INDUCTION OF HISTONE H3K27 METHYLATION BY NON-NATIVE TELOMERE REPEATS IN *NEUROSPORA CRASSA*

by

PEREGRINE D. PAINTER

A THESIS

Presented to the Department of Biochemistry
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2020

An Abstract of the Thesis of

Peregrine D. Painter for the degree of Bachelor of Arts
in the Department of Biochemistry to be taken June 2020

Title: Induction of Histone H3K27 Methylation by Non-native Telomere Repeats in *Neurospora crassa*

Approved: _____*Eric U. Selker, Ph.D.*_____
Primary Thesis Advisor

To live and grow, our bodies require carefully regulated patterns of gene expression to meet the demands of our current developmental state and environment. One way our body accomplishes this regulation is through chromatin modification. Of particular interest to this paper is the methylation of lysine 27 on histone H3 (H3K27me). The di-/trimethylation of H3K27 (H3K27me2/3) is a known contributor to gene silencing and is deposited by Polycomb Repressive Complex 2 (PRC2), a complex found in some form in most eukaryotic organisms. If H3K27me is not patterned appropriately, developmental errors and disease can occur. The factors governing recruitment and activation of PRC2 remain largely unknown and understanding how the body regulates PRC2 is integral to developing treatments for the diseases featuring aberrant PRC2 activity. Telomeres are long stretches of short repeated sequences, $(TTAGGG)_n$ in vertebrates and *Neurospora crassa*, that protect the ends of DNA during replication. It has been shown that telomere repeats are able to induce subtelomeric H3K27me2/3 at their native locations and ectopic H3K27me2/3 if inserted internally into the genome of the fungus *N. crassa*, a relatively simple eukaryotic organism that sports PRC2. In this study I sought to characterize what feature of the telomeric DNA sequence is responsible for the induction of de novo H3K27me2/3. I suspected the ability of telomeres to form G-quadruplex (G4) as a likely mechanism. I designed DNA constructs featuring non-native telomere repeats and inserted them at the *csr*-1 locus of *N. crassa*. My results suggest that the ability to form G4 structure is not alone sufficient for internal repeats to induce ectopic H3K27me2/3.

Acknowledgements

I would like to express my great appreciation and thanks for Dr. Eric Selker. Without his unwavering guidance and support, I would not have had the opportunity to pursue research in a field so sublimely aligned with my interests. Thank you for seeing my potential and providing for me a healthy environment in which to grow as a scientist. Words cannot begin to express how much admiration and gratitude I have towards my immediate mentor, Dr. Kevin McNaught. I joined the Selker lab while I was relatively inexperienced in a molecular biology lab, yet he taught me patiently and with humor right from day one. His passion for this research was infectious and I am grateful to have worked with such an encouraging and knowledgeable mentor.

I would like to thank Dr. David Garcia for acting as my second reader, his expertise was indispensable for refining this thesis into a project I can be proud of. I would also like to extend my warm thanks to Dr. Barbara Mossberg for acting as my CHC representative. Poetry incarnate, her encouragement is boundless and equally strong through both think and thin. Thank you for helping me see beauty at every scale.

To all the other members of the Selker lab, Dr. Jeanne Selker, Dr. William Storck, Dr. Tish Wiles, Dr. Vincent Bicocca, and more, a sincere thank you for weaving such a wonderful lab environment and for providing your advice and ideas on numerous occasions. Thank you to my previous mentors, Dr. Bill Messer and Dr. Aurora Ginzburg, for fostering my love of research. Thank you to the CHC for allowing me to pursue a thesis in the first place, along with the opportunity to become a scholar in the true sense of the word. Finally, I am grateful to my friends and family for their continued support of all my endeavors and for making it all worthwhile.

Table of Contents

List of Figures

List of Tables

Introduction

Preamble

Human beings are incredibly complex organisms. Our development requires a symphony of biochemical processes to be successful. Luckily, our bodies have reliable systems for regulating gene expression to make sure we have the right parts at the right times to make healthy humans more often than not. Unfortunately, we are not perfect, and our molecular machinery can make mistakes. These mistakes can be unnoticeable in the best of cases but can lead to disease, cancer, and/or death in others. Understanding how our bodies conduct this symphony of regulatory processes and what can go wrong is the first step in being able to find mistakes and correct them before they become a problem.

Introduction to epigenetics/chromatin

A copy of all the DNA that provides instructions for the growth, life, and death of a human is contained in virtually all cells of the body. However, the vast majority of those genetic instructions are not needed at any given moment. If one gets a sunburn, one's body will begin to carry out a subset of those instructions that will regenerate one's skin cells and repair DNA damaged by UV light exposure. But if one had stayed indoors, that response would not have been warranted and therefore carrying out "repair" would be a waste of resources. So how does the body keep some genes free to be used and the rest packed up and compact for if/when they are needed? In eukaryotes, the regulation of gene expression is accomplished, in part, through the packaging of DNA into chromatin. If not packaged, the human genome would be nearly two meters

long. All of that must fit into a single cell, which is typically on the order of 100 µm, so packaging is necessary. When packed by the formation of “chromatin” the length of DNA is reduced by ~40,000 fold, making it much easier to handle. At the most basic level, chromatin is made up of “nucleosomes”. Sections of chromosomal DNA 146 base pairs (bp) in length are wound around a complex of eight proteins, called histones, forming DNA-protein complexes that are referred to as nucleosomes (Figure 1).

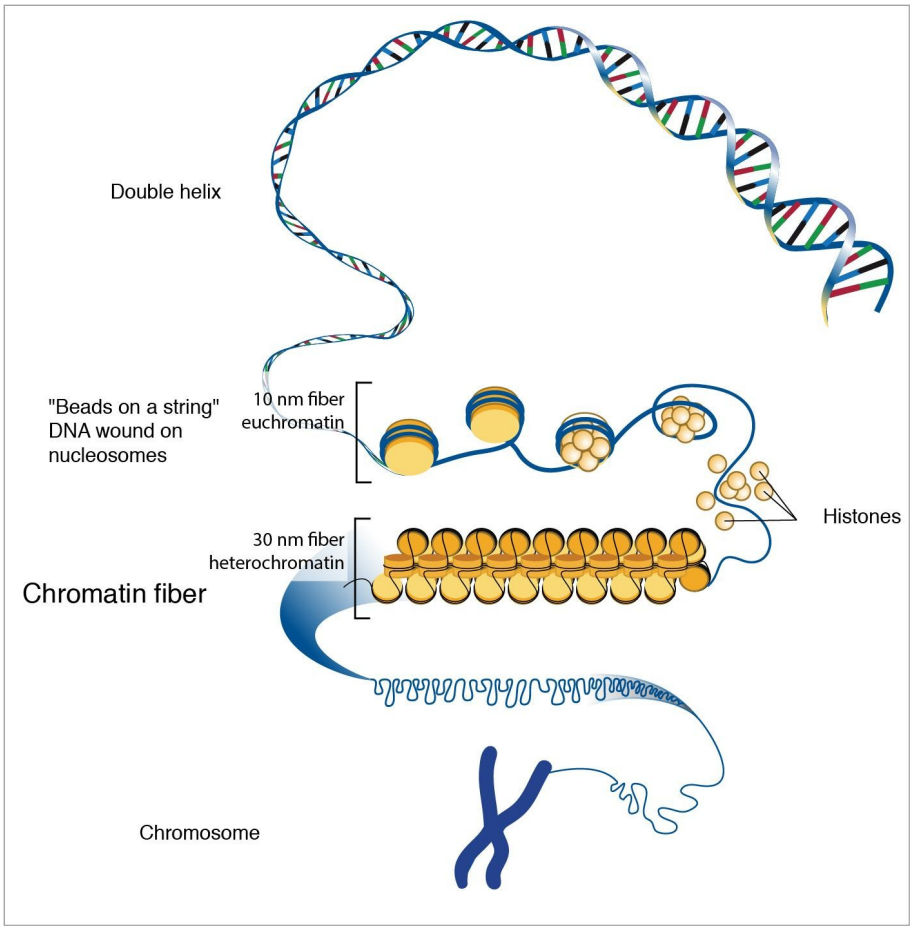

Figure 1. DNA and chromatin formation. Simplified illustration of nucleosomes and chromatin formation in the context of primary through tertiary DNA packing. Notable omissions are histone tails and differentiated histone subunits. When comparing the 10 and 30 nm fiber, notice that less of the DNA is accessible in the denser fiber and therefore is less likely to be actively transcribed. Figure adapted from an illustration from NIH[1].

Wound in such a fashion around the “spools” of nucleosomes, the DNA resembles a “beads-on-a-string” motif called “euchromatin” and still can be transcribed.

