Polymerase pausing induced by sequence-specific RNA binding protein drives heterochromatin assembly

Jahan-Yar Parsa¹, Selim Boudoukha¹,4, Jordan Burke¹,4, Christina Homer¹ & Hiten D. Madhani¹,2,3

¹Dept. of Biochemistry and Biophysics
University of California
San Francisco, CA 94158

²Chan-Zuckerberg Biohub
San Francisco, CA 94158

³hitenmadhani@gmail.com

⁴equal contributions

Running title: RNAPII pauses induce heterochromatin

Key Words: Heterochromatin; Seb1; Polymerase Pausing; Centromere;
In *S. pombe*, transcripts derived from the pericentromeric *dg* and *dh* repeats promote heterochromatin formation via RNAi as well as an RNAi-independent mechanism involving the RNAPII-associated RNA-binding protein Seb1 and RNA processing activities. We show that Seb1 promotes long-lived RNAPII pauses at pericentromeric repeat regions and their presence correlates with the heterochromatin-triggering activities of the corresponding *dg* and *dh* DNA fragments. Globally increasing RNAPII stalling by other means induces the formation of novel large ectopic heterochromatin domains. Such ectopic heterochromatin occurs even in cells lacking RNAi. These results uncover Seb1-mediated polymerase stalling as a signal necessary for heterochromatin nucleation.
INTRODUCTION

Packaging of pericentromeric DNA into heterochromatin is crucial for genome stability, development and health, yet its endogenous triggers remain poorly understood (Allshire and Madhani 2018). A defining feature of pericentromeric heterochromatin is histone H3 lysine 9 methylation (H3K9me) (Rea et al. 2000; Bannister et al. 2001; Lachner et al. 2001). Investigations of the fission yeast, *Schizosaccharomyces pombe* (*S. pombe*), have played a major role in our understanding of this type of repressive chromatin. In *S. pombe*, transcripts derived from the pericentromeric *dg* and *dh* repeats during S phase (Volpe et al. 2002; Chen et al. 2008; Kloc and Martienssen 2008) promote heterochromatin formation through two parallel pathways. The first is an RNAi-dependent mechanism involving recruitment of the Clr4 H3K9 methyltransferase complex (CLR-C) via the Argonaute-containing RITS complex (Motamedi et al. 2004; Noma et al. 2004; Verdel et al. 2004; Sugiyama et al. 2005; Buhler et al. 2006; Bayne et al. 2010).

A second, less-characterized pathway does not require RNAi factors. Components that impact this mechanism include the RNAPII-associated RNA-binding protein Seb1, the repressor/remodeler complex SHREC, the nuclear 5’-3’ exonuclease Dhp1 (Rat1/Xrn2 in *S. cerevisiae*), the nuclear 3’-5’ exonuclease Rrp6, and the RNA export factor Mlo3/Yra1 (Sadaie et al. 2004; Sugiyama et al. 2007; Reyes-Turcu et al. 2011; Marina et al. 2013; Chalamcharla et al. 2015; Lemay et al. 2016; Tucker et al. 2016; Wittmann et al. 2017). Seb1, Dhp1, and SHREC promote RNAi-independent pericentromeric
heterochromatin assembly at pericentromeric regions (Marina et al. 2013; Chalamcharla et al. 2015; Tucker et al. 2016), while Rrp6 and Mlo3 inhibit it (Reyes-Turcu et al. 2011). Other fungi, such as *Neurospora crassa* and *Cryptococcus neoformans*, as well as somatic mammalian cells, do not require RNAi for heterochromatin assembly (Freitag et al. 2004; Wang et al. 2010; Chan and Wong 2012); thus, poorly-understood RNAi-independent mechanisms are important to investigate. In this paper, we show that Seb1, whose role in heterochromatin we established in prior work (Marina et al. 2013), displays extensive binding to *dg* and *dh* repeat RNAs and that it promotes long-lived pausing by RNAPII. We show that inducing pausing by other means can trigger the assembly of ectopic heterochromatin domains independently of RNAi. The cleavage-polyadenylation machinery antagonizes silencing, suggesting a possible mechanism by which heterochromatin is limited at mRNA-coding genes despite the presence of termination-associated pauses. These findings establish a role for Seb1-dependent RNAPII pausing in promoting the formation of repressive chromatin structures.
RESULTS

Nascent Elongating Transcript Sequencing reveals a role for Seb1 in the pausing of RNA Polymerase II

To understand how Seb1 interfaces with the transcription of dg and dh repeats to promote heterochromatin, we employed a previously identified viable, heterochromatin-defective allele, seb1-1 (Marina et al. 2013). When combined with mutants in the RNAi machinery, seb1-1 eliminates pericentromeric heterochromatin, while the corresponding single mutants decrease H3K9me, indicative of partially redundant pathways (Marina et al. 2013). We examined transcription of heterochromatin at single-nucleotide resolution and tested the impact of the seb1-1 allele using Nascent Elongating Transcript Sequencing (NET-Seq) (Churchman and Weissman 2011). To analyze the intrinsic transcriptional properties of heterochromatic sequences prior to the establishment of heterochromatin assembly, we used the clr4Δ mutant, which lacks H3K9me and displays full derepression of most silenced chromatin regions. We compared this strain to a clr4Δ seb1-1 double mutant to assess the impact of seb1-1. We first examined the effect of seb1-1 on transcription of non-heterochromatic regions (Figure 1). Initial inspection revealed numerous genes with a decreased peak density at 5’ regions in the double mutant with increased peak density upstream of annotated cleavage-polyadenylation sites (often called Transcription end sites or TESs) (see Figure 1A-B for examples). To analyze these trends genome-wide, travelling ratios were computed on replicate data to assess relative polymerase progression for the 500 bp segment downstream of the transcription start site (TSS; 5’ traveling ratio) and for the 500 bp
segment upstream of annotated TES (3’ traveling ratio) (Figure 1C; see Methods). A lower travelling ratio in mutant vs. wild-type implies lower pausing over the region examined in the mutant and vice-versa for higher ratios. Iterative K-means clustering revealed three groups (Figure 1D) two of which, representing 77% of genes in our dataset, are significantly impacted by the seb1-1 allele (Figure 1D Clusters I and II; Supplemental Fig. S1). The seb1-1 mutation causes a reduced median 5’ traveling ratio and an increased median 3’ traveling ratio for both clusters (Figure 1D -- clusters I and II; Figure 1E – top and middle panels) while no significant changes were observed for the third cluster (Figure 1E – Cluster III, bottom panel; see Supplemental Fig. S1 for p values). These data indicate that the seb1-1 allele leads to decreased RNAPII pausing at gene 5’ ends with an associated increased 3’ signal; the latter may be due to polymerase release from upstream pauses.

