



ARTICLE



<https://doi.org/10.1038/s41467-020-16280-y>

OPEN

# Conserved protein Pir2<sup>ARS2</sup> mediates gene repression through cryptic introns in lncRNAs

Gobi Thillainadesan<sup>1</sup>, Hua Xiao<sup>1</sup>, Sahana Holla<sup>1</sup>, Jothy Dhakshnamoorthy<sup>1</sup>, Lisa M. Miller Jenkins<sup>2</sup>, David Wheeler<sup>1</sup> & Shiv I. S. Grewal<sup>1</sup>  

Long non-coding RNAs (lncRNAs) are components of epigenetic control mechanisms that ensure appropriate and timely gene expression. The functions of lncRNAs are often mediated through associated gene regulatory activities, but how lncRNAs are distinguished from other RNAs and recruit effector complexes is unclear. Here, we utilize the fission yeast *Schizosaccharomyces pombe* to investigate how lncRNAs engage silencing activities to regulate gene expression in cis. We find that invasion of lncRNA transcription into the downstream gene body incorporates a cryptic intron required for repression of that gene. Our analyses show that lncRNAs containing cryptic introns are targeted by the conserved Pir2<sup>ARS2</sup> protein in association with splicing factors, which recruit RNA processing and chromatin-modifying activities involved in gene silencing. Pir2 and splicing machinery are broadly required for gene repression. Our finding that human ARS2 also interacts with splicing factors suggests a conserved mechanism mediates gene repression through cryptic introns within lncRNAs.

<sup>1</sup>Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. <sup>2</sup>Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. ✉email: [grewals@mail.nih.gov](mailto:grewals@mail.nih.gov)

lncRNAs dynamically regulate gene expression during development and in response to environmental conditions<sup>1–4</sup>. Defects in gene regulation by lncRNAs are frequently linked to diseases including cancer<sup>1,2</sup>. In many cases, lncRNAs govern gene expression by directing chromatin-modifying enzymes and other factors<sup>2</sup>. This function of lncRNAs is mediated via their associated proteins, but the mechanisms by which the lncRNAs selectively engage gene regulatory activities have remained largely unknown.

*S. pombe* is a powerful genetic model system for studying lncRNAs and their roles in the regulation of gene expression. In addition to numerous annotated lncRNAs, several RNA processing factors that are missing in budding yeast are conserved from *S. pombe* to higher eukaryotes. Many lncRNAs control gene expression in response to environmental and developmental signals<sup>5–10</sup>, including *cis*-acting lncRNAs that regulate the expression of nearby genes. Examples of regulatory lncRNAs include the *pvt* lncRNA that represses the acid phosphatase *pho1* gene in the presence of phosphate, and the *nam1* lncRNA that silences the mitogen-activated protein kinase *byr2* gene essential for sexual differentiation<sup>6,7,11</sup>. Transcription termination and degradation of the lncRNAs prevents them from invading and repressing downstream genes<sup>7,11–14</sup>. However, under specific growth conditions, readthrough transcription of lncRNAs leads to repression of downstream genes<sup>15</sup>. Underscoring a direct role, cells defective in lncRNA production show de-repression of target genes<sup>6–8,11,12</sup>. Although these and other lncRNAs play a critical role in mediating gene repression, the exact mechanism is not understood.

RNA processing factors that process diverse RNA species have been implicated in both posttranscriptional and transcriptional silencing<sup>16</sup>. RNAi machinery processes transcripts into small RNAs (siRNAs), but is also critical for targeting chromatin-modifying activities, such as factors involved in heterochromatin assembly<sup>17,18</sup>. The components of the RNAi pathway include the RNA-induced transcriptional silencing complex (RITS: Ago1, Chp1, and Tas3), the RNA-directed RNA polymerase complex (RDRC: Cid12, Hrr1 and Rdp1), and Dicer (Dcr1)<sup>17–21</sup>. In addition to playing a prominent role in processing centromeric repeat transcripts, RNAi targets various other loci, including retrotransposons, sexual differentiation genes, and genes encoding transmembrane proteins<sup>22</sup>.

