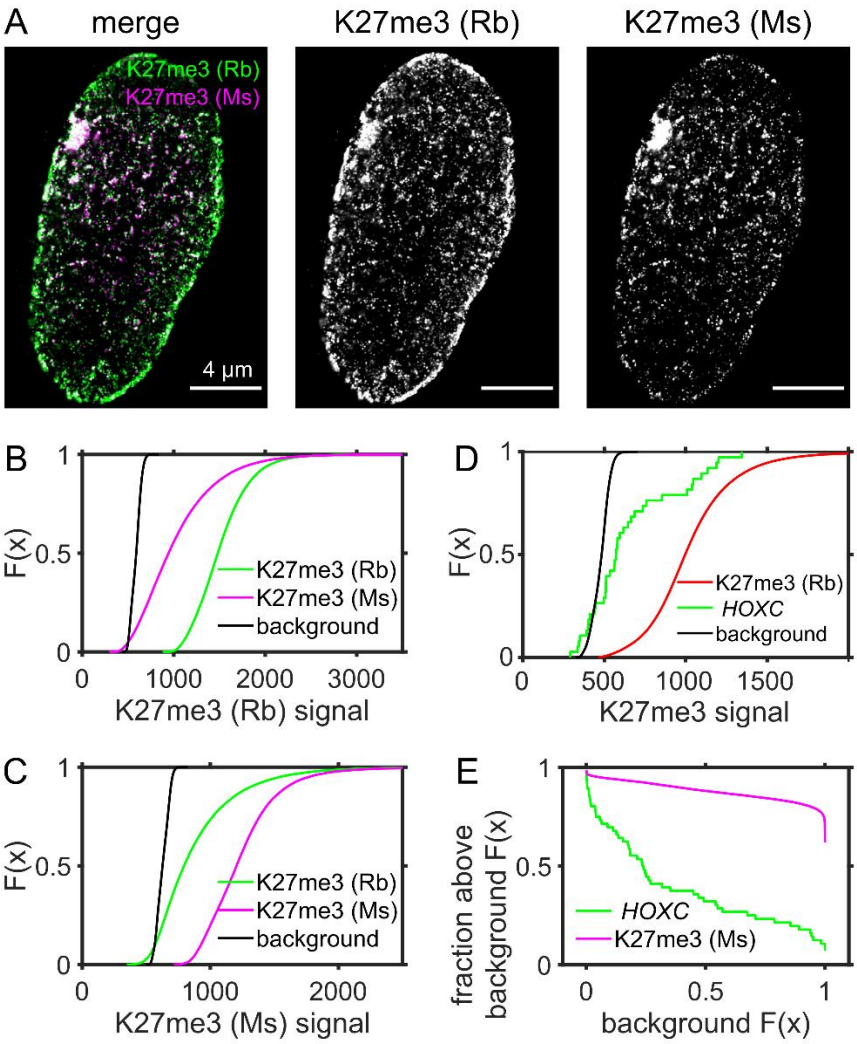
Supplementary Figure 2.15 **SCEPTRE compares H3K4me3 or H3K27me3 signals between alleles of one of multiple genes in the same cell.**

(A-B) Fluorescence signal of either H3K4me3 (K4me3) in **A.**, or H3K27me3 (K27me3) in **B.**, in *MYL6*, *HOXC* or *LINC-PINT* alleles from the same cell (one locus from each cell containing 2-4 loci is randomly assigned as allele A, and another one as allele B). Black lines represent the threshold "on" level for each histone mark fluorescence signal. The correlation coefficient (*r*) for each set is shown in the top-right corner of each plot.



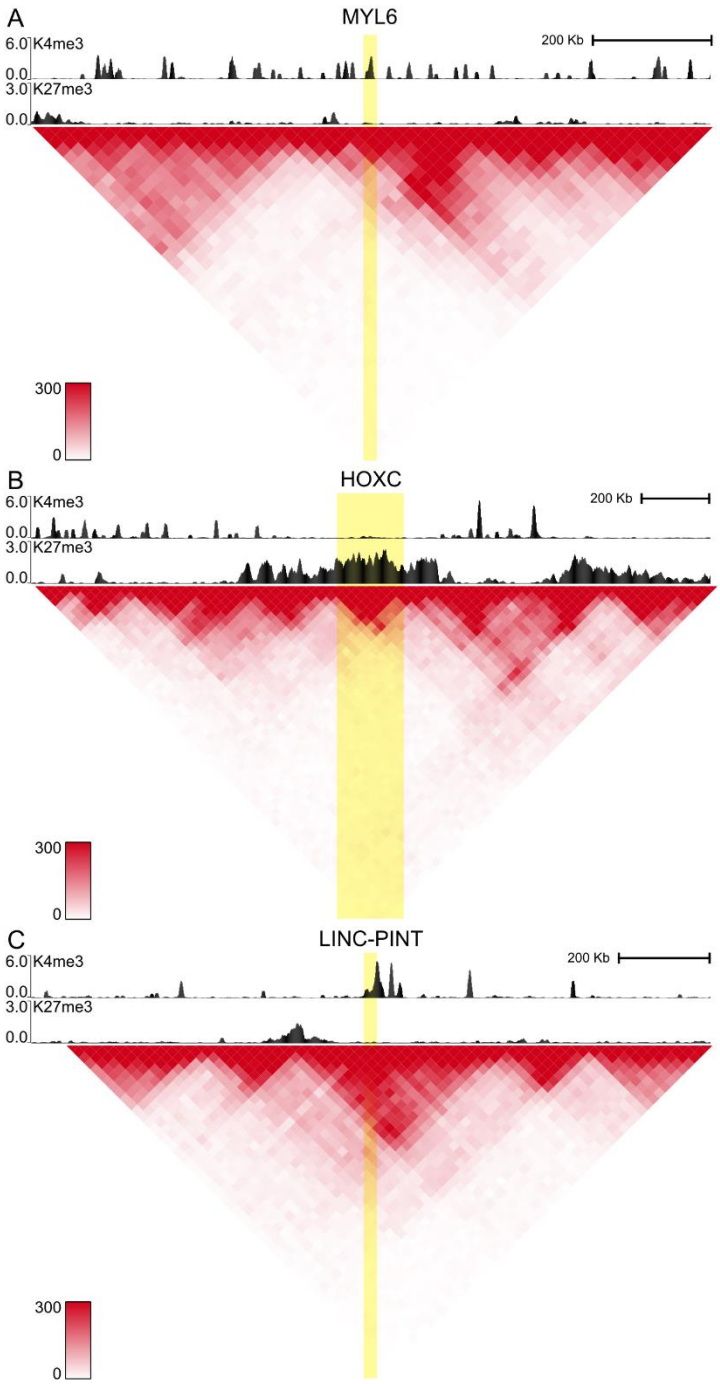
Supplementary Figure 2.16 **SCEPTRE compares H3K4me3 or H3K27me3 signals between alleles from different genes in the same cell.**

(A-B) Comparison of the fluorescence signal (arb. = arbitrary units) of either H3K4me3 (K4me3) in **A.,** or H3K27me3 (K27me3) in **B.**, between randomly selected alleles of *MYL6*, *HOXC* and/or *LINC-PINT* within the same cell. Black lines represent the threshold "on" level for each histone mark fluorescence signal. The correlation coefficient (*r*) for each set is shown in the top-right corner of each plot.

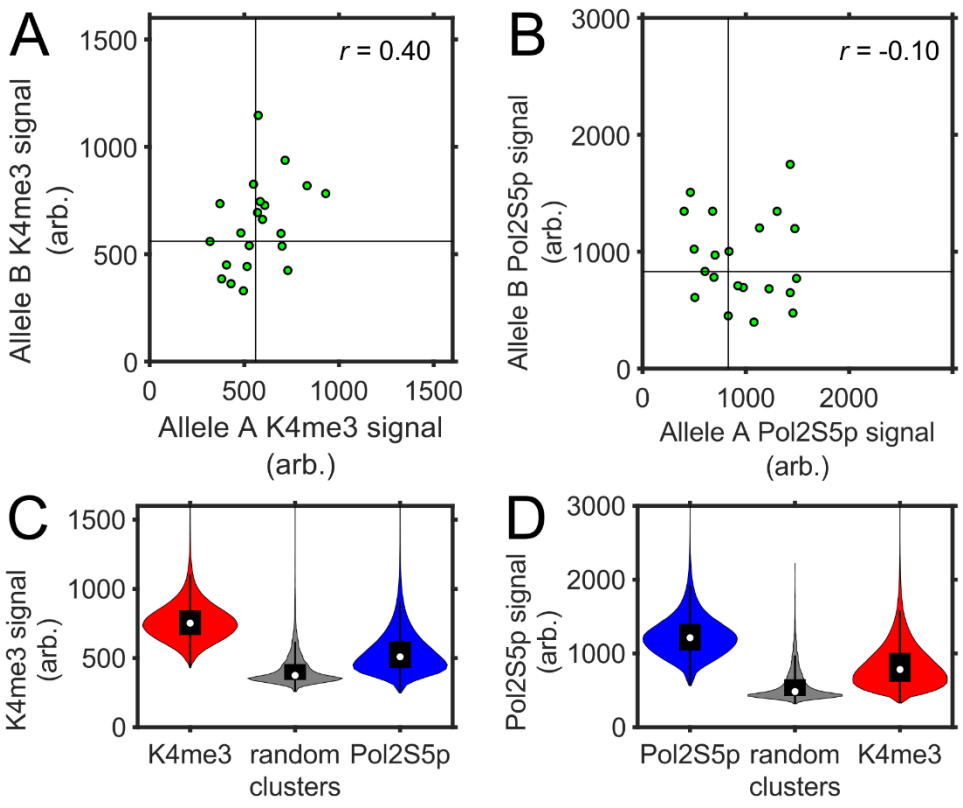


Supplementary Figure 2.17 **Dual labeling of H3K27me3 marks with two different antibodies provides information on the detection efficiency of each antibody.**

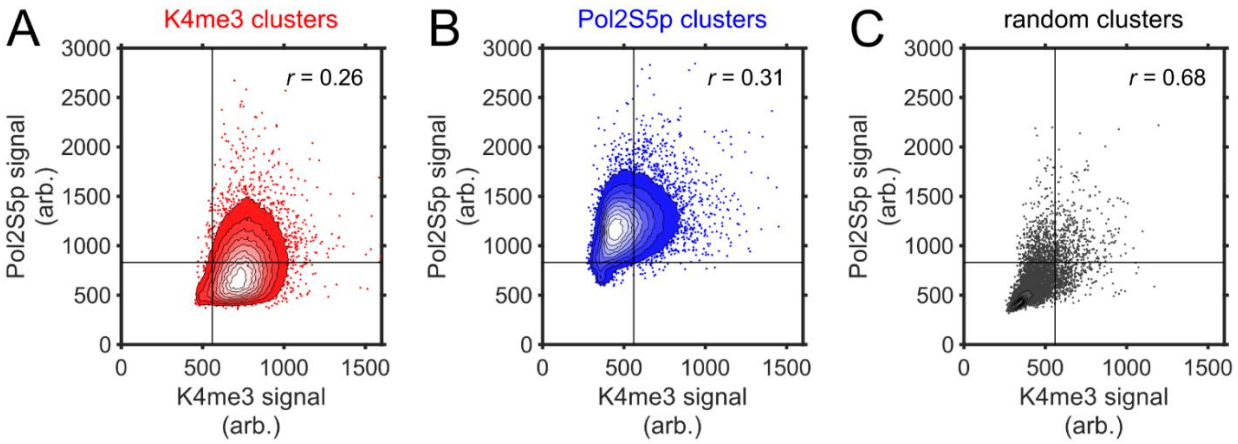
(A) An expanded RPE1 cell with H3K27me3 (K27me3) immunolabeled simultaneously by two different antibodies: Rb×H3K27me3 (Rb, green) and Ms×H3K27me3 (Ms, magenta). **(B-C)** Empirical cumulative distribution functions (F(x)) for the fluorescence signal of Rb×H3K27me3 in **B.**, or Ms×H3K27me3 in **C.**, for Rb×H3K27me3 clusters (green), Ms×H3K27me3 clusters (magenta), or background (black) from the same group of cells as the one in **A. (D)** Empirical cumulative distribution functions (F(x)) for the fluorescence signal of Rb×K27me3 in *HOXC* clusters (green), Rb×K27me3 clusters (red), or background (black) from the RPE expanded cells in **figure 2.5C. (E)** The fraction of either *HOXC* clusters (green) from **figure 2.5C,** or the Ms×H3K27me3 clusters (magenta) from the same group of cells as the one in **A.**, that are above the Rb×H3K27me3 fluorescence signal of background F(x).



Supplementary Figure 2.18 **Analysis of Hi-C for targeted genomic regions in RPE 1 cells.**
Hi-C data, previously published,[127] along with H3K4me3 (K4me3) and H3K27me3 (K27me3) CUT&RUN normalized counts for *MYL6* **(A)**, *HOXC* **(B)** and *LINC-PINT* **(C)** targeted regions (highlighted). Heat map score between 0 - 300 reads in 25 kb bins.

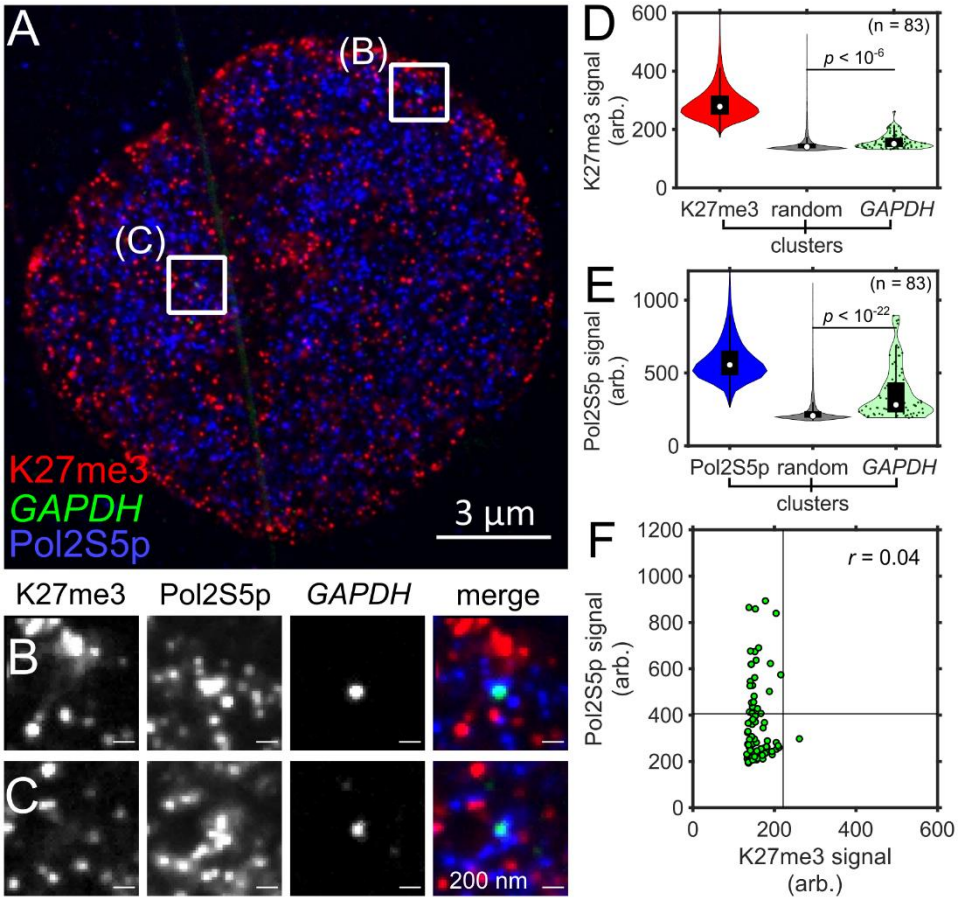


Supplementary Figure 2.19 **SCEPTRE compares H3K4me3 and paused RNA polymerase II signals between different *GAPDH* alleles in the same cell, or between immunolabeled cluster distributions.** **(A-B)** Fluorescence signal (arb. = arbitrary units) of either H3K4me3 (K4me3) in **A.**, or paused RNA polymerase II (Pol2S5p) in **B.**, in *GAPDH* alleles within the same cell from the data set in **figure 2.6** (one locus from each cell containing 2-4 loci is randomly assigned as allele A, and a second locus as allele B). Black lines represent the threshold "on" level for each fluorescence signal. The correlation coefficient (*r*) is shown on the top-right corner of each plot. **(C-D)** Fluorescence signal (arb. = arbitrary units) for either H3K4me3 in **C.**, or paused RNA polymerase II in **D.**, for each distribution of H3K4me3 (red), paused RNA polymerase II (blue) and randomly selected regions (random, gray) clusters within the cells in **figure 2.6**.