A direct role for Seb1 in RNAP II pausing and heterochromatin assembly in the pericentromeric dh and dg repeats

Our prior RIP-qPCR analysis indicates that Seb1 functions directly in heterochromatin assembly by binding pericentromeric dg and dh repeat transcripts (Marina et al. 2013). Further, comparison of the transcriptomes of WT and seb1-1 using RNA-seq revealed no significant changes (p<0.01 and |log₂(fold change)| >1) in the transcript levels of known silencing factors (Supplemental Fig. S2 and Supplemental Table S1). To assess direct interactions of Seb1 with pericentromeric RNA at single nucleotide resolution and across the entirety of the dg and dh regions, we conducted PAR-CLIP in replicate on a clr4Δ strain. Computational analysis identified statistically significant Seb1 PAR-CLIP
read clusters (Corcoran et al. 2011) and confirmed direct binding of Seb1 to pericentromeric transcripts (Figure 2A and B – top panels; Supplemental Fig. S3A-C) via a motif described previously by others for non-heterochromatic sites bound by Seb1 (Lemay et al. 2016; Wittmann et al. 2017) (UGUA; DREME motif analysis (Bailey 2011); p=2.9e-9; E=7.4e-7). Our analysis of Seb1 PAR-CLIP read clusters for coding genes recapitulates published work and is not discussed here further (Lemay et al. 2016; Wittmann et al. 2017).

To compare the binding of Seb1 across transcript classes, we computed the fraction of RNA covered by Seb1 PAR-CLIP read clusters. We observed a ~12-fold higher PAR-CLIP cluster coverage for pericentromeric repeat intervals than for coding gene intervals (Figure 2D and Supplemental Fig. S4A). Non-coding RNAs display the highest coverage at a mean level ~100-fold higher than that of coding-genes (Supplemental Fig. S4B).

We next examined the NET-Seq profiles of pericentromeric heterochromatin sequences of clr4Δ and clr4Δ seb1-1 strains (replicate experiments were conducted). Pericentromeric regions harbor detectable NET-Seq signal in clr4Δ cells despite a low level of polymerase engagement at any single nucleotide (Figure 2A and B – middle panels). The signal overlaps with regions of high Seb1 PAR-CLIP cluster coverage (Figure 2A and B – top panels). Notably, a handful of discrete peaks indicative of pausing were observed, and the largest were lost in the seb1-1 mutant for both antisense and sense signals (Figure 2A and B – bottom panels and Supplemental Table
RNAPII densities at the pericentromeric regions are comparable between $\text{clr4}\Delta$ and $\text{clr4}\Delta \text{seb1-1}$ strains as assessed by ChIP-seq analysis of Rpb3-3xFLAG (Figure 2C and Supplemental Fig. S4C), indicating that decreases in NET-seq peak intensities caused by the $\text{seb1-1}$ mutation are not trivially explained by loss RNAPII recruitment ($n.b.$ rare pauses may not impact overall PolII density as measured by ChIP-seq). Notably, calculation of polymerase dwell times (Larson et al. 2014) at centromeres and across the genome and genotypes revealed that Seb1-dependent pauses are significantly longer-lived on average than Seb1-independent pauses (Figure 2E and Supplemental Fig. S4D). These data reveal detectable Seb1-dependent RNAPII pauses in pericentromeric sequences.

Previous studies identified two segments of pericentromeric DNA that can trigger heterochromatin [L5 (Partridge et al. 2002; Volpe et al. 2003) and Frag1 (Marina et al. 2013) respectively]. Frag1 defined a segment that requires both RNAi and Seb1 for its activity (Marina et al. 2013). To compare activity of pericentromeric fragments to their transcription properties described above, we extended this analysis using a system we employed previously (Marina et al. 2013). The cen1R region was divided into nine overlapping fragments (Figure 2F and Supplemental Table S3). Each fragment was placed downstream of an $\text{adh1}^+$ promoter ($\text{padh1}^+$) in either forward or reverse orientation, and upstream of a transcription terminator. This insert was then placed downstream of $\text{ura4}^+$ (Figure 2G). Silencing of $\text{ura4}^+$ was determined using YS-FOA plates, which selects for $\text{ura4}^+$ repression. The insert of Fragment 1 (Frag1) displays silencing activity; this construct was used previously to isolate the $\text{seb1-1}$ mutant and
was shown to require the padh1\(^+\) promoter for silencing activity (Marina et al. 2013). Three additional fragments exhibit strong silencing activity and each is functional in only one orientation (Figure 2H; Frag2A, 8S and 9S). Thus, these pericentromeric regions harbor a transcription-dependent, orientation-specific signal capable of triggering silencing. To examine the relationship of these regions to those that display detectable Seb1-dependent pauses, we identified clusters of NET-seq peaks (see Methods) and computed the total read density of these clusters within each fragment. A comparison of clr4\(^\Delta\) to clr4\(^\Delta\) seb1-1 strains revealed significant correlation ($\chi^2=12.6$, p<0.001) (Figure 2H – Frag1A, Frag2A, and Frag9S; Supplemental Fig. S5A-C). The exception, Frag8S, displays silencing activity but no detectable Seb1-dependent NET-seq peak clusters (although it does display Seb1-dependent NET-seq signal; Figure 2H and Supplemental Fig. S5D). The heterochromatin assembly activity of this fragment may be pause-independent or, the relevant RNAPII pauses may be below the sensitivity of NET-seq. Introduction of the seb1-1 allele into these reporter strains (Frag2A, 8S and 9S) by homologous recombination leads to reduced silencing of ura4\(^+\) (Figure S5E). The residual activity is likely due to the parallel RNAi pathway (Marina et al. 2013). These data indicate that Seb1 directly recognizes dg and dh RNAs, induces detectable pausing in centromere fragments, and promotes the silencing activity directed by these sequences.