Additionally, *S. pombe* contains conserved machinery that promotes degradation of transcripts by the 3'→5' exonuclease Rrp6<sup>6,23,24</sup>. MTREC (Mtl1-Red1 core) is composed of the Mtr4-like RNA helicase Mtl1 and the zinc finger protein Red1 and serves as the molecular hub of an RNA processing network<sup>6,25</sup> related to NEXT and PAXT in mammals<sup>26</sup>. MTREC and its associated factors preferentially target transcripts containing hexameric DSR (determinant of selective removal) elements, which are bound by a YTH family RNA-binding protein Mmi1<sup>23,27</sup>. Mmi1 physically interacts with the Erh1 protein to form a complex referred to as EMC (Erh1-Mmi1 Complex). EMC recruits MTREC to meiotic genes to prevent their untimely expression during vegetative growth, in addition to targeting *cis*-acting lncRNAs including *pvt* and *nam1*<sup>6,7,12,28,29</sup>. Mmi1 also mediates recruitment of the cleavage and polyadenylation factor (CPF) complex, which acts together with Rrp6 to trigger transcription termination and degradation of lncRNAs, thus preventing them from invading and repressing downstream genes<sup>7,11–14</sup>. Despite these studies, a major unanswered question is how lncRNAs mediate gene repression.

In this study, we demonstrate that besides MTREC, *cis*-acting lncRNAs show enrichment of the highly conserved Pir2 protein (ARS2 in mammals). Remarkably, lncRNAs contain cryptic introns that provide a scaffold for splicing factors and Pir2, which

are required for lncRNA-mediated gene repression. Our analyses show that the Pir2-splicing-machinery recruits silencing effector complexes to aid in the repression of target gene loci. We also find that ARS2 associates with splicing factors in human cells, suggesting that human ARS2 functions similarly to connect regulatory RNAs to gene silencing activities.