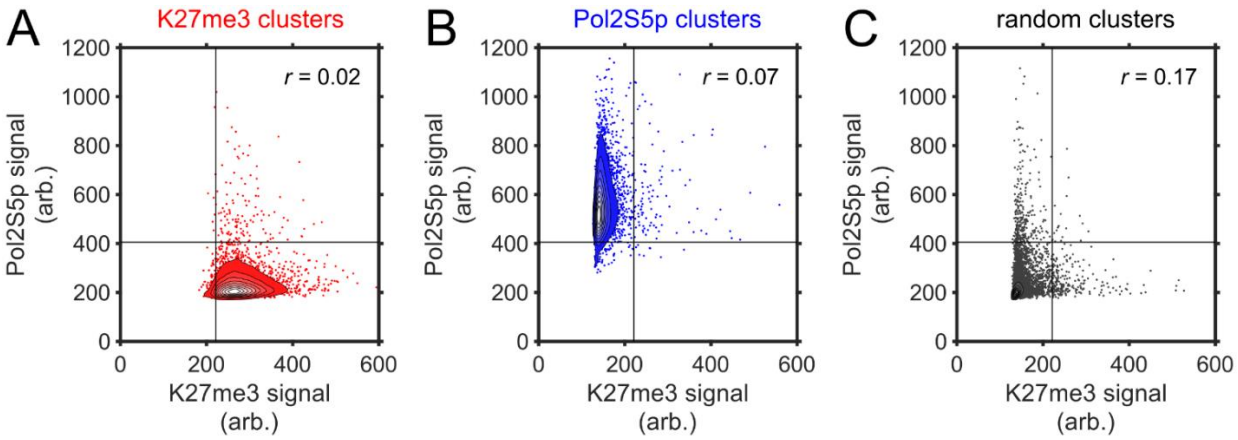
Supplementary Figure 2.20 **SCEPTRE compares H3K4me3 and paused RNA polymerase II signals within segmented immunostained and random clusters.**

(A-C) Contours for the fluorescence signal (arb. = arbitrary units) frequency of H3K4me3 (K4me3) and paused RNA polymerase II (Pol2S5p) in the cluster sets of H3K4me3 (red) in **A.**, paused RNA polymerase II (blue) in **B.**, and randomly selected regions (random, gray) in **C.** Straight black lines represent the threshold "on" level for each fluorescence signal. Contours have uniformly spaced steps ranging from 0.1 to 0.9 frequency and represent all clusters obtained for cells in **figure 2.6**. The remaining scatter in **A.** and **B.** is a 100-fold downsample of the original data by random selection for plot representation purposes. Correlation coefficients (*r*) for each data set, which are calculated before downsampling, are shown in the top-right corner of each plot.



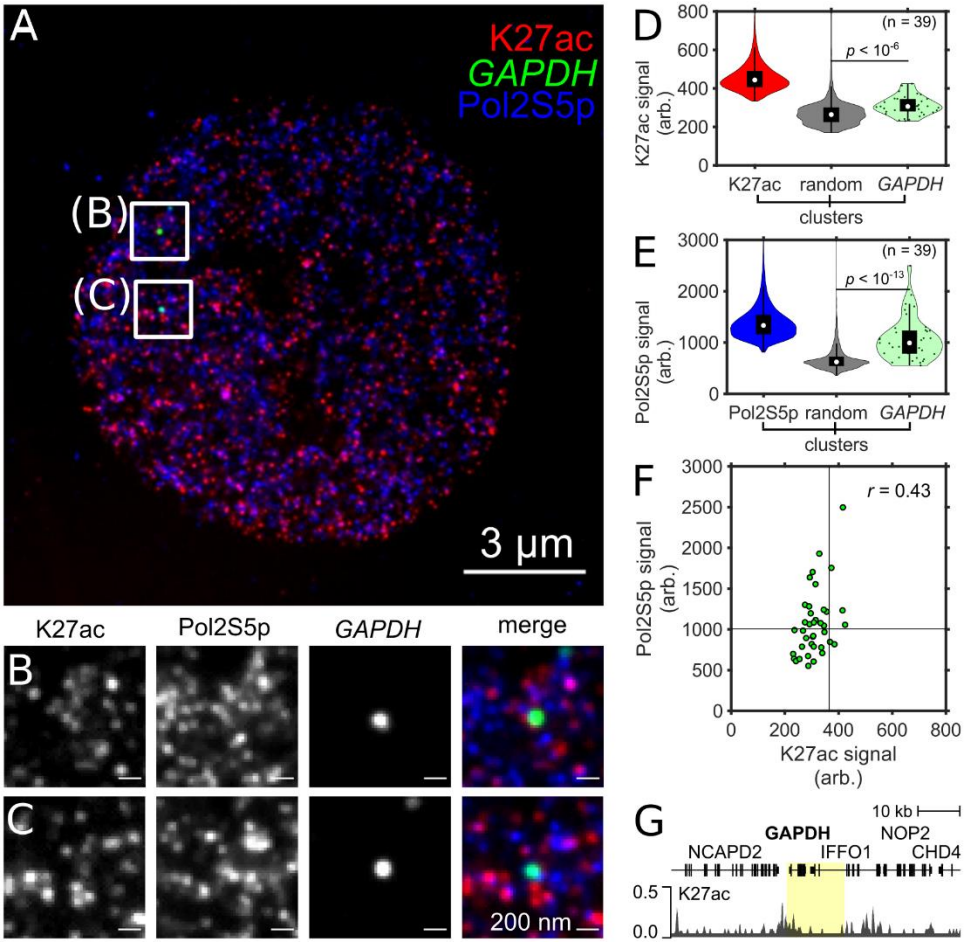
Supplementary Figure 2.21 **SCEPTRE distinguishes between H3K27me3 and paused RNA polymerase II signals at a single genomic region.**

(A) An expanded RPE1 cell with immunolabeled H3K27me3 (K27me3, red) and paused RNA polymerase II (Pol2S5p, blue), and FISH-labeled *GAPDH* (green). **(B-C)** Zoomed in views of the approximate center plane of an image stack for each *GAPDH* allele in the cell seen in **A.** **(D)** Distributions of H3K27me3 fluorescence signal (arb. = arbitrary units) within H3K27me3, randomly selected regions (random) and *GAPDH* clusters. **(E)** Distribution of paused RNA polymerase II fluorescence signal within paused RNA polymerase II, randomly selected regions and *GAPDH* clusters. **(F)** H3K27me3 and paused RNA polymerase II fluorescence signals within *GAPDH* clusters (green). Black lines represent the threshold "on" level for each fluorescence signal. Cluster numbers for **D.** and **E.** are K27me3 = 174072, Pol2S5p = 213724, random = 6099, *GAPDH* = 83. Significance determined by a right-tailed Wilcoxon rank-sum test of fluorescence signals in *GAPDH* against random cluster distributions. All scale bars are in pre-expansion units.



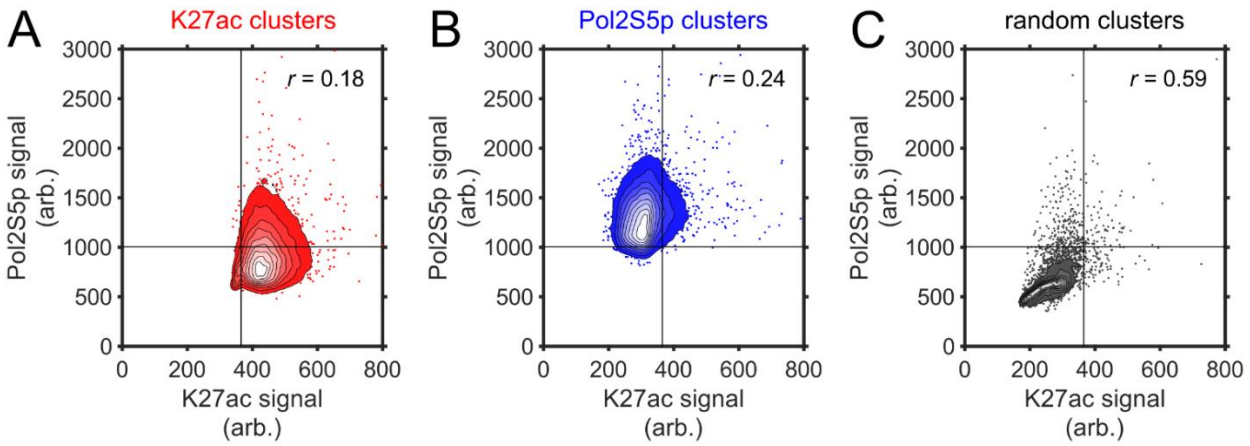
Supplementary Figure 2.22 **SCEPTRE compares H3K27me3 and paused RNA polymerase II signals within segmented immunostained and random clusters.**

(A-C) Contours for the fluorescence signal (arb. = arbitrary units) frequency of H3K27me3 (K27me3) and paused RNA polymerase II (Pol2S5p) in the cluster sets of H3K27me3 (red) in **A.**, paused RNA polymerase II (blue) in **B.**, and randomly selected regions (random, gray) in **C.** Straight black lines represent the threshold "on" level for each fluorescence signal. Contours have uniformly spaced steps ranging from 0.1 to 0.9 frequency and represent all clusters obtained for cells in **supplementary figure 2.18**. The remaining scatter in **A.** and **B.** is a 100-fold downsample of the original data by random selection for plot representation purposes. Correlation coefficients (*r*) for each data set, which are calculated before downsampling, are shown in the top-right corner of each plot.



Supplementary Figure 2.23 **SCEPTRE compares H3K27ac and paused RNA polymerase II signals at a single genomic region.**

(A) An expanded RPE1 cell with immunolabeled H3K27ac (K27ac, red) and paused RNA polymerase II (Pol2S5p, blue), and FISH-labeled *GAPDH* (green). **(B-C)** Zoomed in views of the approximate center plane of an image stack for each *GAPDH* allele in the cell seen in **A. (D)** Distributions of H3K27ac fluorescence signal (arb. = arbitrary units) within H3K27ac, randomly selected regions (random) and *GAPDH* clusters. **(E)** Distribution of paused RNA polymerase II fluorescence signal within paused RNA polymerase II, randomly selected regions and *GAPDH* clusters. **(F)** H3K27ac and paused RNA polymerase II fluorescence signals within *GAPDH* clusters (green). Black lines represent the threshold "on" level for each fluorescence signal. (**G**) CUT&RUN normalized counts for H3K27ac marks in RPE1 cells for the FISH targeted *GAPDH* region (highlighted). Cluster numbers for **D.** and **E.** are K27ac = 82644, Pol2S5p = 153482, random = 3815, *GAPDH* = 39. Significance determined by a right-tailed Wilcoxon rank-sum test of fluorescence signals in *GAPDH* against random cluster distributions. All scale bars are in pre-expansion units.



Supplementary Figure 2.24 **SCEPTRE compares H3K27ac and paused RNA polymerase II signals within segmented immunostained and random clusters.**

(A-C) Contours for the fluorescence signal (arb. = arbitrary units) frequency of H3K27ac (K27ac) and paused RNA polymerase II (Pol2S5p) in the cluster sets of H3K27ac (red) in **A.**, paused RNA polymerase II (blue) in **B.**, and randomly selected regions (random, gray) in **C.** Straight black lines represent the threshold "on" level for each fluorescence signal. Contours have uniformly spaced steps ranging from 0.1 to 0.9 frequency and represent all clusters obtained for cells in **supplementary figure 2.20**. The remaining scatter in **A.** and **B.** is a 100-fold downsample of the original data by random selection for plot representation purposes. Correlation coefficients (*r*) for each data set, which are calculated before downsampling, are shown in the top-right corner of each plot.

Supplementary Table 2.1 **Summary of sample preparation and imaging conditions.**

Figure	Fixation	Primary ab(s).	Secondary ab(s).	Post-fixation	DNA FISH conditions	other stain:	Imaging
2.3	10% PFA 10 min, RT	Hu × cen. (5 µg/mL)	D × Hu, AT488 (~13 dp, 2 µg/mL)	none	Protocol: 2-step* 200 nM alphasat probe 200 nM alphasat adapter 600 nM AF647 reporter A Denaturation: 95 °C, 15 min Hybridization: 37 °C, 30 min	Hoechst (1 µg/mL) 10 min	Microscope: LSC Image thickness: 4.3 µm buffer: 0.2× SSC filter: 1
2.4	EtOH:MeOH 6 min, -20 °C	Rb × H3K4me3 (2 µg/mL) Ms × H3K27me3 (2 µg/mL)	D × Rb, AF568 (~3.5 dp, 2 µg/mL) D × Ms, AF488 (~2.5 dp, 3 µg/mL)	4% PFA 10 min, RT	Protocol: Single-step** 200nM *GAPDH* set 250nM AT647N reporter B Denaturation: 90 °C, 2.5 min Hybridization: 42 °C, ON	None	Microscope: SDC Image thickness: 211 nm buffer: water filter: none
2,5A	4% PFA 10 min, RT	Rb × H3K4me3 (2 µg/mL)	D × Rb, AF488 (~2.5 dp, 3 µg/mL)	none	Protocol: Single-step** ~4 µM oligo pool*** 240 nM *MYL6* Adapter 250 nM AF750 reporter C 250 nM *LINC-PINT* adapter 250 nM AF647 reporter D 1.25 µM *HOXC* adapter 1.25 µM AT565 reporter E Denaturation: 90 °C, 2.5 min Hybridization: 42 °C, ON	None	Microscope: SDC Image thickness: 261 nm buffer: ALOX**** filter: none
2.5C	Same as Fig. 2.5A	Rb × H3K27me3 (2 µg/mL)	Same as Fig. 4A	none	same as fig. 4A	None	Same as Fig. 4A
2,6	Same as Fig. 2.4	Rb × H3K4me3 (2 µg/mL) Ms × Pol2S5p (2 µg/mL)	D × Rb, AF568 (~2.7 dp, 3 µg/mL) D × Ms, AF488 (~2.5 dp, 3 µg/mL)	Same as Fig. 3	Protocol: Single-step** 100 nM *GAPDH* set 100 nM AT647N reporter B Denaturation: 90 °C, 2.5 min Hybridization: 42 °C, ON	None	Same as Fig. 3
2.7	Same as Fig. 2.4	Same as Fig. 3	D × Rb, AF568 (~3.3 dp, 2 µg/mL) D × Ms, AF488 (~2.5 dp, 2 µg/mL)	none	none	None	Microscope: Widefield Image thickness: 1 µm buffer: 0.2× SSC filter: 1
2.8	Same as Fig. 2.3	Same as Fig. 2	Same as Fig. 2	none	none	Same as Fig. 2	Microscope: LSC Image thickness: 5.3 µm buffer: water filter: 1
2.10A	Same as Fig. 2.5A	Same as Fig. 4A	Same as Fig. 4A	none	Protocol: Single-step** 200 nM *GAPDH* set 300 nM AT565 reporter B Denaturation: 92.5 °C, 10 min Hybridization: 37 °C, ON	none	Microscope: LSC Image thickness: 225 nm buffer: water filter: none

2.10B	Same as Fig. 2.5A	Same as Fig. 4B	Same as Fig. 4A	none	Same as Sup. Fig. 3A	none	Microscope: SDC Image thickness: 206 nm buffer: water filter: none
2.11	Same as Fig. 2.4	Ms × H3K4me3 (2 µg/mL) Rb × H3K27me3 (2 µg/mL)	Same as Fig. 3	Same as Fig. 3	Same as Fig. 3	none	Same as Fig. 3
2.12	Same as Fig. 2.4	Rb × H3K4me3 (2 µg/mL) Ms × H3K4me3 (2 µg/mL)	Same as Fig. S1	Same as Fig. 3	None	None	Same as Fig. S4B
2.13	Same as Fig. 2.4	Rb × H3K27me3 (2 µg/mL) Ms × H3K27me3 (2 µg/mL)	Same as Fig. S1	Same as Fig. 3	None	None	Same as Fig. S4B
2.19	Same as Fig. 2.4	Rb × H3K27me3 (2 µg/mL) Ms × Pol2S5p (2 µg/mL)	Same as Fig. 5	Same as Fig. 3	Protocol: Single-step** 100 nM *GAPDH* set 120 nM AT647N reporter B Denaturation: 90 °C, 2.5 min Hybridization: 42 °C, ON	none	Same as Fig. 3
2.21	Same as Fig. 2.4	Rb × H3K27ac (2 µg/mL) Ms × Pol2S5p (2 µg/mL)	Same as Fig. 5	Same as Fig. 3	Same as Sup. Fig. 12	Same as Fig. 2	Same as Fig. 3

* 2-step protocol: alpha-satellite probe is hybridized first after denaturation, and after the sample is washed, adapter and reporter probes are hybridized in a second step.

