Theses and Dissertations   Graduate School

2024

Mathematical Analysis of Eukaryotic Pericentromere

Puranjan Ghimire
Virginia Commonwealth University

COPYRIGHT PAGE

Mathematical Analysis of Eukaryotic Pericentromere

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of

Philosophy at Virginia Commonwealth University.

By

Puranjan Ghimire
M.S. in Physics, 2017
Virginia Commonwealth University

Advisor

Richard Inho Joh, Ph.D.
Assistant Professor, Department of Physics

Committee Members

Jason Reed, Ph.D.,
Professor, Department of Physics

Daeha Joung
Assistant Professor and Associate Director of the Nanoscience & Nanotechnology PhD
Program, Department of Physics

Rebecca L. Heise, Ph.D.
Inez A. Caudill, Jr. Distinguished Professor and Chair, Department of Biomedical Engineering

Virginia Commonwealth University
Richmond, Virginia
August 2024

Acknowledgement

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DNA   **Deoxyribonucleic Acid**

RNA   **Ribonucleic Acid**

ncRNA   **Noncoding RNA**

lncRNA   **Long noncoding RNA**

RNAi   **RNA interference**

H3K9me   **Methylation at 9th lysine of 3rd histone protein**

H3K9ac   **Acetylation at 9th lysine of 3rd histone protein**

WT   **Wild Type**

ODE   **Ordinary Differential Equation**

DNAme   **DNA methylation**

QSSA   **Quasi Steady State Approximation**

M repeat   **Methylated repeat**

U repeat   **Unmethylated repeat**

CN   **Copy Number**

rUM   **Rate of conversion of U to M per minute**

HSAT   **Human Satellite Repeat**

Abstract

The centromere is an essential component of the chromosomal structures of eukaryotes, which is required for proper chromosomal segregation and stability. Centromere is often surrounded by regions called pericentromere, which are made up of multiple self-repeating DNA elements with very low transcriptional activities. These repeated sequences in pericentromere are called pericentromeric repeats. The organization of pericentromeric repeats varies across species from around 10 in fission yeast to thousands in higher eukaryotes like human. These pericentromeric repeating DNA sequences organize themselves within a nucleus in densely packed structures, called heterochromatin, where genes tend to be transcriptionally inactive or silenced. The heterochromatin assembly requires silencing, but how silencing works across various pericentromeric repeats with various sequences remains unanswered.

Although the sequences and length of these repeats vary from organism to organism, their presence across many organisms is a conserved feature of eukaryotes. This suggests that the repetitive nature of pericentromeric sequences is a functional feature of the centromere, which may be under selective pressure. In this dissertation, I will discuss mathematical models that I have developed to quantify gene silencing in fission yeast and human. My model, which takes copy number as an explicit parameter, predicts that the pericentromere is silenced only if there are many copies of repeats.

RNA interference (RNAi) is one of the known mechanisms that is conserved in many eukaryotes and is responsible for heterochromatin assembly in pericentromere. RNAi is the canonical pathway for the assembly and maintenance of pericentromeric gene silencing. My model predicts that pericentromere becomes bistable or desilenced if the copy number of repeats is reduced. This suggests that the copy number of pericentromeric repeats alone can determine the fate of heterochromatin silencing in fission yeast. Through sensitivity analysis, I identified parameters that favor bistability and desilencing. Stochastic simulation shows that faster cell division and noise favor the desilenced state. These results show the unexpected role of pericentromeric repeat copy number in gene silencing and provide a quantitative basis for how the copy number allows or protects repetitive and unique parts of the genome from heterochromatin silencing, respectively.

In addition, I developed a mathematical model for humans, where there are several redundant heterochromatin factors and processes for silencing. Pericentromeric repeats in higher eukaryotes are highly diverse even in one organism, and there are various repeats in pericentromere, different in each chromosome. Pericentromeric heterochromatin is often misregulated in human diseases, with the expansion of pericentromeric repeats in human solid cancers. Recent studies suggest that human satellite II repeats in cancer increase by the integration of RNA-derived cDNA by the process called reverse transcription. This indicates that the copy number of human satellites is highly dynamic with the transcriptional activity. In this thesis, I will discuss two types of mathematical models, the first one is ODE-based, second one is the "polymer chain model", to describe the dynamic nature of HSATII repeats in human pericentromere. Both models predict that HSATII repeat number expansion happens as a result of reduced methylation

at those repeats, with the ODE model also suggesting the possibility of the coexistence of higher and lower steady states of HSATII copy numbers. "Polymer chain model" enables us to understand the spatial structure of repeats, like the length of methylated repeat domain in different conditions.

In conclusion, my study highlights the important role of pericentromeric repeat copy numbers, a conserved chromosomal feature in eukaryotes. Future studies may reveal how different underlying silencing mechanisms contribute to various structures across species.

Vita

Puranjan Ghimire earned his bachelor's degree in physics from Tribhuvan University, Nepal, in 2009. He then obtained a master's degree in physics and applied physics from Virginia Commonwealth University in 2017. He has pursued a PhD in Nanoscience and Nanotechnology at VCU's physics department, focusing on the mathematical analysis of the pericentromeric region of eukaryotic organisms.

Chapter 1: Introduction

1.1. Chromatin

Chromatin is a complex of DNA and proteins that form the chromosomes in the cells of humans and other higher organisms. Key proteins, primarily histones, package a large amount of genomic DNA into a compact structure within the cell nucleus. Histones are essential for providing structural support to chromosomes. Each chromosome consists of a long DNA molecule that must fit into the cell nucleus, achieved by the DNA wrapping around histone complexes to form a more compact shape (Fig 1.1). Histones also play a critical role in regulating gene expression.(1)

The total length of DNA in a cell is about 5 to 6 feet, and it must be orderly packed into the nucleus, about 10% of cell volume (2). This is done by the DNA wrapping around histone proteins, forming a "beads on a string" structure known as nucleosomes (Fig 1.1), which are the basic repeating units of chromatin within the nucleus. In humans, approximately six feet of DNA must be compacted into a nucleus smaller than the diameter of a human hair, and nucleosomes are central to this process. A single nucleosome comprises about 150 base pairs of DNAs wrapped around a core of histone proteins. To form a chromosome, nucleosomes fold and condense the DNA further. These nucleosomes coil and gather to form fibrous material called chromatin (Fig 1.1). Chromatin fibers can unwind to allow DNA replication and transcription. (1)

1.2.   Histone Epigenetic modification

Histones are an essential group of proteins for the structural integrity of chromosomes as DNA wraps around histone protein complexes, creating a compact chromosome structure. The four types of histones—H2A, H2B, H3, and H4—form octamers, which, in turn, along with DNA, create nucleosomes. These nucleosomes coil into a spiral structure known as a solenoid, with additional H1 proteins linking each nucleosome to maintain the overall chromatin structure (Fig 1.1). Thus, histones are crucial for organizing and compacting the genome within a cell. (3)

Nucleosomes are the building blocks of chromosomes, and RNA and non-histone proteins can also bind to nucleosomes and DNA. All these factors form chromatin, the DNA protein complex that forms chromosomes. Chromatin exists in two states: euchromatin, which is open and accessible for transcription, and heterochromatin, which is densely packed and transcriptionally inactive. These chemical modifications can transform DNA between its active (euchromatin) and inactive (heterochromatin) forms (4). The "histone code" hypothesis posits that DNA transcription is primarily regulated by these post-translational modifications, allowing phenotypic changes without altering the genetic code, thereby controlling gene expression (5).

Each histone has an N terminus (amino end) and a C terminus (carboxy end), with tails protruding from the chromatin surface. These tails, comprising about 25-30% of the histone's mass, are sites for chemical modifications (3). Lysine amino acids in these tails are the primary sites for histone methylation and acetylation. Histone methylation, which involves adding a methyl group to a lysine, typically represses gene activity and supports the heterochromatin state (4). In contrast,

histone acetylation, which adds an acetyl group to a lysine, usually activates gene expression and supports the euchromatin state (4).
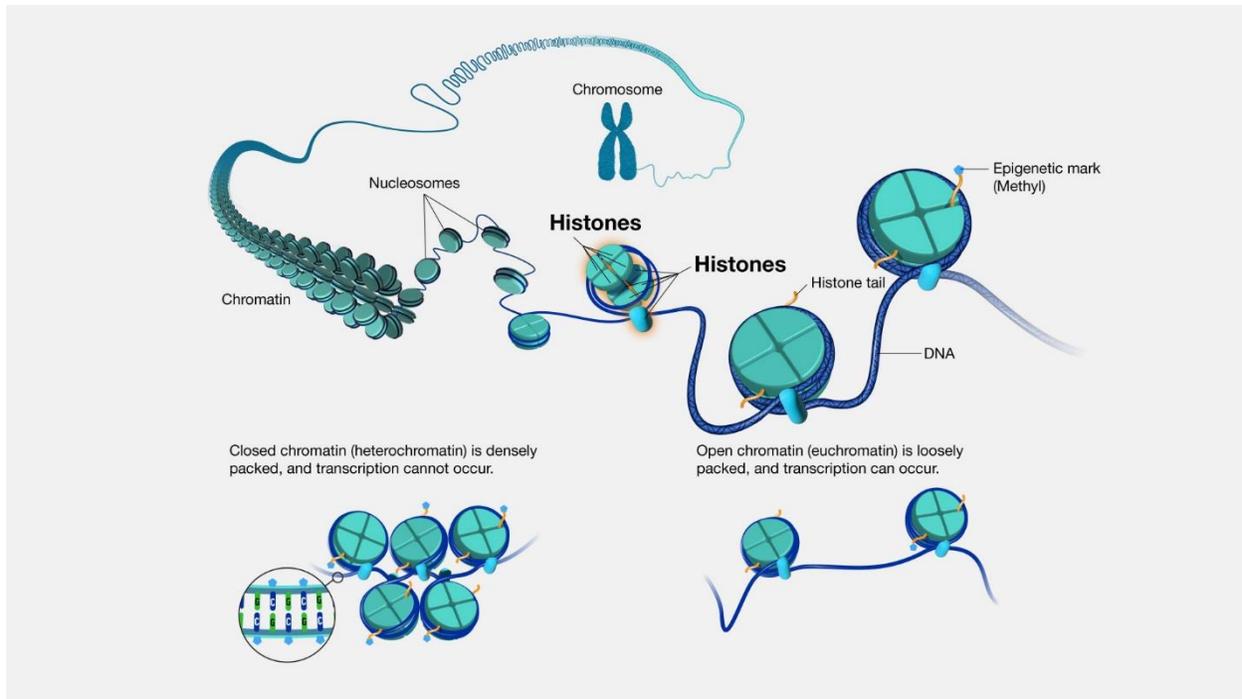
Fig 1.1 Schematic diagram of Chromatin, nucleosome, histone, euchromatin, and heterochromatin (1)

1.3.   Euchromatin and Heterochromatin

Chromosomes have two distinct regions based on their structure and function: euchromatin and heterochromatin. Heterochromatin is a highly condensed, gene-poor, and transcriptionally inactive part of the genome, while euchromatin is less condensed, gene-rich, and transcriptionally active (6). Nucleosome modifications are essential to differentiating between these regions. Heterochromatin is typically marked by hypoacetylation of histones, methylation of 9th lysine of

3rd histone (7–10), and, in some organisms, DNA methylation (11,12). On the other hand, euchromatin is characterized by acetylation of histones H3 and H4 and methylation of 4th lysine of 3rd histone (13–15). Heterochromatin is predominantly found in regions such as telomeres and pericentromeres, rich in repetitive DNA sequences (16–20). A characteristic feature of heterochromatin is its ability to self-sustain its dense structure and transcriptionally silent state in a site-specific manner across chromosomes throughout the cell cycle and across generations.

1.4.   DNA Epigenetic modification

DNA can also be chemically modified, and a methyl group (CH3) is added to a specific nucleotide sequence in methylation at the fifth position if cytosine in CpG dinucleotides (5mC). This DNA methylation also represses gene expression. (3)

Upstream of transcription start sites, within the promoter regions of DNA, there are areas known as "CpG islands," which are clusters of cytosine and guanine dinucleotides. When methylation occurs within promoter regions, it prevents transcription factors from recognizing and binding to the promoter sequence. As a result, DNA methylation at the promoter region silences the transcription of specific genes. (3)

1.5.   Centromere

Eukaryotic DNA, which is packaged into chromosomes, must be accurately replicated and segregated during cell division. This process hinges on a specialized region of a chromosome, known as centromeres, which ensures chromosomes are correctly segregated. The centromere recruits a proteinaceous macromolecular structure called kinetochore, which connects to the

microtubules of the mitotic and meiotic spindles. Consequently, centromeres and kinetochores play crucial roles in chromosome segregation. (21)

1.6.   Pericentromeric repeats

Surrounding a core centromere in eukaryotes is the pericentromeric region, or pericentromere. Pericentromeric regions consist of highly repetitive DNA sequences called pericentromeric repeats. In eukaryotic cells, these repeats are arranged in a dense chromatin structure called heterochromatin, which is crucial for proper chromatin segregation and maintaining genome integrity (22–24). The pericentromeric repeats vary widely in their copy number from 10s to over 1000, depending on the organism. Fission yeast, a unicellular eukaryote, contains 10-15 copies of dg and dh repeats across its three chromosomes (25). A DNA sequence in the human pericentromeric region can be categorized into two types based on the arrangement of repeating units: tandem repeats and interspersed repeats (26). Interspersed repeats, also known as transposons, include both DNA and RNA transposons (27). Tandem repeats refer to a sequence where basic repeating units are connected head-to-tail, forming an array (28). In the human centromere, tandem repeat encompasses various types such as α, β, γ-satellites, as well as satellites I, II, III (HSAT I, II, III), along with macro and mega satellites, which have variable number of sizes and lengths. For example, α-satellites are typically 171 bp long and cover a region spanning between 0.2 to 0.8 mega bases in the centromere. Another satellite repeat HSATII ranges from 23 to 220 base pairs in length and span between 11-70 kilobases in the pericentromere (26).

1.7.   Motivation for the study of eukaryotic pericentromere

Even though the length and sequence of pericentromeric repeats vary across organisms, their conserved presence (29,30) across many organisms suggests that there may be functional importance of the presence of multiple pericentromeric repeats. These repeats can change depending on the environmental condition, and the copy number of these repeats can increase or decrease (31,32). How these repeats contribute to the heterochromatin function and chromosomal stability/instability is still poorly understood. In this thesis, I will discuss how gene silencing and genomic instability depend on the copy number of repeats using the fission yeast and humans as model organisms.

The fission yeast, S. Pombe, is one of the most well-studied model organisms for pericentromeric heterochromatin assembly (33–36), with many pathways conserved in mammals. It possesses a conserved RNA interference pathway (RNAi) with RNA-dependent RNA polymerase (Rdp1) complex and Argonaute protein (Ago1), which recruits the only H3K9 methyl transferase Clr4 to pericentromeric regions along with chromatin remodeling complexes (37,38). Its reduced redundancy compared to higher eukaryotes and smaller copy number (10 to 15) makes fission yeast an ideal model organism for studying the function of pericentromeric repeats.

In mammalian cells, pericentromeres are also strongly repressed by H3K9me3 and DNA methylation. Pericentromeric noncoding RNAs (ncRNAs) often emerge in response to stress (39,40), indicating they might play a role in the adaptive process. In cancer, pericentromeric repeats (HSATII) are significantly upregulated in many epithelial and solid tumor cells, suggesting a widespread alteration in epigenetic heterochromatin silencing (32). Additionally, the copy

number of HSATII increases in cancers through reverse transcription and integration of satellite repeat RNAs, which is associated with poor survival among pancreatic patients (31).

In my thesis, I have developed mathematical models to investigate how the copy number of pericentromeric repeats influences gene silencing in two model organisms: fission yeast and humans. For fission yeast, my model incorporates the RNAi pathway for pericentromeric gene silencing, with repeat copy number explicitly included as a parameter. The model explores HSATII copy number expansion for humans through the reverse transcription of RNA and its integration into the pericentromeric region (31). The following chapters will present a detailed mathematical analysis and the predicted outcomes for fission yeast and human pericentromeric regions.

Chapter 2: Mathematical Model for the Role of Multiple Pericentromeric Repeats on Heterochromatin Assembly*[1]

2.1. Introduction

The eukaryotic centromere is comprised of a core region that is flanked by multiple repetitive DNA sequences, called pericentromeric repeats. Pericentromeric repeats in eukaryotic cells are organized in a dense chromatin structure known as heterochromatin which is vital for the establishment and maintenance of chromosomal stability. For example, pericentromeric heterochromatin is essential for accurate chromatin segregation and genome stability (22,23). Heterochromatin assembly is also required for gene silencing and repression of recombination among repeats (41). The de-repression of pericentromeric repeats in human cancers (e.g., BRCA1-mutated breast cancer (32,42), and tumor formation caused by the forced transcription of pericentromeric satellite RNAs in mice (43) suggest that the transcriptional silencing of pericentromeric repeats is also crucial for preventing tumorigenesis. Heterochromatin is often characterized by DNA methylation, repressive histone modifications, and hypo-acetylation of histones. One of the hallmarks of heterochromatin formation in eukaryotic cells is the methylation of lysine 9 (K9) of histone 3 (H3) (9,44,45). This chromatin mark and the associated enzymes are

---

conserved from a unicellular eukaryote, like the fission yeast, to a complex mammal, like a human (46).

In the fission yeast, there is one histone methyltransferase for H3K9, Clr4, which facilitates H3K9 mono- (me1), di- (me2), and tri- (me3) methylation. The nucleation of pericentromeric heterochromatin and its spreading by various chromatin-dependent mechanisms is extensively studied in fission yeast where the RNA interference (RNAi) pathway is predominantly responsible for pericentromeric heterochromatin assembly (35,47–49). RNAi recruits Clr4 via the RNA-induced transcriptional silencing (RITS) complex physically interacting with Clr4 (37,38,50,51). RITS is a small interfering RNA (siRNA)-containing effector complex, consisting of three proteins, the sole Argonaute homolog (Ago1), the GW domain protein, Tas3, and a chromodomain-containing protein, Chp1. RITS interacts with H3K9me via chromodomain protein Chp1 and nascent RNA via Ago1 (35,47–49). In addition to the interaction with Clr4, RITS can also interact with RNA-dependent RNA polymerase complex (RDRC) (37,52), which together with Dicer (Dcr1) amplifies and processes nascent long non-coding RNAs (lncRNAs) into siRNAs creating a positive feedforward loop, where siRNAs increase H3K9me, and H3K9me stimulate the subsequent synthesis of siRNAs (35).

Although the silencing of pericentromeric repeats is a conserved feature of eukaryotes, there is a considerable variation among those repeats in terms of their lengths, constituting sequences, and organization (53). However, the presence of multiple repeats, which flank the central core is a conserved feature (54). This suggests that the repetitive nature of pericentromeres may be an evolutionary conserved feature with a functional consequence. Here, we hypothesize that the copy number of the repeats is critical for pericentromeric gene silencing. Experimental

manipulation and quantification of copy number-dependent gene silencing are difficult because of the presence of identical repeats in the pericentromeric region. In this work, we developed a mathematical model for copy-number-dependent gene silencing in pericentromeric repeats, and our simple model supports that multiple copies are required for gene silencing. We have analyzed the robustness of model parameters through sensitivity analysis and stochastic simulation using the Gillespie algorithm. Our results suggest that the copy number of pericentromeric repeats alone can significantly impact gene silencing, which may affect genomic stability and human diseases.

2.2.   Results

2.2.1.   Mathematical modeling of gene silencing at pericentromeric repeats

To investigate how the copy number of pericentromeric repeats affects gene silencing, we developed a mathematical model incorporating lncRNA, siRNA, and H3K9 methylation as key parameters involved in the establishment of heterochromatin. A schematic representation depicting a simplified molecular pathway used for mathematical modeling is shown in Fig 2.1, and the parameters are shown in Table S2.2. Briefly, the model tracks the level of RNA (pericentromeric lncRNA), siRNA, and H3K9 methylation which provide the specificity needed for the recruitment of downstream complexes. These molecules also create H3K9me and siRNA feedforward loops, critical for the establishment of heterochromatin at pericentromeres (See 2.4 Materials and Methods).
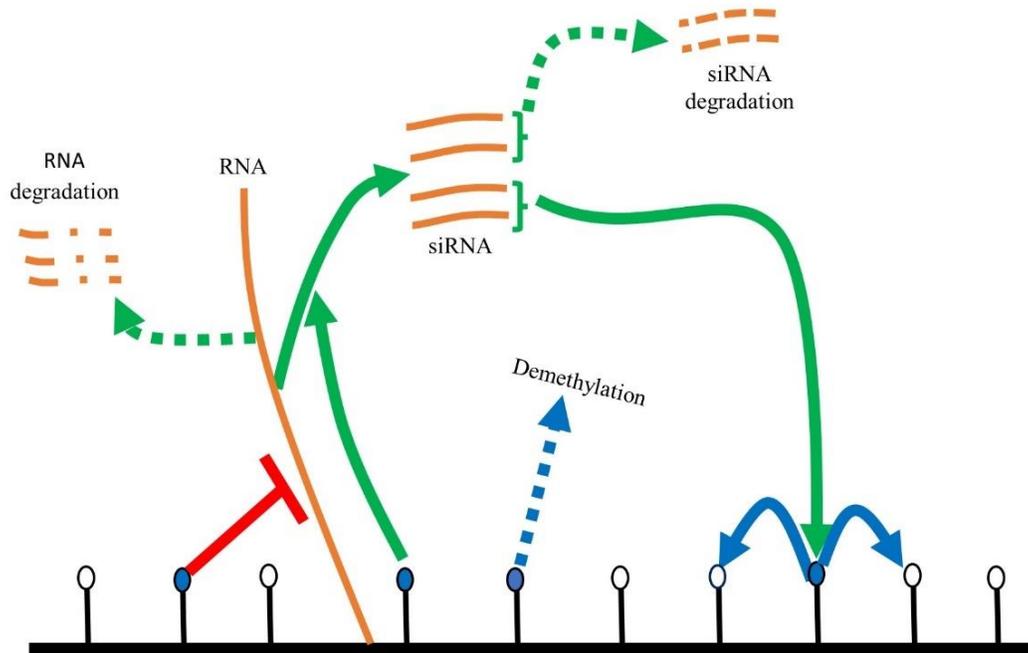
Fig 2.1. Schematic diagram for mathematical modeling of siRNA-mediated gene silencing. Blue and white circles represent methylated and unmethylated histones, respectively. siRNA can recognize nascent lncRNAs, which can lead to H3K9me, which can spread to nearby histones. H3K9me in turn represses the transcription, and RNA is either degraded or turned into siRNA.

RNA: According to the nascent transcript model, lncRNAs provide a platform to facilitate siRNA-mediated H3K9me with RITS and Clr4 complexes (37,38,51). Their transcription and recognition by RITS complex nucleate heterochromatin at centromeres. The lifetime of lncRNA was assumed to be 50min (Table S2.2), based on the experimental data in yeast (55,56). RNA can be degraded or turned into siRNAs in a methylation-dependent manner.

siRNA and H3K9me amplification loops: RITS contains Chp1, a chromodomain protein, capable of binding to H3K9me. This tethers the RITS-RDRC-Dcr1 siRNA biogenesis pathway

(assembled on heterochromatin lncRNAs) to heterochromatin via siRNA base-paring interactions with the complementary nascent pericentromeric lncRNAs. Also, Clr4 physically interacts with RITS (50) and this interaction is critical for targeting Clr4 to heterochromatin. Once at heterochromatin, Clr4-mediated H3K9 methylation can spread in cis, and iterative cycles of siRNA generation, and RITS-Clr4-mediated H3K9 methylation amplify the H3K9me and siRNA signals at heterochromatin. H3K9me, in turn, represses the transcription (57), but even a small number of RNA can be efficiently turned into siRNA.