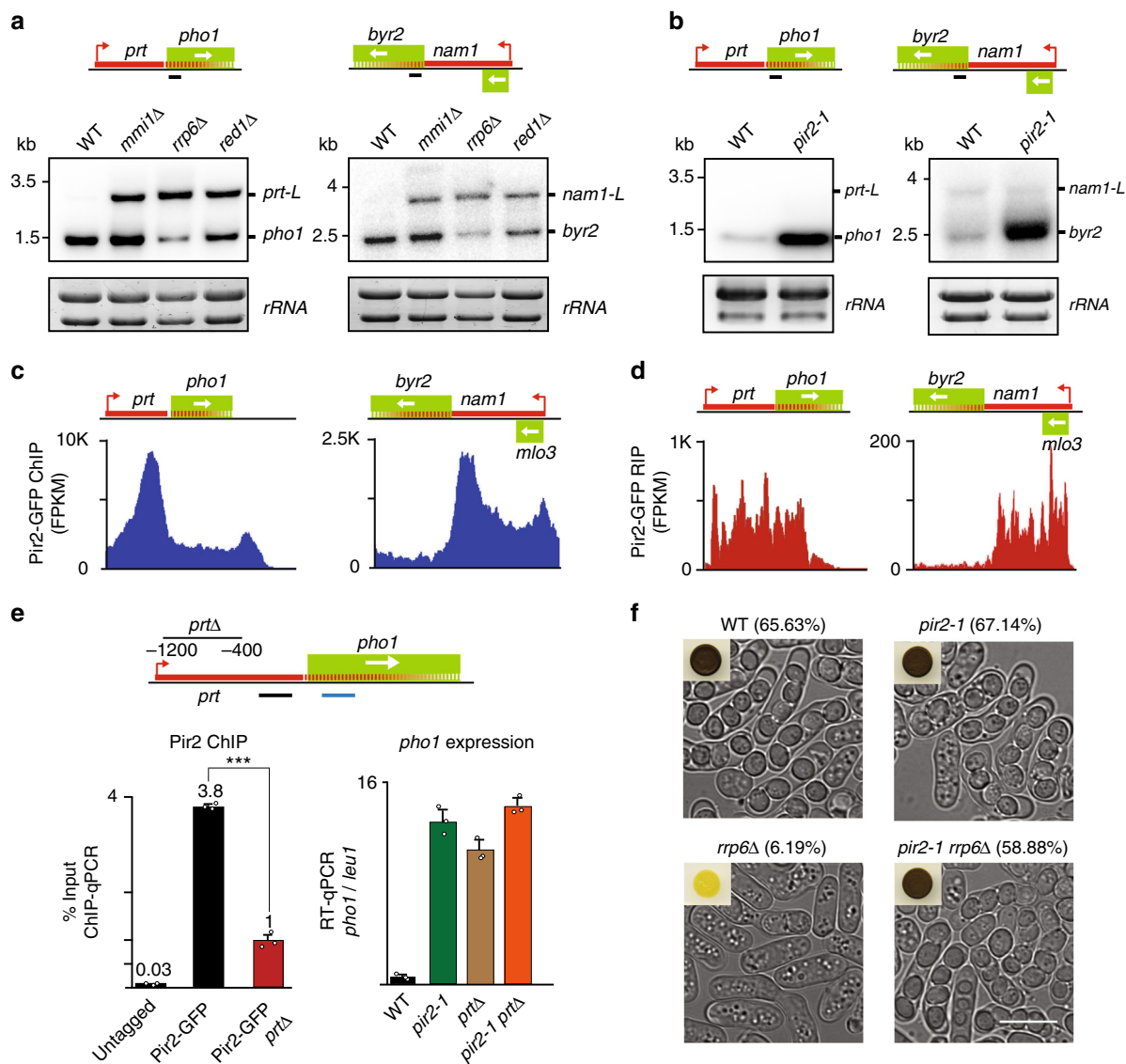
## Results

**Pir2 is required for lncRNA-mediated repression.** We investigated if MTREC and its associated factors, including the Pir2<sup>ARS2</sup> protein<sup>6,25</sup>, are required for repression of *pho1* and *byr2* by lncRNA. Pir2<sup>ARS2</sup> is an essential protein implicated in various aspects of RNA metabolism<sup>29,30</sup>. Loss of the MTREC subunit Red1 resulted in the accumulation of longer readthrough transcripts (referred to as *pvt-L* and *nam1-L*) (Fig. 1a), as was also observed in *mmi1Δ* and *rrp6Δ* cells (Fig. 1a and Supplementary Fig. 1a)<sup>6,7,11</sup>. By contrast, a mutation in *pir2* (*pir2-1*)<sup>29</sup> did not affect the levels of *pvt* and *nam1* lncRNAs (Fig. 1b). Surprisingly, *pir2-1* showed a drastic upregulation of *pho1* and *byr2* genes as compared to wild-type (WT) (Fig. 1b), similar to the effect observed upon deletion of the lncRNA (Supplementary Fig. 1b)<sup>6,7,11</sup>. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) confirmed Pir2 enrichment at lncRNAs, including *pvt* and *nam1* (Fig. 1c and Supplementary Fig. 1c). At *nam1*, Pir2 enrichment encompassed the upstream *mlo3* locus. Moreover, RNA immunoprecipitation sequencing analysis (RIP-seq) showed that Pir2 binds to the lncRNAs (Fig. 1d and Supplementary Fig. 1d). Consistently, deletion of *pvt* abolished Pir2 localization at the target locus (Fig. 1e). Together, these results suggest that lncRNAs recruit Pir2 to repress their downstream genes. Supporting the function of Pir2 and lncRNA in the same pathway, we found no additive effect on *pho1* expression in the *pir2-1 pvtΔ* double mutant when compared to the effect in the single mutants (Fig. 1e).

