Supplementary Information for

Repetitive DNA promotes RNAi-mediated heterochromatin formation via an anti-silencing factor in fisson yeast.

Takahiro Asanuma, Soichi Inagaki, Tetsuji Kakutani, Hiroyuki Aburatani, Yota Murakami

Correspondence to: yota@sci.hokudai.ac.jp

This PDF file includes:

Supplementary Discussions 1 to 3
Supplementary Fig. 1

Other Supplementary Information for this manuscript include the following:

Supplementary Table 1 [Schizosaccharomyces pombe strains used in this study.]
Supplementary Table 2 [Primers used in this study.]
Supplementary Data 1 [Sequence data of RACE products]

Supplementary Discussion 1

5'/3'RACE analysis of *cenH* and *Kint2::ura4*$^{+}$

In this study, *cenH* at the mating type locus was chosen as a model for *dg*/*dh* elements because, although it is highly homologous to pericentromeric *dg*/*dh* elements, *cenH* has a specific insertion that enables us to distinguish it from other *dg*/*dh* elements. Furthermore, in addition to the RNAi pathway, the DNA-binding proteins Atf1/Pcr1 also recruit Clr4 and maintain heterochromatin at the mating type locus[38]. Epe1 is localized to heterochromatin through its interaction with H3K9me binding protein Swi6/HP1. Therefore, this redundant maintenance of heterochromatin at the mating type locus enables the detection of transcriptions induced by Epe1 in RNAi-defective mutants without being affected by disruption of heterochromatin. In RNAi-defective mutants, ncRNA expression becomes detectable, because it evades degradation by RNAi (Extended Data Fig. 3b).

In the absence of heterochromatin (Extended Data Fig. 3d, *clr4Δ*), dominant major TSSs were detected on both forward and reverse strands of *cenH*, and minor widespread TSSs sandwiched between them were also detected on both strands. In wild-type cells, it was impossible to clone 5'-RACE products of the *cenH* ncRNA due to its extremely low expression levels. Therefore, as stated above, TSSs activated in the presence of heterochromatin were identified using RNAi-defective *dcr1Δ* cells. Notably, in the presence of heterochromatin, transcriptions from widespread TSSs were promoted (Extended Data Fig. 3d, *dcr1Δ*), causing smearing of 5'RACE products when analyzed by agarose gel electrophoresis (Extended Data Fig. 3b, *dcr1Δ*). This activation of widespread TSSs is dependent on Epe1, because additional deletion of *epe1*$^{+}$ dramatically abrogated their activation, resulting in a convergent band (Extended Data Fig. 3b*, dcr1Δepe1Δ*). A single deletion of *epe1*$^{+}$ also resulted in dramatic repression of the widespread TSSs and left only the major TSSs (Extended Data Fig. 3d, *epe1Δ*). These remaining major TSSs may be counterparts of a promoter previously identified at the pericentromere that is resistant to silencing by heterochromatin[10,36], and explain the ncRNA transcription observed in the absence of Epe1 (Extended Data Fig. 3b). On the other hand, consistent with the result of CAGE-seq, Epe1 OP caused hyper-activation of widespread TSSs (Extended Data Fig. 3d, Epe1 OP), resulting in obvious smearing when products were analyzed by agarose gel electrophoresis (Extended Data Fig. 3b, Epe1 OP).

We also performed 3'RACE analysis to identify transcription termination sites (TTSs) of *cenH* ncRNAs and *Kint2::ura4*$^{+}$ (Extended Data Fig. 3d,e). We found that TTSs of *Kint2::ura4*$^{+}$, which was expressed by Epe1 OP in the presence of heterochromatin, corresponded to those observed in *clr4Δ* (Extended Data Fig. 3e). This result indicates that Epe1 induces transcription of a reporter gene embedded within heterochromatin using inherent TTSs, as is the case with TSSs (See main text). On the other hand, we identified multiple TTSs in *cenH* ncRNA for both forward and reverse strands, as previously reported at pericentromeric *dg*/*dh* elements by PolyA-seq[39]. Compared with its effect on TSSs, the presence or absence of Epe1 did not significantly affect the distribution of TTSs (Extended Data Fig. 3d), suggesting that Epe1 primarily affects the activity of TSSs in *dg*/*dh* elements.