** single-step protocol: All probes are hybridized together in one step.

*** 4 µM oligo pool is assumed to contain ~180 nM *MYL6* probe set, ~244 nM *LINC-PINT* probe set and ~1.2 µM *HOXC* probe set.

**** ALOX buffer contains: 10 mM Tris buffer (pH 8) with 1 mM Methyl viologen, 1 mM Ascorbic acid, 2% (v/v) MeOH, ~30 units/mL alcohol oxidase and 0.2% (w/v) catalase.

Additional notes: Thickness in the imaging column refers to the thickness of the displayed data in terms of the pixel sizes set in pre-expansion units. Filter refers to the number of pixels used in applying a 3D median filter to the image data set before being displayed in the figure.

Acronyms: PFA=Paraformaldehyde; RT=room temperature (~22 °C); ab=antibody; Hu=Human; Rb=Rabbit; Ms=Mouse; D=Donkey; dp=dyes per protein; ON=Overnight (~18 hours); AF=Alexa Fluor; AT=ATTO-TEC; alphasat=alpha-satellite; LSC=Laser Scanning Confocal; SDC = Spinning Disk Confocal.

Supplementary Table 2.2 **Summary of image processing and analysis conditions.**

Figure	nuclear mask channel	gaussian smooth (SD)	contrast adjustment threshold	Binarization method	FISH size filter (voxels)
2.3	Hoechst	2	nuclear: 1 α-centromere: 2 anti-centromere: 5	Otsu	α-cen. size: 20 – 10000
2.4	H3K27me3	1	nuclear: 3 *GAPDH*: 10 H3K4me3: 3 H3K27me3: 2	Laplace	*GAPDH*: ≥20
2.5A	H3K4me3	1	nuclear: 2.5 *MYL6*: 10 *HOXC*: 10 *LINC-PINT*: 10 H3K4me3: 2.5	Laplace	*MYL6*: ≥20 *HOXC*: ≥20 *LINC-PINT*: ≥20
2.5C	H3K27me3	1	nuclear: 3 *MYL6*: 10 *HOXC*: 10 *LINC-PINT*: 10 H3K27me3: 3	Laplace	*MYL6*: ≥20 *HOXC*: >50 *LINC-PINT*: >50
2.6	Pol2S5p	1	nuclear: 3 *GAPDH*: 15 H3K4me3: 5 Pol2S5p: 5	Laplace	*GAPDH*: >50
2.10A	H3K4me3	1	nuclear: 4 *GAPDH*: 15 H3K4me3: 3	Laplace	*GAPDH*: >80
2.10B	H3K27me3	1	nuclear: 3: *GAPDH*: 15 H3K27me3: 3	Laplace	*GAPDH*: >80
2.11	H3K27me3	1	Nuclear: 3 *GAPDH*: 10 H3K4me3: 3 H3K27me3: 3	Laplace	*GAPDH*: >50
2.13	H3K4me3 (Rb)	1	Nuclear: 2 H3K4me3 (Rb): 3 H3K4me3 (Ms): 3	Laplace	none
2.17	H3K27me3 (Rb)	1	Nuclear: 2 H3K27me3 (Rb): 3 H3K27me3 (Ms): 3	Laplace	None
2.20	H3K27me3	1	Nuclear: 3 *GAPDH*: 15 H3K27me3: 9 Pol2S5p: 9	Laplace	*GAPDH*: >75
2.22	Hoechst	1	Nuclear: 1 *GAPDH*: 10 H3K27ac: 2 Pol2S5p: 4	Laplace	*GAPDH*: ≥20

Acronym: SD=standard deviation

Additional notes: Contrast adjustment threshold represents the number of third quartiles above the median of an image stack histogram used to establish the threshold for clipping during contrast adjustment (see Materials and Methods for more details).

Supplementary Table 2.3 **Transcription levels for genes found within DNA FISH-labeled genes profiled by SCEPTRE.**

DNA FISH-labeled region	gene within labeled region	Transcription level (FPKM)*
GAPDH	*GAPDH*	1618.65
	IFFO1	3.00
MYL6	*MYL6*	928.92
	*MYL6*B	19.73
	SMARCC2	10.09
HOXC	*HOXC*4	0.00
	*HOXC*5	0.01
	*HOXC*6	0.00
	*HOXC*8	0.00
	*HOXC*9	0.00
	*HOXC*10	0.00
	*HOXC*11	0.00
	*HOXC*12	0.00
	*HOXC*13	0.00
	HOTAIR	0.00
	HOXC-AS1	0.00
	HOXC-AS2	0.00
	HOXC-AS3	0.00
LINC-PINT	*LINC-PINT*	2.59

*Results obtained in previous RNA-seq study performed on RPE1 cells.[126]

CHAPTER 3: COMBINING FISH WITH NOVEL LABELING AND IMAGING STRATEGIES

3.1 PREFACE:

The ability of ExM to preserve labeled samples and structures allows for broader capabilities than colocalization of immunolabeled protein structures to genomic DNA. For this chapter, I would like to highlight my contributions to two different projects, where I explored other ways to use FISH with expansion microscopy. In the first project, I combined DNA FISH with a novel method, known as fluorescence labeling of abundant reactive entities (FLARE), which reveals structural features of a cell or tissue by conjugating fluorophores to abundantly available functional groups within each sample. In the second project, I helped design and implement a FISH library for detection of RNA transcripts in expanded cells to test the capabilities of a homebuilt spinning disk confocal in detecting single molecule structures within a sample.

At the time that I was developing SCEPTRE, other members of the Vaughan group were exploring a novel method for labeling functional groups that were abundantly present across a whole cell or tissue sample. For example, Lysine amines on proteins can be conjugated to fluorophores functionalized with a N-hydroxysuccinimidyl ester group, while carbohydrate hydroxyl groups can be oxidized into reactive aldehydes that will react with fluorophores presenting a hydrazide group.[142] . As with the same principles of H&E (hematoxylin and eosin)[143] stains, these non-specific labels can generate feature-rich stains that reveal critical details about a tissues structure and diseased state. This is even more so with ExM, which provides an improvement in resolution for better detection of relevant features. FLARE was developed and validated by the Vaughan group with this principle in mind.[144,145]

Dr. Min Yen Lee and Dr. Chenyi Mao, the primary developers of the method, wanted to make sure that FLARE, which revealed unique and identifiable structures within each sample, was still

compatible with traditional labeling techniques used in fluorescence microscopy. Although they showed FLARE worked with immunofluorescence, they wondered if it would work with FISH for genomic structural labeling. In Sections 3.2, I will highlight my contribution to the FLARE method in validating its compatibility with DNA FISH. In Section 3.3, I will provide an updated and more thorough protocol that was published later.

In addition to labeling genomic DNA, FISH can also label RNA molecules for counting the transcripts of a gene. Dr. Aaron Halpern developed a simple and affordable scheme to adapt an existing widefield microscope into a spinning disk confocal by introducing a custom-made glass spinning disk into the laser path of the microscope.[146] As part of validating the utility of the homebuilt spinning disk confocal, we wanted to see if the single transcripts of a gene could be imaged and identified with this microscope. In section 3.4, I will talk about my design and implementation of an RNA FISH library for imaging of transcripts with the homebuilt disk confocal in an expanded cell.

3.2 A FEATURE-RICH COVALENT STAINS FOR SUPER-RESOLUTION AND CLEARED-TISSUE FLUORESCENCE MICROSCOPY

3.2.1 ABSTRACT:

Fluorescence microscopy is a workhorse tool in biomedical imaging but often poses substantial challenges to practitioners in achieving bright or uniform labeling. In addition, while antibodies are effective specific labels, their reproducibility is often inconsistent, and they are difficult to use when staining thick specimens. We report the use of conventional, commercially available fluorescent dyes for rapid and intense covalent labeling of proteins and carbohydrates in super-resolution (expansion) microscopy and cleared tissue microscopy. This approach, which we refer to as Fluorescent Labeling of Abundant Reactive Entities (FLARE), produces simple and robust stains that are modern equivalents of classic small-molecule histology stains. It efficiently reveals a wealth of key landmarks in cells and tissues under different fixation or sample processing conditions and is compatible with immunolabeling of proteins and in situ hybridization labeling of nucleic acids.

3.2.2 INTRODUCTION:

Two of the most important developments in fluorescence microscopy over the past one to two decades are super-resolution microscopy, for imaging small features beneath the ~250-nm diffraction limit of visible light, and cleared tissue microscopy, for deep imaging of intact specimens.[147–151] Researchers are now able to routinely measure nanoscale molecule distributions, protein oligomerization, or protein-protein interactions and to determine the large-scale organization of biological specimens. Central to these efforts are a wide range of probes to fluorescently label the sample, with one of the most important classes being fluorescently labeled antibodies. While antibodies are powerful in their ability to specifically label specimens without need for genetic manipulation, they often suffer slow penetration in thick tissues, poor binding in specimens that have been heavily fixed or processed [e.g., formalin-fixed paraffin-embedded (FFPE) tissue], high cost, and inconsistent lot-to-lot reproducibility or commercial availability.[152,153]

Small-molecule histology stains such as H&E (hematoxylin and eosin)[143] are attractive potential alternative labels for super-resolution and cleared tissue microscopy, but they pose several challenges.[154] First, many histology stains are affinity based, rather than covalent, and produce uneven staining particularly for relatively thick specimens, whereas covalent labeling renders the labels compatible with a wider range of sample processing or clearing techniques. Second, nonfluorescent histology stains are poorly suited to volumetric imaging and fluorescence-based super-resolution microscopy methods. Third, classic histology stains have limited flexibility in their spectral properties, whereas modern fluorescent dyes are available in many forms across the visible and near infrared with high quantum yields and good resistance to photobleaching.

To partly address the limitations of antibodies and small-molecule histology stains, we developed simple procedures for using commercially available small-molecule fluorophores to brightly label a wide variety of major organelles and landmark structures in cells and tissues for super-resolution microscopy and cleared tissue microscopy. The idea is to use small molecules to covalently label abundant chemical functional groups on biological samples, instead of specific biomolecules, to reveal the general physiology of the sample. These include the use of amine-reactive fluorophores for measuring distributions of proteins and aldehyde-reactive fluorophores for measuring distributions of oxidized carbohydrates, although other abundant reactive groups (thiols, carboxylates, etc.) may also be suitable. We termed this approach Fluorescent Labeling of Abundant Reactive Entities (FLARE). While the underlying chemical reactions are well known[142,155–160] for instance, when labeling purified proteins,[142] specimen surfaces,[161] or isolated live cells for subsequent tracking,[162] the great utility of this simple labeling approach in super-resolution and cleared tissue microscopy has not been previously recognized. For super-resolution microscopy, we used expansion microscopy (ExM), a recently developed technique that physically expands biological specimens embedded in a swellable polymer.[94,95,163] The physical expansion enables features closer than the ~250-nm diffraction limit of

traditional light microscopy to be resolved in the expanded state on a conventional microscope. For cleared tissue microscopy with unexpanded specimens, we used a variant of the solvent-based tissue clearing method immunolabeling-enabled three-dimensional imaging of solvent-cleared organs (iDISCO)[164] that uses ethyl cinnamate (EC) for index matching.[165]

3.2.3 RESULTS:

As seen in figure 3.1, FLARE can be easily combined with different staining modalities [e.g., fluorescence in situ hybridization (FISH) labeling of nucleic acids], is compatible with a large selection of excellent fluorophores across the visible and near infrared and reveals abundant details in a wide range of samples and sample processing methods.

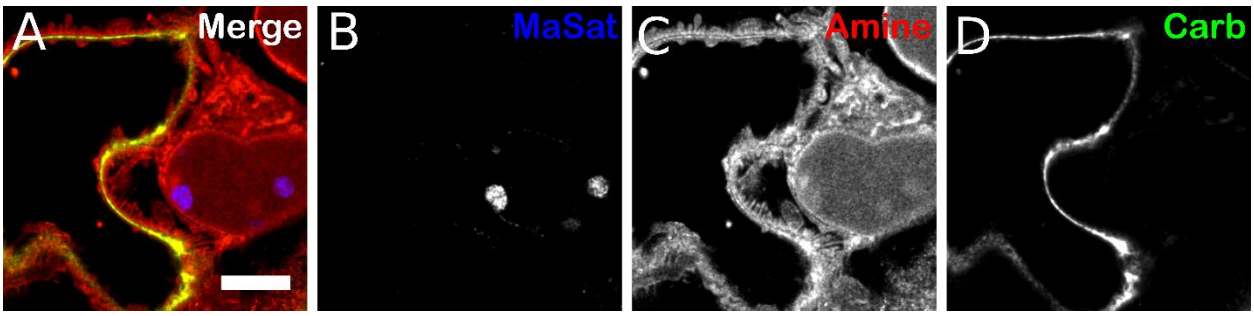


Figure 3.1 **Use of FLARE staining on mouse kidney tissue with DNA FISH protocol.**
(A-D) Confocal microscopy images of an expanded mouse kidney tissue that was stained for carbohydrates, amine, and then DNA FISH against pericentromeric major satellite (MaSat) DNA. **(A)** Multichannel, maximum intensity projection and **(B-D)** corresponding single channel images. Scale bar is 3 µm and is in pre-expansion units.

3.2.2 PROCEDURE:

The DNA-FISH procedure was adapted from Beliveau et al.[25] FLARE-stained and MAP-expanded mouse kidney samples were first incubated in 0.5% Triton X-100 in 1× PBS for 2 hours. Samples were washed once in 1× PBS, then in 2× SSCT (0.1% Tween 20 in 2× SSC), and lastly, in hybridization buffer (50% formamide and 0.1% Tween 20 in 2× SSC) for 10 min each time. Samples were incubated in fresh hybridization buffer for 30 min at 60°C. The hybridization mixture (50% formamide, 10% dextran sulfate, 0.1% Tween 20, 3 mM sodium azide, 100 nM MaSat oligo probe, and 100 nM oligo reporter in 2× SSC)

was preheated to 92.5°C for 5 min and then added to each sample at an approximate 2:1 volume ratio. Samples were denatured at 92.5°C for 10 min and hybridized at 37°C overnight. Samples were washed three times in 2× SSCT, first at 60°C, second at 37°C, and lastly at room temperature, 15 min each time. Samples were stored in 0.2× SSCT at 4°C for at least 1 hour or until needed for imaging (within a week). Before imaging, samples were fully expanded by replacing the sample buffer at least twice with water every 10 min at 4°C.

The MaSat oligo probe (5'-GGAATATGGC GAGAAAACTG AAAATCACGG AATGATACGG CGACCACCGA ACTGCTACAG-3') contains 30 nt of the mouse major satellite repeat DNA sequence [obtained from Lehnertz et al.][166] and 30 nt complimentary to the fluorophore conjugated oligo reporter (5'-/5ATTO647NN/CTGTAGCAGT TCGGTGGTCG CCGTATCATT-3).

3.3 FLUORESCENT LABELING OF ABUNDANT REACTIVE ENTITIES (FLARE) FOR CLEARED-TISSUE AND SUPER-RESOLUTION MICROSCOPY

All material in this sub-chapter is reproduced with permission from:

Min Yen Lee*, Chenyi Mao*, Adam K. Glaser, **Marcus A. Woodworth**, Aaron R. Halpern, Adilijiang Ali, Jonathan T. C. Liu & Joshua C. Vaughan. "Fluorescent Labeling of Abundant Reactive Entities (FLARE) for Cleared-Tissue and Super-Resolution Microscopy" Nat. Protoc. 17, 819-846 (2022) Copyright 2022, The Author(s), under exclusive license to Springer Nature Limited

* These authors contributed equally to this work.