Repeat copy number: There are over 10 copies of pericentromeric repeats in fission yeast. At least one repeat is present on each side of three S. pombe centromeres. The number of repeats varies across chromosomes as chromosome 3 has multiple copies on both sides of the centromere (58). Each copy can have its own transcriptional activity independent of one another. However, Ago1-bound siRNA can target nascent transcripts across chromosomes and exogenous loci (59), and repeats show similar profiles in gene silencing and H3K9 methylation across chromosomes. The copy number is an explicit parameter in our mathematical modeling. Each copy of the pericentromeric repeat can be transcribed independently, and without any feedback, more copies promote more transcription linearly dependent on the copy number.

Histone deacetylase activity. Although our model focuses on the H3K9 methylation, histone deacetylases (HDACs) are critical in the regulation of pericentromere silencing. Sir2 and Clr3 are HDACs in fission yeast, whose deletion leads to a partial loss of silencing (59–65). The main substrate of Clr3 is H3K14ac, but its enzymatic activity is required for heterochromatin silencing via H3K9me (66,67). To incorporate this process, we developed an alternative model (See 2.4 Materials and Methods) to track three H3K9 states: 1) acetylated, 2) unmodified, and 3)

12

methylated. We assume that the new histones are acetylated (68), and acetylated histones need to be unmodified prior to possible methylation (69).

2.2.2. Bistability between silenced and desilenced states is possible at a small copy number

To understand the role of copy number on gene silencing, we first analyzed solutions of the ordinary differential equations (ODEs) (see Eqs (2.1-2.3)) while varying the copy numbers of pericentromeric repeats (CN). Our model assumes that repeats are transcribed and their copy number, an explicit parameter in our model, enhances transcription linearly for a given level of methylation. Fig 2.2A-2.2B depicts the dynamics of methylation from multiple initial conditions for different CNs, and Fig S2.1 shows the results of RNA, siRNA, and H3K9me. Specifically, there can be one (CN = 15) or two (CN = 5) stable steady states depending on the copy number. As described above, H3K9me modulates the lncRNA transcription via the RNAi machinery. If the copy number is low, the amount of siRNA is low, and the system remains desilenced (Fig 2.2A and Fig S2.1). If the copy number is high, enough RNA and siRNA can be generated, which in turn silences the repeats through methylation (Fig 2.2B and Fig S2.1). If H3K9me is low, RNA degradation proceeds via the conventional mechanism with few siRNAs, whereas if H3K9me is high, RNA is predominantly processed into siRNA.
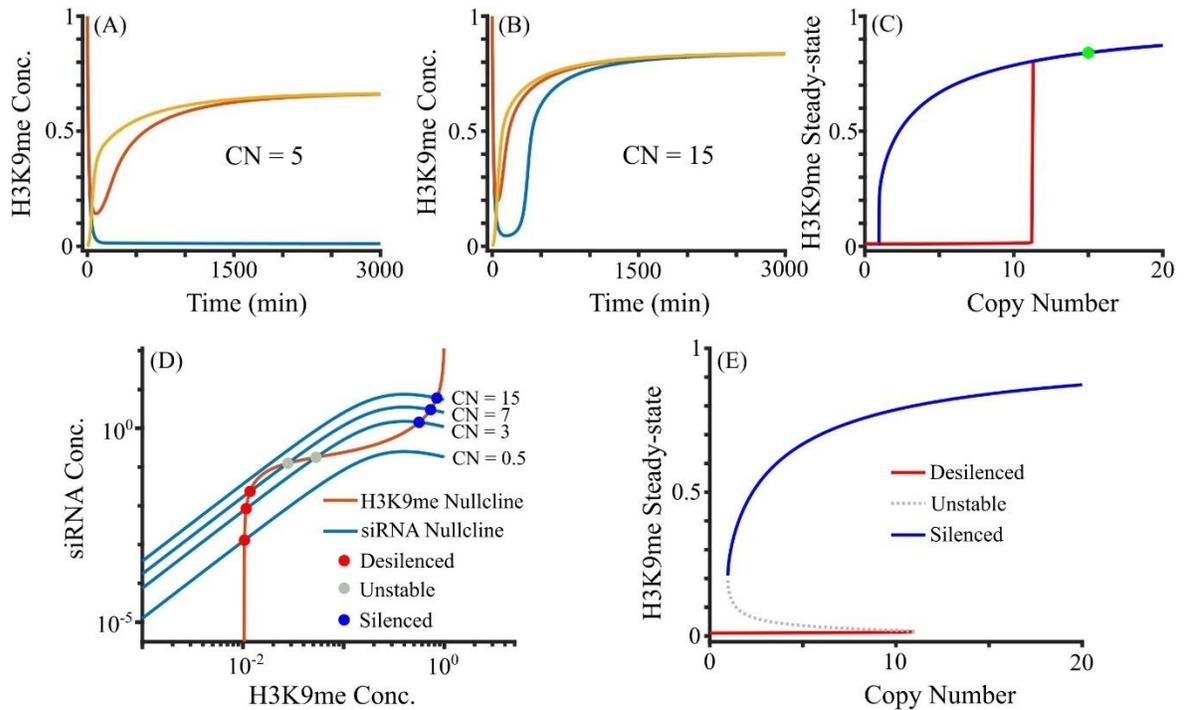
Fig 2.2. Change in the steady-state behavior as a function of repeat copy number. (A-B) Solutions of H3K9me for different CN_s in the full model. Colors indicate the use of different initial conditions. (C) Steady-state H3K9me as a function of copy number. The system is bistable (silenced and desilenced) or monostable. Wild Type (WT) in fission yeast is shown at CN = 15 by a green star. (D-E) Quasi-steady state approximation. (D) Nullclines of siRNA and H3K9me using QSSA. Representative siRNA nullclines at indicated CN_s are shown. (E) Bifurcation diagram showing the bistability with both silenced and desilenced states for smaller copy numbers and monostability at high copy numbers.

To assess the effect of copy number change on gene silencing, we systematically varied the copy number and quantified the number of stable steady states. Fig 2.2C depicts the steady states of methylation for various CNs, where the blue and red curves represent the higher and lower steady states of H3K9me, respectively. Fig. 2.2C shows that for copy numbers up to around 11, the system displays both silenced and desilenced steady states. However, for CN greater than 11, there is only one steady state, which is silenced. Fig S2.2 shows the dynamics from multiple initial conditions as CN increases, where the fraction of desilenced states decreases until the system

14

becomes monostably silenced. The known copy number of lncRNA dg/dh in S. pombe wild-type (WT) cells is around 15, which is represented by the green star in Fig 2.2C, and the pericentromere of WT cells are silenced. Our results suggest that the heterochromatin-mediated silencing is nonlinearly dependent on the repeat copy number, which can change qualitatively from monostability to bistability. Overall, our model demonstrates that the repeat copy number is a critical dynamical parameter of RNAi-mediated silencing at the fission yeast centromeres.

2.2.3. A small change in copy number leads to a big change in steady states of H3K9me

The turnover rate of siRNA is typically lower than RNA due to binding by Argonaute proteins (70,71), which results in slower changes in siRNA than RNA. Because RNA reaches steady-state faster than siRNA, to test how CN modulates the steady states of the system, we used a quasi-steady-state approximation (QSSA), in which we assumed that RNA concentration reaches equilibrium at any given siRNA and H3K9me level. The time evolution of siRNA and H3K9me in the full model is similar to that in the QSSA (Fig S2.3). QSSA reduces three variables in the full model into two variables, which allowed us to plot nullclines in 2-dimensions and thus find the steady-states. Fig 2.2D shows the nullclines from Eqs (2.6-2.7), which represent curves where siRNA and H3K9me concentrations are not changing, respectively. The intersections of these two nullclines are the steady-states, which include both unstable and stable steady-states. The number of steady states can change between 1 and 3, with 1 or 2 stable steady states. The siRNA nullcline is CN-dependent as the nullcline moves up with high CN, and such changes control the number of steady states of the system. The H3K9me nullcline is CN-independent but the nonlinear shape of this nullcline allows it to have one to three intersections. The intersection point with low methylation (< 0.2) is a desilenced steady state, and the intersection point with high methylation

(~ 1) is a silenced steady state. If there are three steady states, the middle steady state is unstable. For copy number 15, the siRNA nullcline intersects the H3K9me nullcline at the silenced state only, and this transition from bistability to monostability is due to the saddle-node bifurcation (72).

Fig 2.2E shows the bifurcation diagram, revealing the stable and unstable steady states as a function of the copy number. The system is bistable for lower copy numbers and becomes silenced via the saddle node around copy number 11. Between copy numbers 11 and 12, the system drastically changes its steady-state dynamics between a single silenced state and the coexistence of both silenced and desilenced states. A big change in steady states of H3K9me with a small change in copy number suggests that the copy number is an important factor for the gene silencing of cells, especially if cells were initially desilenced.

2.2.4. Sensitivity Analysis

To investigate the robustness of silenced and desilenced states in a cell, we systematically varied parameters and tracked how the behavior of the system changes. Most of the parameters in our model are related to the proteins participating in the RNAi pathway of pericentromeric gene silencing. To identify the key parameters in copy number-dependent gene silencing, we tracked steady-state changes when a parameter was varied within 25% of its reference value while keeping all other parameters fixed at their reference values. In all cases, such changes in the parameters can lead to 1) bistability between silencing and desilencing and 2) monostability of silencing or desilencing. Several factors in our model can repress methylation like demethylation and RNA degradation, while other factors like methylation by siRNA, siRNA biogenesis, and spreading of methylation enhance methylation. Fig 2.3A-2.3D shows if the solution is silenced or bistable at

16

the indicated copy number and four representative parameters (see Fig S2.4 for other parameters). Enhancing anti-silencing processes can destabilize silencing and lead to switching from silencing to bistability. For example, a high RNA degradation (δ_1) and demethylation rate (δ_3) facilitate switching from silencing to bistability (Fig 2.3A-2.3B). On the other hand, enhancing H3K9me-associated processes can stabilize silencing. Fig 2.3C-2.3D shows the changes in the steady state behaviors where a high methylation rate (ϵ) and methylation spreading rate (ϕ) favor silencing.

To test how each parameter affects system behavior, we also analyzed how the nullclines change from QSSA when one of the parameters is changed. Fig 2.4 depicts the changes in nullclines due to variations of different parameters. Our model revealed that several parameters affect the H3K9me nullclines. Demethylation rate (δ_3, the activity of demethylases like Epe1), methylation spreading rate (ϕ chromodomain activity of Clr4 and other chromodomain proteins), affect the H3K9me nullcline (Fig 2.4A), and basal methylation rate (ζ, the RITS-independent activity of Clr4) affect the desilenced end of H3K9me nullcline (Fig 2.4B). These parameters are critical for bistability, and this is in agreement with the experimental observation that the loss of Clr4 chromodomain, which is responsible for binding to H3K9me and equivalent to the reduction in ϕ, leads to variegated bistable heterochromatin establishment (73). Change in methylation rate (ϵ, catalytic activity of Clr4) shifts the H3K9me nullcline (Fig 2.4C), and low ϵ can lead to a bistable or monostably desilenced system. On the other hand, transcription rate (α, RNAP activity), and siRNA degradation rate (δ_2, Eri1 activity) affect the siRNA nullcline, which primarily affect the desilenced state similar to changing CN (Fig 2.4D). siRNA biogenesis rate (γ, Dcr1 activity) and RNA degradation rate (δ_1, the activity of ribonucleases like Dhp1) also affect the siRNA nullcline (Fig 2.4E), and H3K9me can increase with a mutation in exoribonuclease

17

Dhp1 (74). Hill coefficient (ρ_2) and half-maximum methyl concentration (κ_2) can modulate the slope of the siRNA nullcline, which also controls the number of steady states (Fig 2.4F-2.4G), and this parameter determines the cooperativity of H3K9me-dependent transcriptional gene silencing. Hill coefficient (ρ_1) and half-maximum methyl concentration (κ_1) slightly affects the siRNA nullcline leading to almost no change in the stability of the system (Fig 2.4H-2.4I), whereas ρ_3 and κ_3 affect the middle section of the siRNA nullcline leading to bistability from monostable state (Fig 2.4J-2.4K). A wide range of Hill coefficients supports bistability (see Fig S2.5). These results suggest that the bistability is not a unique behavior of parameter values, and the system can exhibit 1) bistability between desilenced and silenced states, 2) monostability of the desilenced state, and 3) monostability of the silenced state.
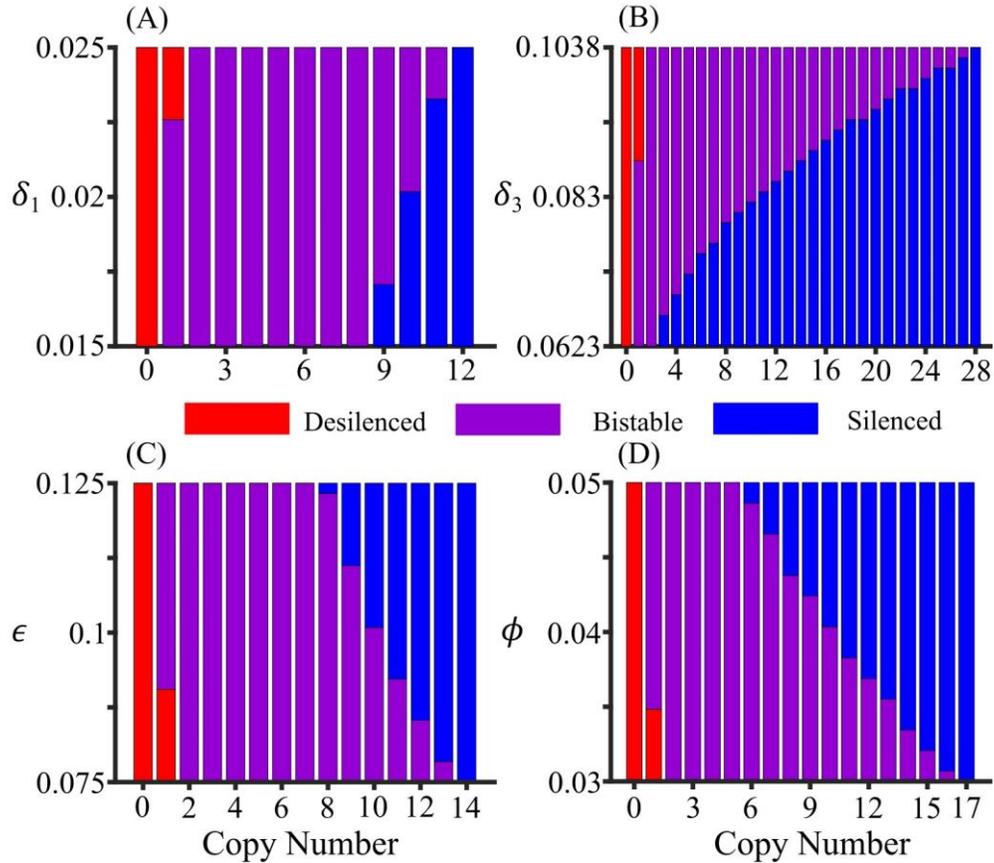
Fig 2.3. Parameter sensitivity analyses within 25% variation on default values of a specific parameter. Colors indicate whether the system is silenced or bistable at the indicated parameter value and copy number. (A) A large RNA degradation rate (δ_1) favors bistability or desilenced state for all copy numbers. All ranges of δ_1 favor bistable state for copy numbers between 2 and 8, favor silenced state for copy number 12 and above, and the range of δ_1 favoring a silenced state increases with an increase in copy number. (B) Like the RNA degradation rate, a higher value of demethylation rate (δ_3) favors bistability or desilenced state. All ranges of δ_3 favor bistable state for copy number 2, favor silenced state for copy number 28 and above, and the range of δ_3 favoring a silenced state increases with the increase in copy number. (C) A small methylation rate (ϵ) favors bistability or desilenced state. All ranges of ϵ favor the bistable state for copy number between 2 and 7, favor the silenced state for copy number 14 and above, and the range of ϵ favoring the bistable state decreases with the increase in copy number. (D) Like methylation rate, a lower value of methylation spreading rate (ϕ) favors bistability or desilenced state. All ranges of ϕ favor the bistable state for copy number 2 to 16, favor the silenced state for copy number 17 and above, and the range of ϕ favoring the bistable state decreases with copy number.
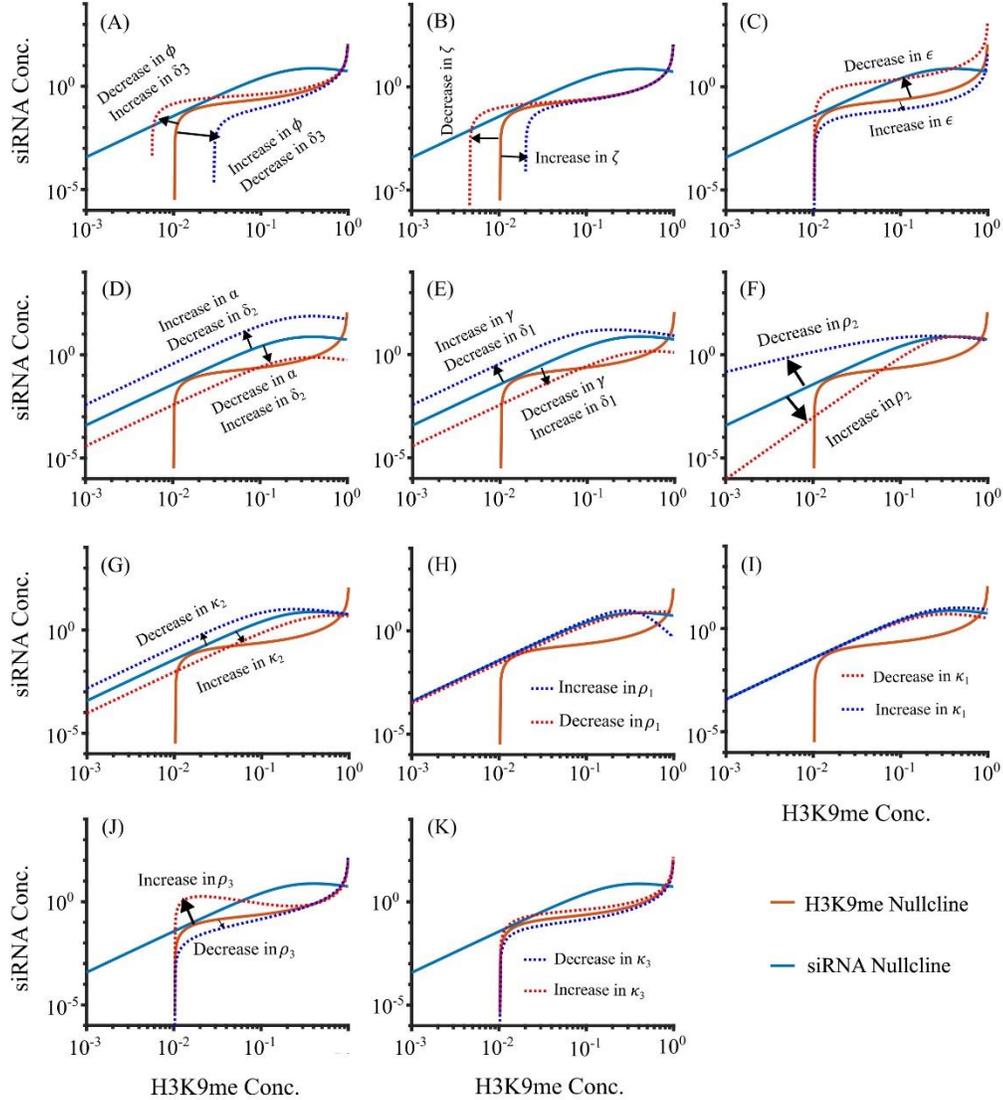
Fig 2.4. Changes in the nullclines as a result of changing parameters using QSSA for Copy number 15. (A) Parameters δ_3 and ϕ affect the stability of the desilenced state by modulating the tail and middle section of the H3K9me nullcline. (B) ζ shifts the tail of the H3K9me nullcline. (C) ϵ modulates the middle and head section of the H3K9me nullcline. (D) α, and δ_2 can shift the siRNA nullcline similar to changing CN. (E) γ and δ_1 shifts the siRNA nullcline. (F) ρ_2 changes the slope of the tail of the siRNA nullcline which affects the stability. (G) κ_2 shifts the tail of the siRNA nullclines. (H) ρ_1 bends the head of the siRNA nullcline with negligible effect on stability. (I) κ_1 slightly bends the head of the siRNA nullcline with no effect on the number of steady states. (J) ρ_3 distorts the shape of the middle section of the H3K9me nullcline affecting the steady states. (K) κ_3 modulates the middle section of the H3K9me nullcline. $\alpha =$ Transcription Rate. $\epsilon =$ methylation rate. $\phi =$ methylation spreading rate. δ_1, δ_2, and $\delta_3 =$ RNA degradation rate, siRNA degradation rate, and demethylation rate respectively. $\gamma =$ siRNA biogenesis rate. $\zeta =$ Basal methylation rate. ρ_1, ρ_2, and $\rho_3 =$ Hill coefficient for transcription, siRNA biogenesis, and methylation by siRNA respectively. κ_1, κ_2, and $\kappa_3 =$ Half maximum methylation for transcription, siRNA biogenesis, and methylation by siRNA respectively.

20

2.2.5. The model supports experimental results

To test if our model can recapitulate experimental observations, we changed our parameter values to match known genetic backgrounds. (Fig 2.5A-2.5B) shows the bifurcation diagram after setting various RNAi-associated parameters to zero. For example, spreading defect ($\phi = 0$) leads to a bistable state while RNA and siRNA degradation defect favors silencing. Our model predicts that at all copy numbers, the system remains silenced if the demethylation rate is zero and desilenced if the methylation (ϵ) and siRNA biogenesis (γ) rates are zero. When the RNA degradation rate (δ_1) is zero, the system is monostable and silenced for a copy number greater than 1. If the siRNA degradation rate (δ_2) is zero, the silenced state is strongly favored for higher copy numbers so that bistability can be observed only for small copy numbers less than 2 (Fig 2.5B). On the other hand, the system is bistable for copy numbers greater than 2 when the methylation spreading rate is zero (Fig 2.5B).

Fig 2.5C-2.5D shows the steady-state H3K9me and RNA concentrations, respectively, for copy number 15 that would correspond to various gene deletions. Our model recapitulates several experimental results. Spreading defect ($\phi = 0$) corresponds to the loss of Clr4 chromodomain or other chromodomain proteins, which leads to variegated bistable heterochromatin establishment (73). Consistent with previous reports our model predicts that loss of H3K9me by, for example, clr4 deletion (equivalent to $\epsilon = 0$) leads to a loss in centromeric silencing (50,51). Moreover, our model correctly predicts that loss of siRNA biogenesis pathway, e.g. by deletion of ago1, chp1, tas3, dcr1, etc. ($\gamma = 0$), leads to a loss of centromeric silencing (37,38,51,75). siRNA degradation defect, $\delta_2 = 0$, is similar to Eri1 deletion, which stabilizes heterochromatin

21

(76). RNA degradation (δ_1) and demethylation (δ_3) defect can be simulated computationally, but these are processes which are not specific to heterochromatin lncRNAs and H3K9me. For example, Epe1 is the primary known demethylase for H3K9me, but its activity affects not only H3K9me but also H3K36me.