**Global enhancement of RNAPII pausing can trigger ectopic heterochromatin assembly**
The shared defect of the viable \textit{seb1-1} allele in both heterochromatin assembly and RNAPII pausing suggests that pausing signals the assembly of heterochromatin. However, given that Seb1 may have other activities impacted by \textit{seb1-1}, we sought an orthogonal test of the role of pausing. Thus, we pursued an alternative strategy of testing whether increasing RNAPII pausing \textit{per se} could be sufficient to trigger heterochromatin assembly and if so whether such an activity required RNAi. We exploited the conserved elongation factor TFIIS which binds paused RNAPII complexes and stimulates RNA hydrolysis by RNAPII, enabling restart (Mayer et al. 2017). It is thought that all genes are subject to this type of rescue mechanism as pausing is a ubiquitous feature of transcription. Two conserved acid residues in domain III of TFIIS are required for catalysis (Jeon et al. 1994). Mutation of these residues to alanine prevents the cleavage of the RNA by RNAPII (Jeon et al. 1994), ultimately resulting in polymerase trapped in a lethal paused/backtracked state (Sigurdsson et al. 2010; Imashimizu et al. 2013). We introduced the corresponding D274A and E275A mutations in the TFIIS gene \textit{tfs1+}, creating a dominant negative allele \textit{tfs1}^{DN} (Figure 3A) (Lemay et al. 2014). To overcome lethality of this allele (Sigurdsson et al. 2010; Imashimizu et al. 2013), we placed it under control of an \textit{nmt1}^{+} thiamine-repressible promoter and inserted it at the \textit{leu1}^{+} locus (Matsuyama et al. 2004), enabling concerted expression of \textit{tfs1}^{+} and \textit{tfs1}^{DN}. We conducted replicate NET-seq analysis on \textit{clr4}\Delta and \textit{clr4}\Delta \textit{tfs1}^{DN} cells under inducing conditions. Genome-wide analysis of 5’ and 3’ traveling ratios indicated TFIIS^{DN}-dependent increases in RNAPII pausing at gene 5’ ends and a more modest effect at 3’ ends (Supplemental Fig. S6A-B).
To test whether stabilizing endogenous RNAPII pauses in this manner triggers de novo ectopic heterochromatin, we performed H3K9me2 ChIP-seq analysis on $tfs1^{DN}$ strains; however, we observed no effects in this background (data not shown). Heterochromatin components are limiting and antagonized by anti-silencing factors, particularly Epe1, which actively removes the H3K9me mark (Zofall and Grewal 2006; Aygun et al. 2013; Audergon et al. 2015). Thus, we constructed epe1Δ $tfs1^{DN}$ mutant strains or epe1Δ strains carrying an integrated vector-only control (epe1Δ-vc), collected multiple strain isolates for each, and performed H3K9me2 ChIP-seq on all isolates (Figure 3B). Because high-level TFIIS$^{DN}$ expression is lethal in epe1Δ cells (Supplemental Fig. S6C – middle panel), experiments were performed in the presence of thiamine, enabling viability (Supplemental Fig. S6C – right panel). Consistent with a slight fitness defect under these conditions (Supplemental Fig. S6C – right panel), RNA-seq analysis revealed that 13% of the transcript pool from the $tfs1$ genes arise from the $tfs1^{DN}$ allele and 87% arise from the wild-type allele when cells are grown in thiamine (Supplemental Fig. S6D), indicating leaky repression. The ChIP-seq data obtained from strain isolates were examined for H3K9me peaks (see Methods). We filtered the results for well-established heterochromatic recruitment sites (including HOODs, Islands, meiotic genes, Epe1-bound genes etc; see Methods), as these genomic regions have an intrinsic propensity (e.g. via RNAi or the RNA elimination machinery) to nucleate H3K9me (Zofall et al. 2012; Yamanaka et al. 2013; Wang et al. 2015).

Remarkably, of 13 isolates derived from epe1Δ $tfs1^{DN}$ parents analyzed by H3K9me ChIP-seq, five separate isolates harbor a distinct ectopic region of heterochromatin,
which we termed Pause-Induced Ectopic heterochromatic Region (PIER) (Figure 3C and d; Supplemental Fig. S7). No novel ectopic heterochromatic loci were observed by whole genome H3K9me2 ChIP-seq in the 15 epe1Δ-vc strains ($\chi^2=7.02$, p=0.0082). PIERs range in size from ~3 to ~15 kb, and each PIER was unique. Three PIERs are bounded by an essential gene on at least one side of the locus suggesting that selection likely prevents observing PIERs that assemble over essential genes (Figure 3D PIERs 2, 3, and 5); this implies that our approach may underestimate the propensity of PIER formation. H3K9me enrichment at known heterochromatin nucleation sites is unrelated to tfs1 genotype and summarized for all strains in Supplemental Fig. S8. These results indicate that TFIIS$^{DN}$-induced RNAPII pausing can be sufficient to nucleate heterochromatin at novel sites.