When it is wound tighter, the DNA with nucleosomes is thought to resemble more of a solid rope, with the wound DNA forming into a tightly-packed solenoid called "heterochromatin," which is minimally/not transcribed.

Epigenetics is the study of mitotically and/or meiotically heritable changes in an organism brought about by modification of gene expression without altering the genetic code itself. This can be thought of as a complex series of switches, turning up or down gene expression in response to a wide variety of stimuli. In humans and many other organisms, DNA modification is essential to development, and errors in DNA modification can lead to developmental disabilities, diseases (such as Leukaemia)[2], and even death. It is therefore important that we understand how our bodies selectively silence and express genes to orchestrate complex tasks, such as development, and what can go wrong in this process so that we can better combat epigenetic diseases.

Chromatin is, in part, regulated by a type of protein modification called methylation, in which methyl groups are added to the tails of histones. Histone tails are strings of amino-acids that extend from the eight proteins that make up the nucleosome. The naming scheme for describing histone modifications, such as **H3K27me3**, is a combination of: the histone tail being modified (H3 is histone subunit 3), the specific residue being modified (K27 is the amino-acid Lysine in the 27th position, from the N-terminus, on the histone tail), and lastly the type and number of modifications (me3 is trimethylation). In some cases, histone methylation has been found to be responsible for the induction of heterochromatin and subsequent silencing of genes in the region.

PRC2 Recruitment

Of particular interest to this investigation is the Polycomb Group (PcG) system. The PcG system is important for the development of higher organisms and errors in this system can lead to cancer and death[3]. PcG proteins form two main complexes: Polycomb Repressive Complex (PRC) 1 and 2. PRC2 is the focus of this paper, as it is enzymatically responsible for mono/di/trimethylation of H3K27 and therefore facultative heterochromatin formation (Figure 2). It is known to be the sole depositor of H3K27me in *Neurospora crassa* (*N. crassa*) and mouse embryonic stem cells[4,5].

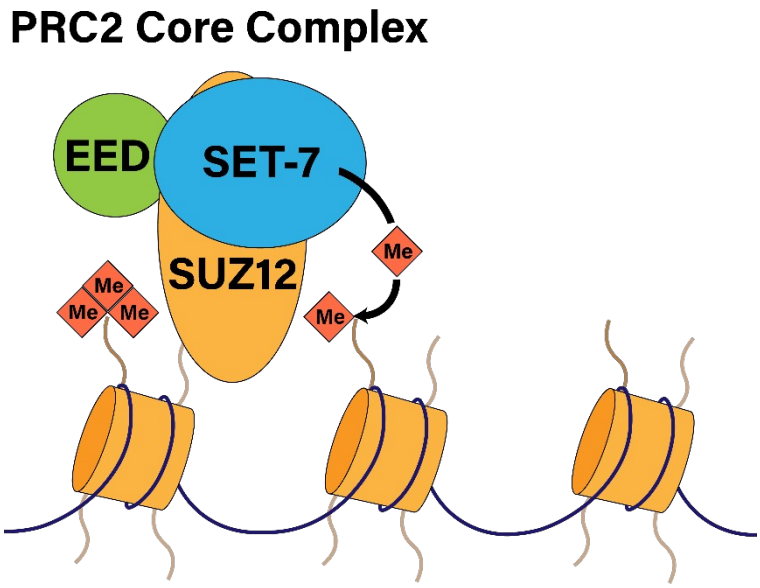

Figure 2. Illustration of the Polycomb Repressive Complex 2 core subunits (*N. crassa*) performing methylation of histone H3K27. DNA is represented as a black line wrapping around histones in orange. Histone tails are depicted in brown. SET-7 is catalytically responsible for the deposition of H3K27 methylation in *N. crassa*.

This methylation is added progressively, which manifests in H3K27me1 formation occurring before K3K27me2 and so on after a round of DNA replication[6]. Each level of methylation appears to have a different function, where H3K27me1 is enriched near actively transcribed genes, H3K27me2 seems to cover many

inter/intragenic regions, and H3K27me3 is found to be enriched at silent genes and often corresponds with sites known to bind PRC2[5,7]. This is logical because the PRC2 complex seems to need progressively more time to perform each stage of methylation, with conversion to H3K27me3 being the most lengthy catalytic process[8]. H3K27me3 recruits PRC1, which ubiquitinates a different histone and further encourages denser chromatin formation[9]. This need for PRC2 to linger in order to catalyze the formation of H3K27me3 spawns the big question in the field of how the body recruits PRC2 to a particular section of DNA long enough to perform its methyl transferase activity.

Core question in modern PRC2 research

"Why this and not that?" is the driving question when it comes to studying genomic patterns of H3K27me. Why do some sections of DNA nearly always get methylated while others do not, and how does the body tell the difference? About ~7% of histone H3 in *N. crassa* has H3K27me3 and many factors have been found to encourage local H3K27me3 deposition. Perhaps the most well-known sites for PRC2 binding, in mammals, are methylated (cytosine pair guanine) CpG islands[10]. CpG islands are DNA sequences with many CpG sites (5'-CG), often with a high CG%. They are frequently found inside human gene promoters and methylated in the promoters of silent genes. PRC2-bound CpG islands typically have a higher CG% compared to those less frequently bound[11]. In RNA, PRC2 showed a high affinity for strands rich in G, as well as RNA constructs that formed G-quadruplex (G4) structures[12].

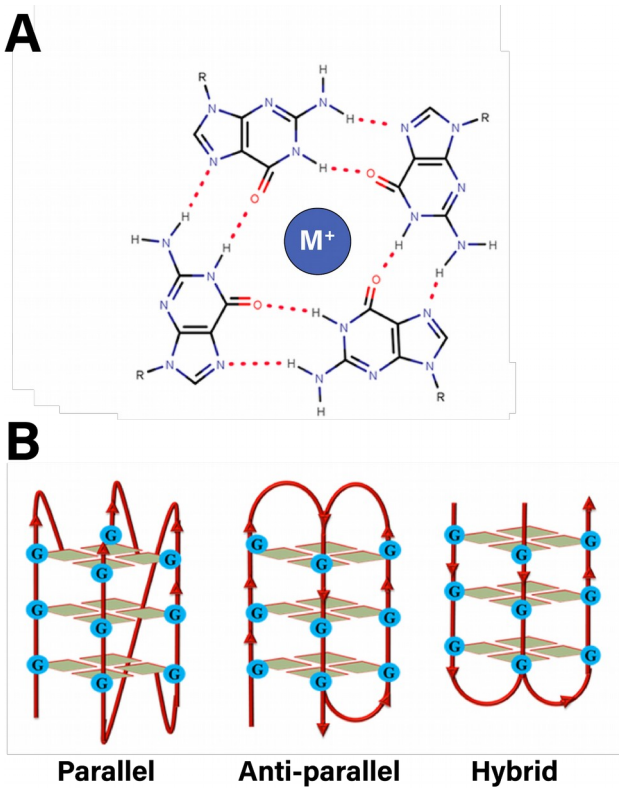

Figure 3. Illustration of DNA secondary structure G-quadruplex (G4). (A) A G-tetrad structure formed around a cation, especially K+. (B) Consecutive G's resulting in stacked G-tetrads referred to as the G4 DNA structure. All illustrations are intramolecular G4 as one would find forming from one segment of dsDNA, though intermolecular G4 is also able to form. Figure adapted from Mishra et al., 2019[13].

G4 is a DNA/RNA secondary structure formed by G-rich strands that have four or more repeated sequences of at least two G bases (e.g. "$[G_3N_4]_4$") G's from up to four different sections of DNA/RNA (can be from the same strand, or up to four different strands, albeit less frequently) can associate (via Hoogsteen hydrogen bonding, which involve non-Watson-Crick pairing) in a stable structure called a G-tetrad. These tetrads can stack if participating strands have runs of G's, forming the extremely stable structure, G4 (Figure 3). This structure forms predictably, especially in uniformly repeated sequences of DNA/RNA that contain stretches of G's. The location of G4 forming regions of many genomes are non-random and correlate with functionally

important regions of the genome involved with replication, gene regulation, and telomere maintenance[14,15].