Supplementary Discussion 2

A truncation analysis of *cenH* fragment using plasmid-based minichromosome

5'RACE analysis revealed that *dg*/*dh* element at the mating type locus (*cenH*) includes Epe1-independent TSSs that are resistant to silencing by heterochromatin as well as Epe1-dependent widespread TSSs (Extended Data Fig. 3). To determine which TSS elements is required for the RNAi-mediated heterochromatin formation, series of truncated *cenH* fragments were cloned into plasmid-based minichromosome[36], and their abilities to establish heterochromatin were verified (Extended Data Fig. 9). As a result, a truncated *cenH* fragment containing only widespread TSSs can establish heterochromatin, indicating that Epe1-independent TSSs, namely, silencing-resistant TSSs are dispensable for the RNAi-mediated heterochromatin formation. On the other hand, *ade6*$^+$ is endogenous euchromatic gene harboring no intron, and obvious antisense transcripts were not detected at *ade6*$^+$ repeat by RT-PCR and Northern blot (Extended Data Fig. 4d,e). Therefore, the establishment of repeat-induced RNAi at *ade6*$^+$x8 allele suggests that, as well as these silencing-resistant TSSs, previously reported other properties of *dg*/*dh* ncRNAs, such as bidirectional transcription[40], secondary structures[41], and splicing[42], are dispensable for the RNAi-mediated heterochromatin formation. These properties may promote the assembly of the RNAi machinery at the *dg*/*dh* ncRNAs more efficiently or be required for primary siRNA production, which was replaced by hairpin RNAs in this study.

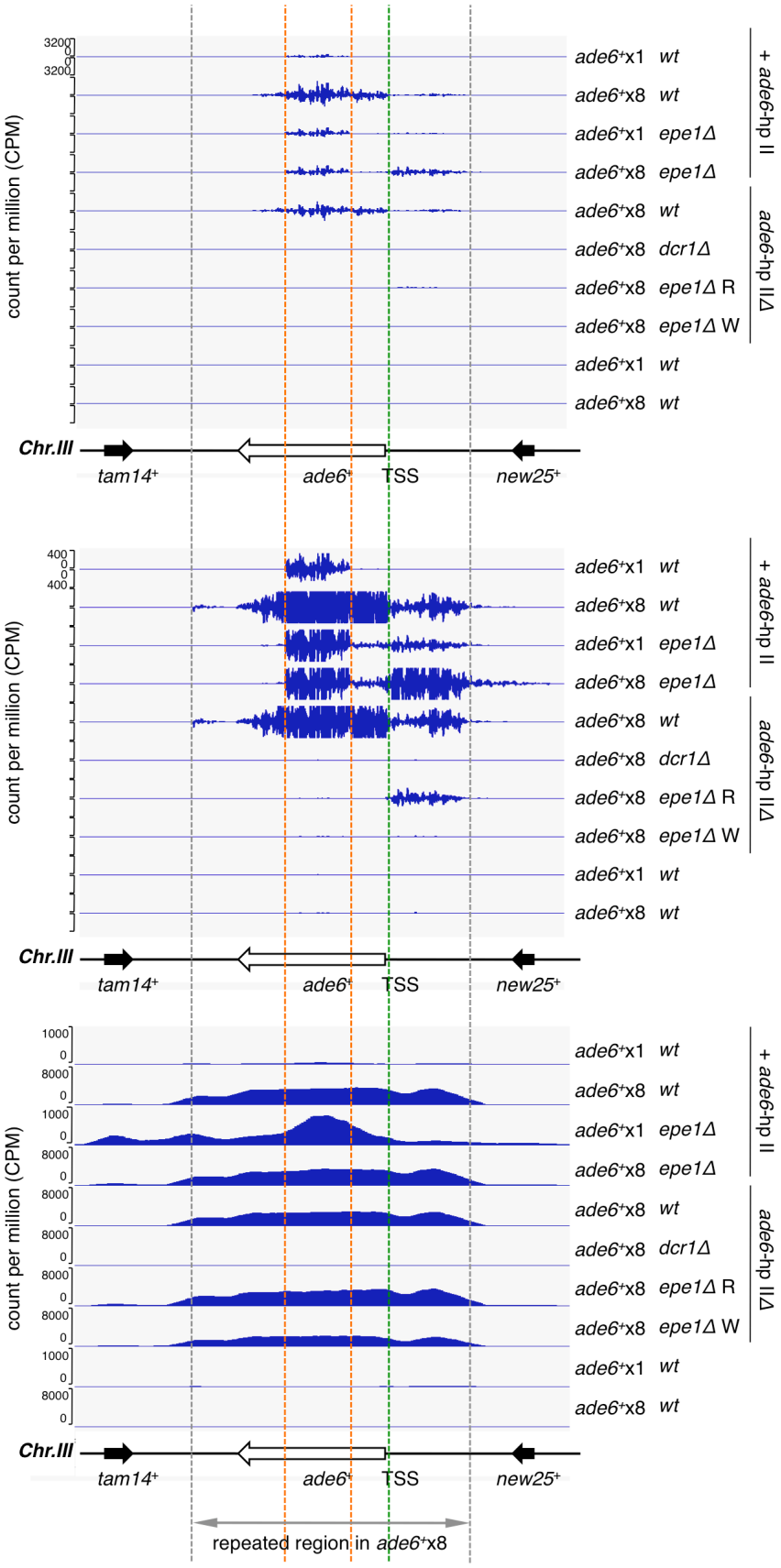
Supplementary Discussion 3

Unexpected small RNA production in the vicinity of robust ectopic heterochromatin