All material in this sub-chapter has been reformatted to adapt to the style of this thesis.

3.3.1. ABSTRACT:

Fluorescence microscopy is a vital tool in biomedical research but faces considerable challenges in achieving uniform or bright labeling. For instance, fluorescent proteins are limited to model organisms, and antibody conjugates can be inconsistent and difficult to use with thick specimens. To partly address

these challenges, we developed a labeling protocol that can rapidly visualize many well-contrasted key features and landmarks on biological specimens in both thin and thick tissues or cultured cells. This approach uses established reactive fluorophores to label a variety of biological specimens for cleared-tissue microscopy or expansion super-resolution microscopy and is termed FLARE (fluorescent labeling of abundant reactive entities). These fluorophores target chemical groups and reveal their distribution on the specimens; amine-reactive fluorophores such as hydroxysuccinimidyl esters target accessible amines on proteins, while hydrazide fluorophores target oxidized carbohydrates. The resulting stains provide signals analogous to traditional general histology stains such as H&E or periodic acid–Schiff but use fluorescent probes that are compatible with volumetric imaging. In general, the stains for FLARE are performed in the order of carbohydrates, amine and DNA, and the incubation time for the stains varies from 1 h to 1 d depending on the combination of stains and the type and thickness of the biological specimens. FLARE is powerful, robust and easy to implement in laboratories that already routinely do fluorescence microscopy.

3.3.2. INTRODUCTION:

Fluorescence microscopy is a technique that is commonly used in the biomedical sciences. It offers the powerful ability to visualize structures or molecules in three dimensions within biological specimens and gives relevant context to the study of the functions of these structures. However, the success of fluorescence microscopy measurements often hinges on the ability of fluorescent labels to detect the objects of interest in a specimen. Immunolabeling of specimens using antibodies is popular due to its high specificity and contrast, but it faces key challenges. [152,153] First, many biological molecules or structures of interest lack good commercially available antibodies that are specific to relevant targets, particularly for specimens that have been heavily fixed or processed. Second, the relatively large size of antibodies (~150,000 Da, or ~15 nm, for an immunoglobulin G) can often make staining of samples >100 μm in thickness very time consuming due to slow diffusion through the sample, with sample preparation

that can take weeks. Fluorescent proteins, which are also popular fluorescent probes, require genetic manipulation and are thus limited to a subset of laboratory model organisms.

To partially address these limitations, we developed a fluorescent labeling method that uses reactive probes to target abundant chemical groups on biological samples. We termed the method FLARE (fluorescent labeling of abundant reactive entities), and we have used this method successfully to label cultured cells, mouse tissues and human tissues.[144]

3.3.3. COMPATIBILITY WITH OTHER PROCEDURES:

There may be structures or molecules that will simply not be possible to label with high contrast using FLARE alone. In such cases, researchers will need to use traditional specific labels, such as immunolabeling of proteins or labeling of nucleic acids by FISH. If desired, however, FLARE can often be combined with the use of these specific labels (e.g., antibodies or DNA oligonucleotides), and the FLARE dyes can be chosen in virtually any spectral region across the visible or near-infrared to accommodate other probes. In terms of antibody labeling, we found that 14 out of 16 tested antibodies were able to label their targets on cells or tissues after FLARE, and that the carbohydrate oxidation step appeared to be the step that perturbed subsequent antibody labeling for 2 out of the 16 tests (amine labeling did not perturb antibody binding in any of the tests).[144] If desired, it is possible to perform immunolabeling before FLARE to avoid oxidation-induced loss of antigenicity,[152] and it may be possible to use milder oxidation that preserves antigenicity, although we have not yet tested this thoroughly. For the combination of FLARE and DNA FISH, we performed FLARE before DNA FISH and achieved the expected results (Figure 3.2b,c), showing that FLARE does not substantially perturb DNA FISH.

The detailed procedures for FLARE that we describe here are compatible with a variety of types of specimens, sample formats and applications, including the examples shown for expanded cells and tissues, cleared thick tissues and thin uncleared tissue sections. Although FLARE is useful for obtaining

relatively general contrast in order to quickly explore the general physiology of specimens, researchers who wish to selectively stain specific structures or molecules are advised to use traditional, specific labels. In those situations, it may nonetheless still be useful to combine FLARE with specific labels to gain the best of both worlds.

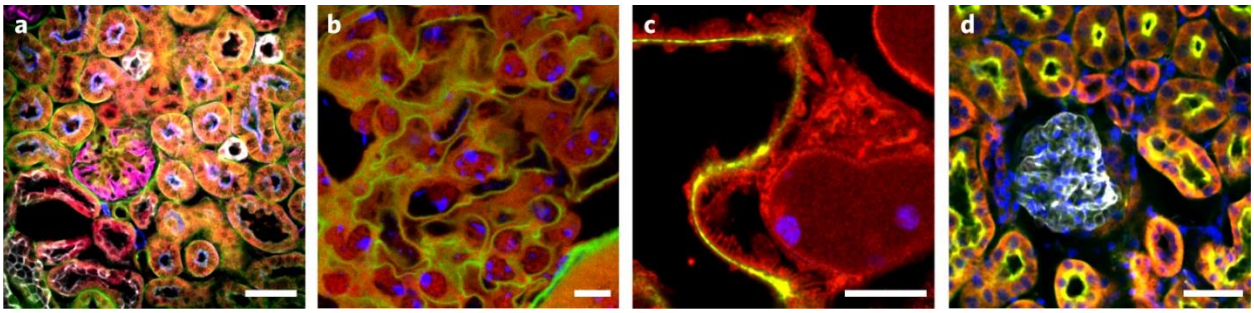


Figure 3.2 **Compatibility of FLARE with diverse procedures.**
FLARE-stained 100-µm-thick mouse kidney sections prepared using various procedures that include two different ExM variants, tissue clearing and the combination with either immunostaining or DNA FISH. **a**, Five-color confocal microscopy image of unexpanded mouse kidney tissue that was stained for carbohydrates (green) and amines (red), and then immunostained for podocalyxin (magenta), aquaporin-1 (blue) and cytokeratin 8+18 (gray). **b**, Confocal microscopy image of unexpanded cleared mouse kidney tissue that was stained for carbohydrates (green), amines (red) and then DNA FISH (blue) against pericentromeric major satellite (MaSat) DNA. **c**, Confocal microscopy image of expanded mouse kidney tissue that was treated by the detergent-based protocol and stained for carbohydrates (green), amines (red) and then DNA FISH (blue) against MaSat DNA. **d**, Confocal microscopy image of enzyme digested expanded mouse kidney tissue that was stained for carbohydrates (green), amines (red) and DNA (blue), and then immunostained for podocalyxin (gray). Scale bars: 50 µm (a,d), 5 µm (b) and 3 µm (c, pre-expansion units). a and c reproduced with permission from ref. 3, AAAS, under a Creative Commons license CC BY-NC 4.0.

3.3.3. MATERIALS:

DNA-FISH on Expanded Kidney tissue General reagents:

- 10 µM Major Satellite (MaSat) DNA oligonucleotide probe (5’-GGAATATGGC GAGAAAACTG AAAATCACGG AATGATACGG CGACCACCGA ACTGCTACAG-3’, IDT).
- 10 µM Fluorescent oligonucleotide reporter probe (5’- /5ATTO647NN/CTGTAGCAGT TCGGTGGTCG CCGTATCATT-3’, IDT). Chemical reagents

● 10× PBS, pH 7.4 (e.g., Fisher Bioreagents # L-5400)

● Sodium azide (NaN3, Fisher Scientific #S227I)

● Deionized water (DI water)

● Triton X-100 (Sigma-Aldrich, #X100)

● Formamide (Fisher Chemical, F84-1) (Caution!)

● 20× SSC (Sigma-Aldrich, S6639)

● 50% OmniPur Dextran Sulfate (EMD Millipore, 3730)

● Tween 20 (Sigma-Aldrich, P9416)

Caution: Researchers should consult the safety data sheet (SDS) for these chemicals prior to use. Formamide is toxic. Procedures using formamide should be performed in a fume hood and researchers are advised to be cautious.

Other materials:

● 1.5 mL microcentrifuge tubes (e.g., Fisher Scientific, 05-408-129)

● Razor blade (American Line, #66-0089) Stock solutions

● 1× PBS (pH 7.4)

● 300 mM sodium azide: o Add 1.95 g of sodium azide to 100 mL of H2O.

● 20% Triton X-100 solution (v/v) o Add 2 mL of Triton X-100 to 8 mL of H2O. o Store at 4 °C until use.

● 20% Tween-20 solution (v/v) o Add 2 mL of Tween-20 to 8 mL of H2O. o Store at 4 °C until use.

● 20× SSC (pH 7.0)

● 2× SSC (pH 7.0)

● 2× SSCT: 2× SSC containing 0.1% Tween 20 (v/v). o Add 5 µL of 20% Tween 20 solution into 995 µL of 2×SSC.

● 20% Triton X-100 solution (v/v) o Add 2 mL of Triton X-100 to 8 mL of H2O. o Store at 4 °C until use.

● DNA-FISH perm solution: 1× PBS containing 0.5% (v/v) Triton X-100.

o Add 25 µL of 20% Triton X-100 solution to 975 µL of 1× PBS. o Store at 4 °C until use.

● Hybridization buffer solution: 2× SSC with 50% (v/v) Formamide and 0.1% (v/v) Tween-20.

o Add 10 µL of 20% Tween-20 solution, 1000 µL of Formamide, 200 µL of 20× SSC to 790 µL of H2O. o Prepare fresh each time.

● Hybridization mixture solution: 2× SSC with 50% Formamide, 10% Dextran Sulfate, 0.1% (v/v) Tween-20, 3mM sodium azide, 100 nM MaSat probe, 100 nM fluorescent reporter probe.

o Add 1 µL of 20% Tween-20 solution, 1 µL of 300 mM sodium azide, 100 µL of Formamide, 20 µL of 20× SSC, 40 µL of 50% dextran sulfate, 2 µL of 10 µM of MaSat probe, 2 µL of 10 µM fluorescent reporter probe, 34 µL H2O.

o Prepare fresh each time.

3.3.4 PROCEDURE:

The following procedure is written as a general procedure for DNA Fluorescence in situ hybridization (FISH) labeling of Major Satellite genomic DNA in MAP-expanded mouse Kidney tissue nuclei. This serves as a starting guide and could be further optimized to target other genomic regions.

Permeabilization and equilibration

•Timing ~3 hours

1. Incubate MAP-expanded mouse Kidney sample in 1×PBS [50 mL per sample] for 10 min.
2. Remove excess buffer and, using a razor blade, slice a small section of MAP expanded mouse kidney sample and weigh the sliced sample [~ 40 mg per gel slice].

▲Critical When removing the excess 1× PBS, avoid aspirating close to the gel, then carefully slice the gelled sample on a clean surface.

3. Prepare a fresh stock of hybridization mixture so that the approximate volume of the mixture is twice that of the sliced sample [> 0.1 mL per gel slice] and place the mixture into a 1.5 mL tube.

4. Place weighed sliced sample into a 1.5 mL tube and incubate in DNA Perm solution [1 mL per tube] for 120 min.

▲Critical Sliced samples must be gently slid into the tube without breaking the gel apart.

5. Carefully remove DNA Perm solution and incubate sample in 1× PBS [1 mL per tube] for 10 min.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

6. Carefully remove 1× PBS and incubate sample in 2× SSCT [1 mL per tube] for 10 min.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

■ Pause Point Able to store gel sample in 2× SSCT for up to a week.

7. Carefully remove 2× SSCT and incubate in hybridization buffer [1 mL per tube] for 10 min at room temperature.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid. DNA denaturation and in situ hybridization •Timing ~20 hr

8. Carefully remove hybridization buffer and add pre-heated hybridization buffer [1 mL per tube], and incubate for 25 min at 60 °C.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

9. Incubate hybridization mixture made in Step 1 for 5 min at 92.5 °C as the sliced sample remains in hybridization buffer for another 5 min at 60 °C.

10. Carefully remove hybridization buffer from the sliced sample, then slide sample with a disposable plastic tip into the preheated tube with hybridization mixture.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

11. Denature sample in hybridization mixture for 10 min at 92.5 °C.
12. Incubate tube overnight (~18 hr) at 37 °C. Wash and expansion •Timing ~1 hr.
13. Carefully remove hybridization mixture and rinse gels with a pre-heated 2× SSCT [0.5 mL per tube].
14. Carefully remove 2× SSCT and incubate in pre-heated 2× SSCT [0.5 mL per tube] for 15 min at 60 °C.

▲Critical To avoid puncturing or ripping of the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

15. Carefully remove 2× SSCT from the tube and incubate in pre-heated 2× SSCT [0.5 mL per tube] for 15 min at 37 °C.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

16. Carefully remove 2× SSCT from the tube and incubate in 2× SSCT [0.5 mL per tube] for 15 min at room temperature.

▲Critical To avoid puncturing or ripping of the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

17. Carefully remove 2× SSCT from the tube and incubate in 0.2× SSCT [0.5 mL per tube] for 15 min at 4°C.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

■Pause Point Able to store gel sample in 0.2× SSCT for up to a week.

18. Carefully remove 0.2× SSCT from the tube and incubate in DI water for 15 min at 4 °C.

▲Critical To avoid puncturing or ripping the gel, aspirate liquid with a micropipette by slowly sliding pipette tip down the inner wall of the tube until reaching the bottom, then slowly aspirate liquid.

19. Proceed to Procedure 1 Step 13 of the main manuscript for sample mounting.

3.4 VERSATILE, DO-IT-YOURSELF, LOW-COST SPINNING DISK CONFOCAL MICROSCOPE

All material in this sub-chapter is reproduced with permission from:

Aaron R. Halpern, Min Yen Lee, Marco D. Howard, **Marcus A. Woodworth**, Philip R. Nicovich, and Joshua C. Vaughan. Biomedical Optics Express Vol. 13, Issue 2, pp. 1102-1120 (2022)

All material in this sub-chapter has been reformatted to adapt to the style of this thesis.

3.4.1 ABSTRACT:

Confocal microscopy is an invaluable tool for 3D imaging of biological specimens, however, accessibility is often limited to core facilities due to the high cost of the hardware. We describe an inexpensive do-it-yourself (DIY) spinning disk confocal microscope (SDCM) module based on a commercially fabricated chromium photomask that can be added on to a laser-illuminated epifluorescence microscope. The SDCM achieves strong performance across a wide wavelength range (~400-800 nm) as demonstrated through a series of biological imaging applications that include conventional microscopy (immunofluorescence, small-molecule stains, and fluorescence in situ hybridization) and super-resolution microscopy (single-molecule localization microscopy and expansion microscopy). This low-cost and simple DIY SDCM is well-documented and should help increase accessibility to confocal microscopy for researchers.

3.4.2 RESULTS FOR RNA FISH WITH DIY SDCM:

We prepared a more challenging specimen for SDCM in which individual *GAPDH* mRNA molecules were labeled by fluorescence in situ hybridization (FISH) in RPE-1 cells prepared using expansion microscopy (ExM) and expanded isotropically three-fold along each of three axes. In ExM, a swellable hydrogel is synthesized within a fixed specimen and then enlarged upon incubation in deionized water so that

features closer than the ~250 nm diffraction limit of light can be resolved in the expanded state.[94,95] Using a 60× 1.27 NA water-immersion objective lens with disk Sector 5 we found that *GAPDH* mRNA were relatively abundant, but were well-resolved thanks to substantial de-crowding achieved by the ~27-fold (volumetric) expansion (Fig. 3.3, A-C). Furthermore, we were able to observe the 3-dimensional structure of bright mRNA clusters within the nucleus that presumably label nascent mRNA near *GAPDH* loci (Fig. 3.3 B-C).