To determine if H3K9me at PIERs lead to repression, we conducted RNA-seq analysis on epe1Δ tfs1$^{DN}$ isolate 2 (containing PIER 2) (Supplemental Fig. S9A and Supplemental Table S4). We observed a significant decrease (p<0.01) in two of the three genes present in PIER2, cta3$^{+}$ and its8$^{+}$ (Supplemental Fig. S9A and Supplemental Table S4). TFIIS$^{DN}$ expression does not alter the expression of known heterochromatin factors (p<0.01, |log$_2$(fold-change)|>1; Supplemental Fig. S9B; Supplemental Table S4).

**PIER induction is independent of RNAi**

Four out of five PIERs contain overlapping convergent genes that have the potential to form double stranded RNA (Figure 3D; PIERs 1, 2, 4, and 5). As double-stranded RNA
can trigger heterochromatin in cis via the RNAi pathway (Simmer et al. 2010), we determined whether the de novo establishment of PIERs requires RNAi. We constructed *epe1Δ ago1Δ* strains by disrupting *ago1* in the *epe1Δ* mutant. Initial ChIP-seq analysis of three isolates revealed two that display ectopic heterochromatin at the *clr4Δ* locus and reduced levels of H3K9me at constitutive heterochromatic loci (Supplemental Fig. S10 and Supplemental Fig. S11B and C top panels). Such adaptive silencing of *clr4Δ* has been described previously in *epe1Δ mst2Δ* strains (Wang et al. 2015) and evidently can occur in strains lacking Epe1 and RNAi. Remarkably, upon integration of *tfs1DN* into these two *epe1Δ ago1Δ* strains, H3K9me2 ChIP-seq revealed establishment of heterochromatin at constitutively silenced loci (e.g. centromeres) and its loss at the *clr4Δ* locus (Supplemental Fig. S11B and C – isolates 1, 3, 4, and 5). This observation indicates that pericentromeric repeats harbor the ability to respond to RNAPII pausing and assemble heterochromatin independently of RNAi. Accompanying these changes, six out of the 12 *epe1Δ ago1Δ tfs1DN* isolates subjected to whole-genome analysis acquired PIERs (Figure 4; Supplemental Fig. S11). Five of six of these PIERs are bounded by an essential gene on at least one side of the region similar to that seen in the *epe1Δ tfs1DN* (Figure 4; PIERs 6-10), consistent with the notion that essential genes limit our ability to observe ectopic heterochromatin. Again, each PIER was unique. Thus, PIERs can be triggered by RNAPII pausing even in cells lacking a functional RITS complex.

The cleavage and polyadenylation machinery antagonizes heterochromatin assembly
To address the role of Seb1 in the formation of PIERs, we interrogated 15 independent isolates of *seb1-1 epe1Δ tfs1DN* for ectopic heterochromatin assembly by whole-genome chromatin immunoprecipitation. A single PIER was observed (Figure 5A), a frequency significantly lower than that produced by *epe1Δ tfs1DN* strains ($\chi^2=4.18$, $p=0.0409$). Two additional regions of ectopic heterochromatin appear evident by visual inspection but were below our cutoffs for H3K9me enrichment (Supplemental Fig. S12 A-C). These data suggest that while frequent PIER formation and high levels of H3K9me accumulation are promoted by Seb1, the *seb1-1* allele is permissive for some degree of H3K9me triggered by *tfs1DN*.

Our data demonstrate that Seb1 promotes RNAPII pausing at centromeres, the major site of constitutive heterochromatin assembly, but also at numerous additional sites (Figure 1). Why then is heterochromatin assembly then restricted to specific sites under standard growth conditions? A potential clue comes from studies that show that Seb1 promotes cleavage-polyadenylation (CPA)-termination and that Seb1 copurifies with the CPA machinery (Lemay et al. 2016; Wittmann et al. 2017). One possibility is that recruitment of the CPA by strong polyadenylation signals might suppress heterochromatin assembly, thereby neutralizing the heterochromatin-promoting effects of Seb1. To test this idea, we utilized a well-characterized temperature sensitive mutant of a CPA factor, *pfs2-11*, which produces transcription read-through past the CPA site under nonpermissive conditions (Wang et al. 2005). Strikingly, under these conditions, we observed increased levels of H3K9me at all centromeres (Figure 5B-D; Supplemental Fig. S13) as well as increased heterochromatin assembly over a 15 kb
subtelomeric region on the right arm chromosome 1 (Figure 5B; Supplemental Fig. S13). Notably, the increase in centeromeric H3K9me is not at the expense of its loss elsewhere as occurs in some mutant backgrounds to due limiting amounts of silencing factors (Allshire and Madhani 2018). These data support a model in which the CPA machinery or its actions suppresses heterochromatin formation at sites of Seb1-promoted CPA (Figure 5E). They are also consistent with the observation that deletion of sequences that include polyadenylation signal enhance heterochromatin assembly on a reporter gene triggered in trans by expression of an artificial hairpin (Yu et al. 2014).
DISCUSSION