Other studies have reported that the silencing in heterochromatin is variegated, suggesting that the system might be bistable. For example, two distinct outcomes of silenced and desilenced populations were observed with the deletion of histone deacetylase Sir2 and HDAC-associated Clr5 (47,77,78). In addition, transient overexpression of Swi6 can alter the epigenetic imprint and switch the desilenced state into the silenced state (79). These observations suggest that our quantitative model can recapitulate several experimental studies on heterochromatin gene silencing and bistability may occur in a wide range of genetic backgrounds.
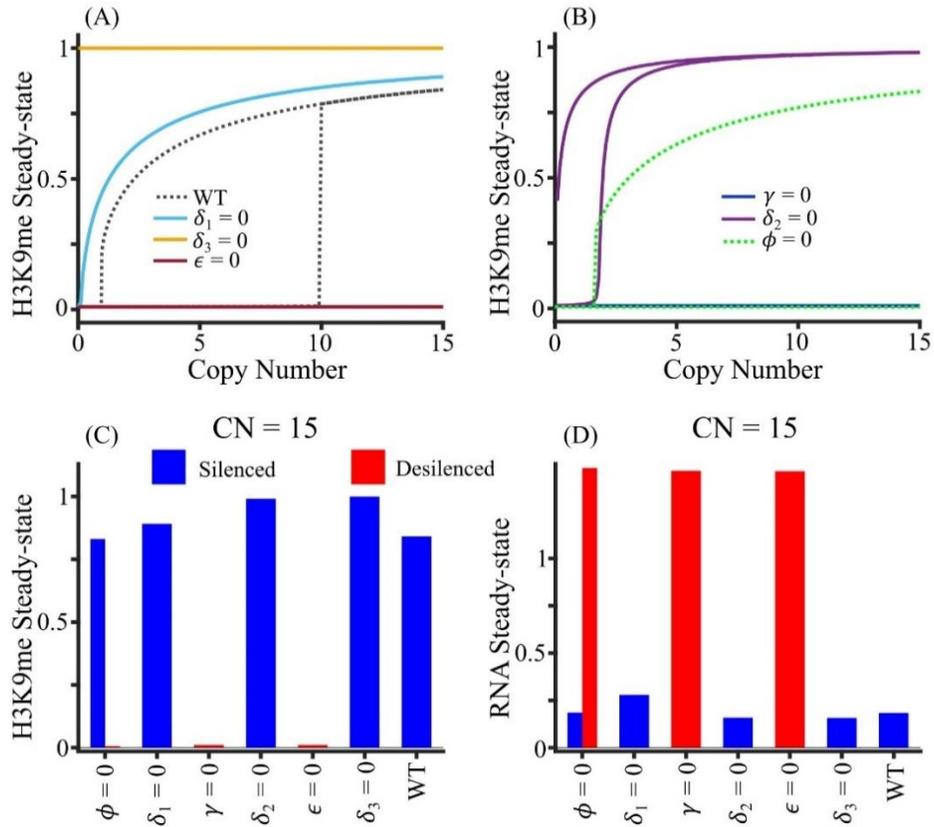
Fig 2.5. Comparison of model prediction for deletion experiments. (A) Bifurcation diagram corresponding to the deletion of the gene encoding RNAi protein, quantified by setting an associated parameter to zero, viz., RNA degradation rate (δ_1), demethylation rate (δ_3 by epe1) and methylation rat (ϵ by clr4). (B) Bifurcation diagrams of siRNA biogenesis (γ by ago1/rdp1/dcr1), siRNA degradation (δ_2 by eri1) and methylation spreading (ϕ by clr4 and other chromodomain proteins) deletions. (C-D) The steady-state concentration of (C) H3K9me and (D) RNA for copy number 15 for deletion of heterochromatin factors. The blue bar represents the silenced state, and the red bar represents the desilenced state.

2.2.6. Stochastic Simulation

ODE solutions represent the average behavior, so we simulated our system stochastically using the Gillespie algorithm (80,81) (See 2.4 Materials and Methods). The maximum number of H3K9me sites was normalized to nC, where n is the average number of H3K9me sites per copy

and C is the copy number, and RNA and siRNA molecules are rescaled 100-fold to represent individual transcripts. Fig 2.6 shows the typical dynamics of the system for copy number 15. To minimize the effect of the initial condition, data were sampled from 4000 to 10000 minutes, and the earlier part was excluded as a burn-in period. Cell division can act as a 2-fold dilution and noise from external fluctuations can affect the abundance of RNA, siRNA, and H3K9me, which is proportional to the amount of RNA, siRNA, and H3K9me in the system. Assuming no cell division and no noise in the system, the system remains silenced at CN = 15 (Fig S2.6A). In a realistic scenario, we considered two additional variations to our model with external noise and cell division. Fig 2.6A and Fig S2.6E-S2.6F show the switching between the silenced and desilenced states where fixed cell division time and noise were incorporated in the simulation.

At each cell division, the number of RNA, siRNA molecules, and H3K9me sites are reduced by half, which Fig S2.6B depicts. Since H3K9me goes down by 50% at each cell division, faster cell division can destabilize silencing which is shown in Fig S2.6. Fraction of silencing at a fixed noise rate (Fig 2.6B) increases with the increase in cell division time, and Fig S2.7A-S2.7B show the distribution of H3K9me over time at different cell divisions and fixed noise rate. With the increase in cell division (growth) rate, distribution is favored toward silencing. Noise can facilitate switching from the desilenced state to the silenced state and vice-versa at different cell division rates (Fig S2.6), but the fraction of silencing decreases with an increase in noise at fixed cell division (Fig 2.6C). Our simulation with different growth rates is in agreement with experimental studies. Non- or slow-growing cells exhibit the spreading of H3K9me into euchromatic loci (82,83), suggesting that heterochromatin becomes more stable with a low cell division rate. Our results suggest that the cell division rate plays a critical role in pericentromere

silencing, which is a physiologically controllable parameter, and its interplay with cellular noise

determines the time evolution of silencing in living cells.
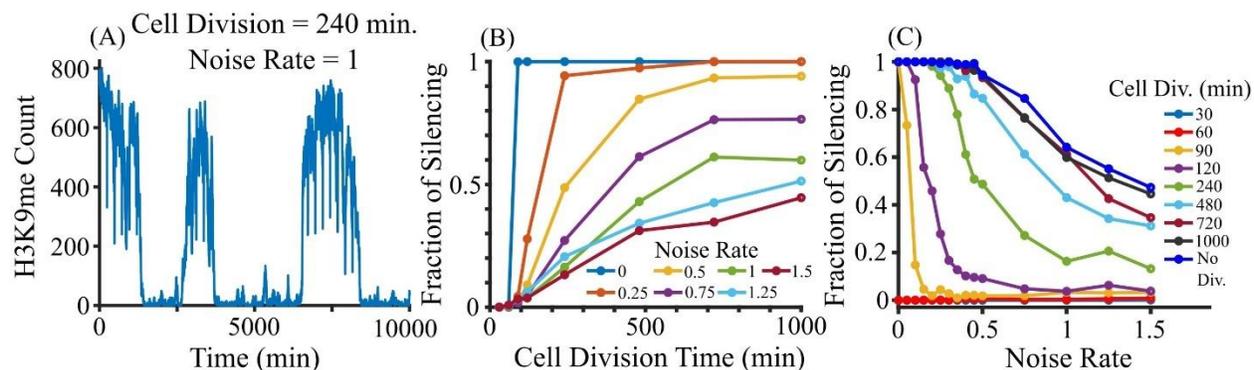
Fig 2.6. Stochastic simulation for copy number 15 using Gillespie algorithm. (A) A stochastic
trajectory of a system of cells with an average cell division time of 240 min, and noise rate of 1. (B) Fraction
of silencing increases with the increase in cell division for a fixed noise rate. (C) The fraction of silencing
decreases with an increase in noise rate for a fixed cell division.

2.2.7. Alternative model with three-state histone modifications

The heterochromatin assembly not only depends on the machinery for histone methylation

but also on histone deacetylases (69,84). Sir2 and Clr3 are known HDACs associated with

heterochromatin assembly. To test the role of HDACs and how HDAC activity shapes the copy-

number-dependent methylation within pericentromere, we also developed an alternative model

with three histone states: acetylated, unmodified, and methylated. Our alternative five-variable

model (with RNA and siRNA) yields result consistent with those of our primary model (See 2.4

Materials and Methods). Fig S2.8 shows the ODE solutions for this alternative model,

demonstrating agreement with our main model's findings. Notably, the existence of two steady

states at copy number 5 (Fig S2.8A-S2.8E) and a single high steady state at copy number 15 (Fig S2.8F-S2.8J) persists in the alternative model. Fig S2.9 illustrates the steady-state levels of H3K9me as a function of copy number, and in line with the main model's result, Fig S2.9A indicates that WT cells are silenced. With a lower deacetylation rate, desilencing is possible consistent with previous experimental data (Fig S2.9B) (65), and methylation is completely lost in the absence of any HDAC activities (Fig S2.9C). These indicate that the HDAC activity is another key component of pericentromeric silencing, and the deacetylation rate can also affect the copy-number-dependent pericentromeric repeat silencing.

2.3.   Discussion

Here we developed a quantitative model for gene silencing and H3K9me using the fission yeast centromere as a model. This region is one of the best-studied models for heterochromatin formation with many experimental studies testing the function of various associated chromatin factors. We used the copy number of repeats as an explicit model parameter and showed that the copy number itself can be a critical model parameter that can alter gene silencing. Depending on the copy number, cells can be bistable, with silenced or desilenced pericentromere, which was shown in certain genetic backgrounds experimentally (47,74,78). Typically, bistability requires nonlinear interactions through various feedback (85) by nonlinear methylation-dependent transcription and siRNA biogenesis in our model. Although we showed how the variation in each parameter affects the steady-state behaviors, the full model in the cell involves many other proteins whose functions are not well known. We assumed that each copy acts independently, but it may be possible that the chromatin state of one repeat preferentially affects the nearest neighbors. In

addition, the boundary elements of heterochromatin can affect each repeat in a distant-dependent manner (86,87). 3D compaction of the genome may further limit enzyme accessibility which facilitates the formation of stable chromatin domains (88), in this case forming a block of silenced repeats.

Our results suggest that RNAi may operate with a threshold copy number for the silencing and target genomic regions with multiple repeats like the pericentromeric repeats, and this leaves the euchromatic loci unaffected by this mechanism. Although few known viruses can infect yeast species, one of the functions of RNAi is a defense against viruses or transposable elements (89,90). RNAi-mediated copy-number-dependent silencing would be effective against viruses or transposable elements if they generate multiple copies. Therefore, unlike sequence-dependent defense mechanisms like CRISPR (91), RNAi-based silencing might work as a natural sensor and a sequence-independent defense system against foreign genetic elements of multiple copies. The threshold copy number and systemic bistability often require strong cooperativity, which in the case of fission yeast is incorporated by the H3K9me-mediated transcriptional gene silencing. There may exist a threshold copy number with weaker cooperativity like RNAi against multiple copies of dsRNA fragments in the absence of chromatin-based feedback. On the other hand, such a threshold copy number also implies that RNAi-mediated silencing is a mechanism to protect the unique euchromatin part of the genome against silencing. We also demonstrate that HDAC activity is required for this silencing, which may be another tunable feature in addition to RNA-mediated silencing.

Our findings underscore the significance of repeat copy numbers in heterochromatin. While manipulating pericentromeric repeats experimentally poses challenges, a few studies have

successfully demonstrated significant alterations in the structure of pericentromeres. Insertion of additional repeat copies with dg-containing minichromosome can lead to the enrichment of H3K9me2 in the exogenous dg copies even in the absence of some heterochromatin factors (59). A recent study showed that cells with fewer dg/dh copies can maintain H3K9me, but genomic instability increases with the loss of pericentromeric repeat copies, suggesting that heterochromatin function may depend on the copy number (92). Our models may be applied to other repeat elements in the genome such as telomeres and rDNA clusters as they may undergo copy-number-dependent chromatin-mediated silencing with additional silencing machineries. Specifically, rDNA clusters are up to 50% different in size from isogenic cells (93), and the disruption of the telomere cap complex can lead to rearrangements among telomeres and rDNAs (94). Such variation in rDNA repeats is due to a high recombination rate, which can be repressed by H3K9me (95,96). Several studies in other organisms have shown that heterochromatin repeats like rDNA may act as a sensor to monitor genome instability and induce senescence (97,98). Unlike pericentromere, cells need some copies of active rDNA, and a dynamic interplay between silencing and recombination can alter the rDNA cluster copy number in adaptation to a fluctuating environment.

The copy number variation within heterochromatin is found in many organisms. In humans, the satellite DNA is organized in long arrays in pericentromeric heterochromatin, and the copy number increases with age (99). Additionally, human satellite II can massively expand in solid tumors, and this progressive elongation of pericentromeric regions is specific to tumors (31,32). Duplication of pericentromeric repeats was also observed in the laboratory mouse, suggesting that such expansion is not restricted to primates (100). Fly genome also shows a

considerable variation in satellite DNA abundance (101) where the change in abundance is correlated in various isolated populations. Therefore, pericentromeric repeats may be highly dynamic and functionally important for cellular homeostasis. Our quantitative model shows that such copy number variation among transcriptionally repressed repeats can alter gene silencing and chromatin state. A more complicated quantitative model would be required to assess the gene silencing on pericentromere in higher eukaryotes with additional regulatory machinery for gene silencing like DNA methylation.

2.4. Materials and Methods

2.4.1. Mathematical modeling

Our mathematical model tracks the concentration of lncRNA, siRNA, and H3K9me (Fig 2.1). The model includes three first-order ODEs.

$$\dot{x}_1 = \frac{C\alpha}{1 + \left(\frac{x_3}{\kappa_1}\right)^{\rho_1}} - \delta_1 x_1 - \frac{\gamma x_1 \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}} \tag{2.1}$$

$$\dot{x}_2 = \frac{\gamma x_1 \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}} - \delta_2 x_2 \tag{2.2}$$

$$\dot{x}_3 = \frac{\epsilon x_2 (1 - x_3) \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}}{1 + \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}} - \delta_3 x_3 + \phi(1 - x_3)x_3 + \zeta(1 - x_3) \tag{2.3}$$

In Eqs. (2.1-2.3), x_1, x_2, and x_3 are the concentrations of RNA transcripts, siRNA, and H3K9me, respectively, and x_3 is normalized with a maximum value of 1. C represents the copy

number of pericentromeric repeats which is an explicit parameter of the model. Each copy can be transcribed independently, and the overall maximum transcription is linearly proportional to C in the absence of methylation. The transcription rate of RNA, as given by the first term in eq. (2.1), decreases nonlinearly with the increase in H3K9me concentration as a Hill function, with Hill coefficient ρ_1, and half-maximum H3K9me concentration κ_1. RNA can be degraded at a rate δ_1, or turn into siRNAs in a methylation-dependent manner. siRNA biogenesis rate is represented as a Hill function which is nonlinearly correlated with methylation, with ρ_2 as the Hill coefficient and κ_2 as the half-maximum methylation. Eq. (2.2) describes the change in siRNA concentration which includes the siRNA biogenesis and degradation marked by biogenesis rate γ and degradation rate δ_2. The first term in eq. (2.3) represents the siRNA-mediated methylation, which is proportional to siRNA concentration and available unmethylated H3K9 $(1 - x_3)$, with the rate that increases non-linearly with H3K9me concentration as a Hill function of the Hill coefficient ρ_3 and half-maximum methylation κ_3. The spreading of methylation is shown by the third term in eq. (2.3), where ϕ is the spreading rate. Methylation other than from the RNAi mechanism is incorporated as basal methylation with rate ζ. δ_3 is the demethylation rate. Here we use $\rho_1 = 1$, $\rho_2 = 2$, and $\rho_3 = 1$, but the bistability occurs for a wide variety of Hill coefficients. Fig. S2.5 shows the effect of Hill coefficients on bistability, and bistability is possible for different Hill coefficients.

2.4.2. Solutions of ODE

The system of Eqs. (2.1-2.3) was solved in MATLAB (R2019b) (102) using an ode45 solver for 24 different initial RNA, siRNA, and H3K9me concentrations. The values of the

parameters used for wild-type cells are listed in Table S2.2, and 24 initial conditions used in the model are listed in Table S2.3.

2.4.3. Bifurcation Diagram

To obtain a bifurcation diagram, Eqs. (2.1-2.3) were solved for different values of copy numbers ranging from 1 to 20, applying the initial conditions listed in Table S2.3 for each copy number to 25,000 minutes. All 24 H3K9me steady states for each copy number were plotted to decide how many clusters they group for Fig. 2.2. The system becomes completely silenced for copy numbers 11 and above. Therefore, the k-means algorithm (103), where k = 2, was applied to divide all steady-states into two clusters and get the average of each cluster. If the absolute difference between the average of one cluster and that of the other is smaller than 0.085, they were monostable and bistable if greater than 0.085. Finally, the average steady state of each cluster was plotted against the corresponding copy number. Fig. S2.2 depicts the fraction of silenced and desilenced states obtained after solving Eqs. (2.1-2.3) for 1000 random initial conditions up to copy number 20. Fractions of both silenced and desilenced states were obtained for copy numbers 1 to 11, where the system is bistable.

2.4.4. Quasi-steady state approximation

Fig. S2.1 shows that RNA reaches steady-state much faster than siRNA, therefore we can approximate the full system where RNA is in equilibrium with respect to siRNA and H3K9me at any given time. This can be done by setting the left-hand side of eq. (2.1) to zero and solving for x_1, we obtain:

$$x_1 = \frac{C\alpha}{1 + \left(\frac{x_3}{\kappa_1}\right)^{\rho_1}} \frac{1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{\delta_1 \left(1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}\right) + \gamma \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}} \tag{2.4}$$

Substituting eq. (4) into eq. (2) leads to:

$$\dot{x}_2 = \frac{\gamma C\alpha \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{\left(1 + \left(\frac{x_3}{\kappa_1}\right)^{\rho_1}\right)\left(\delta_1 \left(1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}\right) + \gamma \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}\right)} - \delta_2 x_2 \tag{2.5}$$

Setting LHS of eq. (2.5) and eq. (2.3) equal to zero gives the equations of siRNA nullclines and H3K9me nullclines respectively.

$$x_2 = \frac{\gamma C\alpha \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{\delta_2 \left(1 + \left(\frac{x_3}{\kappa_1}\right)^{\rho_1}\right)\left(\delta_1 \left(1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}\right) + \gamma \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}\right)} \tag{2.6}$$

$$x_2 = \frac{\left(1 + \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}\right)(\delta_3 x_3 - \phi(1 - x_3)x_3 - \zeta(1 - x_3))}{\epsilon(1 - x_3) \left(\frac{x_3}{k_3}\right)^{\rho_3}} \tag{2.7}$$

Solving Eqs. (2.6-2.7) leads to steady states, both stable and unstable.

32

2.4.5. Sensitivity analysis

To test the sensitivity of a parameter (for example y) on copy number-dependent gene silencing, it was varied in the range $0.75p \leq y \leq 1.25p$, p being its default value, keeping all other parameters to their respective default values. Then we calculated the H3K9me steady states as a function of y for a fixed copy number. This enabled us to visualize the range of values of a particular parameter for which the system was desilenced, bistable, and silenced, and thus the critical y at which the transition happened. This process was repeated for copy numbers 1 to 40 and all the parameters. The result is depicted in the bar diagram (Fig. 2.3 and Fig. S2.4). Each bar shows the range of values for which the system is desilenced, bistable, and silenced for a given copy number. The critical y at which the system changes from bistable to silenced or vice-versa is the border between these two ranges.

2.4.6. Stochastic simulation

To implement the Gillespie algorithm for Eqs. (2.1-2.3) for stochastic simulations (80,81) at the discrete molecular level, the maximum number of H3K9me sites was normalized to nC, where n is the average number of H3K9me sites for methylation per copy, and C is the copy number. Each histone can be independently methylated or demethylated. RNA and siRNA molecules in Eqs. (2.1) and (2.2) were rescaled 100-fold to represent individual transcripts.

Cell division time was fixed, and cell growth rate was assumed to be linear in time, and simulated as $V = Kt + V_0$, where $K = \frac{V_{max} - V_{min}}{Cell\ division\ time}$ and V_0 is the minimum or initial volume after each cell division. Cell volume was increased whenever a reaction occurred, and it was halved

when the time approached the next cell division. The maximum and minimum cell volume was taken to be $\frac{4}{3}$ and $\frac{2}{3}$ respectively so that the average cell volume is 1. Consistent with known biology, RNA, siRNA, and H3K9me were divided into two daughter cells at each cell division. The number of RNA, siRNA molecules, and H3K9me sites was increased or decreased by a step size of 1 after each reaction whose occurrence depends on its rate.

The intrinsic noise was considered as a random walk per particle, whose total rate is proportional to either the number of H3K9me sites or RNA and siRNA molecules in the system. e.g., $r_1 = \beta x_1$, $r_2 = \beta x_2$, and $r_3 = \beta x_3$ are three reaction rates for noise of RNA, siRNA, and H3K9me respectively. β is termed as "Noise Rate" in this paper. Number of x_1, x_2, and x_3 are randomly increased or decreased with step size 1 whenever reactions r_1, r_2, or r_3 are selected respectively.

To calculate the silencing fraction, 50 simulations were run to get the H3K9me count over time at a fixed cell division and noise rate. The average silencing fraction of those 50 simulations was taken and plotted at various noise rates and fixed cell division. For the plot of the probability distribution of H3K9me over time, the H3K9me count was taken from all 50 simulations. Even though the simulation was run for 10000 minutes, data up to 4000 minutes were discarded as burn-in.

2.4.7. Alternative model with two-step processes for methylation

Our main model tracks the level of methylation, but the actual histone states also include acetylation, which is a key step in the heterochromatin assembly. To incorporate this, we have

developed an alternative model that considers two additional variables, specifically acetylated H3K9 and unmodified H3K9 (69,104). In this model, we have assumed that only unmodified H3K9 can be either methylated or acetylated and vice-versa. Also, we assume that new histones are acetylated (105). In this model, x_1, x_2, x_3, x_4, and x_5 represent the RNA and siRNA concentrations, the fraction of methylation, acetylated, and unmodified histones, respectively.

$$\dot{x}_1 = \frac{C\alpha}{1 + \left(\frac{x_3}{\kappa_1}\right)^{\rho_1}} - \delta_1 x_1 - \frac{\gamma x_1 \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}} \tag{2.8}$$

$$\dot{x}_2 = \frac{\gamma x_1 \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}}{1 + \left(\frac{x_3}{\kappa_2}\right)^{\rho_2}} - \delta_2 x_2 \tag{2.9}$$

$$\dot{x}_3 = \frac{\epsilon x_2 x_5 \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}}{1 + \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}} - \delta_3 x_3 + \phi_1 x_3 x_5 + \zeta_1 x_5 - \xi x_3 \tag{2.10}$$

$$\dot{x}_4 = \zeta_2 x_5 - \delta_4 x_4 + \xi - \xi x_4 \tag{2.11}$$

$$\dot{x}_5 = -\left(\frac{\epsilon x_2 x_5 \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}}{1 + \left(\frac{x_3}{\kappa_3}\right)^{\rho_3}} - \delta_3 x_3 + \phi_1 x_3 x_5 + \zeta_1 x_5 \right) - (\zeta_2 x_5 - \delta_4 x_4) - \xi x_5 \tag{2.12}$$

Eqs. (2.8) and (2.9), responsible for tracking the changes in RNA and siRNA concentrations respectively, are consistent with Eqs. (2.1) and (2.2) mentioned in the three-variables model. Eq. (2.10), which details the variation in H3K9me concentration and is mathematically similar to eq. (2.3), distinguishes itself from eq. (2.3) by addressing the methylation acting on unmodified H3K9, denoted by x_5. ξ represents the rate of histone turnover,

and we assume that newly incorporated histones are acetylated (105). ζ_2 and δ_4 represent the rate of acetylation and deacetylation, respectively. Only unmodified histones can be methylated or acetylated.