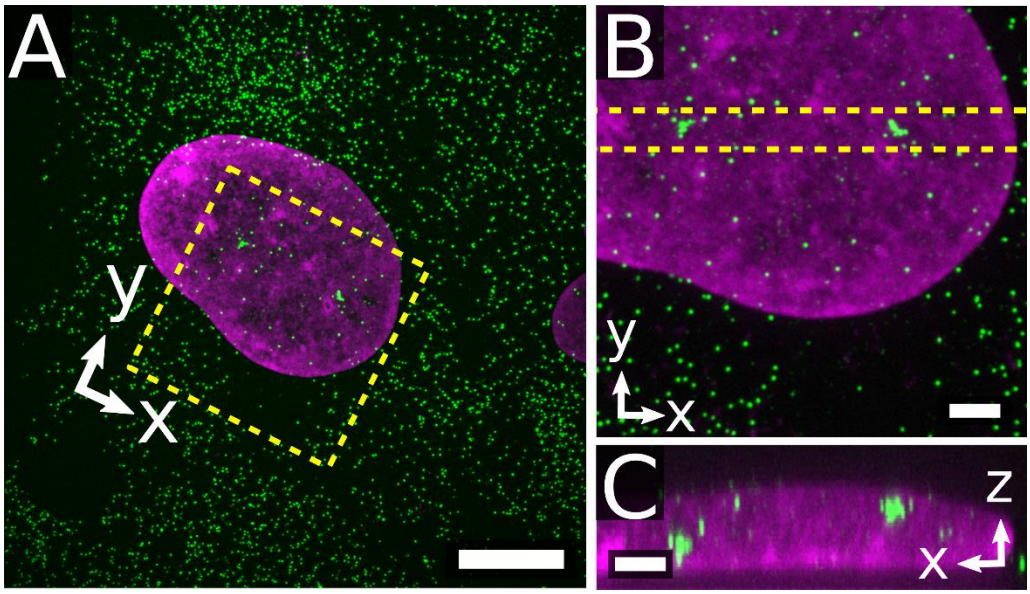


Figure 3.3 **Imaging of RNA FISH-labeled transcripts in an expanded RPE cells using a homebuilt Spinning Disk Confocal.**
(A) Maximum intensity projection of expanded RPE-1 cells stained for *GAPDH* mRNA (green) and DNA (magenta). (B) Zoom-in of the area highlighted in A, showing nascent mRNA clusters, and corresponding transverse projection (C) of the area highlighted in B. Scale bars, 5 µm (A), 2 µm (B,C). Scale bars in expanded samples are given in pre-expansion dimensions.

3.4.2 SAMPLE PREPARATION FOR MRNA FISH LABELING OF EXPANDED CULTURED CELLS:

Single-molecule mRNA FISH imaging was conducted on samples prepared with a slightly modified ExM and FISH protocol. Approximately 70,000 RPE-1 cells were seeded on a 12 mm diameter round coverglass within a well of a 24 well plate and allowed to adhere overnight. The cells were briefly washed in PBS and then fixed in 4% PFA in PBS for 10 min. The cells were extracted in 70% ethanol for 10 min and then equilibrated in wash buffer (2× saline-sodium citrate (SSC), 10% formamide) prior to use.

The sample was treated with a 50 µL of a 125 nM probe set specific to mRNA for *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) in wash buffer containing 10% dextran sulfate overnight at 37 °C and then washed three times for 10 min in excess wash buffer. The sample was treated with a 25 mM methacrylic acid N-hydroxysuccinimidyl (MA-NHS) ester solution in PBS, and then gelled as described below. Gelation was performed by incubating the coverslip containing the sample in monomer solution (1 × PBS, 2 M NaCl, 2.5% acrylamide 0.15% N,N-methylenebisacrylamide, 8.625% sodium acrylate) for 5 min. Meanwhile, 70 µL of monomer containing 0.2% ammonium persulfate and 0.2% tetramethylethylenediamine was placed onto a hydrophobic glass surface and the coverslip was inverted onto the droplet and polymerized in a N_2 atmosphere for 15 min. The sample was allowed to digest overnight in digestion buffer (1× Tris-acetate-EDTA, 0.8 M guanidine hydrochloride, 0.1% TX-100) containing 1% proteinase K. The sample was washed in a large excess of 2× SSC, and then 10 nM of ATTO 565 labeled oligonucleotide readout probe was allowed to hybridize overnight. The sample was washed in a large excess of 2× SSC to remove unhybridized readout probe, 1 µM of TO-PRO-3 iodide (Thermo Fisher, T3605) was added for 30 min, and then the sample was expanded in 0.01× SSC until it reached ∼3× expansion. The sample was mounted on a poly-l-lysine coated coverglass for imaging. The *GAPDH* probe set contained 24 targeting sequences, each of 26-32 nts in length, and was ordered from IDT each normalized to 100 µM concentration.[167] Using an equal component mixture of all sequences, the probe set was enzymatically modified with an amino-11-ddUTP (Lumiprobe, A5040) using a terminal transferase (New England BioLabs, M0315S) as described previously.[168] Briefly, 30 µL of the mixed probe set probe set (10 mM total DNA) was reacted with a 3-fold molar excess of ddUTP using 20 units of enzyme following the enzyme manufacturer's instructions. The modified probe set was then used without further purification and stored at -20 °C. The ATTO 565 labeled reporter oligonucleotide was purchased from IDT. Refer to supplementary spreadsheet 2 for *GAPDH* probe set sequences.

CHAPTER 4: EXAMINING CONTRIBUTORS TO REPRESSIVE MARK HETEROGENEITY AT REPRESSED DEVELOPMENTAL GENES

4.1 PREFACE:

One of the most intriguing observations that arose while developing SCEPTRE was the heterogeneity of marks found at each gene analyzed. Although active histone mark heterogeneity may be explained by marks appearing during unique steps of a gene's transcription cycle,[128] repressive marks are thought to play a much more steady role, since cells must keep certain genes repressed, particularly transcription factors, to maintain their current state. However, we observed heterogeneous levels of H3K27me3 at a repressed developmental gene cluster, with some loci having little to know signal for this canonical repressive mark.[99] We therefore wondered, what factors contribute to repressive mark heterogeneity?

For this last chapter, I will present a new application of the SCEPTRE method for examining three mechanisms that potentially contribute to repressive mark heterogeneity: Cell cycle dynamics, epigenetic state inheritance, and co-repression by alternative marks. Although conventional epigenetic profiling methods, such as ChIP-seq, may address each of these components individually, separating each mechanism's contribution to the overall mark heterogeneity would require testing all three mechanisms at once. For this purpose, I will demonstrate the feasibility of performing SCEPTRE on timelapse imaged cells transduced with a cell cycle reporter system, then propose a follow up experiment that can test the capabilities of the method in distinguishing the extent each mechanism contributes to epigenetic mark heterogeneity.

All materials are adapted from a manuscript in preparation, with the following authors: **Marcus A. Woodworth**, Madeline Wong, Paul Leanza, Chris Kim, Hao Yuan Kueh, and Joshua C. Vaughan.

4.2 INTRODUCTION:

Proper genome organization and regulation is important for a cell's metabolism and development. Nucleosomes, the fundamental unit of chromatin,[14] play an important role in such regulation, from not only restricting access of relevant transcription factors and polymerases to a gene,[38,44,45] but also maintaining histone post-translational modifications that help compact DNA into higher ordered chromatin structures,[43,47,169] and recruit factors relevant to gene regulation.[19] Much research has been dedicated to profiling the distribution of histone modifications, also referred to as histone marks, across the whole genomes of varying cell types.[2,3] Thus, these marks are often referred to as a "code" that defines the epigenetic state of each gene.[20,60,102] However, due to limitations in current methodology, many questions about the impact histone marks have on biological processes remain unanswered, including how abundantly is each histone mark found at a gene, how does each histone mark impact the presence of another, and how well are histone marks levels maintained across multiple generations of cells.

We previously developed a new imaging-based method to quantify histone marks at genomic sites, called Single Cell Evaluation of Post-TRanslational Epigenetic Encoding (SCEPTRE).[99] This method uses the super-resolution technique Expansion Microscopy (ExM)[94,95] to quantify multiple histone marks at *in situ* labeled genomic sites within single cells. By applying SCEPTRE to Retinal Pigment Epithelial (RPE) cells, we observed that clusters of the developmental *HOXC* genes, which are not transcribed in this cell line,[126] presented heterogeneous levels of the repressive mark H3K27me3. This heterogeneity meant that some *HOXC* alleles had little to no detected levels of the repressive mark, despite being silent in this cell type. Given that H3K27me3 is considered a canonical mark for repression of developmental genes,[64,65] we decided to investigate further what mechanisms contributes to this heterogeneity.

Tu further understand the origins of repressive histone mark heterogeneity, we propose to study three potential contributors: Cell cycle, state inheritance and presence of an alternative repressive mark.

In the first case, H3K27me3 and H3K9me3 are diluted due to the delayed deposition of histone marks after transition from the synthesis (S) phase to the G2 phase of the cell cycle.[170] Thus, it is expected that during S phase, H3K27me3 marks would appear lower on any profiled gene. For the second mechanism, repressive mark heterogeneity may be generated at some early stage of a cell population but inherited in later generations. Such inheritance of states have been shown in RNA levels, influencing how cancer cells respond to drug treatments,[171] and the potential bias of cell progenitors towards certain developmental fates over others.[71–73] Lastly, genes that are co-repressed with an alternative mark (e.g. H3K9me3), may present fluctuating levels of each mark so as to maintain repression. Such co-repression has been observed in X-inactivation studies, where H3K27me3 and H3K9me3 both participate in the chromosome repression, forming bordering bands across the X inactivated chromosome.[127,172]

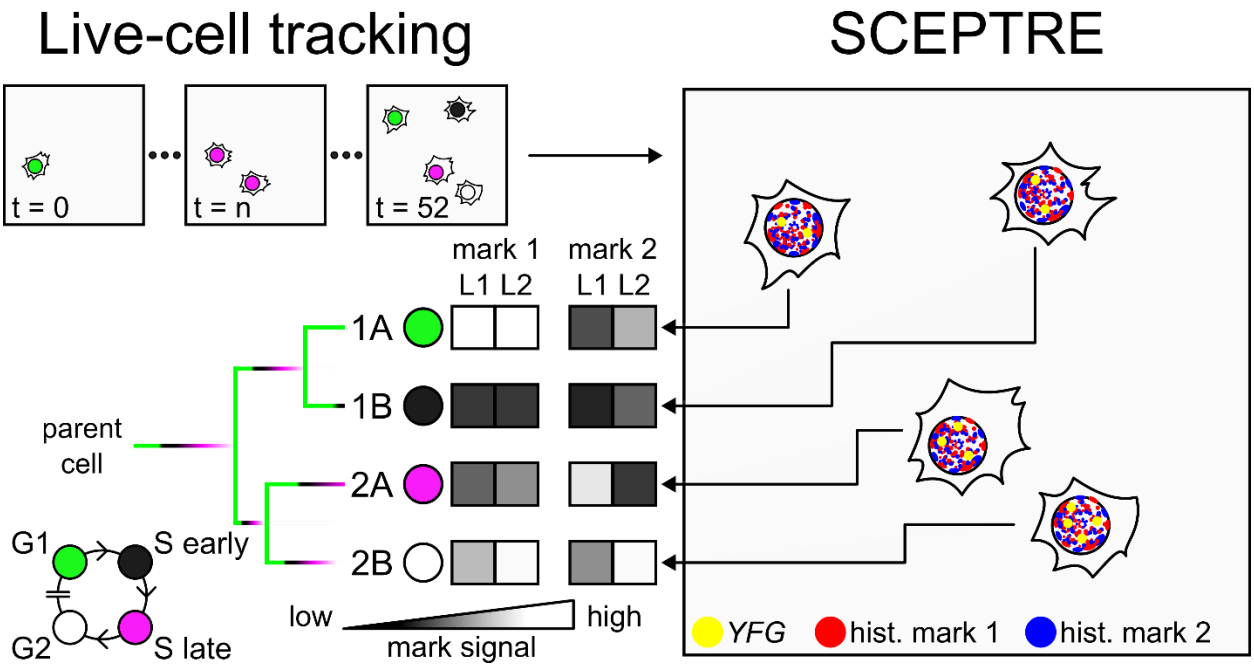


Figure 4.1 **Timelapse imaging of RPE-FUCCI cells followed by epigenetic profiling with SCEPTRE.**
Live RPE-FUCCI cells are first imaged and tracked for multiple divisions over 52 hours. After recording their final cell phase, cells are fixed for SCEPTRE profiling of repressed developmental genes (e.g., *YFG*). These profiles for each gene locus (L1 or L2) are compared to the generational relationship, cell phase, and co-histone mark levels for each cell to distinguish the contribution each mechanism has on epigenetic mark heterogeneity.

Given that the above mechanisms may act independently from each in varying repressive histone mark levels at individual gene alleles, all three must be studied simultaneously to best separate each mechanism's contribution. Although current methods, such as scChIP-seq[75] or MulTI-TAG,[78] may be able to address each mechanism separately, to our knowledge there is no existing method that can investigate them all together. Thus, to address the contribution of cell cycle, epigenetic state inheritance, and the presence of an alternative mark to repressive mark heterogeneity at genomic loci, we combined live-cell tracking of RPE cells with SCEPTRE profiling of repressed developmental genes (figure 4.1).

4.2 PRELIMINARY RESULTS:

4.2.1 CELL CYCLE IMPACT ON H3K27ME3 HETEROGENEITY AT *HOXC*

To investigate if cell cycle dynamics contributes to H3K27me3 heterogeneity at the *HOXC* cluster, we transduced RPE cells with a PCNA-Interacting Protein degron-linked Fluorescent Ubiquitination-based Cell Cycle Indicator (PIP-FUCCI).[173] This minimal FUCCI sensor, compared to the original system,[174] precisely indicates each phase of the cell cycle by either expressing or degrading one of two fluorescent proteins (i.e., mVenus and mCherry) during each phase transition (supplementary figure 4.4). Thus, the FUCCI sensor can report a cell's phase before fixation, allowing us to observe the chromatin state of profiled genes from cells in different phases of the cell cycle.

To test our capability to link repressive mark heterogeneity to cell cycle dynamics, we applied SCEPTRE to PIP-FUCCI transduced RPE cells (RPE-FUCCI) to determine the fluorescence signal of H3K27me3 at the FISH-labeled *HOXC* cluster for cells in G1, early or late S, and G2 phases (Figure 4.2). Given that whole nuclear histone mark level distributions vary between different cell phases (supplementary figure 4.5), we normalized the signal of all segmented clusters by dividing the detected fluorescence signal by the median signal within H3K27me3 clusters in each cell. Although we observed high H3K27me3 signal between all *HOXC* clusters, broad signal heterogeneity is still observed within these clusters after dividing them by cell phase (Figure 4.1B). Despite this observation being similar to previously published

observations,[170] we were surprised to see that H3K27me3 signals at HOXC recover rapidly within the S phase (comparing early and late S results) before cells enter G2 phase. Thus, SCEPTRE demonstrates there is a moderate contribution of cell cycle dynamics towards H3K27me3 heterogeneity at *HOXC* clusters.

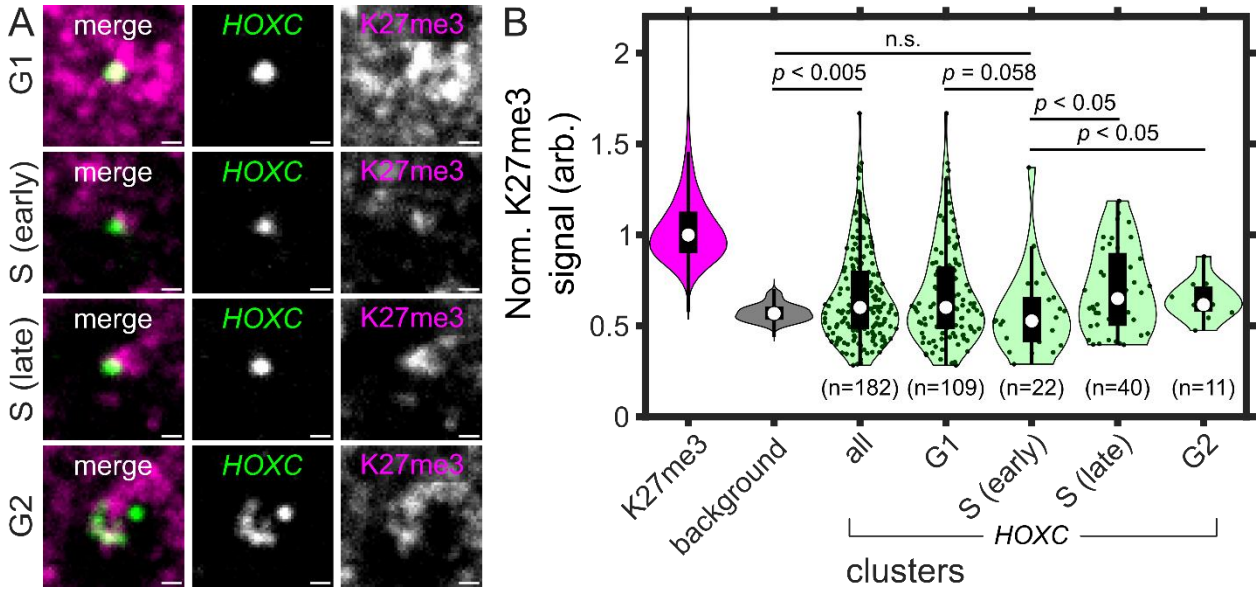


Figure 4.2 **H3K27me3 levels at *HOXC* at various stages of the cell cycle.**
(A) Spinning disk confocal images of the approximate center plane of an image stack for each FISH-labeled *HOXC* cluster (green) and immunostained H3K27me3 clusters (K27me3, magenta) for RPE-FUCCI cells in G1, early S, late S or G2 phase. (B) The distribution of normalized H3K27me3 signal in H3K27me3 clusters, representative background clusters, and at all *HOXC* loci, or at each cell phase (G1, S early, S late and G2). Significance determined by a right-tailed Wilcoxon rank sum test between background clusters and all *HOXC*, G1 vs. early S *HOXC*, late S vs. early S *HOXC*, and G2 vs. early S *HOXC*. All scale bars, in pre-expansion units, are 200 nm.