Our results indicate that Seb1, a conserved RNAPII-associated RNA binding protein that mediates RNAi-independent heterochromatin assembly in S. pombe (Marina et al. 2013), is enriched on pericentromeric ncRNA transcripts relative to coding sequences and promotes long-lived RNAPII pauses. Remarkably, pausing is sufficient to trigger ectopic heterochromatin assembly in an RNAi-independent fashion, indicating that this is a relevant activity of Seb1 in promoting heterochromatin assembly. Binding of Seb1 to euchromatic ncRNAs (e.g. snRNAs) is not associated with detectable heterochromatin assembly, which may be due to high levels of transcription, which induces anti-silencing histone marks and histone turnover, both of which antagonize silencing (Allshire and Madhani 2018). At mRNA-coding genes, heterochromatin assembly suppression by the CPA machinery may also play a role in specificity (Figure 5). Additionally, the repetitiveness of pericentromeric sequences may also contribute to specificity by producing a threshold density of paused polymerases within a discrete genomic interval. Testing these and other possibilities will require the development of tools that enable the programming of pauses of defined length at defined sites and at define levels of transcription. Relevant to the issue of heterochromatin assembly specificity is a recent study (published while this paper was in revision) that reported that the low temperature culturing of S. pombe triggers H3K9me heterochromatin islands across the genome independently of RNAi (Gallagher et al. 2018). Thus, the ability of “euchromatic” sites to assemble facultative heterochromatin is evidently higher than might have been assumed.
Our data are germane to the observation that mutations in the Paf1 complex (Paf1-C), a multifunctional elongation complex that binds cooperatively to RNAPII with TFIIA (Xu et al. 2017), enables synthetic hairpin RNAs to trigger heterochromatin in trans and increases heterochromatin spreading in S. pombe (Kowalik et al. 2015; Sadeghi et al. 2015; Verrier et al. 2015). While the elongation-promoting activity of Paf-C has been suggested to limit heterochromatin by limiting targeting of RITS to the nascent transcript (Kowalik et al. 2015), it may also act via RNAi-independent mechanisms as we find that increased RNAPII pausing can trigger H3K9me independently of RNAi. Seb1-triggered RNAPII pausing may drive heterochromatin assembly by promoting the heterochromatic stalling of replisomes associated with CLR-C through RNAPII-replisome collisions as proposed (Li et al. 2011; Zaratiegui et al. 2011). Analogous concepts have been put forth in S. cerevisiae where tight protein-DNA interactions are sufficient to trigger recruitment of the SIR complex (Dubarry et al. 2011). Consistent with this hypothesis, such transcription-replication conflicts are limited by Paf1-C (Poli et al. 2016), which inhibits heterochromatin assembly, while slowing of replisome progression enhances heterochromatin spread (Singh and Klar 2008; Li et al. 2017). It has also been proposed that the 5’→3’ RNA exonuclease Dhp1 (related to S. cerevisiae Rat1/Xrn2), which is required for RNAi-independent heterochromatin assembly, recruits the silencing machinery via a physical interaction with CLR-C (Chalamcharla et al. 2015; Tucker et al. 2016). Because RNAPII pausing enhances recruitment of Xrn2 (Wagschal et al. 2012; Contreras et al. 2013), and Seb1 copurifies with the Dhp1 (Lemay et al. 2016; Wittmann et al. 2017), Seb1-induced pausing may promote heterochromatin assembly via this
mechanism as well. Our model also readily accommodates genetic observations that null mutants in S. pombe RNAPII elongation factors suppress the H3K9me defect of RNAi mutants (Reyes-Turcu et al. 2011; Sadeghi et al. 2015), as well as analogous observations for mutants in RNA biogenesis factors (Reyes-Turcu et al. 2011) as these factors also promote transcriptional elongation (Luna et al. 2012). Weak cleavage-polyadenylation signals promote heterochromatin assembly (Yu et al. 2014), which is predicted to result in accumulation of paused RNAPII at S. pombe Downstream Pause Elements (Aranda and Proudfoot 1999). Another key factor recruited to pericentromeric regions by Seb1 is remodeling/HDAC complex SHREC (Marina et al. 2013). Thus, Seb1-paused RNAPII may promote heterochromatin assembly through multiple mechanisms. Given the tight coupling of this heterochromatin signal to RNAPII activity, it is tempting to speculate that Seb1-mediated pausing may have evolved from a surveillance mechanism for silencing foreign DNA. Finally, pathogenic triplet repeat expansions in the Friedreich Ataxia gene FXN concomitantly display a block to transcriptional elongation and the appearance of H3K9me on FXN (Punga and Bühler 2010; Li et al. 2015), raising the possibility that pause-induced heterochromatin underlies disease pathogenesis.
**Materials and Methods**

*Yeast strains, plasmids, and media*

A list of all *S. pombe* strains and plasmids used in this study is provided in Supplemental Table S5. Cells were grown at 30°C in synthetic complete medium (SC) with adenine and amino acid supplements with reduced levels of uracil (150mg/L) for PAR-CLIP, or in Edinburgh minimal medium (EMM) supplemented with adenine, uracil, and the appropriate amino acids with or without thiamine (15µM) for NET-seq and ChIP-seq.

*NET-seq*

NET-seq experiments were conducted as previously described (Churchman and Weissman 2012) with minor alterations for *S. pombe*. *S. pombe* cultures were grown in 1L EMM without thiamine to an OD\textsubscript{600} of 0.7 and harvested via filtration and flash frozen in liquid nitrogen. Lysis and immunoprecipitation was conducted as previously described (Shetty et al. 2017). Adaptor ligation was performed using random hexamer-barcoded adaptors. All strains were analyzed in duplicate and sequencing was conducted on a HiSeq 4000 platform.

*RNA-seq*

Strains were grown in YS media + 3% Glucose overnight to OD\textsubscript{600} = 0.7. Cells were harvested by centrifugation, washed twice with ice-cold water and flash-frozen. Pellets were resuspended in 1ml Trizol (Thermo Fisher Scientific, #15596026). 0.5 mm
zirconia-silica beads (BioSpec, #11079105z) were added and lysis was accomplished by three cycles of bead beating for 90 seconds on high (Bead Ruptor 12 Homogenizer, OMNI International). Following centrifugation at 14,000 rpm for 10 min at 4°C, the supernatant was transferred to a microcentrifuge tube, extracted once with chloroform, and precipitated with isopropanol. Following resuspension and re-precipitation with isopropanol, pellets were washed with 75% ethanol and air dried for 30 minutes. Pellets were resuspended in 150µl RNAse-free water.