The parameter referenced and the initial conditions applied for solving ODE Eqs. (2.8-2.12) can be found in Table S2.4 and Table S2.5 respectively. The starting levels of methylated, acetylated, and unmodified H3K9 were selected to ensure their total equals one. The parameters used in Eqs. (2.8-2.9) align with those in Eqs. (2.1-2.2).

Chapter 3: Mathematical Model for the Dynamic Nature of HSATII Repeats in Human Pericentromere

3.1.   Introduction

The centromere is an essential chromosomal region for accurate chromosome segregation during cell division. It consists of the core centromere and the adjacent pericentromeric regions, both of which are defined by distinct structures and repetitive DNA sequences. The human core centromere comprises α-satellite DNA, which is organized in tandem arrays. Each α-satellite unit is approximately 171 base pairs in length, with these arrays spanning from hundreds of kilobases to several megabases (106,107). The total number of α-satellite repeats varies significantly among chromosomes but generally ranges from 0.5 to 5 megabases per chromosome (106,107).

Surrounding the core centromere are the pericentromeric regions, which contain diverse types of repetitive DNA, including satellite DNA, transposable elements, and segmental duplications. Pericentromere allows, with its constitutive heterochromatic structure, proper chromatin segregation through kinetochore formation during cell division (108,109) and is also crucial for maintaining genome integrity (110). Human pericentromeric DNA comprises various classes of tandemly repeated DNA, including β-satellite, and γ-satellite, and satellites I, II, and III (111). These sequences are more heterogeneous than the α-satellite DNA of the core centromere (112). The pericentromeric regions can extend several megabases beyond the core centromere, with the length varying between chromosomes and individuals (107). These regions are vital for chromatin organization, kinetochore stability, and preserving the heterochromatin state necessary

for centromere function (41). These pericentromeric repeats are rich in post-translational modifications of histone tails, such as trimethylation of lysine 9 of histone H3 (23,113), trimethylation of lysine 20 of histone H4 (113,114) and broad histone hypoacetylation (115). H3K9me and H4K20me3 are critical players in pericentromeric heterochromatin assembly in mammals (9,114). Additionally, DNA methylation is associated with heterochromatin formation (116,117), and its loss in the pericentromeric HSATII repeats has been reported in many types of cancers (118).

Histone and DNA methylation are essential for the transcriptional silencing of pericentromeric heterochromatin. Specifically, methylation of histone H3 at lysine 9 (H3K9) and histone H4 at lysine 20 (H4K20) play critical roles (23,113,114). In humans, the methyltransferases SUV39H1 and SUV39H2 are responsible for trimethylating H3K9, a key marker for heterochromatin crucial for silencing pericentromeric repeats (119,120). The H3K9me3 mark deposited by SUV39H1/H2 serves as a binding site for heterochromatin protein 1 (HP1), which is integral to chromatin compaction and the repression of pericentromeric repeats, thus maintaining genomic stability (121–124). The deposition of H4K20me3 on pericentromeric heterochromatin is facilitated by SUV4-20H1 and SUV4-20H2, two SET-containing histone methyl transferases that localize to chromocenters via their interaction with HP1s (114,125).

Another hallmark of mammalian pericentromeric heterochromatin is DNA methylation, which predominantly occurs at cytosine 5 of CpG dinucleotides (5meC) (126). The DNA methyltransferases DNMT3A and DNMT3B are pivotal for the de novo methylation of pericentromeric repeats (127). Conversely, TET enzymes (TET1, TET2, and TET3) are involved in active DNA demethylation by converting 5mC to 5-hydroxymethylcytosine (5hmC) and other

derivatives (128,129). Investigation by Hashimoto et al. (130) demonstrated that TET enzymes are crucial for maintaining DNA methylation homeostasis and that disruptions in their function can lead to aberrant demethylation in pericentromeric regions.

Mis-regulation of chromatin and DNA methylation is associated with various human diseases. For example, SUV39H1 and SUV39H2 have been found to function as both oncogenes and tumor suppressors in different types of cancer (131,132). DNA hypomethylation in the HSATII regions of chromosomes 1 and 16 has been observed in various cancers (116,117), which underscores the importance of proper methylation in pericentromeric regions in carcinogenesis. Hypomethylation can occur very early in cancer and is strongly associated with tumor progression (117). Additionally, overexpression of pericentromeric satellite repeats has been reported in many cancer types (32,133), potentially resulting from the loss of DNA methylation in these regions (134). Notably, DNA methylation of the pericentromeric region is associated with the HP1 proteins and H3K9me3 (123), indicating a complex interplay between DNA methylation and H3K9 methylation. This implies an indirect role of H3K9 methylation in cancer progression.

New discoveries are rewriting the role of pericentromeric regions. Once thought to be transcriptionally silent, these areas are now known to be actively transcribed during cell development to ensure proper chromosome formation (135,136). Additionally, abnormal activation of a specific type of pericentromeric repeat, HSATII, has been linked to cancer (32,137). HSATII consists of short, repetitive 25- to 100-bp consensus sequences organized in long stretches that may span up to thousands of kilobases (26). The misregulation of HSATII RNA transcripts in cancer cells may lead to the dramatic expansion of HSATII through reverse transcription (31), which suggests a surprising link between HSATII RNA and the copy number gain of HSATII

repeats. HSATII RNA can be reverse transcribed into DNA copy, and inserted into the pericentromeric region, thus expanding the HSATII copy number (31). Using reverse transcriptase inhibitors to reduce the reverse transcription of HSATII RNA in an in vivo tumor xenograft model leads to a decrease in HSATII DNA copy number gain and tumor growth (31). In addition, alpha satellites are also associated with chromosomal instability and copy number alteration (32,138), and unique sets of ncRNAs are required for kinetochore assembly and cell-cycle progression (139). These studies suggest the dynamic interplay between human pericentromere and ncRNAs.

The observed increase in copy number gain and DNA hypomethylation in cancer cells suggests a possible link between methylation and copy number variation in HSATII repeats. To explore this connection, we have proposed a mathematical model that considers RNA, H3K9 methylation, and total HSATII repeats as variables. The model uses simple non-linear ODE equations to track changes in these variables over time. Reverse transcription is included as a critical process that introduces new copies in the pericentromere. Our model predicts that decreased methylation activity in HSATII repeats leads to an increase in copy number, and the system can exhibit two possible high and low steady states of copy number. Additionally, we have proposed a polymer chain model where repeats are arranged as monomers in a chain of polymers, and the state of a repeat in the next generation depends on its state in the previous generation. The polymer chain model supports the ODE model's result that decreased methylation activity in HSATII repeats leads to an increase in copy number.

3.2. Results

3.2.1. Mathematical modeling of gene silencing at HSATII pericentromeric repeats

We have developed a mathematical model for the gene silencing in HSATII repeats of human pericentromere. This model keeps track of the number of RNA transcripts associated with HSATII repeats, the number of H3K9 methylated HSATII repeats (M), and the overall total of HSATII repeats (CN). HSATII can be unmodified, H3K9 methylated, or H3K9 acetylated. A schematic representation of the molecular pathway used to describe our mathematical model is shown in Fig 3.1. RNA transcribed from unmethylated HSATII repeats can undergo reverse transcription (31), resulting in additional repeats. Both transcription and reverse transcription activities are reduced by either H3K9me or DNA methylation, which, in our model, is incorporated as being suppressed by M repeat. HSATII copies can be lost due to deletion via recombination, which is also repressed by methylation. The conversion of the unmethylated repeat (U) into the methylated one (M) can be spontaneous or through neighboring interactions. The spontaneous conversion of U to M is due to the histone methyl transferase, e.g., SUV39H1/H2, and the spontaneous conversion of the methylated repeat (M) into unmethylated one (U) is modeled by demethylation. On the other hand, H3K9 methylation can facilitate further methylation, which acts in *cis*. For instance, HP1-HP1 dimers attract SUV39H enzymes (122), which can trimethylate H3K9me1 on neighboring nucleosomes (119,140). Similarly, HP1s recruit DNMT3A and DNMT3B, which catalyze CpG methylation (127). These reactions were modeled as the recruited conversion of U repeats to M.

A feedback loop is generated as abundant RNA can lead to many copies of repeats, and a more significant number of repeats will transcribe more RNA. In this model, each repeat is transcribed independently, and the per-copy transcription is a function of H3K9 methylation.
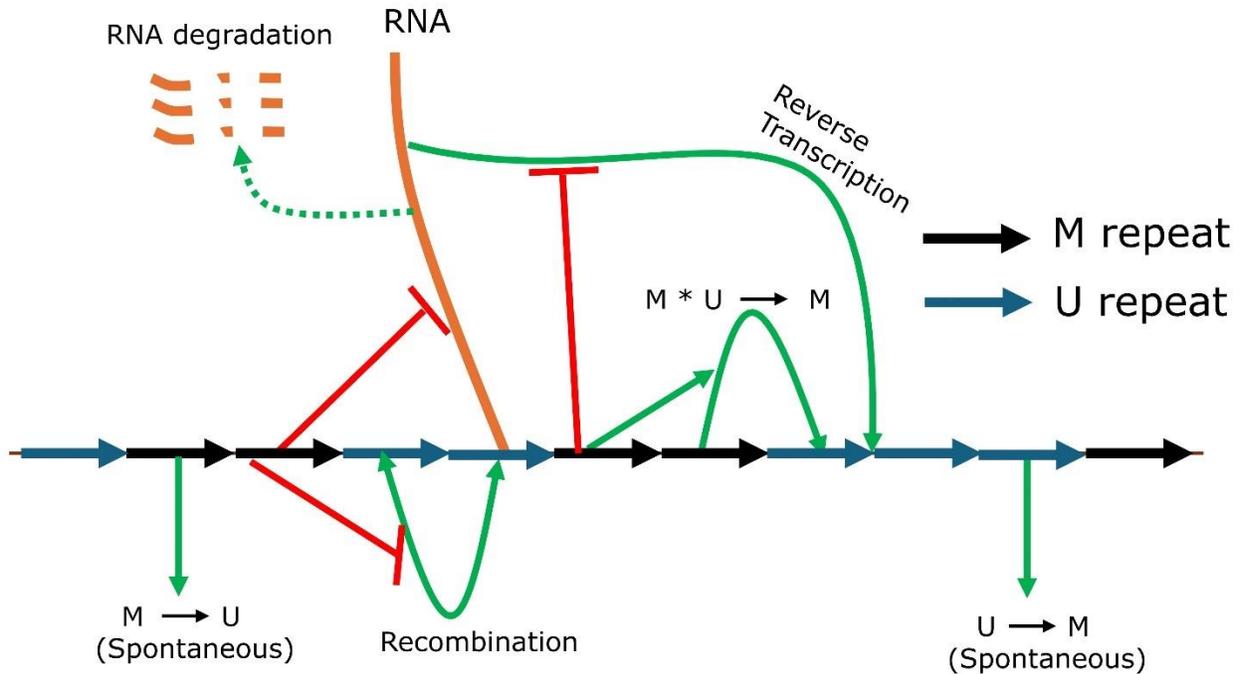
Fig 3.1. The schematic diagram for mathematical modeling of HSATII repeats silencing in humans. HSATII repeats are arranged linearly as blue (unmethylated U) or black arrows (methylated M) in pericentromere. RNA transcription is suppressed by the methylated repeat. RNA can either be degraded or integrated as an unmethylated repeat (U) through reverse transcription. A repeat can be deleted through recombination. M can convert to U spontaneously or by local interaction with M repeat. U can turn into M repeats only spontaneously. Recombination and reverse transcription are suppressed by M. Conversion of U into M through local interaction with M is enhanced by M repeats.

3.2.2. HSATII copy gain with reduced methylation

To quantify the dynamic nature of the copy number of HSATII repeats in the human pericentromere, we solved ODE equations (Eqs. 3.1-3.3) and found their steady states. Fig 3.2 illustrates the ODE solutions of our model equations, and the fraction of methylation is the ratio of methylated HSATII copies among total HSATII copies. When the methylation spreading rate ϕ is high ($\phi = 0.0002$), the default value in our model (see Table S3.1), approximately 80% of repeats are methylated (Fig 3.2A). However, the methylation fraction is lowered to ~ 0.14 % when ϕ is reduced ($\phi = 0.00007$) (Fig 3.2A). This reduction in the methylated repeat fraction also leads to a significant increase in the number of RNA transcripts and an expansion of HSATII (Fig 3.2A). A similar result is obtained when altering spontaneous methylation rate η and demethylation rate μ (Fig 3.2B-3.2C). With a 100-fold decrease in spontaneous methylation rate, the fraction of methylated repeat is reduced to almost zero, RNA transcription increases by almost 13-fold, and repeat expands by nearly 3-fold (Fig 3.2B). Additionally, as the demethylation rate μ is increased from 3.75-fold from its default value, transcription increases by almost 13-fold, and repeats increase by nearly 3-fold (Fig. 3.2C). This suggests that the rate of methylation or demethylation of repeat affects HSATII copy number.
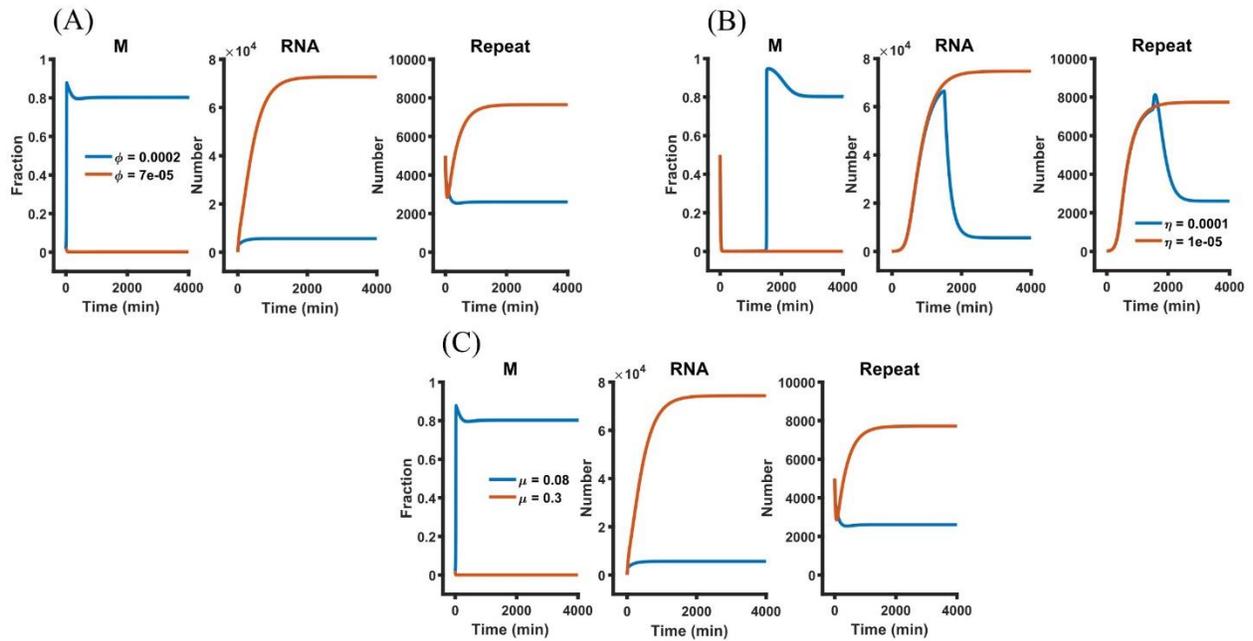
Fig 3.2. Strength of methylation increases the HSATII repeats. (A)-(C) Left panel: Methylation fraction obtained by dividing ODE solution of methylated repeat (eq. 3.2) by the ODE solution of Copy number of repeats (eq. 3.3) for two values of ϕ, η, and μ. Middle Panel: Solutions of RNA equation (eq. 3.1) for two values of ϕ, η, and μ. Right Panel: Solutions of copy number equation (eq. 3.3) for the two values of ϕ, η, and μ.

3.2.3. Potential coexistence of higher and lower steady states of copy number

To understand how each model parameter affects the steady-state dynamics of the HSATII copy number, we estimated the steady states while systematically varying one parameter at a time. First, we varied the methylation spreading rate and tracked the steady states from seven different initial conditions (Fig. 3.3A & Table S3.2 for the list of initial conditions). As the spreading rate decreases, the number of methylated copies decreases, which in turn increases the number of transcripts and HSATII copy numbers. The bifurcation diagram shows that the system can be bistable with two stable steady states with high and low fractions of M repeats. The modulation of

the reverse transcription rate (σ) was qualitatively similar to that of the methylation spreading rate (ϕ), as shown in Fig. 3.3B. These results indicate that when the fraction of methylated repeat is high, RNA and copy number are at lower steady states, and vice versa. This pattern persists in the bistable range of methylated fractions.
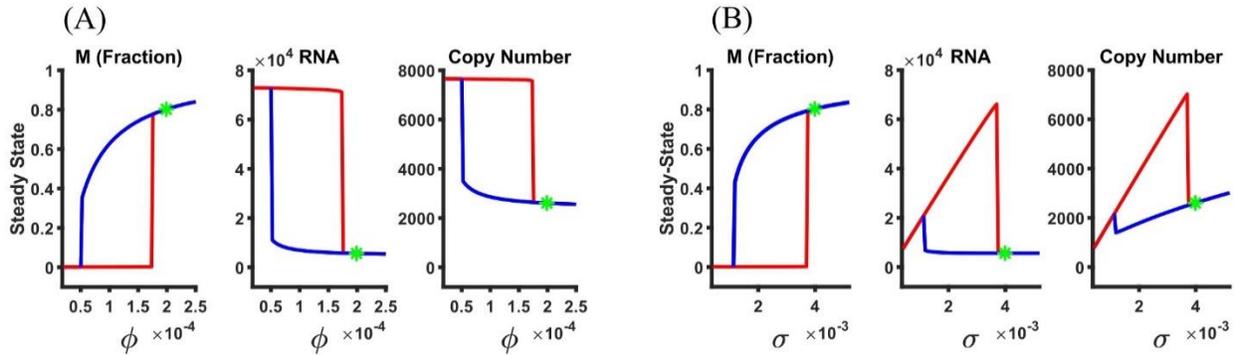
Fig 3.3. Bistability for a range of parameter values. Steady-state solutions for a fraction of methylated repeat, RNA, and copy number for the range of (A) methylation spreading rate (ϕ) and (B) reverse transcription rate (σ). For both parameters, models allow a range of values for lower steady-state, coexistence of both lower and higher steady-state, and higher steady-state. The green star represents the reference value of the parameter.

One notable difference between the steady states of σ and ϕ is that the steady-state copy number is sensitive to σ (Fig 3.3B). This is because the reverse transcription determines the maximal HSATII copy gain, which can further be reduced by H3K9 methylation. The bifurcation diagram for changing other parameters is shown in Fig S3.1.

3.2.4. Analysis of steady states with nullclines

To quantify how each parameter leads to bistability, we analyzed the nullclines of the ODE model after a Quasi Steady State Approximation (QSSA). We performed QSSA by assuming

RNA reaches a steady state faster than other variables to obtain the equations for the M nullcline (eq. 3.5) and CN nullcline (eq. 3.6) (see 3.4 Materials and Methods for details). This can reduce the full 3-dim system into 2-dim, which allows the visual representation of steady states at different parameter values. Fig 3.4A shows M and CN nullclines for three different values of methylation spreading rate (ϕ). CN nullcline, which is independent of ϕ (eq. 3.6), doesn't change with respect to changing ϕ, while M nullcline, which depends on ϕ (eq. 3.5), shifts upward with a decrease in ϕ (Fig. 3.4A). For $\phi = 0.0001$, M nullcline moves upward, increasing the number of intersection pts from 1 to 3, corresponding to higher copy number ~ 7840, middle copy number ~ 7210, and lower copy number ~ 2870 (Fig 3.4A). For a reference ϕ, which is 0.0002, two nullclines intersect at one point corresponding to copy number ~ 2600. When ϕ is further reduced to 0.00003, the CN nullcline moves further upward, this time with only one intersection point corresponding to a higher copy number ~ 7840 (Fig 3.4A). On the other hand, Fig 3.4B shows how the reverse transcription rate σ shapes the CN nullcline. As σ increases, the CN nullcline shifts upward. Fig 3.4B shows that for reference σ, which is 0.004, the CN nullcline and M nullcline intersect at one point corresponding to CN ~ 2600, for $\sigma = 0.002$, CN nullcline shifts downward, leading to three intersection points corresponding to higher copy number ~ 3800, middle copy number ~ 3545, and lower copy number ~ 1770. When σ is reduced to 0.0006, the CN nullcline moves further downward, intersecting at copy number ~ 1150. The nullcline analysis for all the parameters is shown in Fig S3.2. Except for the half maximum number of methylated repeat saturation (κ_4), bistability is observed for all parameters (Fig S3.1 and Fig S3.2). Fig 3.4A shows that decreasing the value of ϕ increases the copy number, while Fig 3.4B shows that decreasing the value of σ decreases the copy number. Fig 3.4C shows a table that lists all the

parameters, which shows how they favor copy number expansion. The left column in the table of

Fig 3.4C lists all the parameters whose increase favors copy number gain, while the right column

lists all the parameters whose reduction favors copy number gain.

(A)

(B)

(C)

Increase in	Decrease in
Transcription Rate (α)	RNA degradation Rate (δ)
Reverse Transcription Rate (σ)	Cooperative conversion rate of M to U (ϕ)
Spontaneous conversion of M to U (μ)	Spontaneous conversion of U to M (η)
Half max. saturation number of M for transcription (κ_1)	Deletion of U repeat by recombination (γ)
Half max. saturation number of M for cooperative methylation (κ_2)	Half max. saturation number of M for copy number recombination (κ_4)
Half max. saturation number of M for reverse transcription (κ_3)	

Fig 3.4. Nullcline Analysis shows the existence of bistability. Nullcline analysis for different values of (A) methylation spreading rate ϕ, (B) reverse transcription rate σ. (A) M nullcline shifts upward when ϕ is reduced, changing the number of intersection points from lower copy number state ($\phi = 0.0002$) to bistable state ($\phi = 0.0001$) to higher copy number state ($\phi = 0.00005$). (B) CN nullcline shifts downward with a decreasing value of σ, changing the number of intersection points from a higher copy number state ($\sigma = 0.004$) to bistable state ($\sigma = 0.002$) to a lower copy number state ($\sigma = 0.0006$). (C) A table that lists parameters that favor in copy number increase or decrease. The left column lists parameters whose increase favors copy number increase, while the right column lists all the parameters whose reduction favors copy number increase.

3.2.5. Sensitivity analysis enables the identification of the range of parameters that favor the state of a higher fraction of M, bistability, and lower fraction of M

To identify the range of parameters that favor the state of a higher fraction of M, lower fraction of M, and bistable, we performed the sensitivity analysis. In one parameter sensitivity analysis, a parameter was varied, and for each value of a parameter, the state of the system was determined (see 3.4 Materials and Methods for details), keeping all other parameters in their reference values for all the parameters except κ_4 (half maximum saturation number of M for copy number recombination) which doesn't show bistability (Fig S3.1 I). Fig 3.5A shows the result of one parameter sensitivity analysis for all the parameters (except κ_4). All parameters exhibit a distinct range of states of a lower fraction of M, bistability, and a higher fraction of M, except the spontaneous conversion rate (η) of M to U, which only shows the bistable and higher state of M regime. For parameters α (transcription rate), κ_1 (half maximum saturation number of M for transcription), κ_3 (half maximum saturation number of M for reverse transcription), ϕ (cooperative conversion rate of U to M), σ (reverse transcription rate), the lower state of the fraction of M is followed by a bistable state which is followed by a higher state of a fraction of M. However, for parameters δ (RNA degradation rate), γ (repeat recombination rate), κ_2 (half maximum saturation number of M for cooperative methylation), and μ (spontaneous conversion rate of M to U), a higher state of the fraction of M is followed by a bistable state which is followed by a lower state of a fraction of M (Fig 3.5A). These results are in line with Fig 3.4C.