Neurospora crassa as a model organism

Neurospora crassa (*N. crassa*) is a filamentous fungus with a haploid genome consisting of 7 chromosomes (Figure 4). It is a common model system for eukaryotic systems due to its ease of propagation in a lab, its genetic and biochemical tractability, and its small and fully sequenced genome. *N. crassa* is an attractive model organism for the study of chromatin modification because it sports DNA methylation and some histone modifications such as H3K27 methylation (deposited by PRC2) unlike other model systems including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Furthermore, H3K27me3 is typically a mark of silent genes in *N. crassa*, and, most importantly, over/under methylation of the genome does not result in gross developmental errors (unlike in humans and other higher organisms)[16]. This allows us to study strains with extreme methylation patterns that would cause other model organisms to be non-viable. The catalytic subunit of PRC2 in *N. crassa* is named SET-7. SET-7 and EZH2 (named so in humans and other higher-order life), are functionally equivalent, both contain the SET methyltransferase domain that deposits H3K27me. SET-7 knockout strains are frequently used as negative controls for H3K27me experiments in *N. crassa*, as without SET-7's catalytic activity PRC2 is non-functional[4]. *N. crassa* additionally has the same telomere repeat sequence as found in humans $(TTAGGG)_n$, albeit much shorter than humans at an average of ~120 bp[17].

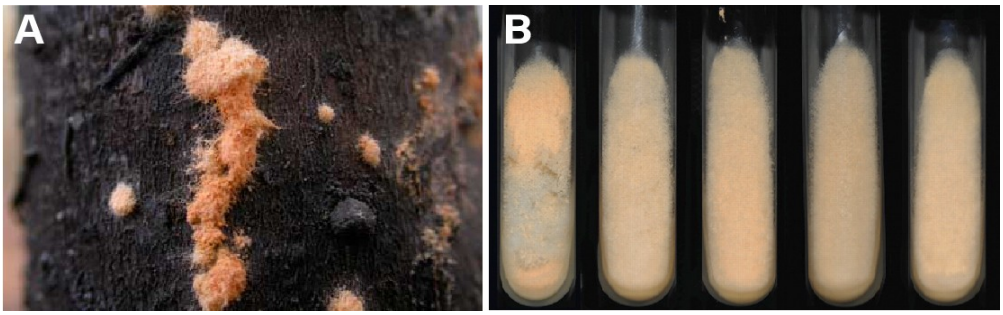

Figure 4. The filamentous fungus *Neurospora crassa*. Pictured in both (A) typical growth conditions in nature and (B) growth on laboratory slants consisting of minimal media. Images adapted from John Taylor of UC Berkeley and Fu et al., 2011[18].

H3K27me3 is typically enriched near DNA telomeres in *N. crassa*. By definition, telomeres are found at the ends of every chromosome in eukaryotic organisms where they act as a protective element for DNA during replication. They are made up of a repeated sequence of DNA, which in vertebrates is 5'-$(TTAGGG)_n$[19]. In humans, telomeres are commonly ~5-15 kb long and commonly form secondary structures. The human telomeric sequence is both rich in G and able to form G4[20], aligning them strongly with factors associated with increased PRC2 affinity. It was found that telomeres and PRC2 associate subunit (PAS) were necessary for H3K27me2/3 of subtelomeric genes[21,22]. Further, internal telomere repeats were sufficient to induce ectopic H3K27me2/3 in *N. crassa*[21]. This finding was of great interest, as telomeres exhibit position-independent H3K27me2/3, and this paper seeks to help characterize what feature of the telomeric DNA sequence is responsible for the induction of *de novo* H3K27me2/3. The possible role of G4 in the induction of H3K27me2/3 is currently unknown. Given the literature on PRC2 recruitment, I hypothesized that internal telomere G4 formation was important for the induction of H3K27me2/3, rather than something inherent to the specific repeated sequence itself. Different organisms have different telomeric sequences, which could affect the possibly

sequence-specific PRC2 recruitment/activation. It has been shown that both *Arabidopsis* (5'-TTTAGGG)$_n$ and *Tetrahymena* (5'-TTGGGG)$_n$ telomeric repeats are capable of forming G4 *in vivo*[23,24], and it follows that these repeats, if inserted internally into the *N. crassa* genome, could also induce H3K27me2/3.

Outline of objectives

Objective 1: Construction of N. crassa *strains containing internal telomere repeats*

Before the effects of these non-native repeats on H3K27me2/3 could be studied, strains containing such repeats had to be created. First, DNA constructs containing the repeats were designed with transformation in mind. In an effort to stay consistent with both natural *N. crassa* telomeres and previous research, all of the DNA constructs contained 20 repeats. In addition to the WT, *Arabidopsis*, and *Tetrahymena* repeats, a negative control for G4 (ΔG4) was created by interrupting the string of G's present in the WT telomere repeats: $(TTAGGG)_{20} \Rightarrow (TTAGcG)_{20}$. By definition, this sequence should not form G4 as it does not contain more than one G in a row. The G ⇒ C swap conserved CG%. The constructs also contained restriction enzyme (RE) cut sites at either ends of the repeats for insertion into targeting plasmids (Figure 5).

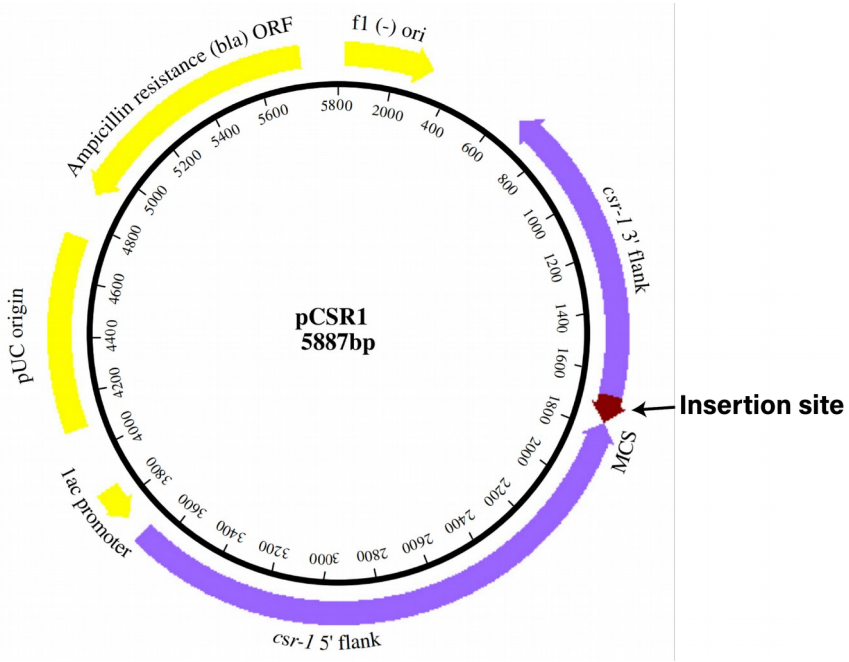

Figure 5. Map of pCSR-1. Important features are the multiple cloning site (MSC) and the Ampicillin resistance gene used to select for successful transformation of the plasmid into *E. coli*. Figure adapted from Bardiya and Shiu, 2007[25].

We considered two possible gene targets for insertion into the *N. crassa* genome, the genes encoding cyclosporine A resistance 1 (*csr*-1) and histidine 3 (*his*-3)[17,21,26]. *His*-3 targeting is the more commonly used method, involving the replacement of a faulty *his*-3 gene with a working one also containing a desired insert, thereby selecting for cells able to grow in the absence of histidine. However, this method has a drawback in that the strain needs to be backcrossed to become homokaryotic (meaning all nuclei are genetically identical). *Csr*-1 does not have this issue, as untransformed nuclei are selected for because they are able to produce cyclosporin A-binding protein and become susceptible to the fungicide. Therefore, *csr*-1 was chosen due to the ease at which homokaryons are obtained. DNA constructs were inserted into the *csr*-1 targeting vector and transformed into *N. crassa* via homologous recombination and selection on cyclosporin.

Objective 2: Assess H3K27me2/3 in constructed strains

Chromatin immunoprecipitation (ChIP) is a method of isolating regions of DNA that are bound by a certain protein. H3K27me2/3 ChIP was used to quantify the degree to which histone residues in certain regions of the *N. crassa* genome were methylated. Both ChIP-quantitativePCR (qPCR) and ChIP-sequencing (ChIP-seq) were utilized for this purpose. The former is quicker and was used to provide an initial check on the quality of ChIP and provide a preliminary check for methylation in the target regions. The drawback of qPCR is that one can only examine discrete regions of DNA that one has primers for, so one will miss out on the broader patterns of H3K27me2/3 that might exist. CHiP-seq is much more resource intensive, as it involves sequencing and counting all of the DNA fragments produced by the CHiP. This gives a complete picture of genomic methylation patterns and offers generally more trustable data. ChIP-seq also was able to confirm one last time that the correct identity and number of repeats had been inserted. CHiP-seq data suggested that G4 was not sufficient for the induction of H3K27me2/3 by internal telomere repeats.