The requirement for Pir2 in mediating the repressive effects of lncRNAs is a highly significant finding. We asked if Pir2 is also required for the repression of *byr2* that is observed upon the accumulation of *nam1* lncRNA in cells lacking Rrp6. Since *byr2* is required for meiotic induction, cells lacking Rrp6 are defective in sporulation (Fig. 1f)<sup>11</sup>. Remarkably, entry into meiosis and sporulation efficiency were restored in *pir2-1 rrp6Δ* cells (Fig. 1f). Similar results were obtained from a qualitative assay in which iodine vapor stains the starch-like compound produced by cells undergoing meiosis a dark brown color. Whereas *rrp6Δ* cells that are defective in meiotic induction due to *byr2* repression stained yellow, the *rrp6Δ pir2-1* double mutant colonies stained dark brown (Fig. 1f). These results support a role for Pir2 in mediating repression of *byr2* by *nam1* lncRNA.

**Pir2-CBC and splicing factors mediate gene repression.** We next tested whether Pir2-associated factors are also required for lncRNA-mediated gene repression. Consistent with co-purification of Pir2<sup>ARS2</sup> with the cap-binding complex (CBC) in *S. pombe* and mammals<sup>25,30–32</sup>, Pir2 co-immunoprecipitated (co-IP) with CBC components (Supplementary Fig. 1e). Moreover, CBC co-fractionated with Pir2 in glycerol gradient analyses (Supplementary Fig. 2). To test whether CBC is required for Pir2-mediated gene repression we constructed a partial loss-of-function mutant allele of the *cbc1* gene (*cbc1-1*), which encodes an essential subunit of CBC. The lncRNA-mediated repression of *pho1* was impaired in *cbc1-1* cells (Supplementary Fig. 1e), suggesting that Pir2 likely acts together with CBC to promote gene repression.

In addition to CBC, Pir2-purified fractions also contain a subset of splicing factors including Cwf10 (EFTUD2 in human), which is a subunit of the U5 small nuclear ribonucleic particle<sup>33</sup>,



**Fig. 1 Pir2 is required for lncRNA-mediated repression of neighboring genes.** **a**, **b** Northern blot analysis of transcripts produced from the *pho1* and *byr2* loci. The black line indicates the position of the radioactive probe. Ribosomal RNA was used as a loading control. Cells were grown in YEA medium. Note that longer exposures were used to detect *prt-L* and *nam1-L* transcripts in **(a)**. As a result, *pho1* and *byr2* bands in WT lanes are darker in **(a)** as compared to **(b)**. **c** ChIP-seq analysis of Pir2-GFP enrichment at *pho1* and *byr2* loci. Source data are provided as a Source data file. **d** RIP-seq analysis of Pir2-GFP at *pho1* and *byr2* loci. **e** ChIP-qPCR analysis of Pir2-GFP (left panel). The region deleted from the *prt* lncRNA (*prtΔ*) is indicated. The amplified region is indicated by a black line. Data are presented as mean values ± SD for *n* = 3 biologically independent samples. Student's *t*-test (two-tailed) was used to calculate *p*-value. Between Pir2-GFP and Pir2-GFP *prtΔ*, *p* = 0.0009 (\*\**p* < 0.001). RT-qPCR analysis of the *pho1* gene (right panel). Transcript levels were normalized to the *leu1* control. The amplified region is indicated by the blue line. Data are presented as mean values ± SD for *n* = 3 biologically independent samples. Mean data distribution is represented by white circles. **f** Representative DIC images of homothallic (*h*<sup>90</sup>) WT and mutant strains. The sporulation frequency is indicated as a percentage (*n* > 1000). Insets show colonies stained with iodine vapor. The white scale bar represents 10 μm.