In addition, we also varied selected pairs of two parameters systematically, and we tracked the steady states, the fraction of M, and the copy number (see 3.4 Materials and Methods for

details). Fig 3.5B shows the system's behavior for various values of ϕ (cooperative conversion rate of U to M) and μ (spontaneous conversion rate of M to U) in terms of low fraction of M, high fraction of M, and bistability. This suggests that the ratio between ϕ and μ determines the system behavior. Likewise, Fig 3.5C illustrates the system's behavior across different values of transcription rate (α) and the reverse transcription rate (σ), highlighting the low fraction of M, high fraction of M, and bistability. This suggests that the product of transcription rate (α) and reverse transcription rate (σ) may determine the dynamics of copy number. These results indicate that bistability may be found with various combinations of parameters.
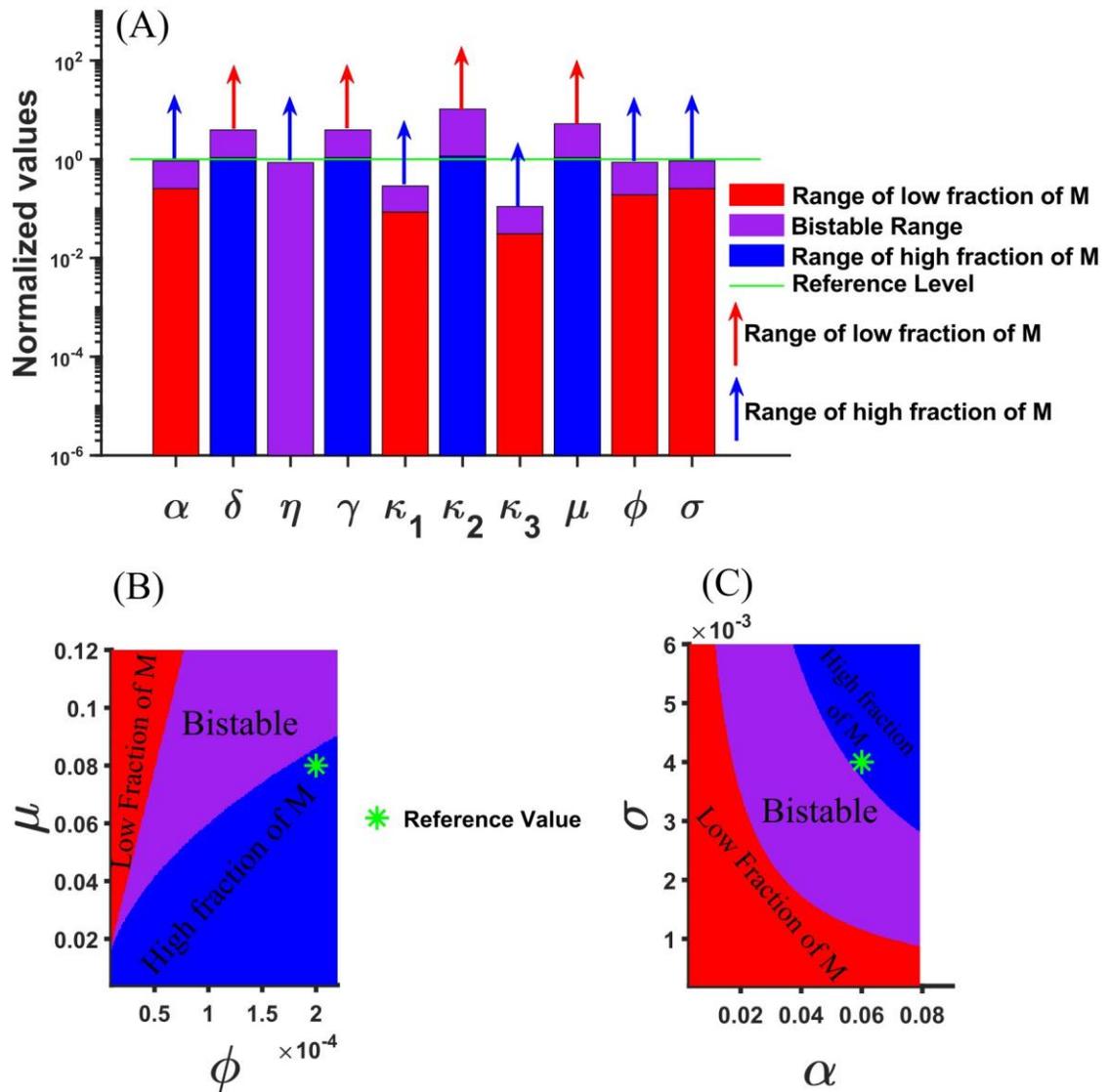
Fig 3.5. Sensitivity Analysis. (A) Range of low fraction of M, bistable, and high fraction of M of a parameter (except κ_4, which doesn't show bistability) while other parameters are kept in their respective reference values for all the parameters. The reference values of parameters are normalized to 1. (B) Range of low fraction of M, bistable, and high fraction of M for the simultaneous variations in methylation spreading rate ϕ and spontaneous conversion of M to U rate μ while all other parameters are kept fixed in their respective reference values. (C) Range of low fraction of M, bistable, and high fraction of M the simultaneous variations in transcription rate α, and reverse transcription rate σ, while all other parameters are kept fixed in their respective reference values.

3.2.6. The polymer chain model also predicts the copy number expansion with weaker local interaction of methylation

The ODE model described the average behavior of the system, but it is possible that HSATII is variable from cell to cell as well as chromosome-to-chromosome variability. To better understand the dynamic nature of human pericentromeric repeat expansion, we developed an alternative polymer chain model (see 3.4 Materials and Methods for details). In this model, each repeat is either unmethylated (U) or methylated (M), and these repeats are arranged as monomers in a polymer chain (see Fig 3.6). Fig 3.6 shows possible reactions at each repeat, and the state is updated in each generation. Each repeat can spontaneously switch to another state (from M to U or U to M) in the next generation. Moreover, a U repeat can transition to an M repeat if it is adjacent to an M neighbor, which we represent as the cooperative conversion rate of U to M (rUM). Additionally, a U repeat can be inserted next to another U repeat in the chain, and a U repeat can be removed from the chain.
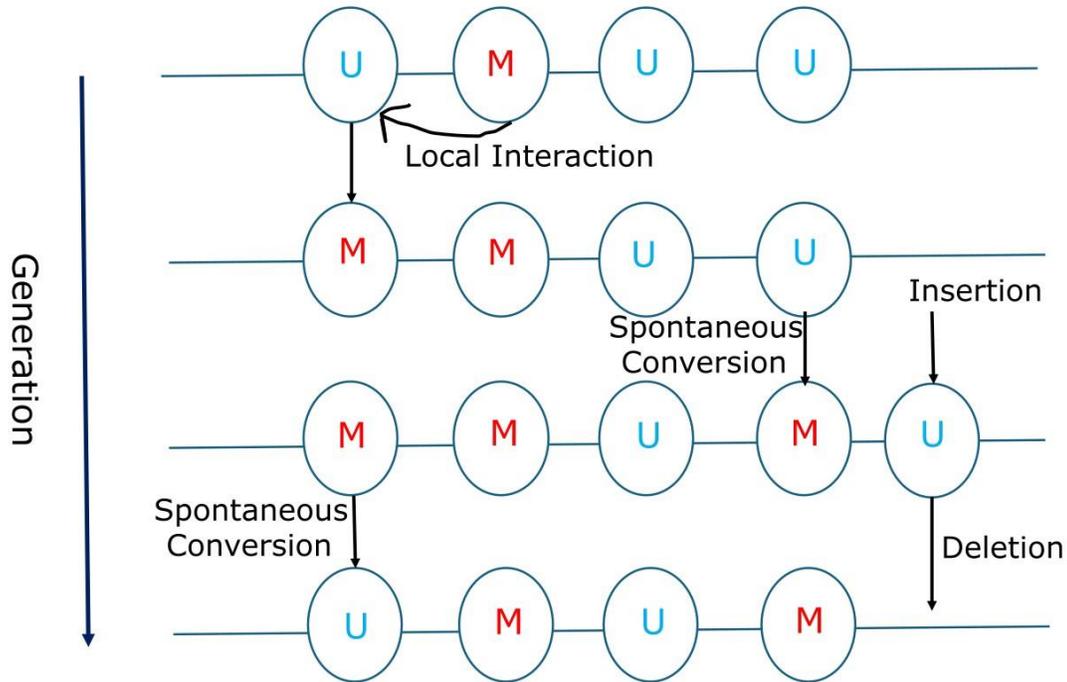
Fig 3.6. Schematic diagram for polymer chain model. The repeats in the chain can be either in a methylated (M) state or an unmethylated (U) state and are arranged as monomers in the polymer chain. A U repeat can convert to an M repeat either by interacting with an adjacent M neighbor or spontaneously. Conversely, an M repeat can revert to a U repeat spontaneously. Additionally, a U repeat can be inserted into the chain, and similarly, a U repeat can be deleted from the chain.

Fig 3.7 illustrates representative evolutions of repeats over 50 generations. When the cooperative conversion rate of U to M (rUM) is low (set at 0.1/min), the repeat copy number stabilizes around approximately 3000 copies, with predominantly methylated repeats (Fig 3.7A). Conversely, with a lower rUM (set at 0.01/min), the copy number reaches around 20000 with mostly unmethylated repeats (Fig 3.7B). In this model, the unmodified and methylated repeats are well mixed, and the overall methylation level determines the number of total copies.
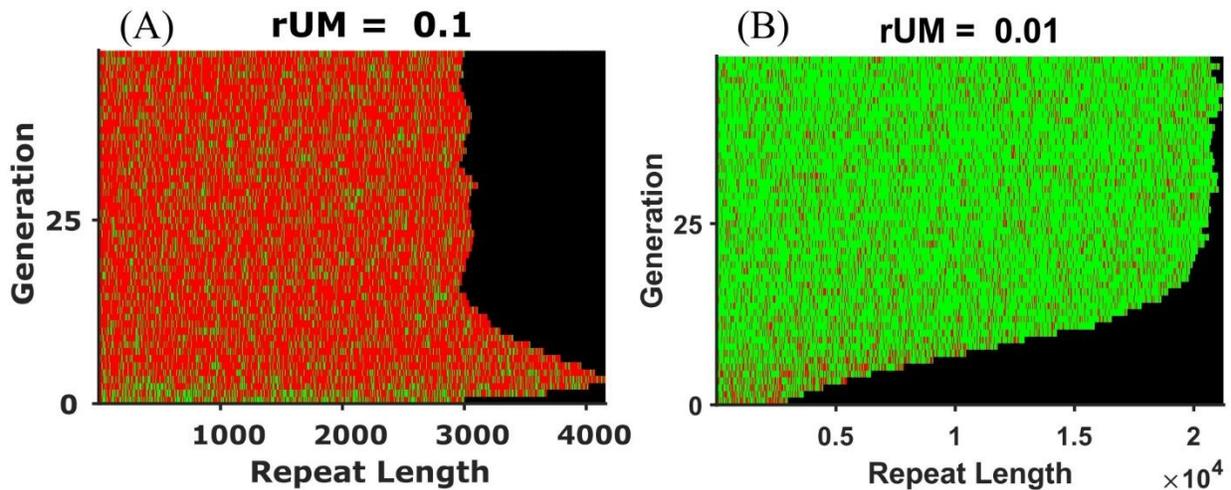
Fig 3.7. Representative diagram of the evolution of repeats. Red represents methylated repeats, and green represents unmethylated repeats. (A) Evolution of repeats with a cooperative conversion rate of U to M at 0.1 per minute. The repeats evolve to saturate at approximately 3000 copies, with a high fraction of methylated repeats. (B) Evolution of repeats with a cooperative conversion rate of U to M at 0.01 per minute. The repeats evolve to saturate at approximately 20000 copies, with a high fraction of unmethylated repeats.

To ask how the rUM facilitates methylation among different copies, we calculated the fraction of M and the copy number for various rUM values. Fig 3.8 presents the distribution of the fraction of M repeats and copy numbers at the 50th generation for various rUM values. The fraction of methylated repeats decreases with decreasing rUM, as the distribution shifts leftward when rUM is reduced from 0.1 per minute, initially slightly but significantly falling below 0.2 when rUM is decreased to 0.01 per minute (Fig 3.8A). The opposite trend is observed for the copy number distribution (Fig 3.8B), which shifts rightward as rUM decreases, concentrating above 20000 copies for an rUM of 0.01 per minute. This increase in copy number with a decrease in the cooperative methylation rate aligns with our ODE model predictions. Furthermore, the dramatic shift in copy number distribution above 20000 from around 5000 when reducing rUM to 0.01 from 0.03 suggests a switch-like behavior in the system, where the system adjusts to either very high or

very low repeat copies depending on whether the perturbation, such as a decrease in rUM, falls below or above its threshold value.
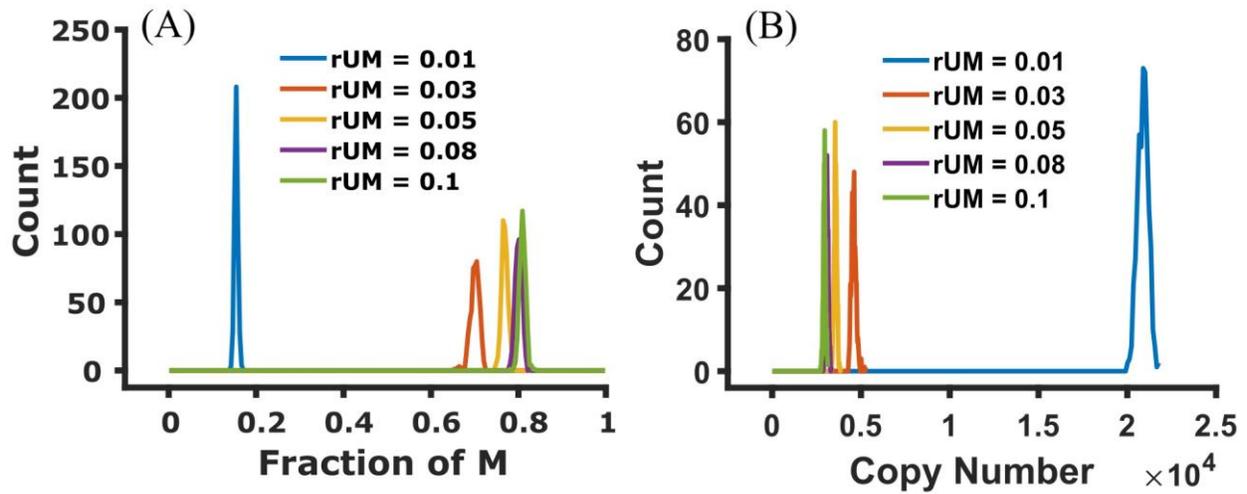
Fig 3.8. Distribution of the fraction of M and copy Number for the 50ᵗʰ generation. (A) Distribution of the fraction of M for the 50th generation at different rates of cooperative conversion of U to M. (B) Distribution of copy numbers for the 50th generation to varying rates of cooperative conversion of U to M. Both (A) and (B) have a sample size of n = 500.

Additionally, the distribution of the median domain size of M repeats indicates that most domain sizes are very small (~2) for lower rUM values. However, larger median domain sizes can occur at higher rUM values (Fig 3.9A). The size distribution of M domains follows a decreasing exponential pattern, meaning that as the domain size increases, its frequency in the pericentromere decreases. This decay rate, derived from fitting the distribution for each rUM, is plotted against the rUM values in Figure 3.9B. It shows that the absolute value of the decay rate of the frequency

of the median size of M domains is very high for an rUM of 0.01, decreases sharply for an rUM of 0.03, and gradually decreases thereafter.
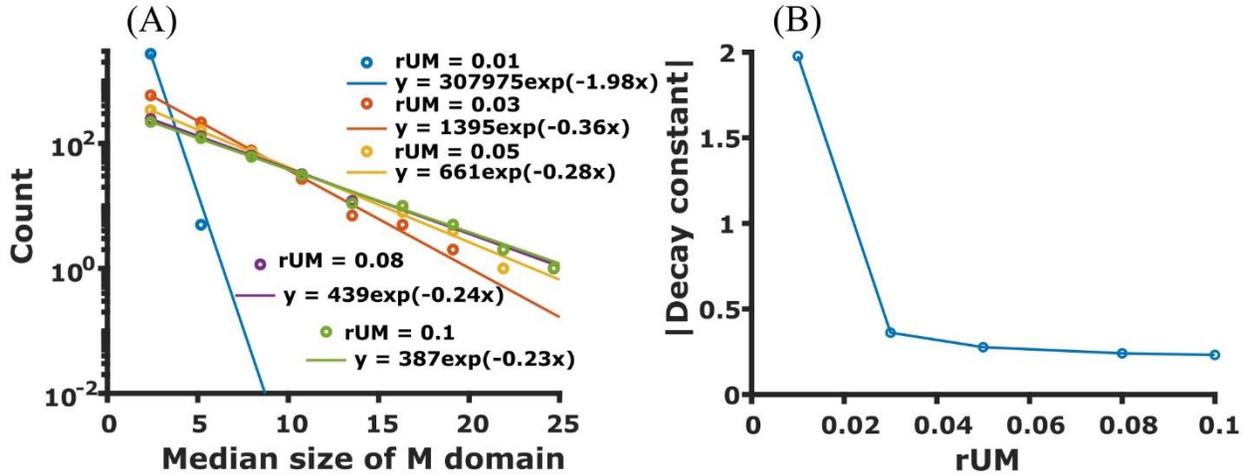
Fig 3.9. Distribution of the median size of the M domain of the 50th generation and the corresponding exponential fit. (A) Distribution of the median size of the M domain at various rates of cooperative conversion from U to M. Open circles represent data points, while solid lines show the exponential fit to these data points. n = 500 (B) Plot of the exponential decay constant against the cooperative conversion rate from U to M.

3.3.   Discussions

The observed rise in copy number gain due to reverse transcription of ncRNA (31) and DNA hypomethylation in cancer cells (130) indicates a potential connection between methylation and copy number variation in HSATII repeats. To investigate this relationship, I developed an ODE-based mathematical model that includes RNA, H3K9me, and total HSATII repeats as variables. Reverse transcription is a critical process in my model, introducing new copies in the pericentromere. My findings predict that reduced methylation activity in HSATII repeats results in an increase in copy number, with the system potentially stabilizing in either a high or low copy

number state. Furthermore, I have introduced a polymer chain model, where repeats are positioned as monomers in a polymer chain, and the state of a repeat in the next generation is influenced by its state in the previous generation. This polymer chain model corroborates the ODE model's prediction that decreased methylation activity in HSATII repeats leads to a higher copy number. By nullcline analysis, I have also identified the parameters that facilitate copy number expansion.

My model can be applied to human transposable elements (TE), which constitute about half of the genome (141) and are mostly class I elements (copy and paste retrotransposons) that use reverse-transcribed RNA intermediates to produce more copies (142). Applying the model to class I human TEs could provide insights into their role in various diseases, including cancer. The big goal of this research is to use copy numbers as a cancer diagnostic tool. My model pinpoints parameters that facilitate copy number expansion. This approach can aid in cancer treatment by targeting specific pathways therapeutically. Using copy numbers as a diagnostic tool could be cost-effective and efficient in preventing cancer metastasis. Future studies might explore targeting HSATII methylation pathways as a therapeutic strategy.

Furthermore, enhancing the polymer chain model by incorporating interchromatin interactions would enable the exploration of repeat dynamics across various chromosomes where HSATII repeats are found.

3.4. Materials and Methods

3.4.1. ODE-based Mathematical modeling

This model keeps track of the number of lncRNAs associated with HSATII repeats, methylated HSATII repeat copy number, and the total HSATII repeat copy number (including methylated and unmethylated forms) using three first-order nonlinear differential equations.

$$\dot{x}_1 = \frac{\alpha x_3}{1 + \frac{x_2}{K_1}} - \delta x_1 \tag{3.1}$$

$$\dot{x}_2 = \frac{\frac{\phi x_2}{K_2} x_2 (x_3 - x_2)}{1 + \frac{x_2}{K_2}} + \eta(x_3 - x_2) - \mu x_2 \tag{3.2}$$

$$\dot{x}_3 = \frac{\sigma x_1}{1 + \frac{x_2}{K_3}} - \frac{\gamma x_3^2}{1 + \frac{x_2}{K_4}} \tag{3.3}$$

In Eqs. (3.1-3.3), x_1, x_2, and x_3 represent the number of lncRNA, methylated HSATII repeats, and the total number of HSATII repeats, respectively, and $x_3 - x_2$ represent the total number of unmethylated repeats. The first term in the first equation shows that RNA transcription is proportional to the total number of repeats at a rate α, which decreases non-linearly as the number of methylated repeats (x_2) increases. Without methylation, transcription increases linearly with the copy number of repeats. The second term in the first equation indicates that RNA degrades linearly at a rate δ. A neighboring interaction between methylated repeat (x_2) and unmethylated repeat ($x_3 - x_2$) can convert an unmethylated repeat into a methylated one, as shown by the first

term in the second equation. The rate ϕ of this reaction is an increasing Hill function that rises with the total number of methylated repeats (x_2) in the system. The second term in eq. (3.2) shows that an unmethylated repeat ($x_3 - x_2$) can spontaneously become a methylated repeat at a rate η. Likewise, a methylated repeat can spontaneously revert to an unmethylated repeat at a rate μ, as shown by the last term in eq. (3.2). This model also introduces reverse transcription of RNA into the genome, and the DNA recombination mechanisms to capture the dynamic nature of human HSATII repeats. The first term in eq. (3.3) shows that reverse transcription is proportional to the total number of transcripts (x_1) at a rate σ, which decreases with methylation according to a decreasing Hill function. The last term in eq. (3.3) indicates that two DNA repeats can recombine, and one can be deleted at a rate γ, which also decreases as the number of methylated repeats increases in the system. κ_1, κ_2, κ_3, and κ_4 represent the half-maximum saturation of methylated repeats in the Hill functions for transcription, neighboring interaction, reverse transcription, and DNA recombination terms in equations (3.1-3.3).

3.4.2. ODE solutions

The system of equations (1-3) was solved in MATLAB (R2023a) (102) using the ode23 solver for seven different initial counts of RNA, methylated repeats (M), and copy numbers (CN) of HSATII repeats. The parameter values used for wild-type cells are provided in Table S3.1, and the seven initial conditions used in the model are listed in Table S3.2.

3.4.3. Bifurcation Diagram

To construct a bifurcation diagram of steady states with respect to parameter values, equations (3.1-3.3) were solved for seven initial conditions up to $10,000$ minutes for all y values in the range $f_1 * p \leq y \leq f_2 * p$, where p is a reference value of the parameter and f_1 and f_2 are factors such that $f_1 < 1$ and $f_2 > 1$. The seven steady states for each y in this range were plotted to determine the number of clusters they formed for Fig 3.3. As the steady states clearly assembled into two groups, the k-means algorithm (103), with $k = 2$, was used to divide all steady states into two clusters and obtain the average of each cluster. If the absolute difference between the averages of the two clusters for the fraction of methylated repeats (M) was less than 0.1, the system was considered monostable; if greater than 0.1, it was considered bistable. Finally, the average steady state of each cluster was plotted against y.