Results

Telomere Repeat DNA Constructs and Targeting

A DNA construct for each of the proposed set of repeats was synthesized with features to ease targeting (listed 5'->3'): The RE site NotI, the 20xRepeat sequence, and a final set of RE sites, PstI and EcoRI. The constructs were 155 bp in length for those including 6 bp telomere repeats and 175 bp in length for the one including 7 bp telomere repeats. Both the constructs and their reverse complements were made by the oligonucleotide synthesis company Integrated DNA Technologies (IDT) as single stranded (ss) DNA. The amount of ssDNA received for each construct was largely on target at 4.0 nmoles per construct, save for the $(TTgGGG)_{20}$ forward primer (FP) with 1.5 nmoles and the $(TTtAGGG)_{20}$ FP with 3.9 nmoles. The quality of these two constructs was unable to be verified by IDT and was estimated to be around 15% and 12.72% the desired construct out of total DNA mass respectively. The forward and reverse ssDNA constructs were combined and heat-annealed to make single pieces of double stranded (ds) DNA.

Having chosen *csr*-1 as the genomic target for the DNA constructs, following previous work[21], the DNA constructs were inserted into a *csr*-1 targeting vector (Figure 5). The *E. coli* strain (E800) containing the pCSR-1 targeting vector was inoculated, grown, and DNA-extracted to obtain a sufficient amount of the desired plasmid. The duplex DNA constructs and pCSR-1 were double digested with NotI-HF and Pst-I, and gel purified. Each digested DNA construct was individually combined with digested pCSR-1 and ligated with T4 ligase. A nice feature of this method of insertion is that the plasmid is unable to ligate to itself due to the mismatched digested ends, so the only

circular plasmids after ligation should be either undigested original plasmid (unlikely due to gel purification beforehand) or the desired DNA construct insert, thereby saving an extra purification step. The plasmid + inserts were subsequently transformed into transformationally competent *E. coli* cells. Successful transformants were identified by plating on ampicillin (amp). Presence of insert was confirmed by digestion of plasmid DNA, a HindIII digest alone and a HindIII + NotI double digest (Figure 6), and electrophoresis with a 2% agarose Tris/Borate/EDTA (TBE) gel, and by Sanger sequencing.

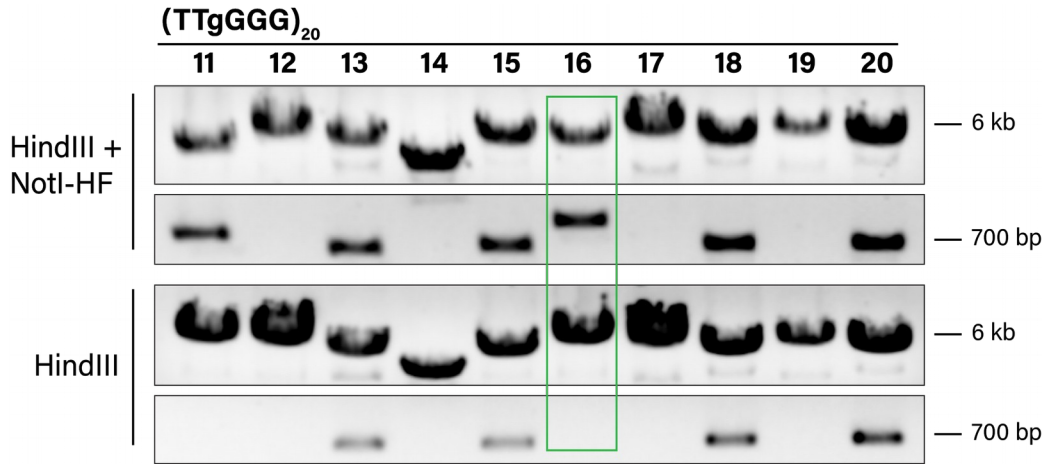

Figure 6. Gel screening for pCSR-1 with insert in *E. coli*. Typical results from a 2% agarose TBE electrophoresis gel screening for successful transformation into *E. coli*. 30 colonies were picked and miniprep for each transformant, and all but $(TTgGGG)_{20}$ came back positive for insert within the first 5 minipreps tested. Minipreps 11-20 $(TTgGGG)_{20}$ were analyzed in this gel after double and single RE digestion as noted on the left. Expected bands: (+) single=5930 bp, double=5162, & 778 bp; (-) single=5200, & 687 bp, double=5150, 687, & 50 bp. 50 bp fragment not was observable. Lanes 11 and 16 were positive for insert, and 16 was used for subsequent experiments after confirmation by sequencing. For reference, 13, 15, 18, and 20 were negative plasmid containing inserted repeats.

Transformation into *N. crassa* by homologous recombination

Transformation into *N. crassa* was conducted according to a typical protocol for the *csr*-1 targeting method[27]. Due to the high rate of false positives in the cyclosporin A-resistance screening for inserts at the *csr*-1 locus of *N. crassa* strains N2818 and N2931,

the transformation protocol and subsequent Southern blot to check for the desired inserts had to be repeated multiple times. If an *N. crassa* cell develops spontaneous resistance to cyclosporin A in an early generation of colony growth, that mutant tends to out-compete any successfully recombined cells, resulting in an entire batch of false positives rather than a few here and there as one might expect. If a strain is especially prone to acquire resistance early, it can result in many fruitless transformation screenings. The false positive rate of strain N2931 was found to be 52.2% in the sum of my experiments. An attempt was made to make a strain of *N. crassa* with a lower rate of false positives through restreaking, which was successful (N7986) (Table 1).

Culture (N2631)	Count (+)	Count (-)	False (+) rate
1a	3	0	**30%**
1b	8	4	
1c	8	4	
2a	5	4	33%
2b	7	2	
2c	4	2	
3a	1	3	59%
3b	3	5	
3c	3	2	

Table 1. False positive rates of restreaked N2931. Strain N2931 was restreaked and 3 colonies were picked (statistically giving us at least one option in the lowest 20% of false positives rates that majority of the time). These were inoculated and plated in technical triplicate (a-c). Colonies were picked and tested for insert presence, positive and negative for insert are the second and third column respectively. False positive rate was calculated via no insert over total colonies tested. Bolded strain was designated as N7986 and used for future transformations.

After two more transformations into the new strain, all DNA constructs were successfully inserted into at least two biological replicates of strain N7986 as confirmed by Southern hybridization (Figure 7) and DNA sequencing. DNA sequencing of the

Arabidopsis and WT repeats proved error prone, returning with high uncertainty across the most of the insert, with over 40% of the segmant unreadable despite the Southern confirming presence of inserts within expected size. The error was hypothesized to be because of G4 formation during the sequencing protocol leading to difficulty reading the nucleotides, so the PCR of the insert region was repeated with 2:3 dNTPs:7-deaza-dGTP in an attempt to help prevent compressions[28]. Sequencing of the 19 strains with inserts revealed the number of inserted non-native telomere repeats ranged from 4 to 22. All four constructs had at least one strain that fell near or above the eight-repeat threshold that was previously found to be sufficient to induce H3K27me3[21].

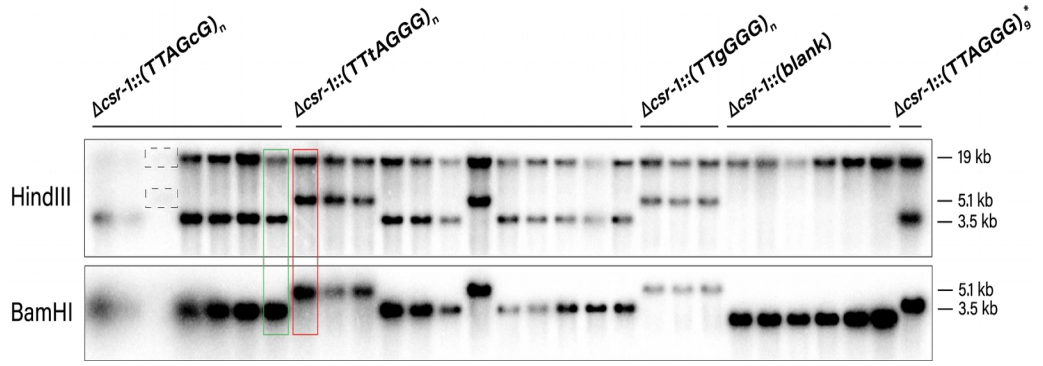

Figure 7. Southern screening for successful *csr*-1 repeat insertion. Southern hybridization using a probe for 5'-csr-1 flank after RE digests indicated on the left. Exact inserted repeat number was unknown and this uncertainty is expressed with "n" replacing a known repeat number in the lane labels. Expected bands: (+) HindIII=18.7, & 3.5 kb, BamHI=3539 bp; (WT) HindIII=18.7, & 5.1 kb, BamHI=5126 bp. Asterisk indicates strain N6382, constructed previously, used as a positive control for presence of $(TTAGGG)_9$ insert at *csr*-1[21]. Black dashed boxes indicate present but hard to digitally view bands. Green and red boxes are examples of a strain positive for insert and a false positive (from cyclosporin A selection) respectively, arbitrarily chosen. Δ*csr*-1::(blank) was linearized pCSR-1, with no insertion, transformed into N2631 and used as a control for the transformation protocol.