Unexpectedly, small RNA-seq analysis revealed that low but significant levels of small RNA are produced upstream of the *ade6*$^+$ TSS in some strains (Supplementary Fig. 1, Upper and Middle). These small RNAs are detected in the *ade6*$^+$x1 strain (*ade6*$^+$x1 *epe1Δ*), suggesting that their production is not as a result of *ade6*$^+$ repetition. Furthermore, these small RNAs were detected even in the absence of *ade6* siRNAs (*ade6*$^+$x8 *epe1Δ* R *ade6*-hp II*Δ*), indicating that the mechanism underlying the production of small RNAs produced upstream of the *ade6*$^+$ TSS is independent of and different from the mechanism responsible for production of *ade6*$^+$ siRNAs. Nonetheless, the RNAi pathway is responsible for the production of these small RNAs, because they were abolished by loss of Dicer (*ade6*$^+$x8 *dcr1Δ ade6*-hp II*Δ*). Notably, all strains producing these small RNAs upstream of the *ade6*$^+$ TSS had high levels of H3K9me in this region (Supplementary Fig. 1, lower). Because H3K9me itself promotes the localization of the RDRC[43,44], these results suggest that transcription coming from the outside region to robust ectopic heterochromatin results in recognition by the RDRC, followed by the production of small RNAs. Indeed, while a significant overlap between these small RNAs and H3K9me was observed upstream of the *ade6*$^+$ TSS, an adjacent *new25*$^+$, a possible source of such transcription, was not covered by H3K9me in these strains. Notably, the presence and absence of these siRNAs correlated with red (*ade6*-repressed) and white (*ade6*-expressing) phenotypes of *ade6*$^+$x8 *epe1Δ* strains (*ade6*$^+$x8 *epe1Δ ade6*-hp II*Δ* cells R or W), although their H3K9me levels on *ade6*$^+$ were comparable. The *ade6*$^+$x8 strains formed white (*ade6*-expressing) colonies even when only one copy of *ade6*$^+$ was derepressed by the loss of H3K9me. On the other hand, the H3K9me domain of *ade6*$^+$x8 *epe1Δ* W cells was comparable to that of R cells in ChIP-seq analysis, because H3K9me levels at each *ade6*$^+$ copy cannot be determined in *ade6*$^+$x8 allele. Therefore, this correlation between the white (*ade6*-expressing) phenotype and the disappearance of small

RNAs may support the hypothesis that the status of H3K9me itself affects the production of small RNAs upstream of the *ade6*$^{+}$ TSS. We assume that *ade6*$^{+}$x8 *epe1Δ* W cells lost H3K9me at the outermost copy of the *ade6*$^{+}$ repeats that will accept transcription from an outside region, resulting in failure to produce these small RNAs.

Supplementary Fig.1



Supplementary Fig. 1 | Small RNA production in the vicinity of ectopic heterochromatin on *ade6*$^{+}$. Small RNA-seq reads mapped in the vicinity of the *ade6*$^{+}$ gene are shown (Upper). An enlarged view is shown in the middle panel. Notations of count per million are omitted for all tracks except the top one. For comparison, the results of ChIP-seq analysis for H3K9me with the same strains are shown in the lower panel. Reads from both of *ade6*$^{+}$x1 or *ade6*$^{+}$x8 cells were mapped on the *ade6*$^{+}$x1 construct to facilitate visualization. Since *ade6*$^{+}$x8 strains have eight copies of the *ade6*$^{+}$ gene, ChIP-seq reads of these strains at *ade6*$^{+}$ are scaled by a factor of eight. Orange dashed lines indicate the region targeted by *ade6*-hp II (*trans*-acting RNAi). The green dashed line indicates the TSS of *ade6*$^{+}$, and gray dashed lines indicate either end of the *ade6*$^{+}$ repeat fragment. The locations and direction of *ade6*$^{+}$ and the neighboring genes, *new25*$^{+}$ and *tam14*$^{+}$, are indicated by arrows below each panel. Note that *ade6*$^{+}$x8 *epe1Δ* R or W indicates whether clones analyzed exhibited Red (*ade6*-repressed) or White (*ade6*-expressing) phenotypes, respectively (see also Fig. 4e). See also Supplementary Discussion 3.