3.4.4. Quasi-steady state Approximation (QSSA)

The bifurcation diagram (Fig 3.3) doesn't confirm the bistability of the system. To determine if bistability can be achieved analytically, we assumed that RNA saturates much faster than methylated repeat (M) and total repeat (CN), meaning RNA is in equilibrium with respect to M and CN at any given time. With this approximation, we can find the steady-state value of RNA by setting the LHS of eq. (3.1) to zero and solving for x_1.

$$x_1^* = \frac{\alpha x_3}{\delta \left(1 + \frac{x_2}{\kappa_1}\right)} \tag{3.4}$$

Setting the LHS of eq. (3.2) to zero and solving for x_3 gives M nullcline

$$x_3 = \frac{\frac{\phi\left(\frac{x_2}{\kappa_2}\right) x_2^2}{1 + \left(\frac{x_2}{\kappa_2}\right)} + \eta x_2 + \mu x_2}{\frac{\phi\left(\frac{x_2}{\kappa_2}\right) x_2}{1 + \left(\frac{x_2}{\kappa_2}\right)} + \eta} \tag{3.5}$$

Substituting x_1^* from eq. (4) into eq. (3) and solving for x_3 gives Copy Number (CN) nullcline.

$$x_3 = \frac{\frac{\sigma\alpha}{\gamma\delta}\left(1 + \frac{x_2}{\kappa_4}\right)}{\left(1 + \frac{x_2}{\kappa_1}\right)\left(1 + \frac{x_2}{\kappa_3}\right)} \tag{3.6}$$

$$x_3 = 0 \tag{3.7}$$

Solving Eqs. (3.5-3.7) leads to steady states, both stable and unstable.

3.4.5. Sensitivity Analysis

One parameter sensitivity analysis: To assess the sensitivity of a parameter (for example y), y was varied in the range $f_1 * p \leq y \leq f_2 * p$, where p is the reference value, and $f_1 < 1$ and $f_2 > 1$ are factors, while other parameters were kept at their reference values. The steady states of the fraction of methylated repeat (M) were then calculated as a function of y. This allowed us to

visualize the range of values for a particular parameter where the system was at low steady state, bistable, and at high steady state, identifying the critical y at which the transition occurred. This result is illustrated in the bar diagram (Fig 3.5A). Each bar shows the range of values where the system is at low steady state, bistable, and at high steady state for a specific parameter. The critical y at which the system transitions from one state to another, and vice versa, is the boundary between these ranges.

Two parameters sensitivity analysis: Two parameters, y and z, were varied in the ranges $f_1 * p \leq y \leq f_2 * p$, and $g_1 * q \leq z \leq g_2 * q$, where p and q are reference values of the two chosen parameters, and $f_1, g_1 < 1$ and $f_2, g_2 > 1$ are multiplying factors. Equations (3.1-3.3) were solved for seven initial conditions (Table S3.2) to determine if the fraction of methylated repeats reached a lower steady-state, higher steady-state, or bistable state for each possible pair of (y, z) within these ranges, keeping all other parameters at their reference values. Finally, a grid of all possible (y, z) pairs within the specified ranges was created. Each point in the grid is marked with a specific color indicating whether the system is desilenced, bistable, or silenced. This allowed for visualization of the regions in two dimensions where the system is in a particular state, as illustrated in Fig 3.5B and C. The figure shows three distinct regions of stability, with clear transitions from one state to another and vice versa.

3.4.6. Polymer chain model

In this model, each repeat is either in an unmethylated (U) or methylated (M) state and is arranged as monomers in a polymer chain (see Fig 3.6). Each repeat can spontaneously convert to the other state (M to U or U to M) in the next generation. Additionally, a U repeat can convert to

an M repeat with the aid of its immediate M neighbor if present. A U repeat can also be inserted next to another U repeat in the chain, and similarly, a U repeat can be deleted from the chain.

In this model, we assume that a reaction event occurs in the next generation. Conversion events and insertion/deletion events are treated as two distinct types of reactions. The probability of a repeat conversion in the next generation is proportional to the ratio of the conversion rate of that reaction to the total conversion rate of the current generation. A repeat can either convert to another state or remain unchanged. To account for the possibility of no conversion, a fixed reaction rate is included in the total conversion rate calculation. Similarly, for insertion/deletion events, a U repeat can be inserted, deleted, or remain unchanged. The total reaction rate for insertion/deletion events is the sum of the insertion rate, deletion rate, and the rate of nothing happening. The probability of insertion or deletion of a U repeat in the next generation is proportional to the ratio of the insertion or deletion rate to the total insertion/deletion rate.

Each nth repeat in the current generation is evaluated to determine if it favors a particular reaction. For instance, if the nth repeat is M, spontaneous and cooperative conversion of U to M cannot occur for that M repeat, and thus, these reactions are set to zero for that repeat. A U repeat can undergo cooperative conversion to M if at least one of its immediate neighbors is in the M state. Additionally, a U repeat can only be inserted next to another U repeat, and only U repeats can be deleted.

The reaction rates for spontaneous conversions are fixed, while the cooperative conversion rate of U to M scales proportionally to the number of M repeats, following an increasing hill function. The insertion rate of U repeats, which decreases with the number of M repeats following

a decreasing hill function, is proportional to the amount of RNA at steady state, calculated using eq. 3.4. The deletion rate of U repeats, also following a decreasing hill function with the number of M repeats, is proportional to the number of U repeats present in the current generation.

Chapter 4: Conclusion and Future Directions

4.1.   Fission Yeast Model

Our mathematical model in fission yeast is one of the first models that incorporates the copy number of pericentromeric repeats and predicts that the number of repeats can solely govern the dynamics of pericentromeric gene silencing. A higher copy number of repeats favors a silenced state, while reduced copy numbers lead to bistability or desilencing. Results suggest that the repeat copy number is critical for gene silencing, and copy-number-dependent silencing is an effective strategy to repress the repetitive part of the genome and protect the unique part of the genome from heterochromatin silencing. The stochastic simulation highlights how faster cell division and noise can influence silencing states. The model outcomes align well with experimental data, demonstrating the importance of repeat copy numbers in gene silencing.

RNA interference (RNAi) machinery and number of its functions are conserved from unicellular eukaryotes, such as fission yeast, to higher eukaryotes (143). For instance, a recent study determines that the Argonaute (Ago) proteins of the RNAi pathway directly and uniformly repress bidirectional pericentromeric lncRNAs in a Dicer-dependent manner in mouse embryonic and adult stem cell, and upon Ago or Dicer loss, bidirectional pericentromeric lncRNAs are uniformly upregulated (144). Our copy number dependent RNAi model in fission yeast has shown that loss of siRNA biogenesis proteins, e.g., Ago1 or Dcr1, leads to increased transcription of pericentromeric repeats. In the mouse embryonic and adult stem cells, our model can be applied to investigate whether a copy number of repeats can be an effective way to silence mouse pericentromeric region. Similarly, using this model for other eukaryotic organisms with RNAi

mechanisms to see if copy-number-dependent gene silencing is at play would help generalize the findings. While the fission yeast model incorporates known epigenetic modifications, e.g., histone methylation and histone acetylation, further development of the model to include additional histone modifications and their interactions with RNAi mechanisms could provide a comprehensive understanding of heterochromatin regulation. Other findings from this model suggest that cellular noise and cell division influence heterochromatin stability. These aspects can be further explored to understand their effects on gene expression, providing deeper insights into the dynamic nature of gene regulation. Most importantly, one can test our model's prediction that loss of repeat copy will result in loss of silencing in the pericentromeric region.

4.2. Human Model

This study aims to provide insight into the dynamic nature of HSATII repeats in human pericentromere. The ODE-based model and Polymer Chain model developed in this study provide a link between HSATII repeat methylation and variation in copy number. Specifically, the ODE model predicts that HSATII copy number increases with the reduction in the fraction of HSATII repeats. This result is significant as it suggests a connection between DNA hypomethylation and copy number variation in cancer cells. Additionally, the ODE model predicts that it is possible for both higher and lower steady states of HSATII copy number to coexist. This result is supported by the polymer chain model, which also indicates that reduced methylation increases the number of HSATII repeats.

The current model can be made more sophisticated by including histone methylation, DNA methylation, and histone acetylation in HSATII repeats. Unlike the existing model, which only

considers general methylated and unmethylated repeats without distinguishing the type of methylation, this approach would provide greater detail and accuracy. Other future work may include comparative studies across different mammals to see if satellite repeat dynamics is a conserved feature. This approach may provide functional implications of repeat dynamics in genome organization and stability. Most importantly, our model predictions suggest a new approach to using copy numbers as a cancer diagnosis tool. HSATII has been observed to increase in its copy number in cancer (31). The model demonstrates a link between copy number variation and the fraction of methylated HSATII repeats and identifies parameters that promote both expansion and reduction of copy numbers. This can help investigate whether copy number increase in healthy cells leads to cancerous transformations. Using copy numbers as a diagnostic tool could be an affordable and effective method to prevent cancer metastasis. Future studies may also aim to explore the possibility of targeting HSATII methylation pathways as a therapeutic strategy.
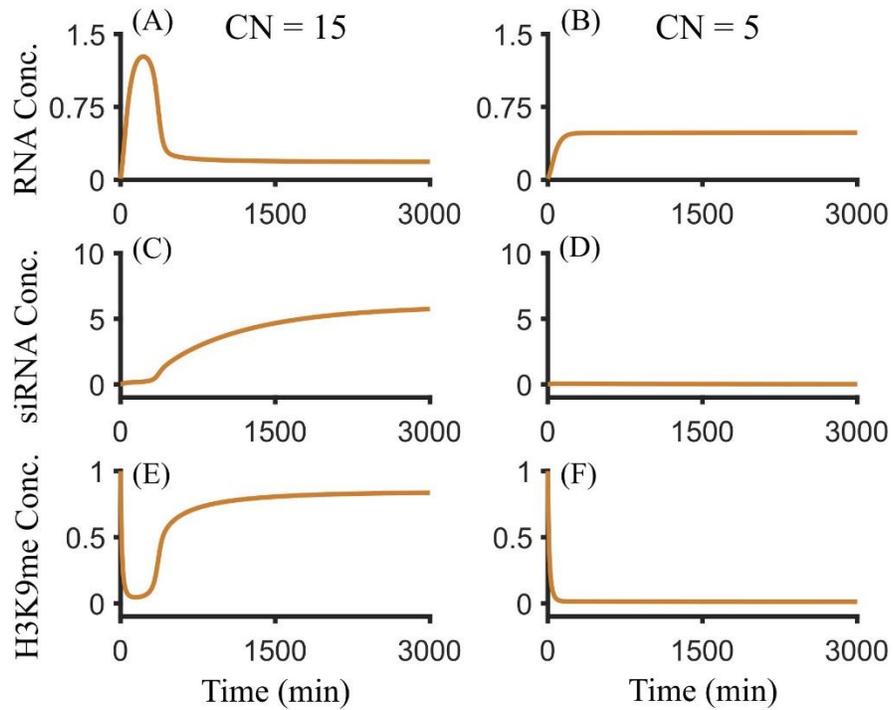
Fig S2.1. **ODE solutions of Eqs. (2.1-2.3).** ODE solution for (A, C, E) CN = 15 and (B, D, F) CN = 5 for the same initial condition. (A, C, E) When CN = 15, the system is silenced, which is represented by high steady-state H3K9me concentration. (B, D, F) When CN = 5, the system is desilenced, which is represented by low steady-state H3K9me concentration.
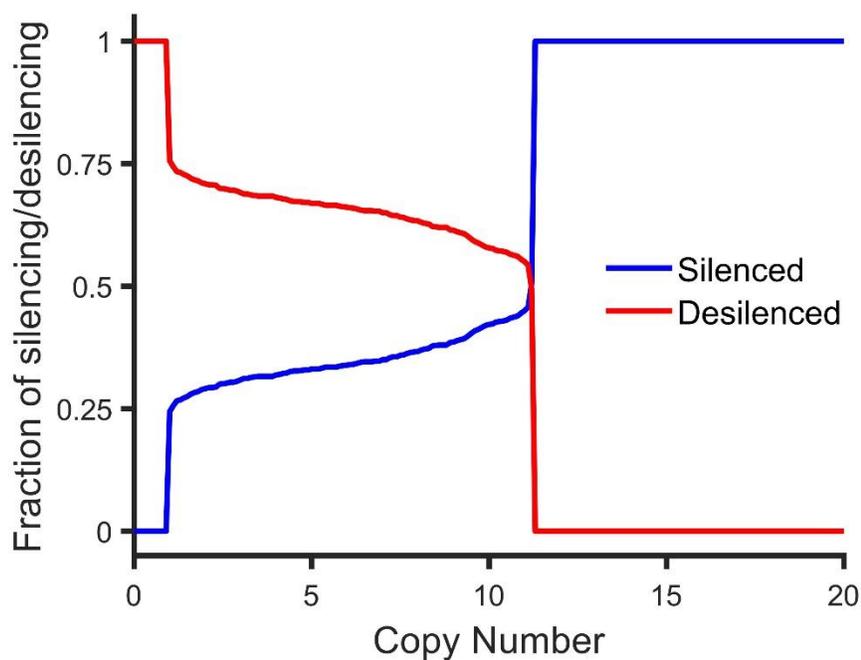
Fig S2.2. Fraction of silenced and desilenced states obtained for 1000 different initial conditions. Bistable region is from CN = 1 to CN = 11.
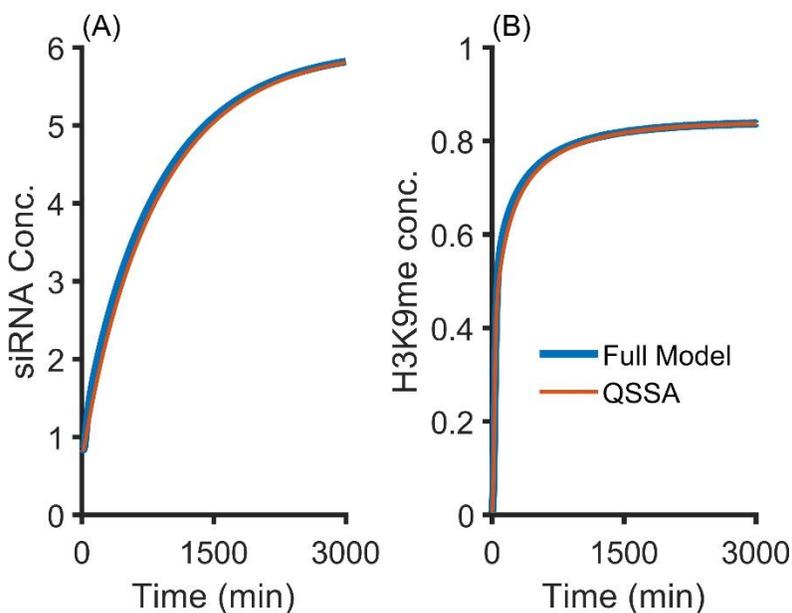
Fig S2.3. Comparison of ODE solutions between Full Model and QSSA. Comparison of time evolutions of (A) siRNA, and (B) H3K9me between full model and QSSA at CN = 15. Both siRNA (A) and (B) H3K9me concentrations evolve at similar rates between the full model and QSSA.
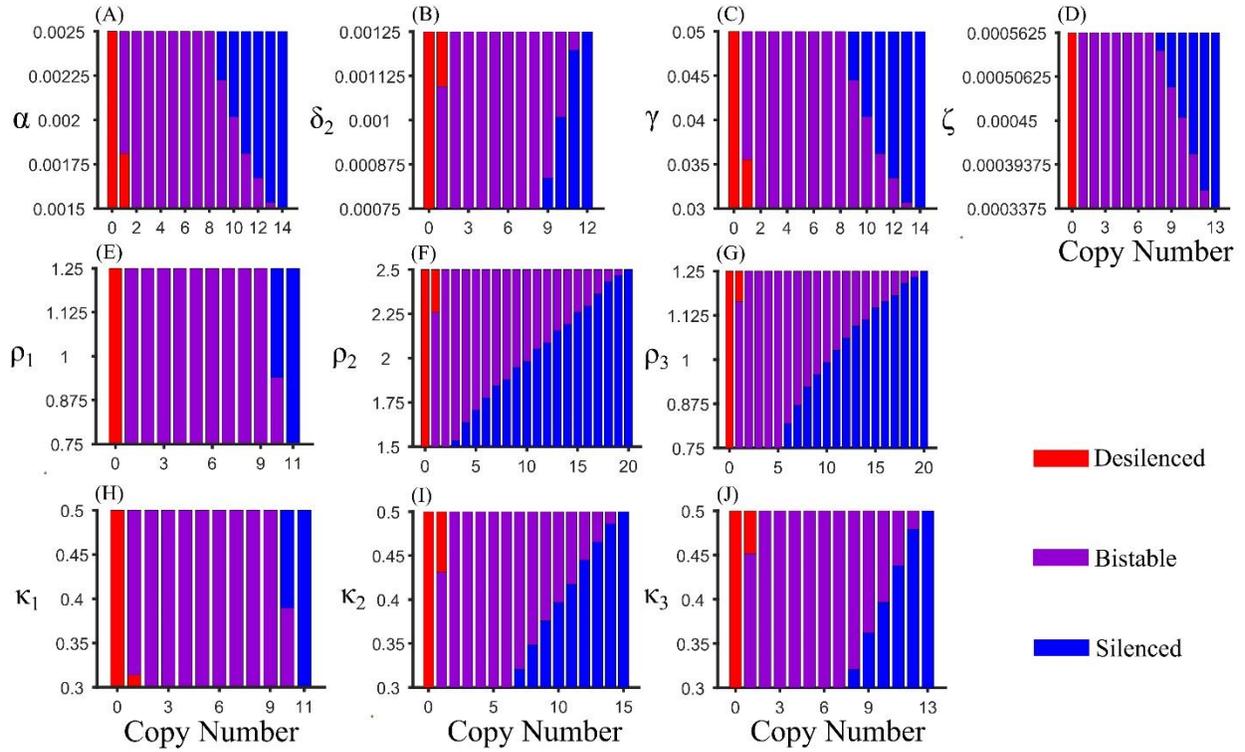
Fig S2.4. Parameter sensitivity analyses within 25% variation on default values of a specific parameter. Range of parameters (within 25% of their default values) facilitating desilenced, silenced, or bistable states at different copy numbers. Other than one parameter, all other parameters are held at the default values. High values of (A) α, (C) γ, (D) ζ, (E) ρ_1, and (H) κ_1 favor silencing, whereas high values of (B) δ_2, (F) ρ_2, (G) ρ_3, (I) κ_2, and (J) κ_3 favor bistability or desilencing. α = Transcription Rate, δ_2 = siRNA degradation rate, γ = siRNA biogenesis rate, ζ = Basal methylation rate, ρ_1, ρ_2, and ρ_3 = Hill coefficient for transcription, siRNA biogenesis, methylation by siRNA respectively. κ_1, κ_2, and κ_3 = Half maximum methylation for transcription, siRNA biogenesis, and methylation by siRNA respectively.
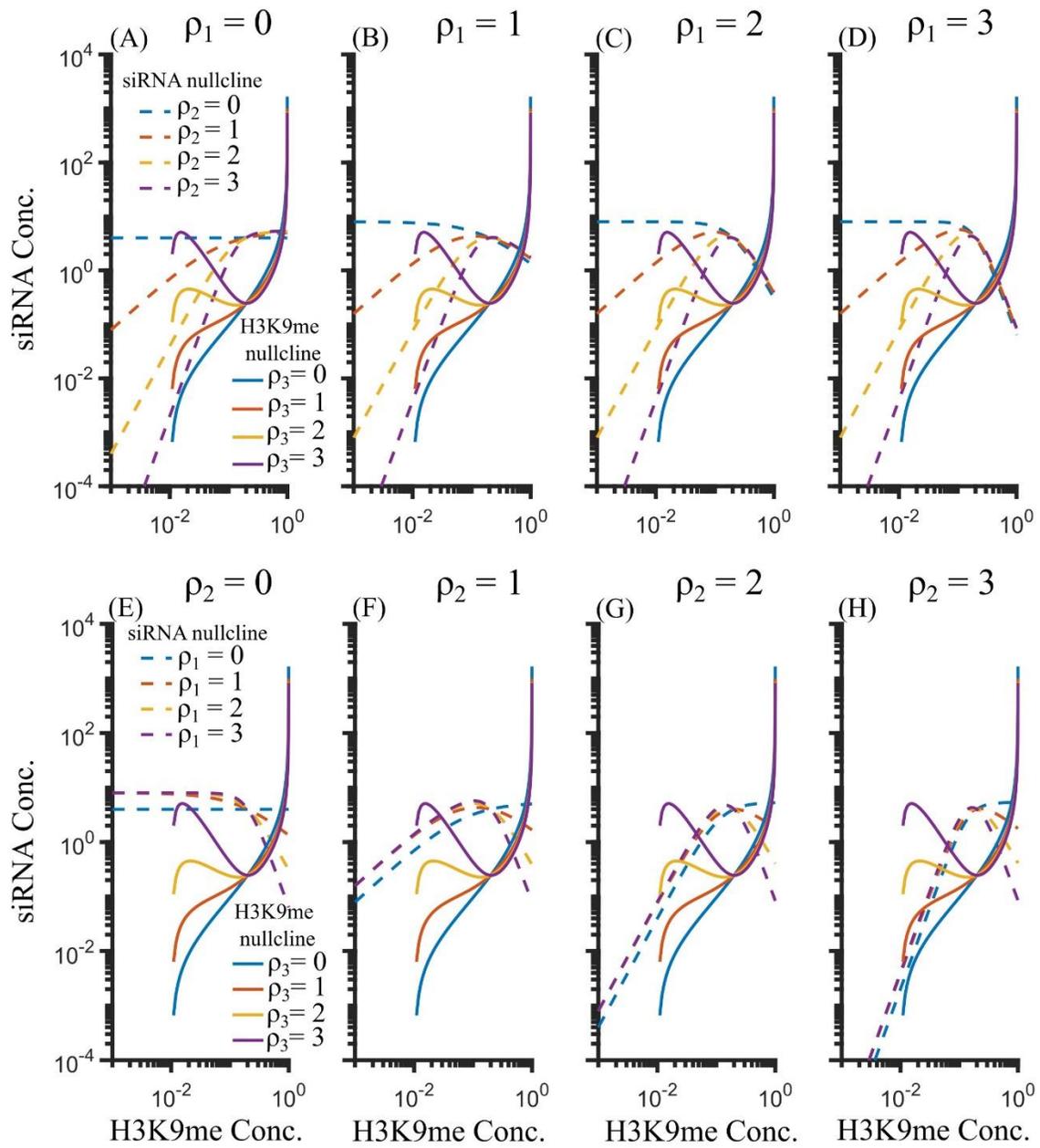
Fig S2.5. Hill coefficient dependence of nullclines. H3K9me nullcline and siRNA nullcline can intersect at three distinct points if $\rho_1 \geq 1$, $\rho_2 \geq 2$, and $\rho_3 \geq 1$ in the reference parameters.
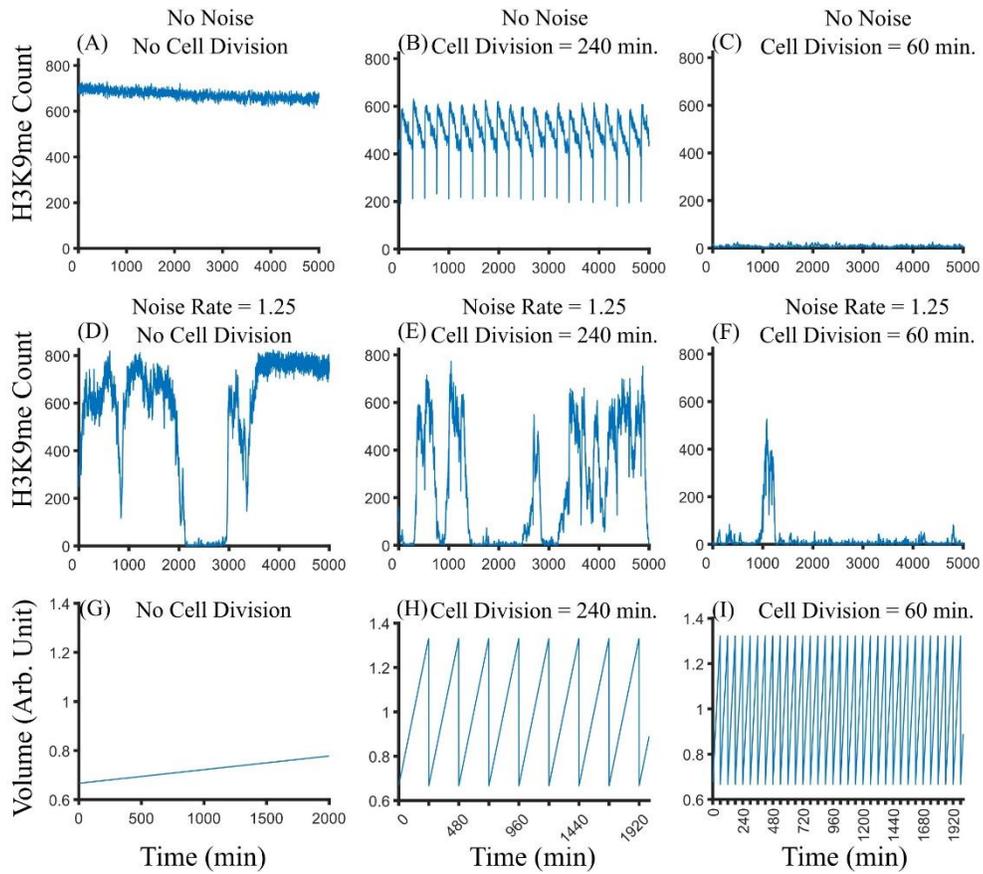
Fig S2.6. Stochastic trajectories for copy number 15 for indicated noise level and cell division. (A) Silenced state is favored for the system with no cell division and noise. (B-C) Cell division can lead to desilencing. (D-F) The H3K9me profile with noise. (G-I) The volume growth with respect to time, for different cell division of the system.
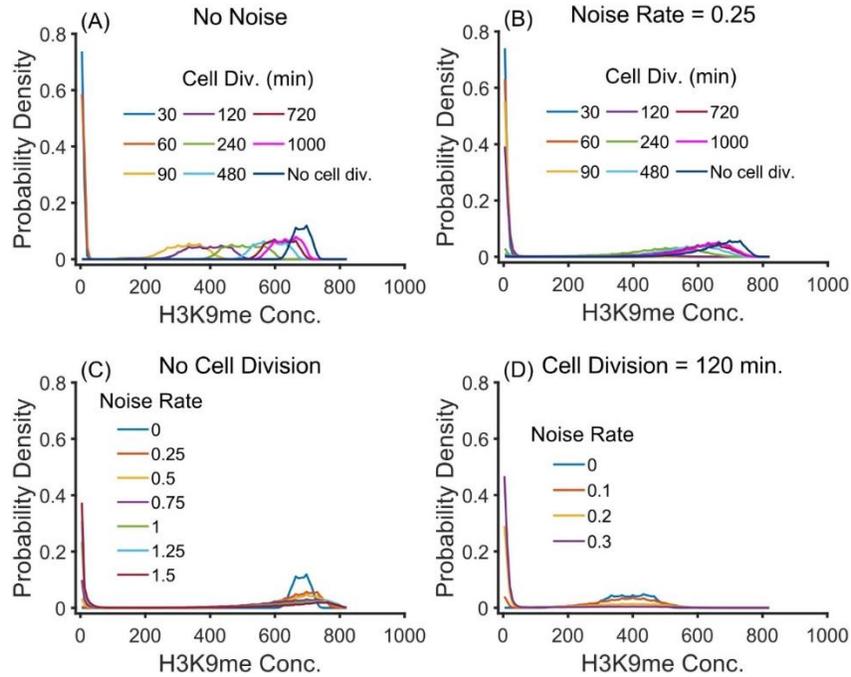
Fig S2.7. The probability density of H3K9me concentration over time at indicated noise rate and cell division for copy number 15. Faster cell division favors the desilenced state for a system (A) without noise and (B) with noise. Increase in noise brings the system down to desilenced state (C) without cell division and (D) with cell division.
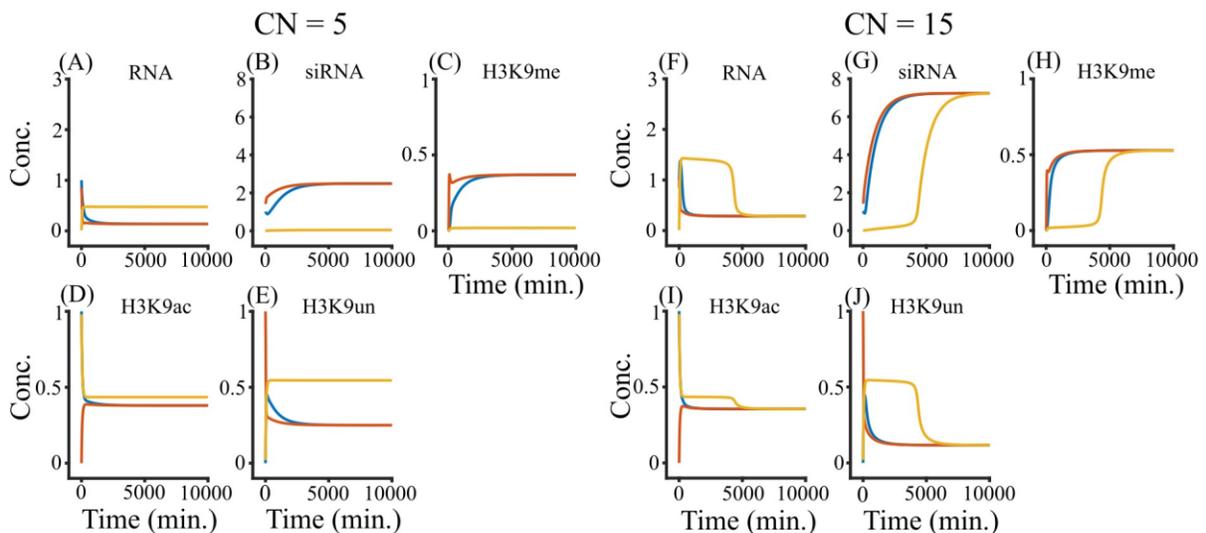
Fig S2.8. Five-variable model ODE solutions. (A-E) ODE solutions when copy number is 5. (F-J) ODE solutions when copy number is 15. Colors indicate the use of different initial conditions.
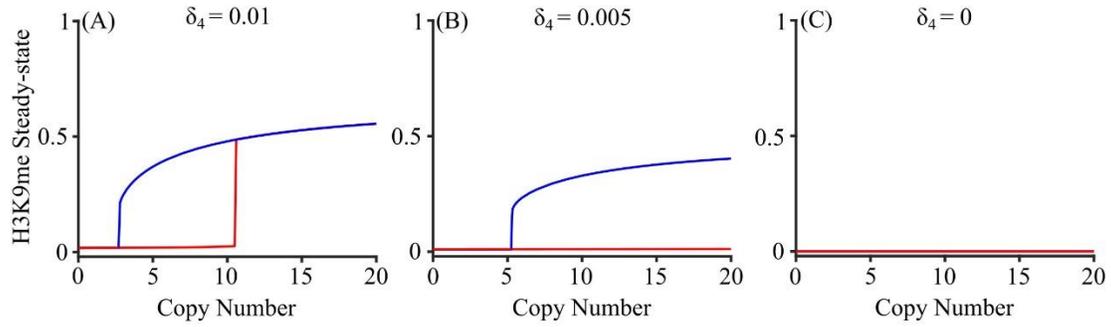
Fig S2.9. Steady-state H3K9me as a function of copy number for a five-variable model for indicated values for deacetylation rate.