ChIP-qPCR of *N. crassa* chromosome regions Tel1L, hH4, 3'-*csr*-1, and 5'-*csr*-1

A H3K27me2/3 ChIP using α-H3K27me2/3 antibody was performed on one strain each of the WT, *Tetrahymena*, and ΔG4 constructs, as well as a previously used Δ*set*-7 strain[22] and a strain with a $(TTAGGG)_{17}$ insert at *csr*-1[21]. Enrichment of

H3K27me2/3 was found at both 5' and 3'-*csr*-1 sites (both ~500 bp away from the insert site) in both WT repeat strains and the *Tetrahymena* telomere repeat strain, with mean enrichments of 6.30, 10.83, and 7.09 fold over hH4 for $(TTAGGG)_{17}$, $(TTAGGG)_{11}$, and $(TTgGGG)_{11}$ respectively (Figure 8). The Δ*set*-7 and $(TTAGcG)_{20}$ insert strains did not display significant enrichment between hH4, a site typically lacking H3K27me2/3, and the *csr*-1 sites (a mean of 1.79 and 1.51 fold respectively), which is consistent with the G4 hypothesis.

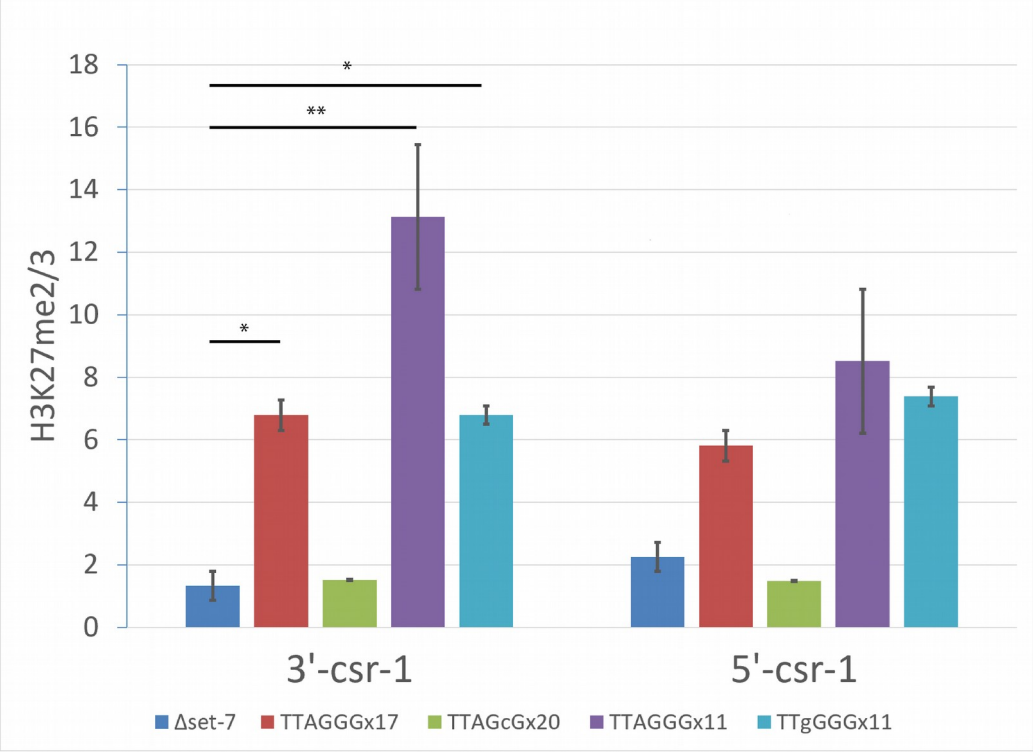

Figure 8. qPCR analysis of H3K27me2/3 ChIP of strains with constructed repeat sequences targeted to *csr*-1 conducted in biological triplicate. Fold increase of H3K27me2/3 in strains at 3'- and 5'-*csr*-1 over hH4 is plotted on the Y-axis. T-Test of both 3' and 5' *csr*-1 strains, from left to right, against Δ*set*-7 gave p=0.066, 0.330, 0.095, and 0.010 respectively (single and double asterisks represent P<0.1 and P<0.025 respectively for both *csr*-1 sites). For statistical information see Table S1.

There was unusually high enrichment in the ΔG4 at hH4 (where I would expect almost none), 0.15% of the input DNA as opposed to ~0.04% for all the other strains (Figure S1). This may have introduced bias leading to a lower perceived enrichment at

csr-1, after the normalization of *csr*-1 to hH4, than was actually present. It is hard to know whether such a bias was present without the evaluating the surrounding levels H3K27me2/3 for context[1].

Predicted G4 structure of internal telomere repeats is not sufficient to induce H3K27me2/3

ChIP-seq was conducted on the $(TTAGGG)_{17}$, $(TTAGGG)_{11}$, $(TTgGGG)_{11}$, $(TTtAGGG)_{22}$, and $(TTAGcG)_{20}$ strains with biological replicates (Figure 9). As the G4 hypothesis would predict, The WT and *Tetrahymena* repeat strains showed new H3K27me2/3 peaks at least 150 kb away from the insertion site and an ~225 kb semi-continuous domain of methylation surrounding the insert as observed previously in a strain of *N. crassa* with a WT telomere insert[21], while the ΔG4 repeat strain displayed a lack of induced H3K27me2/3 around *csr*-1 consistent with wildtype *N. crassa*. This confirmed and extended the ChIP-qPCR data. However, $(TTtAGGG)_{22}$ also displayed an H3K27me2/3 pattern similar to wildtype *N. crassa*. As *Arabidopsis* telomeres are known to form G4 and similar sequences have been used to form G4 *in vitro*, this provides evidence that telomeric G4 alone is not sufficient to induce *de novo* H3K27me2/3 around the insert site.

[1] ChIP-qPCR is like looking out your window to see how much traffic there is. Since you are only seeing the one spot, it is hard to tell if the entire road is backed up or if it is just a random little blip of traffic. CHiP-seq looks at all the traffic on the entire highway, so it is much easier to see the true traffic trends.

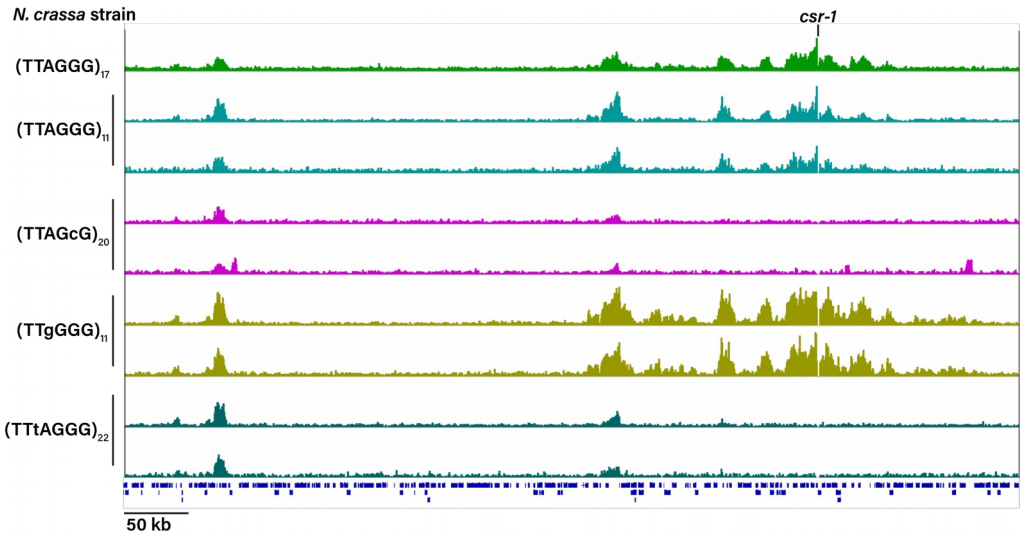

Figure 9. H3K27ChIP-seq of all strains containing constructed telomere repeats. ChIP-seq was conducted in biological replicate, and also included was the the ChIP-seq data for the $(TTAGGG)_{17}$ insert strain created previously (top)[21]. Y-Axis represents DNA reads normalized to a max of 1000 reads. 100 bp sliding windows were used in histogram creation. The bottom biological replicat of the $(TTAGcG)_{20}$ insert strain seemed to have been contaminated by H3K9me3 ChIP from another sequencing lane, resulting in 3 aberrant peaks that match regions usually found to be dense in H3K9me3[29]. Genes are indicated as blue bars at the bottom.