1mg of total RNA was used to isolate mRNA using PolyATtract® Systems III and IV (Promega, #Z5310) according to the manufacturer’s instructions. Input RNA quality and mRNA purity were verified by Bioanalyzer RNA 6000 Pico kit (Agilent, #5067-1513). To address the issue of genomic DNA contamination in RNA samples, we use Zymo RNA Clean & Concentrator Kit 5® (Zymo Research, #11-326) according to the manufacturer’s instruction. A sequencing library was constructed using NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England Biolab, #E740S). Librairies were analyzed for quality and average size on Bioanalyzer High Sensitivity DNA kit (Agilent, #5067-4626). The sequencing was performed on an Illumina HiSeq 4000 platform.

**PAR-CLIP**

PAR-CLIP experiments were conducted as a combination of PAR-CLIP and CRAC protocols (Granneman et al. 2009; Hafner et al. 2010) in two replicates. Cells were grown in 2L of SC medium to an OD$_{600}$ of 0.75. 4-thiouracil (Sigma, #440736-1G) was
added to a final concentration of 1.3 mM for 15 minutes in the dark at 30°C. Samples were immediately crosslinked using the UV Power-Shot Handheld UV Curing System (SPDI UV) at a 365 nm wavelength for 15 minutes while continuously stirred. Samples were collected by filtration, resuspended in 6 mL of buffer TMN150 (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 1.5 mM MgCl₂, 0.5% NP-40, 5 mM beta-mercaptoethanol) and frozen as “yeast popcorn” by dropwise addition to liquid nitrogen. This material was lysed by ball mill (15 Hz, 3 minutes, 5 cycles) (Mixer Mill MM 301, Retsch). CRAC was then performed from this point on as previously described (Granneman et al. 2009) with the exception of gel extraction, which was conducted by electroelution of the gel piece containing the radioactively labeled RNA sample using D-Tube™ Dialyzer Midi tubes (EMD Millipore, #71507-3). Electroelution was carried out in 1X MOPS SDS PAGE Buffer (5 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA) and run for 2 hrs at 100V. The isolated sample was transferred to a fresh tube and 100 ug of Proteinase K (Sigma, #P2308) was added to the sample. Procedures post-Proteinase K treatment were conducted as previously described (Granneman et al. 2009) and samples were sequenced on the HiSeq 4000 platform.

**Spotting Assay**

Strains were grown overnight to saturation and diluted to OD600 of 1. Serial dilutions were performed with a dilution factor of 5. For *ura4* silencing assays, cells were grown on non-selective and 5-fluoroorotic acid (FOA) (2 g/L unless otherwise stated) YS plates at 30°C for 3 days. For *tfs1DN* induction assays, cells were grown on YS, EMM -leu, -thiamine, or EMM -leu +thiamine, for 3 days.
**ChIP-seq**

ChIP-seq was conducted as previously described (Inada et al. 2016). ChIP-Seq immunoprecipitations were performed with 10 µg anti-H3K9me2 (Abcam, #ab1220), and 10 µg anti-FLAG (Sigma, #P3165). Samples were sequenced on a HiSeq 4000 platform.

**NET-seq analysis: Genome alignment**

For NET-seq analysis, adapter sequences (using ATCTCGTAT) were removed and reads were flattened to remove sequence duplicates. Barcoded reads were then mapped to the *S. pombe* genome (Wood et al. 2012) using BOWTIE (Langmead et al. 2009) to align and omit any sequence reads that were misprimed during the reverse transcription step of NET-seq and thus lack a barcode using the following flags: -M1 --best --strata. Unaligned files were collected for further analysis. Barcodes were removed, and the new unique, debarcoded reads were realigned to the genome using the following flags in BOWTIE: -M1 --best --strata.

**NET-seq analysis: Cluster finding**

High-density regions of NET-seq signal were defined across centromeres and coding regions to compare NET-seq density between genotypes. First, NET-seq peaks were discovered by calculating robust Z-scores (based on median and median absolute deviation) from the log2 transform of the number of reads starting at each position in the defined region (centromere fragment or transcript). Positions with a robust Z-score of at
least 2 and at least 10 unique reads were considered peaks. Next, peaks were clustered together using a sliding window (width=50, increment=10). The density of the cluster is calculated as the number of reads in the cluster divided by the size of the cluster in kilobases (kb).

To determine cluster densities for each fragment derived from the right arm of centromere 1, the sum of cluster densities was normalized to the sum of all densities in each sample. Error bars represent the range of two replicates.

**NET-seq analysis: Traveling ratio**

Traveling ratios were calculated for every non-overlapping annotated transcript at least 1000 nt in length according to the Pombase annotation (Wood et al. 2012). Transcripts with fewer than 50 total reads were excluded from the analysis. The traveling ratio was determined for a 0.5 kb window either immediately after the transcription start site (5’ traveling ratio) or immediately before the cleavage polyadenylation site (3’ traveling ratio). Transcripts <1kb in length were omitted from this analysis to ensure the 0.5kb 5’ and 3’ windows used for each traveling ratio do not overlap. Reads were counted in this window and across the entire transcript and then divided by the size of the window or transcript, respectively. Transcripts were clustered using K-means (sklearn.cluster.Kmeans, 3 centroids) based on the travelling ratio at each end of the clr4Δ and clr4Δ seb1-1 mutant on one replicate. P-values for each cluster for the difference between the clr4Δ and clr4Δ seb1-1 traveling ratio distributions were determined by KS test for each pair of replicates. Traveling ratio CDF plots were similar between replicates and a single replicate is presented.
**NET-seq analysis: Dwell time**

Dwell time was determined by normalizing peak height to the average NET-seq signal density of the surrounding 100 nt. NET-seq peaks with at least a two-fold decrease from *clr4Δ* to *clr4Δ seb1-1* in both replicates were considered Seb1-dependent. P-values were determined by KS test.