Silencing	Desilencing
Transcription rate (α)	Half maximum methyl concentration for siRNA biogenesis (κ_2)
SiRNA biogenesis rate (γ)	Hill coefficient for siRNA biogenesis (ρ_2)
Methylation rate (ϵ)	SiRNA degradation rate (δ_2)
Methylation spreading rate (ϕ)	Demethylation rate (δ_3)
Basal methylation rate (ζ)	RNA degradation rate (δ_1)
Half maximum methylation conc. for transcription (κ_1)	Half maximum methylation conc. for methylation (κ_3)
Hill coefficient for transcription (ρ_1)	Hill coefficient for methylation (ρ_3)

Table S2.1. Parameters identified by the analysis that facilitate silencing or desilencing.

Parameters	Values for WT cells
Transcription rate (α)	$0.002 \ mol.min^{-1}$
RNA degradation rate (δ_1)	$0.02 \ min^{-1}$
siRNA biogenesis rate (γ)	$0.04 \ min^{-1}$
Half maximum methylation conc. for transcription (κ_1)	0.4
Hill coefficient for transcription (ρ_1)	1
Half maximum methylation conc. for siRNA biogenesis (κ_2)	0.4
Hill coefficient for siRNA biogenesis (ρ_2)	2
Half maximum methylation conc. for methylation (κ_3)	0.4
Hill coefficient for methylation (ρ_3)	1
siRNA degradation rate (δ_2)	$0.001 \ min^{-1}$
Methylation rate (ϵ)	$0.1 \ min^{-1}$
Demethylation rate (δ_3)	$0.083 \ min^{-1}$
Methylation spreading rate (ϕ)	$0.04 \ min^{-1}$
Basal methylation rate (ζ)	$0.00045 \ min^{-1}$

Table S2.2. Parameters used for solving differential Eqs. (2.1-2.3) for WT cell.

S.N.	Initial condition (RNA, siRNA, H3K9me). Unit (mol) for RNA and siRNA
1	(1, 1, 0)
2	(0.1, 0.1, 0.9)
3	(0, 1, 1)
4	(0.01, 0.01, 0.01)
5	(0, 0.5, 0)
6	(0.5, 0.5, 0.5)
7	(0.85, 1.5, 1)
8	(0.75, 0.35, 1)
9	(1.9, 2, 1)
10	(0, 3, 1)
11	(0.01, 0.01, 0.02)
12	(0.01, 0.01, 0.05)
13	(0.01, 0.01, 0.5)
14	(0.01, 0.01, 1)
15	(0.5, 0.01, 1)
16	(0.5, 0.1, 1)
17	(1,2,1)
18	(0.4, 0.6, 0)
19	(0, 0, 0.1)
20	(0.1, 0, 0)
21	(0.75, 0, 0.75)
22	(0, 0, 1)
23	(0, 4, 0)
24	(1,4,1)

Table S2.3. Initial conditions applied in the simulation (Eqs 2.1-2.3). Methylation concentration is normalized to 1.

Parameters	Values for WT cells
Transcription rate (α)	$0.002 \ mol. \ min^{-1}$
RNA degradation rate (δ_1)	$0.02 \ min^{-1}$
siRNA biogenesis rate (γ)	$0.04 \ min^{-1}$
Half maximum methylation conc. for transcription (κ_1)	0.4
Hill coefficient for transcription (ρ_1)	1
Half maximum methylation conc. for siRNA biogenesis (κ_2)	0.4
Hill coefficient for siRNA biogenesis (ρ_2)	2
siRNA degradation rate (δ_2)	$0.001 \ min^{-1}$
Half maximum methylation conc. for methylation (κ_3)	0.4
Hill coefficient for methylation (ρ_3)	1
Rate of demethylation (δ_3)	$0.045 min^{-1}$
Methylation spreading rate (ϕ)	$0.035 \ mol^{-1}. \ min^{-1}$
Basal methylation rate (ζ_1)	$0.001 \ min^{-1}$
Acetylation rate (ζ_2)	$0.0028 \ min^{-1}$
Rate of deacetylation (δ_4)	$0.01 \ min^{-1}$
Rate of histone turnover (ξ)	$0.005 \ min^{-1}$

Table S2.4. Parameters used for solving five-variable model differential Eqs. (2.8-2.12) for WT cell.

S.N	Initial condition (RNA, siRNA, H3K9me, H3K9ac, H3K9un). Unit (mol)
1	(1, 1, 0, 1, 0)
2	(0.1, 0.1, 0.9, 0.1, 0)
3	(0, 1, 1, 0, 0)
4	(0.01, 0.01, 0.01, 0.99, 0)
5	(0, 0.5, 0.3, 0.3, 0.4)
6	(0.5, 0.5, 0.5, 0, 0.5)
7	(0.85, 1.5, 0, 0, 1)
8	(0.75, 0.35, 0.2, 0.5, 0.3)
9	(1.9, 2, 0.1, 0.35, 0.55)
10	(0, 3, 0.7, 0.15, 0.15)
11	(0.01, 0.01, 0.05, 0.95, 0)
12	(0.01, 0.01, 0, 0.98, 0.02)
13	(0.01, 0.01, 0.26, 0.26, 0.48)
14	(0.01, 0.01, 0.35, 0.35, 0.3)
15	(0.5, 0.01, 0.15, 0.65, 0.2)
16	(0.5, 0.1, 0.25, 0, 0.75)
17	(1, 2, 0.21, 0.43, 0.36)
18	(0.4, 0.6, 0.5, 0.5, 0)
19	(0, 0, 0.7, 0.2, 0.1)
20	(0.1, 0, 0.1, 0.2, 0.7)
21	(0.75, 0, 0.6, 0.25, 0.15)
22	(0, 0, 1, 0, 0)
23	(0, 4, 0, 1, 0)
24	(1, 4, 1, 0, 0)

Table S2.5. Initial conditions applied in the simulation of a five-variable model (Eqs. 2.8-2.12).

Appendix B: Supplementary Information for Chapter 3
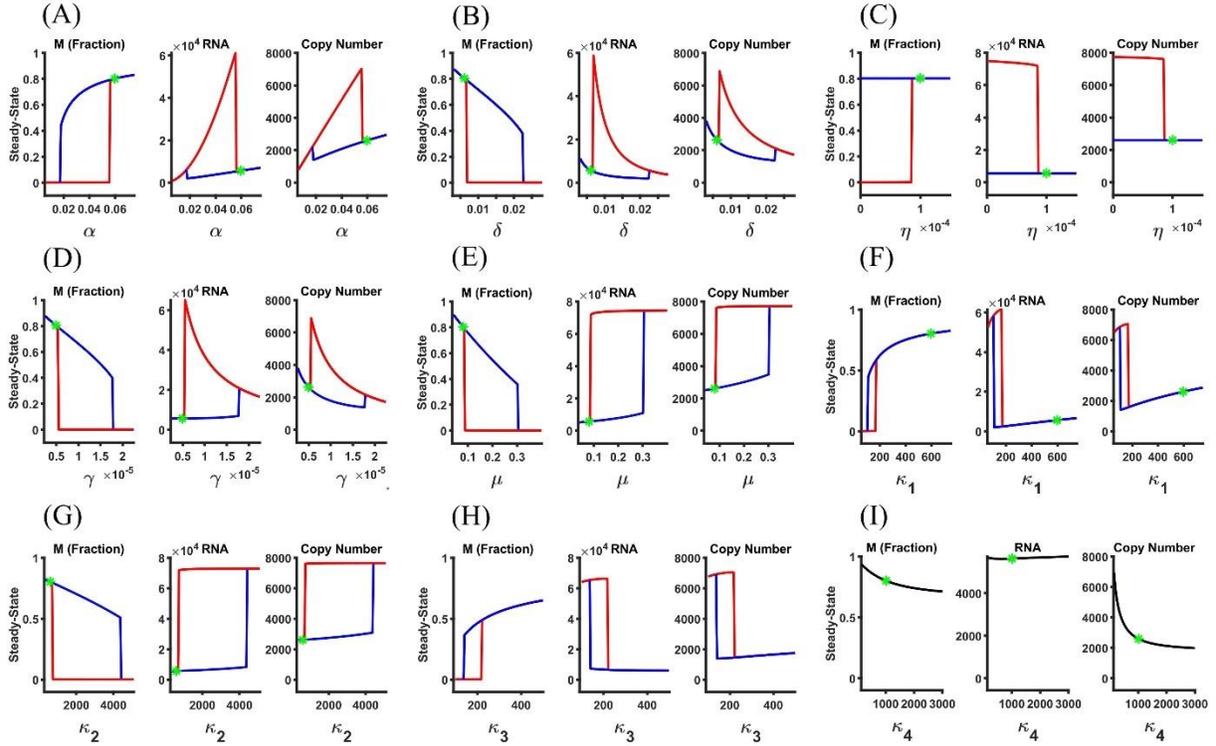
Fig S3.1. Steady-state solutions for fraction of methylated repeat, RNA, and copy number for a range of parameter values. Steady-state solutions for fraction of methylated repeat, RNA, and copy number for the range of (A) transcription rate (α), (B) RNA degradation rate (δ), (C) spontaneous conversion of U to M rate (η), (D) reverse transcription rate (γ), (E) spontaneous conversion of M to U rate (μ), (F) half maximum saturation number of M for transcription (κ_1), (G) half maximum saturation number of M for methylation spreading (κ_2), (H) half maximum saturation number of M for reverse transcription (κ_3), and (I) half maximum saturation number of M for copy number recombination (κ_4). Except κ_4 (I), the model allows a range of values for lower steady-state, coexistence of both lower and higher steady-state, and higher steady-state. The green star represents the reference value of the parameter.
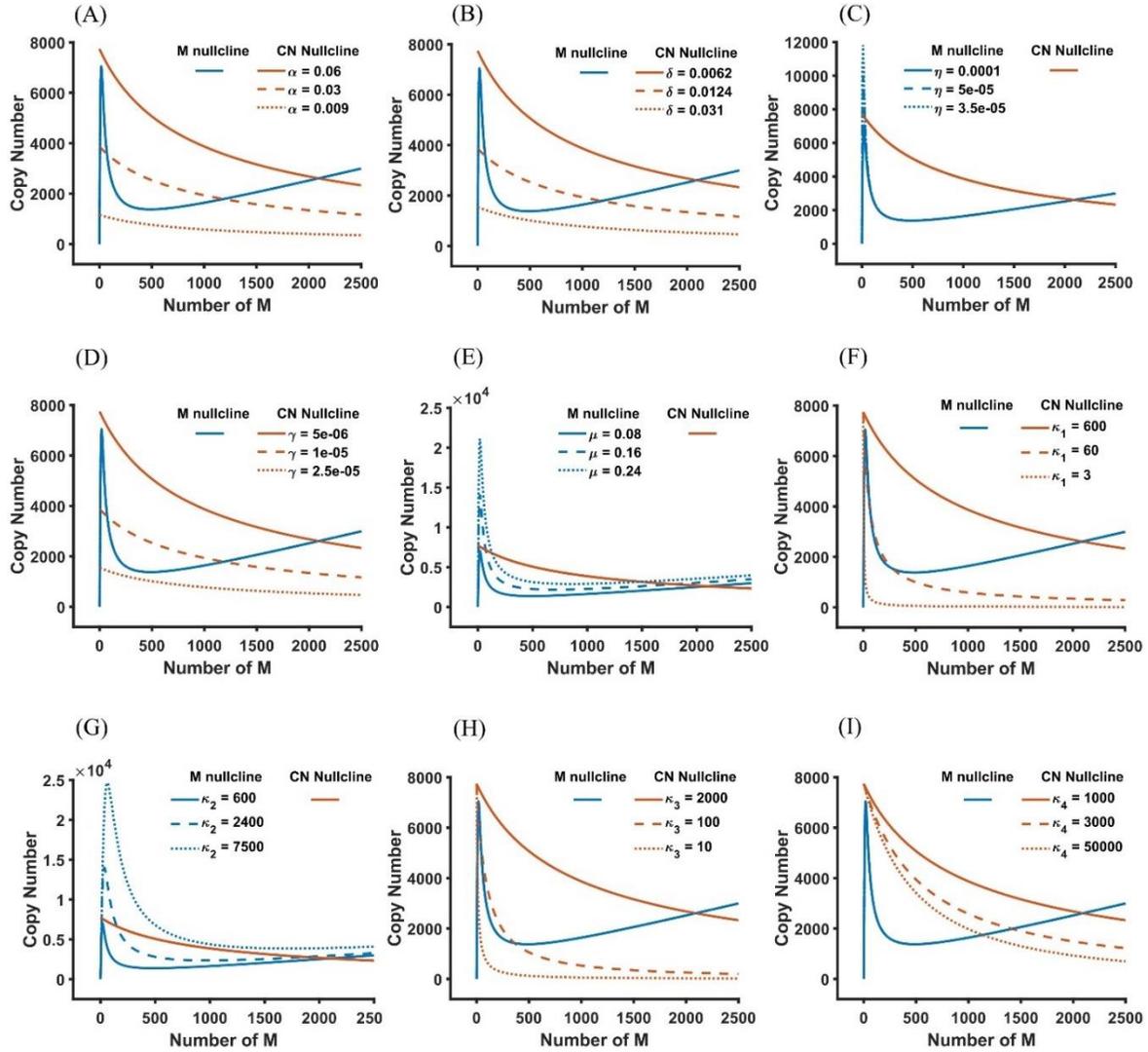
Fig S3.2. Nullcline analysis for different values of (A) transcription rate (α), (B) RNA degradation rate (δ), (C) spontaneous conversion of U to M rate (η), (D) reverse transcription rate (γ), (E) spontaneous conversion of M to U rate (μ), (F) half maximum saturation number of M for transcription (κ_1), (G) half maximum saturation number of M for methylation spreading (κ_2), (H) half maximum saturation number of M for reverse transcription (κ_3), and (I) half maximum saturation number of M for copy number recombination (κ_4). CN nullcline shifts downward and to the left with the decrease in (A) α, (F) κ_1, (H) κ_3 from their reference values (shown as a solid line), changing the number of intersection pts from one to three (bistable state) to one. Decrease in (I) κ_4 also shifts the CN nullcline downward and to the left but the number of intersection points doesn't change. CN nullcline shifts downward and to the left with the increase in (B) δ, and (D) γ changing the number of intersection pts from one to three (bistable state) to one. Decrease in (C) η moves the M nullcline upward, resulting in three intersection points from one intersection point. The shift only happens in the peaked (left) part of the CN nullcline, with no change in the right part (or somehow flat part). Increase in (E) μ, and (G) κ_2 shifts the M nullcline upward and to the right from their reference values (solid line), changing the number of intersection points from one to three (bistable) to one. The solid line represents the nullcline at the reference value in all the figures.

78

Parameters	Values for WT cells	References
Transcription rate (α)	$0.06\ min^{-1}$	*Paolo Maiuri et. al., EMBO Reports, 2011*
Half maximum saturation number of M for transcription (κ_1)	600	
RNA degradation rate (δ)	$0.0062\ min^{-1}$	*Kaiwen Shi, et. al. Plos Comp. Biol. 2021*
Cooperative conversion rate of U to M (ϕ)	$0.0002\ min^{-1}$	
Half maximum saturation number of M for cooperative methylation (κ_2)	600	
Spontaneous conversion of U to M (η)	$0.0001\ min^{-1}$	
Spontaneous conversion rate of M to U (μ)	$0.08\ min^{-1}$	
Reverse transcription rate (σ)	$0.004\ min^{-1}$	
Half maximum saturation number of M for reverse transcription (κ_3)	2000	
Repeat recombination rate (γ)	$0.000005\ min^{-1}$	
Half maximum saturation number of M for reverse transcription (κ_4)	1000	

Table S3.1. Parameter values used for solving differential equations (3.1-3.3)

S. N	Initial Conditions (RNA, M number, Copy Number)
1	(0, 10, 20)
2	(2.04e4, 779, 5212)
3	(10000, 1000, 6000)
4	(20000, 500, 5000)
5	(2000, 1000, 5000)
6	(500, 500, 10000)
7	(0, 2000, 10000)

Table S3.2. Initial Conditions Applied in the Simulation (Eqs. 3.1-3.3)

References:

1. Chromatin [Internet]. [cited 2024 Jul 22]. Available from: https://www.genome.gov/genetics-glossary/Chromatin

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4th ed. Garland Science; 2002.