4.2.2 SCEPTRE PROFILING OF TIMELAPSE RPE-FUCCI CELLS

Next, we investigated if heterogeneous levels of H3K27me3 marks at *HOXC* clusters are inherited across different generations of unique cells. We performed timelapse imaging of the RPE-FUCCI cells for 2-3 divisions, then used SCEPTRE to profile the levels of H3K27me3 at *HOXC* clusters (Figure 4.3). This allowed us to detect and compare the repressive mark levels between HOXC loci of different siblings or whole progeny of each parent cell, in addition to determining their cell cycle phase. Although no

correlation was seen for H3K27me3 signals between randomly selected loci of sibling cells, common repressive mark levels are observed within each lineage, with different signal distributions seen between lineages (supplementary figure 4.6). This suggests that histone mark level inheritance plays are greater role in histone mark heterogeneity compared to cell phases, given that siblings with identical cell phase did not show any greater correlation in H3K27me3 mark levels at HOXC loci compared to siblings with differing cell phases.

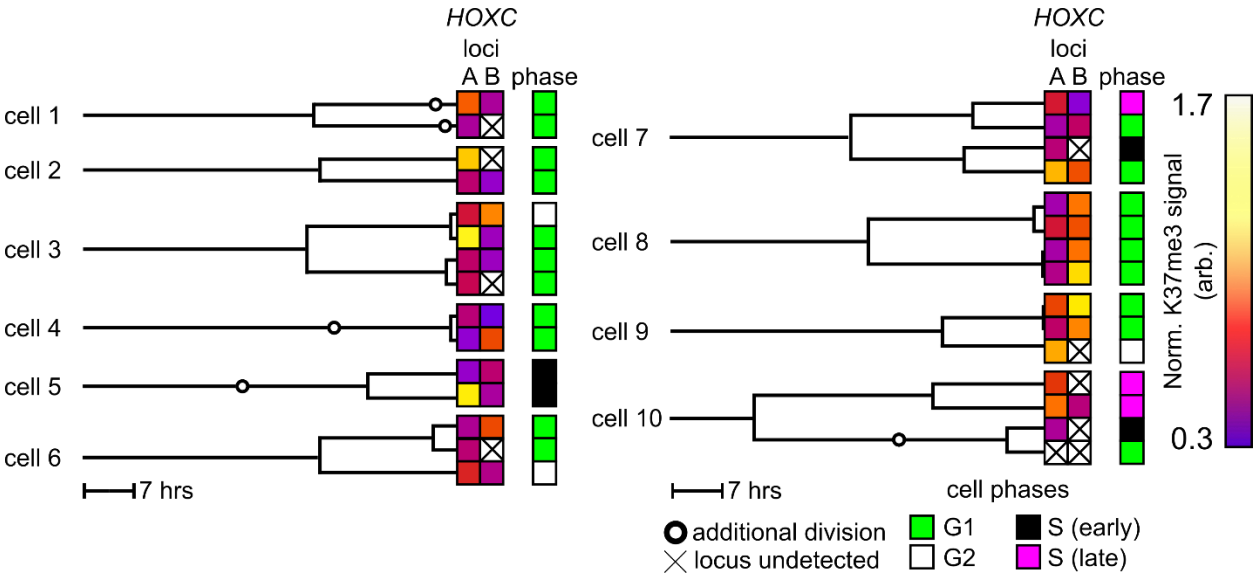


Figure 4.3 **SCEPTRE profiling of H3K27me3 levels at *HOXC* in timelapse imaged RPE-FUCCI cells.**
Cell lineages for 10 different parent cells tracked over 52 hours, along with the normalized H3K27me3 signal for 2 of the *HOXC* loci (randomly assigned as A or B) and the cell state determined by PIP-FUCCI.

4.3 DISCUSSION:

In this preliminary study, we demonstrate that SCEPTRE can interrogate the contribution that various mechanisms have on repressive mark heterogeneity. For the two tested mechanisms, cell cycle and epigenetic state inheritance, cell cycle contributed moderately to the overall heterogeneity of H3K27me3 at the developmental *HOXC* cluster, particularly as cells transitioned from G1 to S phase, while cells originating from the same parent cell-maintained distributions that differed between parent cells. However, given the limited number of cells profiled, and the fact that only one histone mark and developmental gene were targeted, further studies are needed to better quantify the overall influence

each process has on the repressive mark heterogeneity, while taking more mechanisms into consideration. With this in mind, we propose a new set of experiments that would both validate and add to the above preliminary conclusions:

First, perform timelapse imaging on a larger sample, to capture a greater number of cells for each cell phase. Of the 75 cells that were tracked, and SCEPTRE profiled, 65% were in G1, 31% were in S and 4% were in G2. This meant that G1 cells were over-sampled compared to the other phases. Profiling of more cells in S and G2 phase would provide a better comparison between histone mark levels at each phase per profiled gene.

Second, incorporate H3K9me3 signals detection at each targeted gene in addition to H3K27me3 signals. Given that H3K27me3 is not the only mark found on repressed genes, an alternative mark such as H3K9me3 may complement, compete, or independently label a repressed gene. Since staining of H3K9me3 with H3K27me3 is feasible withing these cells (supplementary figure 4.4), visualizing H3K9me3 with H3K27me3 distributions at genomic loci may help explain some of the heterogeneity present at *HOXC* and other developmental genes.

Third, perform SCEPTRE profiling on a broader panel of genes within live-cell tracked RPE-FUCCI cells. Preliminary results are based on a single gene cluster, *HOXC*. Although this region served well for a proof-of-concept, more genes should be targeted to obtain broader conclusions. We therefore have designed and amplified DNA oligopaint FISH probes for a broader panel of genes, including: *SIX6* and *VSX2,* developmental genes repressed during RPE differentiation;[175] and *SGCZ*, which expresses the Zeta-sarcoglycan protein in vasculature that causes muscular dystrophy when mutated.[176] These genes present varying levels of H3K27me3 and H3K9me3 according to ensemble CUT&RUN results (supplementary figure 4.7). Thus, targeting these genes would help delineate the capabilities of SCEPTRE profiling of repressive marks in tracked cells.

We believe that by applying the above changes to our existing experimental set up, we will be able to better clarify what causes repressive mark heterogeneity in state defining genes of developing or diseased cells. Not only will this set up allow for elucidating core biology processes for cell identity maintenance, but also provide other epigenetic researchers with a tool that covers limitations found in current epigenetic profiling methods.

4.4 PROCEDURE:

4.4.1. LIVE-CELL TRACKING

hTERT-RPE1 cells, transduced with a Lentiviral carrying the PIP-FUCCI system (pLenti-PGK-Neo-PIP-FUCCI, addgene), were seeded in an 8-microwell chamber (ibidi) at a density of 2000 cells per well within DMEM growth media (gibco) supplemented with 10% Fetal Bovine Serum (gibco), 1× non-essential amino acids (gibco), and 1× selection antibiotic G418. Cells were incubated at 37°C under 5% carbon dioxide and, using an inverted widefield fluorescence microscope (Leica DMi8) with a 40X air objective, imaged with a sCMOS camera (Photometrics Prime 95B) and a motorized stage (ASI MS-2000). Fluorescent mCherry and mVenus signals, along with Differential Interference Contrast (DIC), were taken of each field of view every 35 minutes for 52 hours.

4.4.2 CELL FIXATION AND IMMUNOSTAINING

After live-cell tracking, cells were fixed in a 1:1 Ethanol/Methanol solution for 6 minutes at -20°C. After fixation, samples were washed three times in 1× Phosphate buffered Saline solution (PBS, Fisher Scientific) supplemented with 3 mM Sodium Azide (Fisher Scientific), then incubated in 0.1% Triton-X 100 (SigmaAldrich) in 1× PBS azide for 10 minutes. Samples were washed three times with 1× PBS azide, then incubated in block solution (10% Bovine Serum Albumin (Rockland Immunochemicals Inc.) in 1× PBS azide) for 1 hour. After blocking, cells were placed in 10 µg/mL of Rabbit anti-H3K9me3 (abcam) and Mouse anti-H3K27me3 (Active Motif) diluted in block. Cells were incubated at Room temperature for 1 hour, then at 4°C overnight.

After overnight incubation, cells were washed three times in block solution, 10 minutes each time, then placed in 6 µg/mL of Donkey anti-Rabbit Alexa Fluor 488 (Jackson ImmunoResearch) and Donkey anti-mouse (Jackson ImmunoResearch) conjugated to Alexa Fluor 568 (SigmaAldrich) diluted in block. Cells were incubated in secondary solution for 2 hours, then washed three times in 1× PBS azide. Samples were post-fixed with 4% paraformaldehyde (PFA, Electron Microscopy Sciences) in 1× PBS for 10 minutes. Samples were then washed three times with PBS azide, then stored at 4°C for ~1 week before proceeding with expansion microscopy.

4.4.3 GELATION, DIGESTION AND EXPANSION OF FUCCI-RPE CELLS

Samples were treated with 5mM methacrylic acid N-Hydroxysuccinimidyl ester (MA-NHS, SigmaAldrich) in 1× PBS azide for 10 minutes. Samples were washed three times with 1× PBS, then incubated in ExM monomer solution (1× PBS with 2 M NaCl, 2.5% (w/w) acrylamide (bio-rad), 0.15% (w/w) N,N'-methylenebisacrylamide (bio-rad) and 8.625% (w/w) sodium acrylate (SigmaAldrich)) for 10 minutes at 37°C. Cells were gelled within the 8-microWell chambers with 193 µL of ExM monomer solution supplemented with 4 µL of 10% (v/v) TEMED (SigmaAldrich) and 3 µL of 10% (w/v) APS (Bio-Rad) solutions per well. The ibidi chamber was placed in a sealable container back-filled with nitrogen gas, then incubated at 37°C for 1 hour.

After gelation, each gel was removed from the chamber by first detaching them from the well walls with a thin micro-pipette tip, adding 10 µL of ExM digestion solution (1× TAE (Fisher Scientific) with 0.5% (v/v) Triton X-100, 0.8 M guanidine HCl (SigmaAldrich) and 8 units/ml proteinase K (Thermo Fisher)), then gently pulling the gel out of the well from the corner with a fine tipped metal tweezers. Gels were deposited into 12-well plates containing 1 mL of ExM digestion solution. Gels were digested overnight (~17 hours) at 37°C. Digested samples were expanded by placing the gels in a 150 cm petri-dish filled with DI water, with the water exchanged four times every 20 minutes. 1mL of 20× Saline

Sodium Citrate buffer (SSC, Thermo Fisher) was added to each sample before storing at 4°C for 1 week before DNA FISH.

4.4.4 DNA FISH FOR SCEPTRE PROFILING

DNA FISH was done using our previous protocol,[99] with the following adaptations: Partially expanded gel samples for tracked RPE-FUCCI cells were sliced into 50 µL (e.g., 2 mm X 2 mm X 5 mm) sections. Each section was incubated in 500 µL of hybridization buffer (2X SSC with 0.1% Tween-20 (SigmaAldrich) and 50% OmniPur Formamide (Millipore-Sigma)) for 10 minutes at RT. Gel sections were then incubated for 60 minutes at 60°C, followed by a denaturing step in 100 µL hybridization mixture (2× SSC with 50% Formamide, 10% OmniPur Dextran Sulfate (Millipore-Sigma), 0.1% Tween-20, 3mM Sodium Azide, 4 µM ssDNA oligo library, and 1.25 µM *HOXC* adapter along with a 1.1× concentration of ATTO647N conjugate ssDNA reporter) for 5 minutes at 95°C. Samples were incubated overnight at 42°C. After hybridization, samples were washed three times, 15 minutes each time, in wash buffer (2× SSC with 0.1% Tween-20) at 60°C, 37°C and finally at RT. Cells were stored in 0.2× SSC with 0.01% Tween-20.

4.4.5 SAMPLE MOUNTING AND IMAGING

DNA FISH-labeled samples were first incubated in DI water for 10 minutes, then immobilized on a poly-L-lysine–coated rectangular no. 1.5 coverslip. All samples were imaged on a homebuilt spinning disk confocal microscope,[146] using a Nikon CFI60 Plan Apochromat 60× 1.27 NA water-immersion objective.

4.4.6 IMAGE PROCESSING AND ANALYSIS

Spinning Disk Confocal images were processed and analyzed with our previously published MATLAB scripts,[99] with the following adaptation: H3K27me3 immunofluorescent signal was rolling ball background subtracted with a gaussian filter, set to 50 standard deviations, before segmentation. Background clusters are defined as all voxels within the nuclear defined region that are outside of the H3k27me3 cluster regions, averaging over groups of 149 voxels (e.g., roughly the size of a FISH-labeled locus). Cell cycle was either visually determined, by comparing mVenus and mCherry intensities at same

contrast across cells, or by taking the mean fluorescence intensity of a 3×3 box at the center of each live tracked cell for each fluorescent protein channel (i.e., mVenus and mCherry). Cell phases were classified, following the PIP-FUCCI signal design:[173] Positive mVenus and negative mCherry represents G1 phase; positive mVenus and positive mCherry represent G2 phase; negative mVenus and negative mCherry represent early S phase; and negative mVenus and positive mCherry represent late S phase.