**RNA-seq analysis**

Analysis was performed using TopHat (Trapnell et al. 2009) and DESeq2 (Love et al. 2014). Changes in transcript expression levels required >2-fold change in mutants compared to wildtype to be considered significantly changed enough to have a functional consequence. Data analysis was performed on 2 replicates per condition.

To determine the fraction of reads derived from the expression of *tfs1DN* we divided the total number of reads that specifically aligned to the mutated region of the *tfs1DN* allele by the total number of reads (both WT and mutant alleles) that aligned to this same region.

**Seb1 PAR-CLIP data analysis by PARalyzer**

For PAR-CLIP analysis, adapter sequences were removed and reads were mapped to the *S. pombe* genome (Wood et al. 2012) using BOWTIE (Langmead et al. 2009), allowing for three mismatches with the following flags: -M1 -v3 --best --strata. Seb1 binding site read clusters were identified with PARalyzer (Corcoran et al. 2011). Reads
of < 20nt were omitted, and read clusters required at least 10 reads and at least two T→C conversions per cluster to be called as a Seb1 binding site. The PARalyzer OUTPUTCLUSTERSFILE file was converted to a genome browser readable file (.bam) for analysis. PAR-CLIP cluster coverage was calculated as the fraction of the interval of interest harbouring covered by a PARalyzer-called PAR-CLIP cluster (centromere arm, coding gene or ncRNA). P-values were determined by KS test.

**DREME motif analysis**

DREME (Bailey 2011) motif discovery for short, ungapped sequences was utilized to find Seb1-specific binding motifs. PAR-CLIP clusters – that were flattened to remove identical, recurrent sequence clusters originating from all three centromeric regions – were subjected to motif analysis. A shuffled sequence set created from the input sequences was utilized as a control.

**ChIP-seq analysis**

ChIP-seq analysis was conducted as previously described (Inada et al. 2016). Briefly, adaptor sequences from ChIP-seq sequencing libraries were removed (using GATCGGAAGA) and reads <20nt were omitted. Reads were aligned to the *S. pombe* genome (Wood et al. 2012) using BOWTIE (Langmead et al. 2009) with the following flags: -M1 --best --strata. Aligned reads were smoothed over a 1kb window.

**ChIP-seq analysis: PIER discovery**
H3K9me ChIP-seq peaks were considered as novel ectopic sites of H3K9me if two criteria were met: 1) H3K9me peaks were ≥3-fold higher than the genome background signal in the isolate, and 2) when normalized to the WCE (denoted “H3K9me enrichment” in the figures), the H3K9me signal at the peak was ≥3-fold higher than the background and parental H3K9me ChIP-seq signals. A curated list of genomic regions previously observed to have a propensity to form heterochromatin in various S. pombe backgrounds (Yamanaka et al. 2012; Zofall et al. 2012; Wang et al. 2015) was generated (Supplemental Table S6). In epe1Δ backgrounds, oscillation and spreading of H3K9me can occur (Trewick et al. 2007); thus, peaks within 10kb of our curated list of H3K9me nucleation sites, or within 10kb of H3K9me regions present in the parental strain, were not counted as novel H3K9me nucleation events.

ChIP-seq analysis: H3K9me levels at HOODs, islands, meiotic genes, and PIERs

For all isolates and whole cell extracts, RPKMs for each region in Supplemental Table S6 and all PIERs were normalized to the RPKM of a 10kb window surrounding each region (5kb upstream and downstream). The ratio of H3K9me enrichment values from isolates to the whole cell extracts and plotted as a heatmap (Supplemental Fig. S8)

Data sets

All available sequencing data sets are listed in Supplemental Table S7 and were deposited in the Gene Expression Omnibus with the SuperSeries accession number GSE114540.
ACKNOWLEDGEMENTS

We thank the members of the Madhani lab for support and scientific discussion, Robin Allshire (The University of Edinburgh) for the gift of the pDUAL-pnmt1+ plasmid, Fred Winston and Ameet Shetty (Harvard Medical School) for sharing their S. pombe NET-seq protocol, Chris Norbury (The University of Oxford) for the gift of the pfs2-11 S. pombe strain, and Jeffrey Hussmann (UCSF) for helpful discussions on data analysis. We thank Diana Marina, Smita Shankar, and Ken Finn for their early contributions to this work. We thank Sigurd Braun, Geeta Narlikar, Robin Allshire, Phillip Dumesic and Bassem Al-Sady for critical review of the manuscript. Research in the Madhani laboratory is supported by grants from the U.S. National Institutes of Health. H.D.M. is a Chan-Zuckerberg Biohub Investigator.

J.-Y.P. and H.D.M. designed the study. J.-Y.P. performed all of the experiments reported in the paper except for the RNA-seq work which was performed by S.B. J.E.B. performed computational analysis and C.H. contributed software for the analysis. J.-Y.P. and H.D.M. wrote the manuscript with input from all authors.
REFERENCES


Contreras X, Benkirane M, Kiernan R. 2013. Premature termination of transcription by RNAP II. Transcription 4: 72-76.