Discussion

Polycomb repressive complexes are essential for the modulation of gene expression and cellular development in higher eukaryotes[30–32]. The way a cell orchestrates patterns of H3K27me, however, is not fully characterized. Recent research with *N. crassa* has suggested that telomeres are a source of position-dependent methylation while also revealing certain loci that seem to induce local H3K27me2/3 regardless of their position in the genome[21]. Telomere repeats are known to induce subtelomeric H3K27me2/3 and in *N. crassa* internal telomere repeats also are sufficient to induce local ectopic H3K27me2/3[21]. This study sought to better characterize what feature of these repeated sequences is able to act as a cis-acting factor for H3K27me2/3 through the insertion of non-native telomere repeats at the *csr*-1 locus in *N. crassa*. The ability to form the G4 DNA secondary structure seemed like a likely candidate to cause this effect, as telomeres in eukaryotes that code for PRC2 tend to be able to form G4 and PRC2 has a greater affinity for G-rich sequences of RNA that can form G4 compared to those that cannot[12,20,23,33]. However ChIP-seq for H3K27me revealed that G4 forming sequences alone were not sufficient to induce H3K27me2/3, as 154 bp of *Arabidopsis* telomere repeats inserted at *csr*-1 were not able to induce the ~225 kb pattern of H3K27me2/3 induced with the WT or *Tetrahymena* repeats. The altered WT repeat strain that was predicted to be unable to form G4 lacked H3K27me2/3 as expected.

The maintenance of native telomeres themselves could provide a potential mechanism for these observations. Telomere repeat-containing RNAs (TERRAs), long

and non-coding, have been shown to be integral for maintenance of telomere length via the inhibition of telomerase and association with telomere repeat-binding factors (TRFs)[34]. In human cells, TERRA generation was found to be necessary for the establishment of a number of heterochromatin markers at the telomeres, including H3K27me2/3, through knockout of TERRA transcription. The same study observed PRC2 core subunits directly interacting with TERRAs[35]. This combined with the research on RNA containing PRC2 binding motifs (such as G4) could suggest that actively transcribed regions containing telomere repeats could preferentially recruit PRC2 to their loci[12]. The predicted structure of *N. crassa* TERRA G4 seems to be highly specific[36], which may explain why any nascent TERRA generated from the *Arabidopsis* internal repeats was not able to associate with PRC2. However, the results of this study (and the PRC2 RNA affinity study) seem to cast some doubt on the TERRA hypothesis if the G4 structure is not dependent on the particular number and sequence of bases forming the loops between each tetrad. Human TRF1 is able to bind directly to genes containing $(TTAGGG)_n$ motifs through the docking of its two DNA binding domains onto both ribose phosphate backbones on the major groove side, which orient the DNA recognition helix of TRF1 to form sequence-specific hydrogen bonds to the row of G-C pairs (G4 and G5 in particular), with additional water-mediated protein-DNA interactions (at many surrounding bases) that were also important for binding affinity[37]. TERRA has been found to bind to internal PRC2 sites dependent on TRF1, which may lead to increased PRC2 recruitment to those genomic sites[38]. The $(TTGGGG)_n$ motif could be similar enough to $(TTAGGG)_n$ to bind TERRA, while the $(TTTAGGG)_n$ repeats might not be able to reproduce the highly specific interactions

required for strong TRF1 binding due to the extra base in the motif. An extension of this research could involve testing *N. crassa* TRF1's binding to various sequences, including the DNA constructs made in this paper.

Another mechanism previously proposed was telobox-mediated PRC2 recruitment. Teloboxes, motifs of the sequence (AAACCCTAA), can be bound by TRFs and mutation of the telobox sequences decreases rather than abolishes TRF binding[39]. Teloboxes paired with GA repeats are sufficient to recruit transcription factors that interact with PRC2 in *Arabidopsis*[33]. The lack of H3K27me2/3 induction from the *Arabidopsis* repeats, rather than just a reduced level of methylation, renders it somewhat unlikely that the mechanism for internal repeated-induced H3K27me2/3 is related to the telobox. The most likely model given current knowledge is TERRA-mediated PRC2 recruitment, whereby native telomere RNAs are able to recruit PRC2 to telomeres and telomere repeat-containing loci. This research could be extended by performing a H3K27me2/3 ChIP on a Δtert (telomerase reverse transcriptase) strain (or with CRISPR-Cas9 as previously described[35]) to disable the ability to generate native TERRAs and see if the induced patterns of methylation from the native/non-native internal repeats are still present. To examine a step in common with both telobox and TERRA-mediated H3K27me, a catalytic inactivation or knockout of key TRFs in the telomere repeat insert strains could be made to elucidate their role in the induced H3K27me3. Testing to see if TRFs were binding to the non-native internal repeats may provide insight into why *Arabidopsis* telomere repeats were found to be insufficient for induction of ectopic H3K27me.

In summary, I established that non-native telomere repeats that have been reported to be able to form G4 are not necessarily sufficient to induce H3K27me2/3; while the G4 structure may be involved, the ability to form G4 does not guarantee PRC2 recruitment.

Materials and Methods

Strains, media, and growth conditions

E. coli strains (Table S6) were selected for plasmid on LB + Ampicillin agar plates and inoculated for DNA preps in 5 mL LB + amp liquid culture. *N. crassa* strains (Table S3) were grown using standard media and conditions, described previously[40,41].

Creation and insertion of DNA constructs into pCSR-1

DNA construct forward and reverse primers (Table S2) were designed in the program ApE and ordered from the oligonucleotide synthesis company Integrated DNA Technologies (IDT) as ssDNA. SsDNAs were resuspended to 50 μM with TE. Forward and reverse primers were combined, 10 μL each, and incubated for 5 minutes at 94°C to anneal and left to cool at room temperature (RT) overnight (o/n). pCSR-1 was DNA prepped from liquid cultures of E800, and both pCSR-1 DNA and duplex DNA constructs were double RE digested o/n with NotI-HF and PstI. Digested pCSR-1 and duplex DNA were purified by 0.8% agarose TBE gel and extracted using the Monarch DNA Gel Extraction Kit (NEB T1020L). DNA concentrations were confirmed by broad range Qubit (ThermoFisher). Digested DNA constructs were each combined with digested pCSR-1 and ligated with T4 ligase.

Transformation into *E. coli*

DNA constructs ligated into pCSR-1 were transformed into chemically competent DH5α *E. coli* cells as described previously[42], 45s at 42°C, 2 mins on ice, and

then an hour of agitated recovery at 37°C with SOC media. Cells were then plated 1:10 and 1:1 on LB + Amp. Picked colonies were inoculated and DNA prepped. Presence of insert and repeat number was confirmed by EcoRI-HF and NotI-HF double digest ran on 2% agarose TBE gel, and by sequencing (Sequetech) of *csr*-1 (PCR with flanking primers) with primer 6569 (Table S5).

Transformation into *N. crassa*

Plasmids p3411, p3416, p3418, and p3419 (Table S4) were linearized with DraIII-HF and gel purified on 0.8% agarose TBE. Gel extracted linearized plasmids were individually transformed into *N. crassa* strains N2631 and N7986 (Table S3) as previously described[27], save for the plating on minimal media + cyclosporin A (CSA) plates with FGS, where 10 mL TOP agar (50x Vogels, 1.5% agar, dH_2O, 1x FGS, 1x CSA) was mixed with transformed cells in recovery media before plating ⅓ of the mixture on each of 3 of the min CSA selection plates. Colonies were picked after 2-5 days and inoculated in liquid culture. Genomic DNA was extracted as previously described[43]. Presence of insert and repeat number was confirmed by Southern hybridization as previously described[44], and by sequencing of *csr-1* (PCR) with primer 6569. Study-specific Southern features were digesting genomic DNA with HindIII and BamHI-HF (individually) and using the 5'-*csr*-1 flank PCR product labeled with ^{32}P as a probe. PCR for sequencing (primers 6569 and 6644) was repeated with 2:3 dNTPs:7-deaza-dGTP in order to prevent sequencing errors.