3. Shahid Z, Simpson B, Miao KH, Singh G. Genetics, Histone Code. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Jul 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK538477/

4. Jeffries MA. Epigenetic editing: How cutting-edge targeted epigenetic modification might provide novel avenues for autoimmune disease therapy. Clinical Immunology. 2018 Nov 1;196:49–58.

5. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000 Jan;403(6765):41–5.

6. Huisinga KL, Brower-Toland B, Elgin SCR. The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma. 2006 Apr 1;115(2):110–22.

7. Jenuwein T, Allis CD. Translating the Histone Code. Science. 2001 Aug 10;293(5532):1074–80.

8. Nakayama J ichi, Rice JC, Strahl BD, Allis CD, Grewal SIS. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science. 2001 Apr 6;292(5514):110–3.

9. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000 Aug;406(6796):593–9.

10. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. The EMBO Journal. 1994 Aug;13(16):3822–31.

11. Keshet I, Lieman-Hurwitz J, Cedar H. DNA methylation affects the formation of active chromatin. Cell. 1986 Feb 28;44(4):535–43.

12. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008 Jun;9(6):465–76.

13. Grunstein M. Yeast Heterochromatin: Regulation of Its Assembly and Inheritance by Histones. Cell. 1998 May 1;93(3):325–8.

14. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation Between Histone Lysine Methylation and Developmental Changes at the Chicken β-Globin Locus. Science. 2001 Sep 28;293(5539):2453–5.

15. Noma K ichi, Allis CD, Grewal SIS. Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries. Science. 2001 Aug 10;293(5532):1150–5.

16. Blasco MA. The epigenetic regulation of mammalian telomeres. Nat Rev Genet. 2007 Apr;8(4):299–309.

17. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO Journal. 2005 Feb 23;24(4):800–12.

18. Schoeftner S, Blasco MA. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. The EMBO Journal. 2009 Aug 19;28(16):2323–36.

19. Schueler MG, Sullivan BA. Structural and Functional Dynamics of Human Centromeric Chromatin. Annual Review of Genomics and Human Genetics. 2006 Sep 22;7(Volume 7, 2006):301–13.

20. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007 Apr;8(4):272–85.

21. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol. 2016 Jan;17(1):16–29.

22. Dillon N. Heterochromatin structure and function. Biology of the Cell. 2004;96(8):631–7.

23. Peters AHFM, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, et al. Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. Cell. 2001 Nov 2;107(3):323–37.

24. Rhind N, Gilbert DM. DNA Replication Timing. Cold Spring Harb Perspect Biol. 2013 Aug 1;5(8):a010132.

25. Grewal SIS, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007 Jan;8(1):35–46.

26. Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. Commun Biol. 2023 Sep 19;6(1):1–21.

27. Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes & Genetic Systems. 2019;94(6):233–52.

28. Genovese LM, Geraci F, Corrado L, Mangano E, D'Aurizio R, Bordoni R, et al. A Census of Tandemly Repeated Polymorphic Loci in Genic Regions Through the Comparative Integration of Human Genome Assemblies. Front Genet [Internet]. 2018 May 2 [cited 2024 Jul 22];9. Available from: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2018.00155/full

29. Cechova M, Harris RS, Tomaszkiewicz M, Arbeithuber B, Chiaromonte F, Makova KD. High Satellite Repeat Turnover in Great Apes Studied with Short- and Long-Read Technologies. Molecular Biology and Evolution. 2019 Nov 1;36(11):2415–31.

30. Zaratiegui M, Irvine DV, Martienssen RA. Noncoding RNAs and Gene Silencing. Cell. 2007 Feb 23;128(4):763–76.

31. Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, et al. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. Proceedings of the National Academy of Sciences. 2015 Dec 8;112(49):15148–53.

32. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, et al. Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. Science. 2011 Feb 4;331(6017):593–6.

33. Goto DB, Nakayama J ichi. RNA and epigenetic silencing: Insight from fission yeast. Development, Growth & Differentiation. 2012;54(1):129–41.

34. Hoffman CS, Wood V, Fantes PA. An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. Genetics. 2015 Oct 1;201(2):403–23.

35. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015 Feb;16(2):71–84.

36. Spatz A, Borg C, Feunteun J. X-Chromosome Genetics and Human Cancer. Nat Rev Cancer. 2004 Aug;4(8):617–29.

37. Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D. Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs. Cell. 2004 Dec 17;119(6):789–802.

38. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SIS, et al. RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Science. 2004 Jan 30;303(5658):672–6.

39. Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, et al. Stress-induced transcription of satellite III repeats. Journal of Cell Biology. 2003 Dec 29;164(1):25–33.

40. Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, et al. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Research. 2008 Feb 1;36(2):423–34.

41. Ellermeier C, Higuchi EC, Phadnis N, Holm L, Geelhood JL, Thon G, et al. RNAi and heterochromatin repress centromeric meiotic recombination. Proceedings of the National Academy of Sciences. 2010 May 11;107(19):8701–5.

42. Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, et al. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature. 2011 Sep;477(7363):179–84.

43. Zhu Q, Hoong N, Aslanian A, Hara T, Benner C, Heinz S, et al. Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Molecular Cell. 2018 Jun 7;70(5):842-853.e7.

44. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001 Mar;410(6824):120–4.

45. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001 Mar;410(6824):116–20.

46. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007 Feb 23;128(4):693–705.

47. Alper BJ, Lowe BR, Partridge JF. Centromeric heterochromatin assembly in fission yeast—balancing transcription, RNA interference and chromatin modification. Chromosome Res. 2012 Jul 1;20(5):521–34.

48. Allshire RC, Ekwall K. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. Cold Spring Harb Perspect Biol. 2015 Jul 1;7(7):a018770.

49. Martienssen R, Moazed D. RNAi and Heterochromatin Assembly. Cold Spring Harb Perspect Biol. 2015 Aug 1;7(8):a019323.

50. Gerace EL, Halic M, Moazed D. The Methyltransferase Activity of Clr4Suv39h Triggers RNAi Independently of Histone H3K9 Methylation. Molecular Cell. 2010 Aug 13;39(3):360–72.

51. Noma K ichi, Sugiyama T, Cam H, Verdel A, Zofall M, Jia S, et al. RITS acts in cis to promote RNA interference–mediated transcriptional and post-transcriptional silencing. Nat Genet. 2004 Nov;36(11):1174–80.

52. Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SIS. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to

siRNA production. Proceedings of the National Academy of Sciences. 2005 Jan 4;102(1):152–7.

53. Brändle F, Frühbauer B, Jagannathan M. Principles and functions of pericentromeric satellite DNA clustering into chromocenters. Seminars in Cell & Developmental Biology. 2022 Aug 1;128:26–39.

54. Thakur J, Packiaraj J, Henikoff S. Sequence, Chromatin and Evolution of Satellite DNA. International Journal of Molecular Sciences. 2021 Jan;22(9):4309.

55. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. Cell. 2014 Feb 13;156(4):812–24.

56. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012 May 1;22(5):885–98.

57. Motamedi MR, Hong EJE, Li X, Gerber S, Denison C, Gygi S, et al. HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms. Molecular Cell. 2008 Dec 26;32(6):778–90.

58. Li PC, Green MD, Forsburg SL. Mutations Disrupting Histone Methylation Have Different Effects on Replication Timing in S. pombe Centromere. PLOS ONE. 2013 May 1;8(5):e61464.

59. Buscaino A, Lejeune E, Audergon P, Hamilton G, Pidoux A, Allshire RC. Distinct roles for Sir2 and RNAi in centromeric heterochromatin nucleation, spreading and maintenance. The EMBO Journal. 2013 May 2;32(9):1250–64.

60. Grewal SIS, Klar AJS. Chromosomal Inheritance of Epigenetic States in Fission Yeast During Mitosis and Meiosis. Cell. 1996 Jul 12;86(1):95–101.

61. Yamada T, Fischle W, Sugiyama T, Allis CD, Grewal SIS. The Nucleation and Maintenance of Heterochromatin by a Histone Deacetylase in Fission Yeast. Molecular Cell. 2005 Oct 28;20(2):173–85.

62. Thon G, Cohen A, Klar AJ. Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of Schizosaccharomyces pombe. Genetics. 1994 Sep 1;138(1):29–38.

63. Ekwall K, Ruusala T. Mutations in rik1, clr2, clr3 and clr4 genes asymmetrically derepress the silent mating-type loci in fission yeast. Genetics. 1994 Jan 1;136(1):53–64.

64. Shankaranarayana GD, Motamedi MR, Moazed D, Grewal SIS. Sir2 Regulates Histone H3 Lysine 9 Methylation and Heterochromatin Assembly in Fission Yeast. Current Biology. 2003 Jul 15;13(14):1240–6.

65. Alper BJ, Job G, Yadav RK, Shanker S, Lowe BR, Partridge JF. Sir2 is required for Clr4 to initiate centromeric heterochromatin assembly in fission yeast. The EMBO Journal. 2013 Aug 28;32(17):2321–35.

66. Watts BR, Wittmann S, Wery M, Gautier C, Kus K, Birot A, et al. Histone deacetylation promotes transcriptional silencing at facultative heterochromatin. Nucleic Acids Research. 2018 Jun 20;46(11):5426–40.

67. Zofall M, Sandhu R, Holla S, Wheeler D, Grewal SIS. Histone deacetylation primes self-propagation of heterochromatin domains to promote epigenetic inheritance. Nat Struct Mol Biol. 2022 Sep;29(9):898–909.

68. Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen ON, et al. Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 2015 Mar 15;29(6):585–90.

69. Dodd IB, Micheelsen MA, Sneppen K, Thon G. Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. Cell. 2007 May 18;129(4):813–22.

70. Pisacane P, Halic M. Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi. Nat Commun. 2017 May 25;8(1):15332.

71. Hoerter JAH, Krishnan V, Lionberger TA, Walter NG. siRNA-Like Double-Stranded RNAs Are Specifically Protected Against Degradation in Human Cell Extract. PLOS ONE. 2011 May 27;6(5):e20359.

72. ScienceSites. Steven Strogatz. 2014 [cited 2024 Jul 23]. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Available from: https://www.stevenstrogatz.com/books/nonlinear-dynamics-and-chaos-with-applications-to-physics-biology-chemistry-and-engineering

73. Akoury E, Ma G, Demolin S, Brönner C, Zocco M, Cirilo A, et al. Disordered region of H3K9 methyltransferase Clr4 binds the nucleosome and contributes to its activity. Nucleic Acids Research. 2019 Jul 26;47(13):6726–36.

74. Tucker JF, Ohle C, Schermann G, Bendrin K, Zhang W, Fischer T, et al. A Novel Epigenetic Silencing Pathway Involving the Highly Conserved 5'-3' Exoribonuclease Dhp1/Rat1/Xrn2 in Schizosaccharomyces pombe. PLOS Genetics. 2016 Feb 18;12(2):e1005873.

75. Provost P, Silverstein RA, Dishart D, Walfridsson J, Djupedal I, Kniola B, et al. Dicer is required for chromosome segregation and gene silencing in fission yeast cells. Proceedings of the National Academy of Sciences. 2002 Dec 24;99(26):16648–53.

76. Iida T, Kawaguchi R, Nakayama J ichi. Conserved Ribonuclease, Eri1, Negatively Regulates Heterochromatin Assembly in Fission Yeast. Current Biology. 2006 Jul 25;16(14):1459–64.

77. Freeman-Cook LL, Gómez EB, Spedale EJ, Marlett J, Forsburg SL, Pillus L, et al. Conserved Locus-Specific Silencing Functions of Schizosaccharomyces pombe sir2+. Genetics. 2005 Mar 1;169(3):1243–60.

78. Hansen KR, Hazan I, Shanker S, Watt S, Verhein-Hansen J, Bähler J, et al. H3K9me-Independent Gene Silencing in Fission Yeast Heterochromatin by Clr5 and Histone Deacetylases. PLOS Genetics. 2011 Jan 6;7(1):e1001268.

79. Nakayama J ichi, Klar AJS, Grewal SIS. A Chromodomain Protein, Swi6, Performs Imprinting Functions in Fission Yeast during Mitosis and Meiosis. Cell. 2000 Apr 28;101(3):307–17.

80. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics. 1976 Dec 1;22(4):403–34.

81. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977 Dec 1;81(25):2340–61.

82. Joh RI, Khanduja JS, Calvo IA, Mistry M, Palmieri CM, Savol AJ, et al. Survival in Quiescence Requires the Euchromatic Deployment of Clr4/SUV39H by Argonaute-Associated Small RNAs. Molecular Cell. 2016 Dec 15;64(6):1088–101.

83. Gallagher PS, Larkin M, Thillainadesan G, Dhakshnamoorthy J, Balachandran V, Xiao H, et al. Iron homeostasis regulates facultative heterochromatin assembly in adaptive genome control. Nat Struct Mol Biol. 2018 May;25(5):372–83.

84. O'Kane CJ, Hyland EM. Yeast epigenetics: the inheritance of histone modification states. Bioscience Reports. 2019 May 7;39(5):BSR20182006.

85. Qian H. Cooperativity in Cellular Biochemical Processes: Noise-Enhanced Sensitivity, Fluctuating Enzyme, Bistability with Nonlinear Feedback, and Other Mechanisms for Sigmoidal Responses. Annual Review of Biophysics. 2012 Jun 9;41(Volume 41, 2012):179–204.

86. Braun JE, Huntzinger E, Fauser M, Izaurralde E. GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. Molecular Cell. 2011 Oct 7;44(1):120–33.

87. Lan F, Zaratiegui M, Villén J, Vaughn MW, Verdel A, Huarte M, et al. S. pombe LSD1 Homologs Regulate Heterochromatin Propagation and Euchromatic Gene Transcription. Molecular Cell. 2007 Apr 13;26(1):89–101.

88. Abdulla AZ, Vaillant C, Jost D. Painters in chromatin: a unified quantitative framework to systematically characterize epigenome regulation and memory. Nucleic Acids Research. 2022 Sep 9;50(16):9083–104.

89. Muhammad T, Zhang F, Zhang Y, Liang Y. RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. Cells. 2019 Jan;8(1):38.

90. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philosophical Transactions of the Royal Society B: Biological Sciences. 2008 Oct 16;364(1513):99–115.

91. Horvath P, Barrangou R. CRISPR/Cas, the Immune System of Bacteria and Archaea. Science. 2010 Jan 8;327(5962):167–70.

92. Gu X, Ye T, Zhang XR, Nie L, Wang H, Li W, et al. Single-chromosome fission yeast models reveal the configuration robustness of a functional genome. Cell Reports [Internet]. 2022 Aug 23 [cited 2024 Jul 23];40(8). Available from: https://www.cell.com/cell-reports/abstract/S2211-1247(22)01054-3

93. Pasero P, Marilley M. Size variation of rDNA clusters in the yeasts Saccharomyces cerevisiea and Schizosaccharomyces pombe. Molec Gen Genet. 1993 Jan 1;236(2):448–52.

94. Yamamoto I, Nakaoka H, Takikawa M, Tashiro S, Kanoh J, Miyoshi T, et al. Fission yeast Stn1 maintains stability of repetitive DNA at subtelomere and ribosomal DNA regions. Nucleic Acids Research. 2021 Oct 11;49(18):10465–76.

95. Jeffares DC, Rallis C, Rieux A, Speed D, Převorovský M, Mourier T, et al. The genomic and phenotypic diversity of Schizosaccharomyces pombe. Nat Genet. 2015 Mar;47(3):235–41.

96. Jeffares DC. The natural diversity and ecology of fission yeast. Yeast. 2018;35(3):253–60.

97. Kobayashi T. Ribosomal RNA gene repeats, their stability and cellular senescence. Proceedings of the Japan Academy, Series B. 2014;90(4):119–29.

98. Yokoyama M, Sasaki M, Kobayashi T. Spt4 promotes cellular senescence by activating non-coding RNA transcription in ribosomal RNA gene clusters. Cell Reports [Internet]. 2023 Jan 31 [cited 2024 Jul 23];42(1). Available from: https://www.cell.com/cell-reports/abstract/S2211-1247(22)01845-9

99. Ershova ES, Malinovskaya EM, Konkova MS, Veiko RV, Umriukhin PE, Martynov AV, et al. Copy Number Variation of Human Satellite III (1q12) With Aging. Front Genet [Internet].

2019 Aug 7 [cited 2024 Jul 23];10. Available from: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00704/full

100. Thomas JW, Schueler MG, Summers TJ, Blakesley RW, McDowell JC, Thomas PJ, et al. Pericentromeric Duplications in the Laboratory Mouse. Genome Res. 2003 Jan 1;13(1):55–63.

101. Wei Y, Yang CR, Wei YP, Zhao ZA, Hou Y, Schatten H, et al. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. Proceedings of the National Academy of Sciences. 2014 Feb 4;111(5):1873–8.

102. MathWorks - Makers of MATLAB and Simulink [Internet]. [cited 2024 Jul 23]. Available from: https://www.mathworks.com/

103. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability [Internet]. Oakland, CA, USA; 1967 [cited 2024 Jul 21]. p. 281–97. Available from: https://books.google.com/books?hl=en&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=Classification+and+analysis+of+multivariate+observations.+MacQueen+1967&ots=nQUcH2LhnM&sig=tS7ZpgnF68Z_9dUeNWzO27n-DAY

104. Angel A, Song J, Dean C, Howard M. A Polycomb-based switch underlying quantitative epigenetic memory. Nature. 2011 Aug;476(7358):105–8.

105. Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen ON, et al. Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 2015 Mar 15;29(6):585–90.

106. Hayden KE, Strome ED, Merrett SL, Lee HR, Rudd MK, Willard HF. Sequences Associated with Centromere Competency in the Human Genome. Molecular and Cellular Biology. 2013 Feb 1;33(4):763–72.

107. Willard HF. Centromeres of mammalian chromosomes. Trends in Genetics. 1990 Jan 1;6:410–6.

108. Bierhoff H, Postepska-Igielska A, Grummt I. Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics. 2014 Jan 1;9(1):53–61.

109. Vos LJ, Famulski JK, Chan GKT. How to build a centromere: from centromeric and pericentromeric chromatin to kinetochore assemblyThis paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. Biochem Cell Biol. 2006 Aug;84(4):619–39.

110. Brown SW. Heterochromatin. Science. 1966 Jan 28;151(3709):417–25.

111. Francastel C, Magdinier F. DNA methylation in satellite repeats disorders. Essays in Biochemistry. 2019 Aug 6;63(6):757–71.

112. Dimitri P, Junakovic N. Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. Trends in Genetics. 1999 Apr 1;15(4):123–4.

113. Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, et al. Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. Molecular Cell. 2003 Dec 1;12(6):1591–8.

114. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 2004 Jun 1;18(11):1251–62.

115. 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. 1992 Mar 1;101(5):322–32.

116. Carone DM, Lawrence JB. Heterochromatin instability in cancer: From the Barr body to satellites and the nuclear periphery. Seminars in Cancer Biology. 2013 Apr 1;23(2):99–108.

117. Ehrlich M. Dna Hypomethylation In Cancer Cells. Epigenomics. 2009 Dec 1;1(2):239–59.

118. Mahmood N, Rabbani SA. DNA Methylation Readers and Cancer: Mechanistic and Therapeutic Applications. Front Oncol [Internet]. 2019 Jun 11 [cited 2024 Jul 17];9. Available from: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2019.00489/full

119. Peters AHFM, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AAHA, Perez-Burgos L, et al. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. Molecular Cell. 2003 Dec 1;12(6):1577–89.

120. Rao VK, Pal A, Taneja R. A drive in SUVs: From development to disease. Epigenetics. 2017 Mar 4;12(3):177–86.

121. Aagaard L, Laible G, Selenko P, Schmid M, Dorn R, Schotta G, et al. Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. The EMBO Journal. 1999 Apr;18(7):1923–38.

122. Yamamoto K, Sonoda M. Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. Biochemical and Biophysical Research Communications. 2003 Feb 7;301(2):287–92.

123. Lehnertz B, Ueda Y, Derijck AAHA, Braunschweig U, Perez-Burgos L, Kubicek S, et al. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current Biology. 2003 Jul 15;13(14):1192–200.

124. O'Carroll D, Scherthan H, Peters AHFM, Opravil S, Haynes AR, Laible G, et al. Isolation and Characterization of Suv39h2, a Second Histone H3 Methyltransferase Gene That Displays Testis-Specific Expression. Molecular and Cellular Biology. 2000 Dec 1;20(24):9423–33.

125. Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callén E, et al. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev. 2008 Aug 1;22(15):2048–61.

126. Probst AV, Almouzni G. Pericentric heterochromatin: dynamic organization during early development in mammals. Differentiation. 2008 Jan 1;76(1):15–23.

127. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell. 1999 Oct 29;99(3):247–57.

128. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science. 2011 Sep 2;333(6047):1303–7.

129. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Molecular Cell. 2011 May 20;42(4):451–64.

130. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Research. 2012 Jun 1;40(11):4841–9.

131. Chu Y, Chen Y, Guo H, Li M, Wang B, Shi D, et al. SUV39H1 regulates the progression of MLL-AF9-induced acute myeloid leukemia. Oncogene. 2020;39(50):7239–52.

132. Li B, Zheng Y, Yang L. The Oncogenic Potential of SUV39H2: A Comprehensive and Perspective View. J Cancer. 2019 Jan 1;10(3):721–9.

133. Ferreira D, Meles S, Escudeiro A, Mendes-da-Silva A, Adega F, Chaves R. Satellite non-coding RNAs: the emerging players in cells, cellular pathways and cancer. Chromosome Res. 2015 Sep 1;23(3):479–93.

134. Eymery A, Horard B, Atifi-Borel ME, Fourel G, Berger F, Vitte AL, et al. A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Research. 2009 Oct 1;37(19):6340–54.

135. Casanova M, Pasternak M, El Marjou F, Le Baccon P, Probst AV, Almouzni G. Heterochromatin Reorganization during Early Mouse Development Requires a Single-Stranded Noncoding Transcript. Cell Reports. 2013 Sep 26;4(6):1156–67.

136. Probst AV, Okamoto I, Casanova M, Marjou FE, Baccon PL, Almouzni G. A Strand-Specific Burst in Transcription of Pericentric Satellites Is Required for Chromocenter Formation and Early Mouse Development. Developmental Cell. 2010 Oct 19;19(4):625–38.

137. Saksouk N, Simboeck E, Déjardin J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics & Chromatin. 2015 Jan 15;8(1):3.

138. Ichida K, Suzuki K, Fukui T, Takayama Y, Kakizawa N, Watanabe F, et al. Overexpression of satellite alpha transcripts leads to chromosomal instability via segregation errors at specific chromosomes. International Journal of Oncology. 2018 May 1;52(5):1685–93.

139. McNulty SM, Sullivan LL, Sullivan BA. Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Developmental Cell. 2017 Aug 7;42(3):226-240.e6.

140. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol. 2004 Apr;5(4):296–305.

141. Jurka J. Repbase Update: a database and an electronic journal of repetitive elements. Trends in Genetics. 2000 Sep 1;16(9):418–20.

142. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007 Dec;8(12):973–82.

143. Gutbrod MJ, Martienssen RA. Conserved chromosomal functions of RNA interference. Nat Rev Genet. 2020 May;21(5):311–31.

144. Sandoval R, Dilsavor CN, Grishanina NR, Patel V, Zamudio JR. Mammalian RNAi represses pericentromeric lncRNAs to maintain genome stability [Internet]. bioRxiv; 2024 [cited 2024 Jul 23]. p. 2024.05.09.593425. Available from: https://www.biorxiv.org/content/10.1101/2024.05.09.593425v2