Figure 1. Seb1 controls polymerase progression. A-B) NET-seq signatures for clr4Δ (black) and clr4Δ seb1-1 (red) strains for two genes (SPAC1B3.12c and SPAC31G5.17c). C) Traveling ratios at the 5’ and 3’ regions of genes. D) K-means clustering of NET-seq 5’ and 3’ traveling ratios (Cluster I, n=373; Cluster II, n=911; Cluster III, n=383). NET-seq replicates are represented for each clr4Δ and clr4Δ seb1-1. E) Cumulative distribution function (cdf) plots of 5’ (left column) and 3’ (right column) traveling ratios for each cluster from (D) comparing clr4Δ (black) to clr4Δ seb1-1 (red) strains in each plot. KS tests were conducted for p-values (Supplemental Fig. S1).
Figure 2. Seb1 directly binds to pericentromeric transcripts and induces RNAPII pausing in silencing-promoting segments. A-B) Seb1 PAR-CLIP read clusters (log2 reads) and NET-seq peaks (RPM) aligned to the right arm of centromere 1 in clr4Δ (PAR-CLIP) or comparing clr4Δ (black) and clr4Δ seb1-1 (red) strains (NET-seq). A) Reads clusters/peaks aligning to the antisense transcript. B) Read clusters/peaks aligning to the sense transcript. Arrows indicate locations of Seb1-dependent NET-seq peak clusters identified computationally (see Methods). C) ChIP-seq data of Rpb3-3xFLAG for the right arm of centromere 1. Comparing RNAPII enrichment in clr4Δ (black) and clr4Δ seb1-1 (red). D) Seb1 PAR-CLIP read cluster coverage at the centromeres compared to coding genes. E) RNAPII dwell time analysis for centromeres and coding genes comparing Seb1-dependent (dark grey) and Seb1-independent (light grey) pauses (**p=0.0068; ***p=3.6e-59) F) Illustration of the right arm of centromere 1 and nine overlapping fragments analyzed G) Illustration of the reporter construct utilized to determine the silencing capacity of centromere fragments (F). adh1+ promoter (padh1+); a bidirectional terminator (term); B-boxes boundary element (B); nourseothricin-resistance gene (natR). Construct was placed downstream of ura4+. H) Silencing assays for each fragment in antisense (A) or sense (S) orientations (spotting assay). Cells were plated on non-selective YS medium (N/S) and YS medium supplemented with 5-FOA (5-FOA). Controls for plating were a strain encoding a single functional ura4+ gene placed in the innermost repeats of centromere 1 (imr::ura4+), and a strain with the construct in (G) containing no fragment (No Frag). Ratios of clr4Δ/clr4Δ seb1-1 for NET-seq clusters in each fragment in antisense and sense transcription units (bar graph). Value of 1 represents no change in NET-seq signal (dotted line); a value >1
represents NET-seq clusters in $clr4\Delta$ that are reduced in $clr4\Delta\ seb1-1$. Silencing-competent fragments are orange; silencing-deficient fragments are in dark grey; n.p. denotes no peaks.
Figure 3. TFIIS\textsuperscript{DN} induces ectopic heterochromatin formation. A) Representation of the domains present in TFIIS. Two acidic residues in domain III (denoted by *) were mutated to alanine residues to produce TFIIS\textsuperscript{DN}. B) Pipeline for isolating TFIIS\textsuperscript{DN}
expressing cells and analysis of genome wide H3K9me2. *S. pombe* *epe1Δ* strains were transformed with *tfs1DN* controlled by nmt1+ thiamine repressible promoter (*pDUAL-pnmt1+nmt1DN*) or a vector control (*pDUAL-pnmt1+vc*) and selected on EMM -leu +thiamine plates. 13 isolates for each *epe1Δ* *tfs1DN* and 15 for *epe1Δ* vector control strains were collected. ChIP-seq for H3K9me2 was conducted and analyzed for each isolate as well as a parental strain for each set of isolates. C) Genome wide representation of H3K9me2 ChIP-seq enrichment for *epe1Δ* *tfs1DN* Isolate 1. PIER = Pause-Induced Ectopic heterochromatic Region. D) Genome browser images of PIERs 1 through 5 that were observed from *epe1Δ* *tfs1DN* isolates. Each plot contains the H3K9me2 enrichment of the parental *epe1Δ* strain (blue) and the *epe1Δ* *tfs1DN* strain (orange). Genome features are displayed above each browser image. Essential coding gene = black arrow; coding gene = grey arrow; ncRNA = white arrow; tRNA = black bar.
Figure 4. TFIIIS\textsuperscript{DN}-induce ectopic heterochromatin is RNAi-independent. Genome browser images of PIERs 6 through 11 that were observed from \textit{epe1\Delta ago1\Delta tfs1\textsuperscript{DN}} isolates. \textit{epe1\Delta ago1\Delta} strains were transformed with \textit{pDUAL-pnmt1\textsuperscript{+}-tfs1\textsuperscript{DN}} and H3K9me2 ChIP-seq was conducted. Genome-wide H3K9me2 enrichment was analyzed and 6 PIERs (PIER 6 though 11) were observed from \textit{epe1\Delta ago1\Delta tfs1\textsuperscript{DN}}. Each plot contains the H3K9me2 enrichment of the parental \textit{epe1\Delta ago1\Delta} strain (blue) and the \textit{epe1\Delta ago1\Delta tfs1\textsuperscript{DN}} strain (orange). Genome features are displayed above each browser image. Essential coding gene = black arrow; coding gene = grey arrow; ncRNA = white arrow; tRNA = black bar.
**Figure 5.** Seb1 promotes PIER formation and the CPA machinery limits heterochromatin assembly. A) Graph of the single PIER from seb1-1 epe1Δ tfs1DN isolates. The plot depicts the H3K9me2 enrichment of the parental seb1-1 epe1Δ strain (blue) and the seb1-1 epe1Δ tfs1DN strain (orange). Genome features are displayed above the image. Essential coding gene = black arrow; coding gene = grey arrow. B-D) Image of H3K9me enrichment at the telomeres and centromere of chromosome I (B), chromosome II (C), and the centromere III (D) in wild type (blue) and pfs2-11 (orange) mutant backgrounds after temperature shift. E) Model depicting activation of heterochromatin assembly by Seb1-dependent RNAPII pausing and its repression by the CPA machinery.