H3K27me2/3 chromatin immunoprecipitation (ChIP)

ChIP was performed as previously described[41,45], using H3K27me2/3 antibody (Active Motif 39536), eluted from beads with 125 µL TES (10 mM Tris-HCl, 1 mM EDTA, 1% sodium dodecyl sulfate) at 65°C, and purified using the Qiagen MinElute Kit. Purified DNA was used for subsequent qPCR or prepared for sequencing by DNA Library Prep (NEB) as previously described[41]

ChIP-qPCR and ChIP-seq

ChIP-qPCR was performed as previously described[41] using the first eight primers of Table S5, which included regions Tel1L, hH4, and the 5' & 3' flanks of *csr*-1. H3K27me2/3 levels were normalized to background at hH4. Mapping and visualization of ChIP-seq data was conducted as previously described[41,45]; aligned with corrected *N. crassa* OR74A (NC12) genome (Bowtie2), normalized using Reads Per Kilobase per Million (RPKM) mapped reads, and a 25 base bin size.

Glossary

Chromatin — complex of DNA and protein (histones) found in eukaryotic cells. Packages and compresses DNA into a more manageable size, as well as helping to prevent DNA damage and regulating gene expression.

Competent (cells) — an acute state where a cell is able to take up exogenous DNA from its environment. This can be induced e.g. by electroporation in the case of *N. crassa*.

Complex — multiple individual parts (subunits) making one functional unit, like two halves of a pair of scissors coming together to make a "scissor complex." Complexes, in this context, involve the association of two or more protein, DNA, and/or RNA subunits to make one whole unit.

Euchromatin — DNA wound on nucleosomes to form a "beads-on-a-string" type of appearance with a fiber width of 10 nm. This DNA is lightly packaged and is still accessible for gene expression.

Filamentous (fungi) — growth pattern that resembles a branching structure, with thin "branches" elongating and forming as the fungi grows. Results in that typical "fuzzy" appearance one associates with bread molds and other fungi.

Genome — the genetic material of an organism. All the code contained in every cell that makes you you, a dolphin a dolphin, and yeast a yeast.

Heterochromatin — Euchromatin that is further wound into a higher-order structure of a solenoid, creating a highly dense 30-100 nm fiber. This DNA is highly packaged and is seldom expressed (though still accessible in some cases).

Histone — A protein spool for DNA made up of eight histone subunits, an octamer made of 4 histone dimers. When in its octomeric form, DNA can be wound around to form a nucleosome. Histones are the essential protein component of Chromatin.

Nucleosome — A segment of ~146 bp of DNA wrapped around a histone complex. Often it is compared to a spool of thread, with the DNA as the thread and histone octamer as the spool. A single, long, DNA strand can have many nucleosomes.

Residue — a single subunit that makes up a polymer, such as nucleic acids in the case of DNA and amino acids in the case of a polypeptide/protein.

Terminus — the N- and C-terminus correspond to the amine end (N) and carboxylic acid end (C), respectively, of a polypeptide strand. Used for orientation similarly to the 5'- and 3'-ends of DNA.

Transcription — the act of an RNA polymerase creating a strand of RNA from a DNA template, the first step in gene expression. This RNA can then act as a template to make various proteins. If a gene is being highly transcribed, it is said to be "highly expressed." If it is not readily being transcribed, the gene is said to be "silenced."

Transformation — genetic alteration of a cell from the intake of genetic material from the cell's environment. The cell taking up the DNA has to be in a state of competence. Once taken up, the DNA is now apart of the cell and will be replicated and transcribed like its own, which is handy for selection as one can have a cell take up a gene that confers resistance to an antibiotic.

Supplemental Information

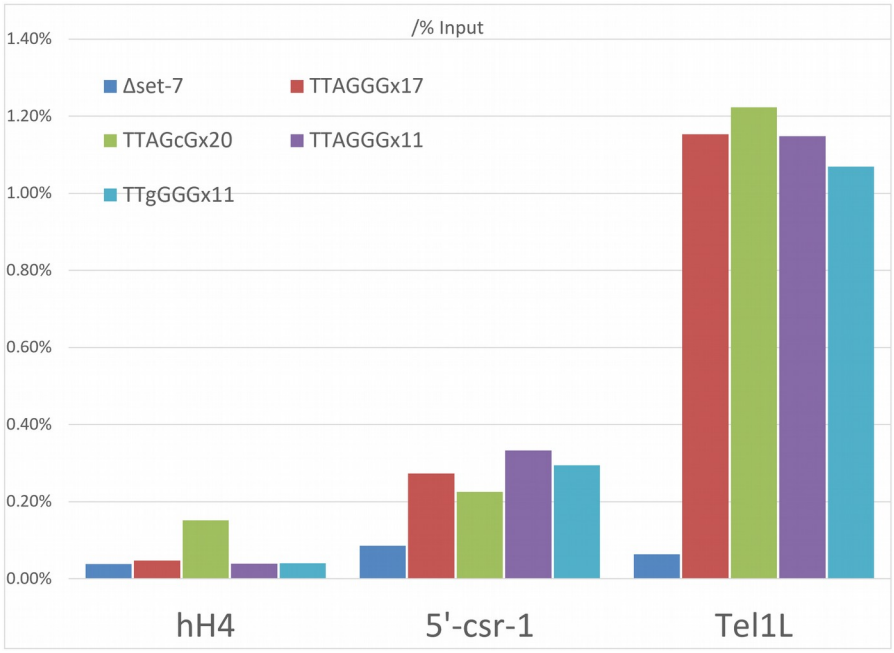

Figure S1. qPCR data normalized to the percent enrichment from the input DNA. Note the aberrant enrichment at hH4 in the $(TTAGcG)_{20}$ strain compared to the others. 3'-*csr*-1 data were also recorded, but not displayed for the sake of image clarity. It was consistent with 5'-*csr*-1.

Strain/hH4	AVG	SD	CI
Δ-set7	1.7905	0.4655	0.6451
TTAGGGx17	6.2980	0.4892	0.6780
TTAGcGx20	1.5051	0.0163	0.0226
TTAGGGx11	10.8257	2.3112	3.2031
TTgGGGx11	7.0917	0.2947	0.4085

Table S1. Statistical analysis of qPCR data. Mean, standard deviation, and confidence interval (95%) were calculated.

Table S2. DNA Telomere Repeat Constructs.		
Primer	Description	Sequence (5' ⇒ 3')
6559	TTAGGG ultramer FP	ATGCATGCgcggccgcTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGctgcagaattcATGCATGC

6560	TTAGGG ultramer RP	GCATGCATgaattctgcagCCCTAAgcggccgcGCATGCAT
6571	TTAGcG ultramer FP	ATGCATGCgcggccgcTTAGcGctgcagaattcATGCATGC
6572	TTAGcG ultramer RP	GCATGCATgaattctgcagCgCTAAgcggccgcGCATGCAT
6573	TTgGGG ultramer FP	ATGCATGCgcggccgcTTgGGGctgcagaattcATGCATGC
6574	TTgGGG ultramer RP	GCATGCATgaattctgcagCCCcAAgcggccgcGCATGCAT
6575	TTtAGGG ultramer FP	ATGCATGCgcggccgcTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGTTtAGGGctgcagaattcATGCATGC
6576	TTtAGGG ultramer RP	GCATGCATgaattctgcagCCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAACCCTaAAgcggccgcGCATGCAT

Table S3. *N. crassa* strains.

Strain	Genotype
N2718	*mat a; Δmus-52::hph+*
N2931	*mat a; Δmus-52::bar*$^+$
N4730	*mat A; Δset-7::bar*$^+$
N6382	*mat a; Δmus-52::bar*$^+$*; Δcsr-1::(TTAGGG)*$_9$
N6986	*mat a; Δmus-52::bar*$^+$*; Δcsr-1::(TTAGGG)*$_{17}$
N7986	*mat a; Δmus-52::bar*$^+$ (restreak of N2931)
N8028	*mat a; Δmus-52::bar*$^+$*; Δcsr-1::(TTAGcG)*$_{20}$
N8029	*mat a; Δmus-52::bar*$^+$*; Δcsr-1::(TTAGcG)*$_{20}$
N8030	*mat a; Δmus-52::bar*$^+$*; Δcsr-1::(TTtAGGG)*$_{22}$

N8031	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{22}$
N8032	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{22}$
N8033	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{n}$
N8034	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{n}$
N8035	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{22}$
N8036	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{n}$
N8037	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{n}$
N8038	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTtAGGG)*$_{n}$
N8127	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{6}$
N8128	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{11}$
N8129	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{6}$
N8130	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{8}$
N8131	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{4}$
N8132	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTAGGG)*$_{5}$
N8133	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTgGGG)*$_{11+}$ (sequencing inconclusive, 11 or 12)
N8134	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTgGGG)*$_{5}$
N8135	*mat a; Δmus-52::bar*$^{+}$*; Δcsr-1::(TTgGGG)*$_{n}$

Table S4. Plasmids.