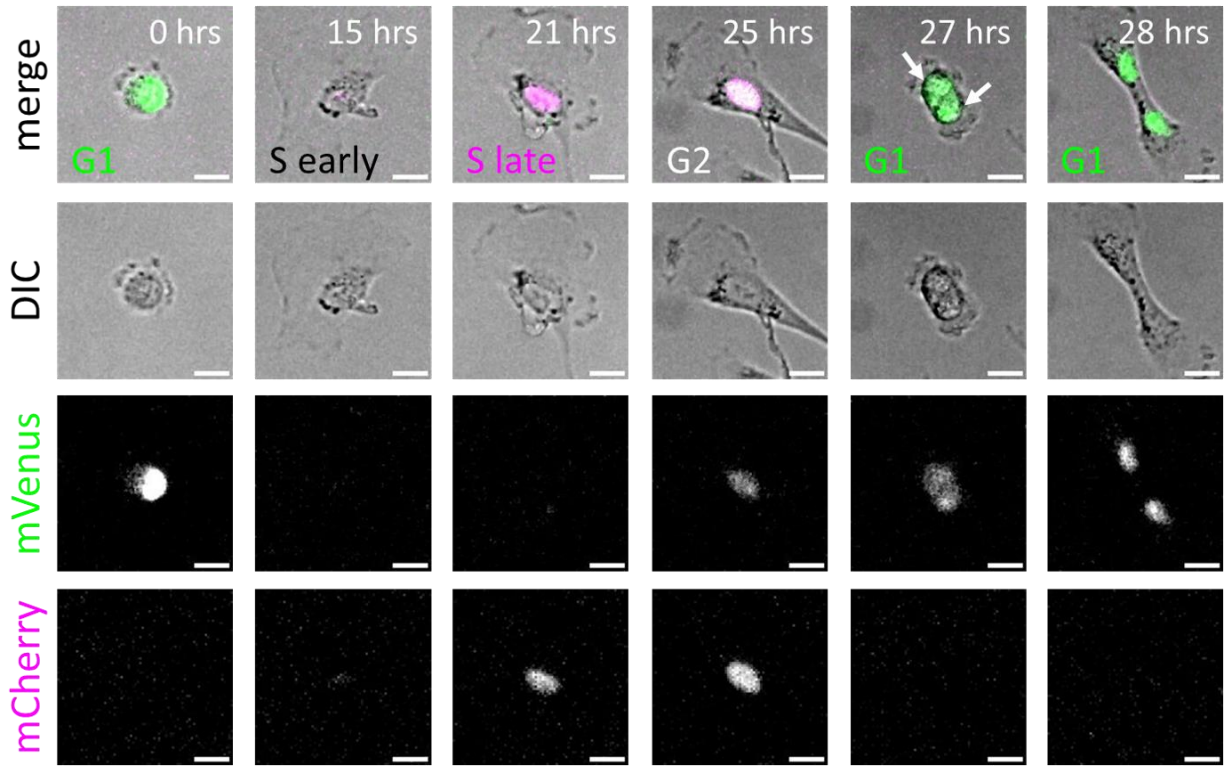
4.4.8 CUT&RUN FOR H3K27ME3 AND H3K9ME3

CUT&RUN followed by sequencing[110,111] results for H3K27me3 in RPE cells were obtained from our previously published results, while CUT&RUN results for H3K9me3 in RPE1 cells were performed using the same specifications as the H3K27me3 run, with the exception of using Rabbit anti-H3K9me3 (abcam) as the targeting antibody.[99]

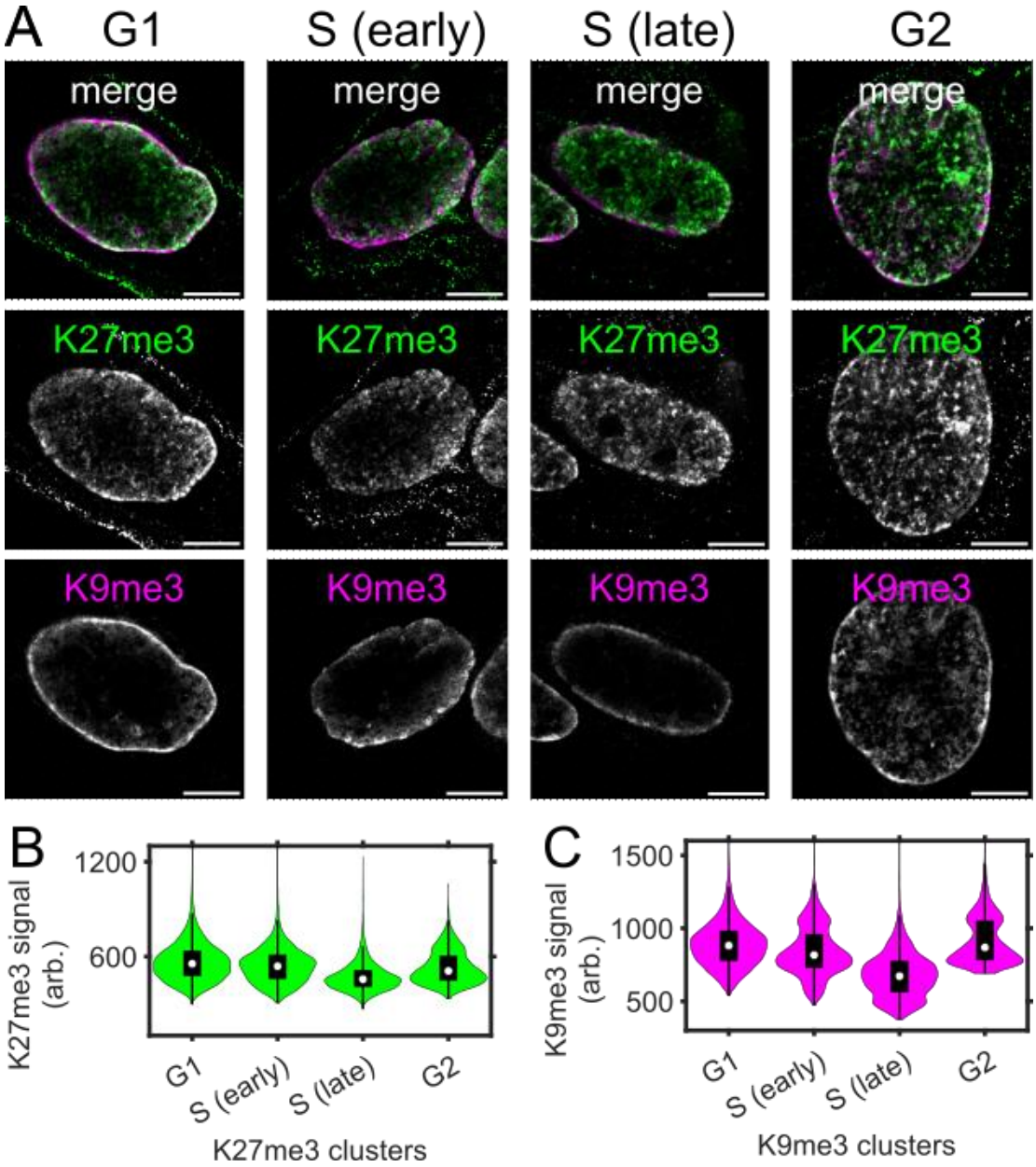
4.4.7 STATISTICAL ANALYSES

For Figure 4.2, significance for *HOXC* H3K27me3 signal distributions between background and all, G1 and early S, Late S and early S, G2 and early S clusters were determined using a Wilcoxon ranksum test in MATLAB. Pearson correlation coefficients were determined between: H3K27me3 signals at randomly selected loci, one from each sibling cell; H3K27me3 signals at randomly selected loci, one from each sibling cell with identical cell phase; H3K27me3 signals at randomly selected loci, one from each sibling cell with different cell phase. In all correlation cases, no significant correlation was detected.

4.5 SUPPLEMENTARY FIGURES:

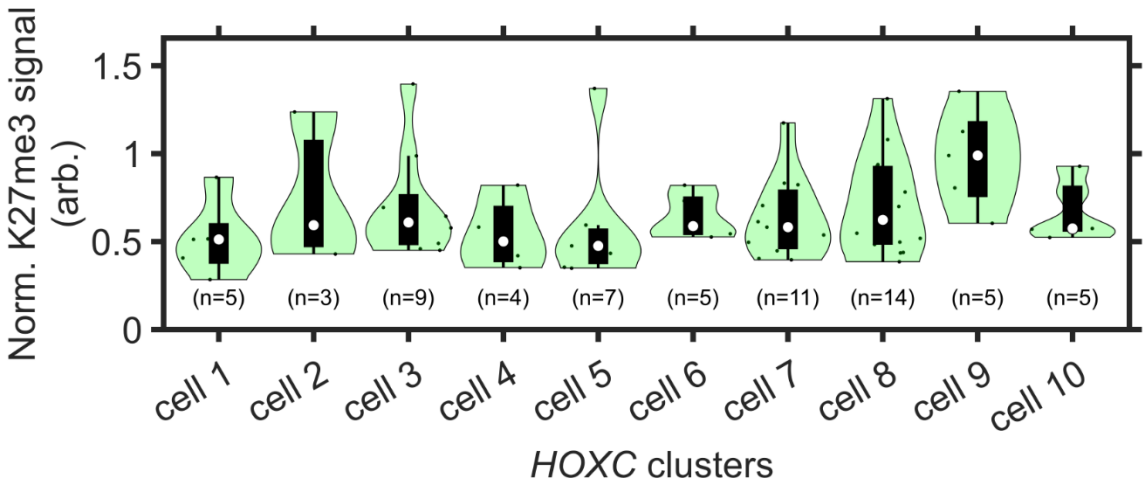


Supplementary Figure 4.4 **Time lapsed images of PIP-FUCCI transduce RPE cells.**
Example images of an RPE-FUCCI cell tracked over 52 hours, expressing mVenus (green) and mCherry (magenta) depending on its cell phase. At the beginning of tracking (0 hrs), the cell is in G1 phase with positive mVenus and negative mCherry. At 15 hrs, the cell is in early S phase, having degraded its mVenus signal (negative) and with no positive mCherry yet. At 21 hrs, the cell is in late S phase, having negative mVenus and positive mCherry. At 25 hours, the cell is in G2 phase with both fluorescent proteins positive. At 27 hours, the cell has divided, and the sibling cells degrade the mCherry signal while maintaining the mVenus signal, as they start in G1 again. A median filter with a radius of 0.5 was applied to each image. Scale bar: 16 µm.



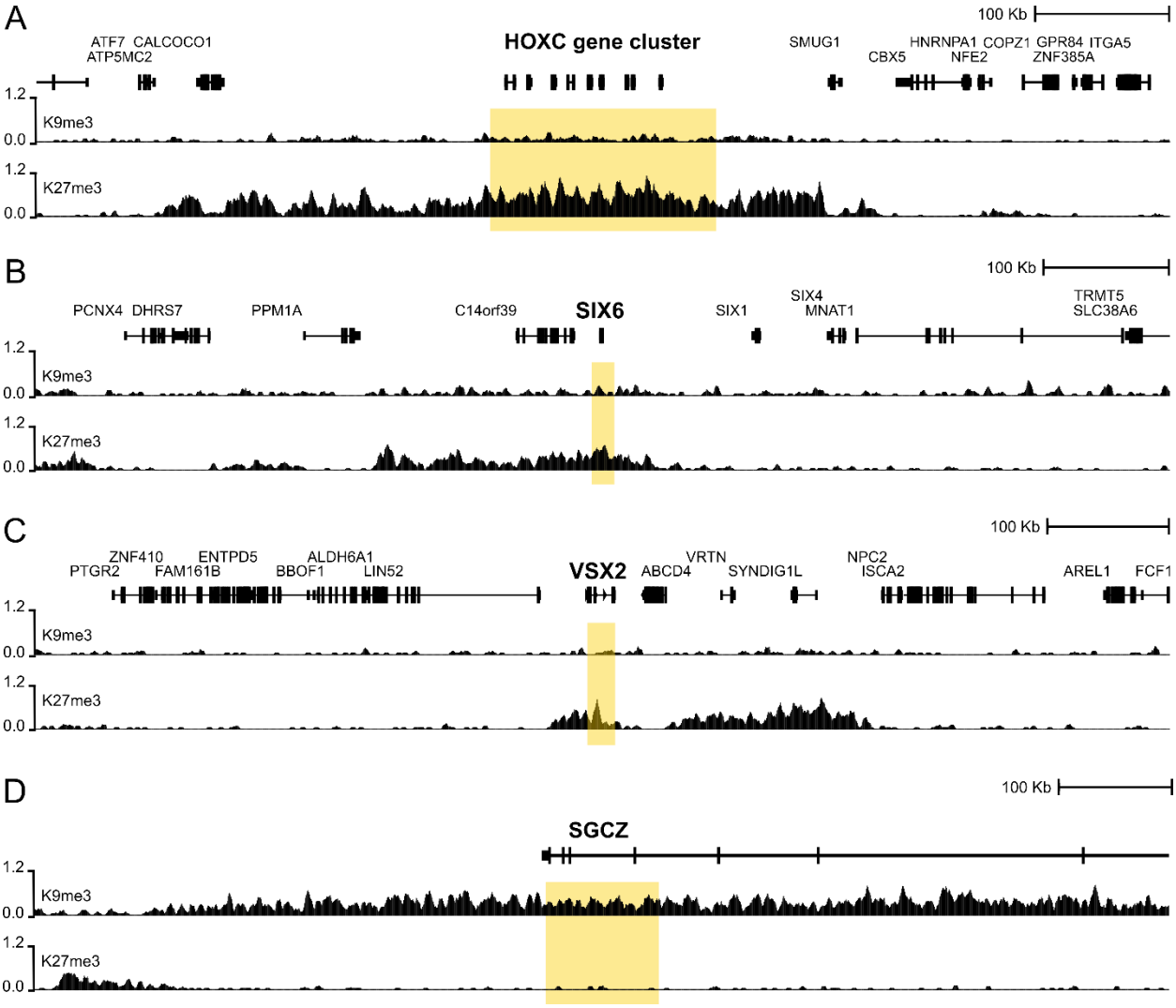
Supplementary Figure 4.5 **Varying levels of H3K27me3 or H3K9me3 cluster intensities across separate phases of the cell cycle**.

(A) Representative image of a 250 nm section of RPE1-FUCCI cells stained for H3K27me3 (K27me3, Green) and H3K9me3 (K9me3, magenta) in either G1, S (early) S (late) or G2 phases. (B) Mean fluorescence H3K27me3 signal for H3K27me3 clusters in cells seen in G1, S (early), S (late) or G2 phases. (C) Mean fluorescence H3K9me3 signal for H3K9me3 clusters in cells seen in G1, S (early), S (late) or G2 phases. Scale bars, in pre-expansion units, are 4 µm.



Supplementary Figure 4.6 **Varying distributions of Normalized H3K27me3 signal across progeny for 10 different parent cells.**

Parent cells are the same cells represented in figure 4.3.



Supplementary Figure 4.7 **Varying densities of H3K9me3 and H3K27me3 marks at a panel of genes.** (A-D) CUT&RUN followed by sequencing normalized counts for H3K9me3 (top) and H3K27me3 (bottom) marks in RPE1 cells for the FISH-targeted *HOXC* (A), *SIX6* (B), *VSX2* (C) and *SGCZ* (D) regions (highlighted).

REFERENCES:

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Bernstein, B. E., Meissner, A. & Lander, E. S. The Mammalian Epigenome. *Cell* **128**, 669–681 (2007).
3. Rivera, C. M. & Ren, B. Mapping Human Epigenomes. *Cell* **155**, 39–55 (2013).
4. Qu, H. & Fang, X. A Brief Review on the Human Encyclopedia of DNA Elements (ENCODE) Project. *Genomics Proteomics Bioinformatics* **11**, 135–141 (2013).
5. Ludwig, C. H. & Bintu, L. Mapping chromatin modifications at the single cell level. *Development* **146**, (2019).
6. Flemming, W. *Zellsubstanz, Kern und Zelltheilung*. (Vogel, 1882).
7. Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* **2**, 72–75 (2001).
8. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell Biol.* **4**, 809–814 (2003).
9. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
10. Johns, E. W. The Histones, their Interactions with DNA, and Some Aspects of Gene Control. in *Ciba Foundation Symposium - Homeostatic Regulators* 128–143 (John Wiley & Sons, Ltd, 1969). doi:10.1002/9780470719695.ch8.
11. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc. Natl. Acad. Sci.* **51**, 786–794 (1964).
12. Itzhaki, R. F. Studies on the accessibility of deoxyribonucleic acid in deoxyribonucleoprotein to cationic molecules. *Biochem. J.* **122**, 583–592 (1971).
13. Clark, R. J. & Felsenfeld, G. Structure of Chromatin. *Nature. New Biol.* **229**, 101–106 (1971).

14. Kornberg, R. D. & Lorch, Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell* **98**, 285–294 (1999).

15. Olins, A. L. & Olins, D. E. Spheroid Chromatin Units (v Bodies). *Science* **183**, 330–332 (1974).

16. Woodcock, C. L. F., Safer, J. P. & Stanchfield, J. E. Structural repeating units in chromatin: I. Evidence for their general occurrence. *Exp. Cell Res.* **97**, 101–110 (1976).

17. Norton, V. G., Marvin, K. W., Yau, P. & Bradbury, E. M. Nucleosome linking number change controlled by acetylation of histones H3 and H4. *J. Biol. Chem.* **265**, 19848–19852 (1990).

18. Jeppesen, P., Mitchell, A., Turner, B. & Perry, P. Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. *Chromosoma* **101**, 322–332 (1992).

19. Turner, B. M. Decoding the nucleosome. *Cell* **75**, 5–8 (1993).

20. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074–1080 (2001).

21. Dijk, E. L. van, Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **30**, 418–426 (2014).

22. Bradbury, S. Landmarks in biological light microscopy. *J. Microsc.* **155**, 281–305 (1989).

23. Levsky, J. M. & Singer, R. H. Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.* **116**, 2833–2838 (2003).

24. Volpi, E. V. & Bridger, J. M. FISH glossary: an overview of the fluorescence in situ hybridization technique. *BioTechniques* **45**, 385–409 (2008).

25. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci.* **109**, 21301–21306 (2012).

26. Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).