Plasmid	Description
p3021	pCSR1, plasmid #665 from FGSC, Bardiya and Shiu (2006)
p3411	$(TTAGGG)_{20}$ insert into p3021 via Not1-HF & Pst1
p3416	$(TTAGcG)_{20}$ insert into p3021 via Not1-HF & Pst1
p3417	$(TTAGcG)_{20}$ insert into p3021 via Not1-HF & Pst1
p3418	$(TTtAGGG)_{22}$ insert into p3021 via Not1-HF & Pst1
p3419	$(TTgGGG)_{26}$ insert into p3021 via Not1-HF & Pst1
p3426	pCSR1+$(TTAGGG)_{20}$ via NotI-HF & PstI
p3427	pCSR1+$(TTAGGG)_{20}$ via NotI-HF & PstI
p3428	pCSR1+$(TTgGGG)_{26}$ via NotI-HF & PstI
p3429	pCSR1+$(TTgGGG)_{26}$ via NotI-HF & PstI

Table S5. Primers.

Primer	Description	Sequence (5’ ⇒ 3’)
3565	Tel1L FP	AGCGTTCAAATGCCGTGACCTGT
3566	Tel1L RP	GCTTTCGTCCACGGAGAAACGC
4082	hH4 FP	CATCAAGGGGTCATTCAC
4083	hH4 RP	TTTGGAATCACCCTCCAG

5351	5' *csr-1* flank FP	TTTGGTCCCTACCCCAGACA
5352	5' *csr-1* flank RP	CCATTGACGACATTGCGGAG
5353	3' *csr-1* flank FP	CGCCGTTAATGCAGTTGTGAT
5354	3' *csr-1* flank RP	CCCCAGCAACTGCGTCTATT
6569	pCSR-1 sequencing FP	TCAGTTTACGTGATCTCATCAACCATCAG
6570	pBM61 sequencing primer	GTGTCTTTTAGCATTCAGACCCCATTAG
6644	pCSR-1 sequencing RP	GCTCCGCAATGTCGTCAATGG

Table S6. *E. coli* strains.

Strain	Description
E800	pCSR-1 in DH5a, amp
E977	p3426 in DH5a, amp
E978	p3427 in DH5a, amp
E979	p3428 in DH5a, amp
E980	p3429 in DH5a, amp

Bibliography

1. Austin, C. Chromatin. *Genome.gov* https://www.genome.gov/genetics-glossary/Chromatin.

2. Yang, X., Wong, M. P. M. & Ng, R. K. Aberrant DNA Methylation in Acute Myeloid Leukemia and Its Clinical Implications. *Int J Mol Sci* **20**, (2019).

3. Conway, E., Healy, E. & Bracken, A. P. PRC2 mediated H3K27 methylations in cellular identity and cancer. *Current Opinion in Cell Biology* **37**, 42–48 (2015).

4. Jamieson, K., Rountree, M. R., Lewis, Z. A., Stajich, J. E. & Selker, E. U. Regional control of histone H3 lysine 27 methylation in Neurospora. *PNAS* **110**, 6027–6032 (2013).

5. Højfeldt, J. W. *et al.* Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nature Structural & Molecular Biology* **25**, 225–232 (2018).

6. Alabert, C. *et al.* Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nature Cell Biology* **16**, 281–291 (2014).

7. Ferrari, K. J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell* **53**, 49–62 (2014).

8. Zee, B. M., Britton, L.-M. P., Wolle, D., Haberman, D. M. & Garcia, B. A. Origins and Formation of Histone Methylation across the Human Cell Cycle. *Mol Cell Biol* **32**, 2503–2514 (2012).

9. Sanz, L. A. *et al.* A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *EMBO J* **27**, 2523–2532 (2008).

10. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).

11. Wachter, E. *et al.* Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife* **3**, e03397 (2014).

12. Wang, X. *et al.* Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* **65**, 1056-1067.e5 (2017).

13. Mishra, S. K. *et al.* Characterization of highly conserved G-quadruplex motifs as potential drug targets in Streptococcus pneumoniae. *Scientific Reports* **9**, 1791 (2019).

14. Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. *PLOS Computational Biology* **6**, e1000861 (2010).

15. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627–8637 (2015).

16. Aramayo, R. & Selker, E. U. Neurospora crassa, a model system for epigenetics research. *Cold Spring Harb Perspect Biol* **5**, a017921 (2013).

17. Wu, C. *et al.* Characterization of Chromosome Ends in the Filamentous Fungus Neurospora crassa. *Genetics* **181**, 1129–1145 (2009).

18. Fu, C. *et al.* Identification and Characterization of Genes Required for Cell-to-Cell Fusion in Neurospora crassa. *Eukaryotic Cell* **10**, 1100–1109 (2011).

19. Meyne, J., Ratliff, R. L. & Moyzis, R. K. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc Natl Acad Sci U S A* **86**, 7049–7053 (1989).

20. Neidle, S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *The FEBS Journal* **277**, 1118–1125 (2010).

21. Jamieson, K. *et al.* Telomere repeats induce domains of H3K27 methylation in Neurospora. *eLife* **7**, e31216 (2018).

22. McNaught, K. J., Wiles, E. T. & Selker, E. U. Identification of a PRC2 accessory subunit required for subtelomeric H3K27 methylation in Neurospora. *Molecular and Cellular Biology* (2020) doi:10.1128/MCB.00003-20.

23. Demkovičová, E. *et al.* Telomeric G-Quadruplexes: From Human to Tetrahymena Repeats. *Journal of Nucleic Acids* https://www.hindawi.com/journals/jna/2017/9170371/abs/ (2017) doi:10.1155/2017/9170371.

24. Zhang, L., Tamura, K., Shin-ya, K. & Takahashi, H. The telomerase inhibitor telomestatin induces telomere shortening and cell death in Arabidopsis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1763**, 39–44 (2006).

25. Bardiya, N. & Shiu, P. K. T. Cyclosporin A-resistance based gene placement system for Neurospora crassa. *Fungal Genet. Biol.* **44**, 307–314 (2007).

26. Nowrousian, M. *Fungal Genomics*. (Springer Science & Business Media, 2014).

27. Margolin, B., Freitag, M. & Selker, E. Improved plasmids for gene targeting at the his-3 locus of Neurospora crassa by electroporation. *Fungal Genetics Reports* **44**, 34–36 (1997).

28. Musso, M., Bocciardi, R., Parodi, S., Ravazzolo, R. & Ceccherini, I. Betaine, Dimethyl Sulfoxide, and 7-Deaza-dGTP, a Powerful Mixture for Amplification of GC-Rich DNA Sequences. *The Journal of Molecular Diagnostics* **8**, 544–550 (2006).

29. Gessaman, J. D. & Selker, E. U. Induction of H3K9me3 and DNA methylation by tethered heterochromatin factors in Neurospora crassa. *PNAS* **114**, E9598–E9607 (2017).

30. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16**, 643–649 (2015).

31. Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **43**, 31–37 (2017).

32. Mozgova, I. & Hennig, L. The polycomb group protein regulatory network. *Annu Rev Plant Biol* **66**, 269–296 (2015).

33. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nature Genetics* **49**, 1546–1552 (2017).

34. Wang, C., Zhao, L. & Lu, S. Role of TERRA in the Regulation of Telomere Length. *Int J Biol Sci* **11**, 316–323 (2015).

35. Montero, J. J. *et al.* TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin. *Nat Commun* **9**, 1548 (2018).

36. Martadinata, H. & Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K+ Solution. *Biochemistry* **52**, 2176–2183 (2013).

37. Court, R., Chapman, L., Fairall, L. & Rhodes, D. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep* **6**, 39–45 (2005).

38. Marión, R. M. *et al.* TERRA regulate the transcriptional landscape of pluripotent cells through TRF1-dependent recruitment of PRC2. *eLife* **8**, e44656 (2019).

39. Zhou, Y. *et al.* Telobox motifs recruit CLF/SWN–PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nature Genetics* **50**, 638–644 (2018).

40. Davis, R. H. *Neurospora: Contributions of a Model Organism*. (Oxford University Press, 2000).

41. Wiles, E. T. *et al.* Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *bioRxiv* 868117 (2019) doi:10.1101/868117.

42. Froger, A. & Hall, J. E. Transformation of Plasmid DNA into E. coli Using the Heat Shock Method. *J Vis Exp* (2007) doi:10.3791/253.

43. Pomraning, K. R., Smith, K. M. & Freitag, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**, 142–150 (2009).

44. Miao, V. P. W., Freitag, M. & Selker, E. U. Short TpA-rich segments of the ζ-η region induce DNA methylation in Neurospora crassa11Edited by K. Yamamoto. *Journal of Molecular Biology* **300**, 249–273 (2000).

45. Jamieson, K. *et al.* Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res.* **26**, 97–107 (2016).