27. Moffitt, J. R. & Zhuang, X. RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). in *Methods in Enzymology* vol. 572 1–49 (Elsevier, 2016).

28. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

29. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49 (2019).

30. Ren, B. *et al.* Genome-Wide Location and Function of DNA Binding Proteins. *Science* **290**, 2306–2309 (2000).

31. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

32. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

33. Bernstein, B. E. *et al.* Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse. *Cell* **120**, 169–181 (2005).

34. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

35. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

36. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).

37. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).

38. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

39. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532–537 (1984).

40. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).

41. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73**, 1897–1901 (1976).

42. Song, F. *et al.* Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **344**, 376–380 (2014).

43. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **160**, 1145–1158 (2015).

44. Polach, K. J. & Widom, J. Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation. *J. Mol. Biol.* **254**, 130–149 (1995).

45. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2012).

46. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin Compaction by a Polycomb Group Protein Complex. *Science* **306**, 1574–1577 (2004).

47. Xu, J. *et al.* Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells. *Cell Rep.* **24**, 873–882 (2018).

48. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

49. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).

50. Dekker, J. & Heard, E. Structural and Functional Diversity of Topologically Associating Domains. *FEBS Lett.* **589**, 2877–2884 (2015).

51. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

52. Nozaki, T. *et al.* Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol. Cell* **67**, 282-293.e7 (2017).

53. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).

54. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).

55. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).

56. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41 (2000).

57. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

58. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

59. Tian, Z. *et al.* Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* **13**, R86 (2012).

60. Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).

61. Verdone, L., Caserta, M. & Di Mauro, E. Role of histone acetylation in the control of gene expression. *Biochem. Cell Biol. Biochim. Biol. Cell.* **83**, 344–353 (2005).

62. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).

63. Heintzman, N. D. & Ren, B. Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.* **19**, 541–549 (2009).

64. Ringrose, L. & Paro, R. Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).

65. Steffen, P. A. & Ringrose, L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* **15**, 340–356 (2014).

66. Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).

67. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46 (2007).

68. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631–644 (2008).

69. Bryder, D., Rossi, D. J. & Weissman, I. L. Hematopoietic Stem Cells. *Am. J. Pathol.* **169**, 338–346 (2006).

70. Kueh, H. Y. *et al.* Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nat. Immunol.* **17**, 956–965 (2016).

71. Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic Transformations of Genome-wide Epigenetic Marking and Transcriptional Control Establish T Cell Identity. *Cell* **149**, 467–482 (2012).

72. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).

73. Yu, V. W. C. *et al.* Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell* **167**, 1310-1322.e17 (2016).

74. Garber, M. *et al.* A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Mol. Cell* **47**, 810–822 (2012).

75. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

76. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1–10 (2019).

77. Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazzio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol. Cell* **81**, 4736-4746.e5 (2021).

78. Meers, M. P., Llagas, G., Janssens, D. H., Codomo, C. A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. *Nat. Biotechnol.* 1–9 (2022) doi:10.1038/s41587-022-01522-9.

79. Mongelard, F., Vourc'h, C., Robert-Nicoud, M. & Usson, Y. Quantitative assessment of the alteration of chromatin during the course of FISH procedures. 7.

80. Cremer, M. *et al.* Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes. in *The Nucleus: Volume 1: Nuclei and Subnuclear Components* (ed. Hancock, R.) 205–239 (Humana Press, 2008). doi:10.1007/978-1-59745-406-3_15.

81. Bailey, S. M., Goodwin, E. H. & Cornforth, M. N. Strand-specific fluorescence in situ hybridization: the CO-FISH family. *Cytogenet. Genome Res.* **107**, 14–17 (2004).

82. Brown, J. M., De Ornellas, S., Parisi, E., Schermelleh, L. & Buckle, V. J. RASER-FISH: non-denaturing fluorescence in situ hybridization for preservation of three-dimensional interphase chromatin structure. *Nat. Protoc.* **17**, 1306–1331 (2022).

83. Wang, Y. *et al.* Genome oligopaint via local denaturation fluorescence in situ hybridization. *Mol. Cell* **81**, 1566-1577.e8 (2021).

84. Krufczik, M. *et al.* Combining Low Temperature Fluorescence DNA-Hybridization, Immunostaining, and Super-Resolution Localization Microscopy for Nano-Structure Analysis of ALU Elements and Their Influence on Chromatin Structure. *Int. J. Mol. Sci.* **18**, 1005 (2017).

85. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci.* **112**, 11870–11875 (2015).

86. Geng, Y. & Pertsinidis, A. Simple and versatile imaging of genomic loci in live mammalian cells and early pre-implantation embryos using CAS-LiveFISH. *Sci. Rep.* **11**, 12220 (2021).

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

88. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung: I. Die Construction von Mikroskopen auf Grund der Theorie. *Arch. Für Mikrosk. Anat.* **9**, 413–418 (1873).

89. Streibl, N. Three-dimensional imaging by a microscope. *JOSA A* **2**, 121–127 (1985).

90. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).

91. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).

92. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).

93. Hell, S. W. Microscopy and its focal switch. *Nat. Methods* **6**, 24–32 (2009).

94. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).

95. Chozinski, T. J. *et al.* Expansion microscopy with conventional antibodies and fluorescent proteins. *Nat. Methods* **13**, 485–488 (2016).

96. Wassie, A. T., Zhao, Y. & Boyden, E. S. Expansion microscopy: principles and uses in biological research. *Nat. Methods* **16**, 33–41 (2019).

97. Chen, F. *et al.* Nanoscale imaging of RNA with expansion microscopy. *Nat. Methods* (2016) doi:10.1038/nmeth.3899.

98. Zhao, Y. *et al.* Nanoscale imaging of clinical specimens using pathology-optimized expansion microscopy. *Nat. Biotechnol.* (2017) doi:10.1038/nbt.3892.

99. Woodworth, M. A. *et al.* Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy. *Nucleic Acids Res.* 15 (2021) doi:https://doi.org/10.1093/nar/gkab423.

100. Ng, K. K. *et al.* A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *eLife* **7**, e37851 (2018).

101. Woodworth, M. A., Kueh, H. Y. & Vaughan, J. C. Multiplexed, single-cell profiling of histone modifications with SCEPTRE. (2021).

102. Prakash, K. & Fournier, D. Evidence for the implication of the histone code in building the genome structure. *Biosystems* **164**, 49–59 (2018).

103. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).

104. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

105. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

106. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).

107. Kind, J. *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192 (2013).

108. Chaumeil, J., Okamoto, I. & Heard, E. X-Chromosome Inactivation in Mouse Embryonic Stem Cells: Analysis of Histone Modifications and Transcriptional Activity Using Immunofluorescence and FISH. in *Methods in Enzymology* vol. 376 405–419 (Academic Press, 2003).

109. Nehmé, B., Henry, M. & Mouginot, D. Combined fluorescent in situ hybridization and immunofluorescence: Limiting factors and a substitution strategy for slide-mounted tissue sections. *J. Neurosci. Methods* **196**, 281–288 (2011).

110. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

111. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted *in situ* genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).

112. Beliveau, B. J. *et al.* In Situ Super-Resolution Imaging of Genomic DNA with OligoSTORM and OligoDNA-PAINT. in *Super-Resolution Microscopy* 231–252 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-7265-4_19.

113. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

114. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

115. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

116. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).

117. Beliveau, B. J. *et al.* OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc. Natl. Acad. Sci.* **115**, E2183–E2192 (2018).

118. Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* (2019) doi:10.1038/s41592-019-0404-0.

119. Kubalová, I. *et al.* Prospects and limitations of expansion microscopy in chromatin ultrastructure determination. *Chromosome Res.* **28**, 355–368 (2020).

120. Barber, R. D., Harmer, D. W., Coleman, R. A. & Clark, B. J. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol. Genomics* **21**, 389–395 (2005).

121. Lodish, H. *et al.* Oxidation of Glucose and Fatty Acids to CO2. *Mol. Cell Biol. 4th Ed.* (2000).

122. Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S. & Di Croce, L. The Bivalent Genome: Characterization, Structure, and Regulation. *Trends Genet.* **36**, 118–131 (2020).

123. Kinkley, S. *et al.* reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4 + memory T cells. *Nat. Commun.* **7**, 12514 (2016).

124. Bora, P. S., Bora, N. S., Wu, X., Kaplan, H. J. & Lange, L. G. Molecular cloning, sequencing, and characterization of smooth muscle myosin alkali light chain from human eye cDNA: homology with myocardial fatty acid ethyl ester synthase-III cDNA. *Genomics* **19**, 186–188 (1994).

125. Lai, K.-M. V. *et al.* Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. *PLOS ONE* **10**, e0125522 (2015).

126. Harenza, J. L. *et al.* Transcriptomic profiling of 39 commonly-used neuroblastoma cell lines. *Sci. Data* **4**, 170033 (2017).

127. Darrow, E. M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci.* **113**, E4504–E4512 (2016).

128. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone marks in the 'drivers seat': functional roles in steering the transcription cycle. *Trends Biochem. Sci.* **42**, 977–989 (2017).

129. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).

130. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).

131. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18**, 263–273 (2017).

132. Nguyen, H. Q. *et al.* 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).

133. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol.* **4**, e309 (2006).

134. Bartman, C. R. *et al.* Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. *Mol. Cell* **73**, 519-532.e4 (2019).

135. Pease, N. A. *et al.* Tunable, division-independent control of gene activation timing by a polycomb switch. *Cell Rep.* **34**, (2021).

136. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).

137. Kahn, T. G. *et al.* Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res.* **44**, 10132–10149 (2016).

138. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* **14**, 593–599 (2017).

139. Halpern, A. R., Alas, G. C. M., Chozinski, T. J., Paredez, A. R. & Vaughan, J. C. Hybrid Structured Illumination Expansion Microscopy Reveals Microbial Cytoskeleton Organization. *ACS Nano* **11**, 12677–12686 (2017).

140. Nozawa, R.-S. *et al.* Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat. Struct. Mol. Biol.* **20**, 566–573 (2013).

141. Kang, H. *et al.* Dynamic regulation of histone modifications and long-range chromosomal interactions during postmitotic transcriptional reactivation. *Genes Dev.* **34**, 913–930 (2020).

142. Hermanson, Greg. *Bioconjugate Techniques, 2nd Edition*.

143. *Theory and practice of histological techniques*. (Elsevier Churchill Livingston, 2013).

144. Mao, C. *et al.* Feature-rich covalent stains for super-resolution and cleared tissue fluorescence microscopy. *Sci. Adv.* **6**, eaba4542 (2020).

145. Lee, M. Y. *et al.* Fluorescent labeling of abundant reactive entities (FLARE) for cleared-tissue and super-resolution microscopy. *Nat. Protoc.* **17**, 819–846 (2022).

146. Halpern, A. R. *et al.* Versatile, do-it-yourself, low-cost spinning disk confocal microscope. *Biomed. Opt. Express* **13**, 1102–1120 (2022).

147. Richardson, D. S. & Lichtman, J. W. Clarifying Tissue Clearing. *Cell* **162**, 246–257 (2015).

148. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Biol.* **18**, 685–701 (2017).

149. Sigal, Y. M., Zhou, R. & Zhuang, X. Visualizing and discovering cellular structures with super-resolution microscopy. *Science* **361**, 880–887 (2018).

150. Tainaka, K., Kuno, A., Kubota, S. I., Murakami, T. & Ueda, H. R. Chemical Principles in Tissue Clearing and Staining Protocols for Whole-Body Cell Profiling. *Annu. Rev. Cell Dev. Biol.* **32**, 713–741 (2016).

151. Schermelleh, L. *et al.* Super-resolution microscopy demystified. *Nat. Cell Biol.* **21**, 72–84 (2019).

152. Schnell, U., Dijk, F., Sjollema, K. A. & Giepmans, B. N. G. Immunolabeling artifacts and the need for live-cell imaging. *Nat. Methods* **9**, 152–158 (2012).

153. Bradbury, A. & Plückthun, A. Reproducibility: Standardize antibodies used in research. *Nature* **518**, 27–29 (2015).

154. Elfer, K. N. *et al.* DRAQ5 and Eosin ('D&E') as an Analog to Hematoxylin and Eosin for Rapid Fluorescence Histology of Fresh Tissues. *PLOS ONE* **11**, e0165530 (2016).

155. Ornstein, L., Mautner, W. & Davis, B. new horizons in fluorescence microscopy. *1957* **24**, 1066.

156. Weinblatt, F. M., Shannon, W. A. & Seligman, A. M. A new fluorescent method for the demonstration of macromolecular aldehydes. *Histochemistry* **41**, 353–359 (1975).

157. Lomant, A. J. & Fairbanks, G. Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). *J. Mol. Biol.* **104**, 243–261 (1976).

158. Bragg, P. D. & Hou, C. Subunit composition, function, and spatial arrangement in the Ca2+- and Mg2+-activated adenosine triphosphatases of Escherichia coli and Salmonella typhimurium. *Arch. Biochem. Biophys.* **167**, 311–321 (1975).

159. Baskin, J. M., Dehnert, K. W., Laughlin, S. T., Amacher, S. L. & Bertozzi, C. R. Visualizing enveloping layer glycans during zebrafish early embryogenesis. *Proc. Natl. Acad. Sci.* **107**, 10360–10365 (2010).

160. Saitoh, Y. *et al.* Three-dimensional reconstruction of living mouse liver tissues using cryotechniques with confocal laser scanning microscopy. *J. Electron Microsc. (Tokyo)* **59**, 513–525 (2010).

161. Hill, D. E., Fetterer, R. H. & Urban, J. F. Biotin as a probe of the surface of Ascaris suum developmental stages. *Mol. Biochem. Parasitol.* **41**, 45–52 (1990).

162. Swirski, F. K. *et al.* A Near-Infrared Cell Tracker Reagent for Multiscopic In Vivo Imaging and Quantification of Leukocyte Immune Responses. *PLoS ONE* **2**, e1075 (2007).

163. Ku, T. *et al.* Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nat. Biotechnol.* **34**, 973–981 (2016).

164. Renier, N. *et al.* iDISCO: A Simple, Rapid Method to Immunolabel Large Tissue Samples for Volume Imaging. *Cell* **159**, 896–910 (2014).

165. Klingberg, A. *et al.* Fully Automated Evaluation of Total Glomerular Number and Capillary Tuft Size in Nephritic Kidneys Using Lightsheet Microscopy. *J. Am. Soc. Nephrol.* **28**, 452–459 (2017).

166. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).

167. Tsanov, N. *et al.* smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* gkw784 (2016) doi:10.1093/nar/gkw784.

168. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).

169. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e21 (2019).

170. Alabert, C. *et al.* Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* **29**, 585–590 (2015).

171. Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-00837-3.

172. Chadwick, B. P. & Willard, H. F. Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc. Natl. Acad. Sci.* **101**, 17450–17455 (2004).

173. Grant, G. D., Kedziora, K. M., Limas, J. C., Cook, J. G. & Purvis, J. E. Accurate delineation of cell cycle phase transitions in living cells with PIP-FUCCI. *Cell Cycle* **17**, 2496–2516 (2018).

174. Sakaue-Sawano, A. *et al.* Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132**, 487–498 (2008).

175. Bharti, K. *et al.* A Regulatory Loop Involving PAX6, MITF, and WNT Signaling Controls Retinal Pigment Epithelium Development. *PLOS Genet.* **8**, e1002757 (2012).

176. Wheeler, M. T., Zarnegar, S. & McNally, E. M. ζ-Sarcoglycan, a novel component of the sarcoglycan complex, is reduced in muscular dystrophy. *Hum. Mol. Genet.* **11**, 2147–2154 (2002).

VITA

Marcus Woodworth lived most of his childhood in Armenia, Colombia, South America with his two parents, David and Margot, and two brothers, Davis and Lucas. As part of a missionary family, he grew up participating in preaching, music, and theatrical performances with Christian narratives. This continued, even after graduating from high school at the top of his class, as he spent a year and a half at a bible institute in Argentina, leading outreach drama teams to perform in parks, theaters, and churches around Buenos Aires during each weekend. He moved to California, where in 2016, he completed his B.S. in Chemistry with honors at the University of California, Irvine. Soon after that, he married the love of his life, Cara, with whom he now has two daughters with. As Marcus pursued his PhD degree in Chemistry, he was awarded a Ruth L. Kirschstein Pre-Doctoral National Research Service Award while working under the mentorship of Prof. Joshua Vaughan and Prof. Hao Yuan Kueh.