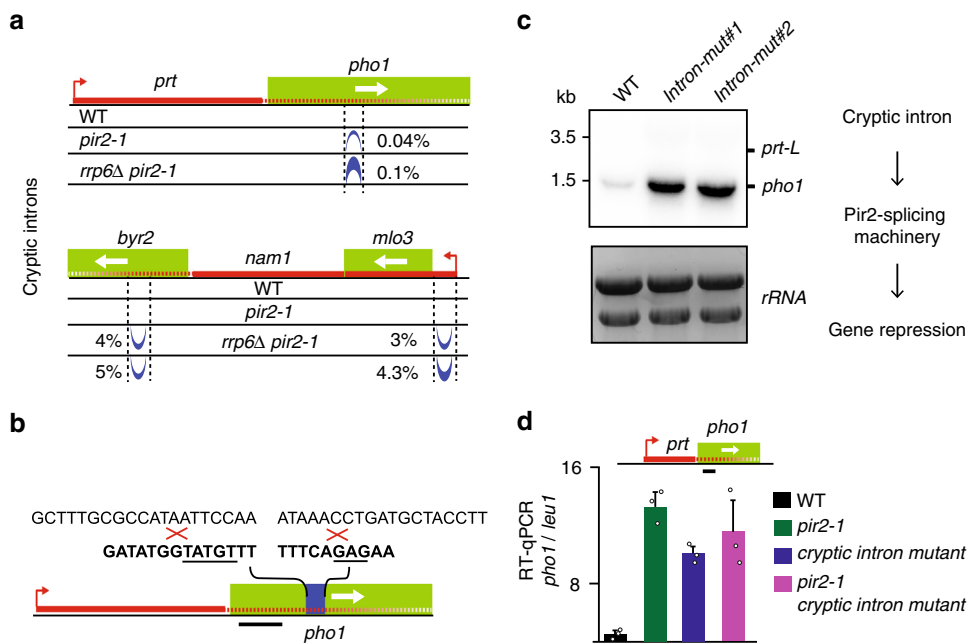
and Cwf21 (SRRM2 in humans)<sup>25</sup>. Biochemical analyses showed that Pir2 indeed forms a complex with splicing factors. We confirmed their association by co-IP (Fig. 2a) and also found that Pir2 co-eluted with a subfraction of Cwf10 in a glycerol gradient (Fig. 2b). Cwf10 eluted in two major fractions, indicating the presence of a smaller complex containing Pir2 and a second larger complex likely representing the active spliceosome. To confirm this, we determined the elution profiles of the splicing protein Cdc5, a core component of the active spliceosome (part of the Nineteen complex; NTC), and the associated factor Spp42 (Prp8)<sup>34–36</sup>. Indeed, Cdc5 and Spp42 eluted as a larger complex corresponding to the active spliceosome and were not found in the smaller fraction with Pir2 (Supplementary Fig. 2).

Interestingly, our glycerol gradient analysis showed exclusive co-elution of Cwf21 and Pir2. Together, these results suggest that Pir2 forms a complex with splicing factors that are not part of the active spliceosome.

To determine whether Pir2 acts together with splicing factors to promote gene repression by lncRNAs, we performed northern blot analysis. A significant increase in the level of both *pho1* and *byr2* mRNAs in *cwf10-1* as compared to WT confirmed that the splicing machinery indeed affects the expression of genes repressed by lncRNAs (Fig. 2c). We then performed epistasis analysis to test if Pir2 and the splicing machinery are components of the same silencing pathway. We found no cumulative increase in the expression of genes repressed by lncRNA in the *pir2-1*







**Fig. 3 The cryptic intron within lncRNA promotes downstream gene repression.** **a** RNA-seq TopHat splice junctions showing cryptic introns. Numbers indicate the percentage of reads spliced. **b** Schematic showing splice site mutations in the cryptic intron of *prt-pho1*. The deleted sequence is in bold and the inserted sequence is shown above. 5' and 3' splice sites are underlined. **c** Northern blot analysis of transcripts produced from the *pho1* gene in WT and strains carrying the cryptic intron splice site mutations. The probe is shown in **(b)** as a black line. Source data are provided as a Source data file. Cells were grown in YEA medium. **d** RT-qPCR analysis of *pho1* expression, normalized to *leu1*. The amplified regions are indicated by the black line. Data are presented as mean values  $\pm$  SD for  $n = 3$  biologically independent samples. Mean data distribution is represented by white circles.

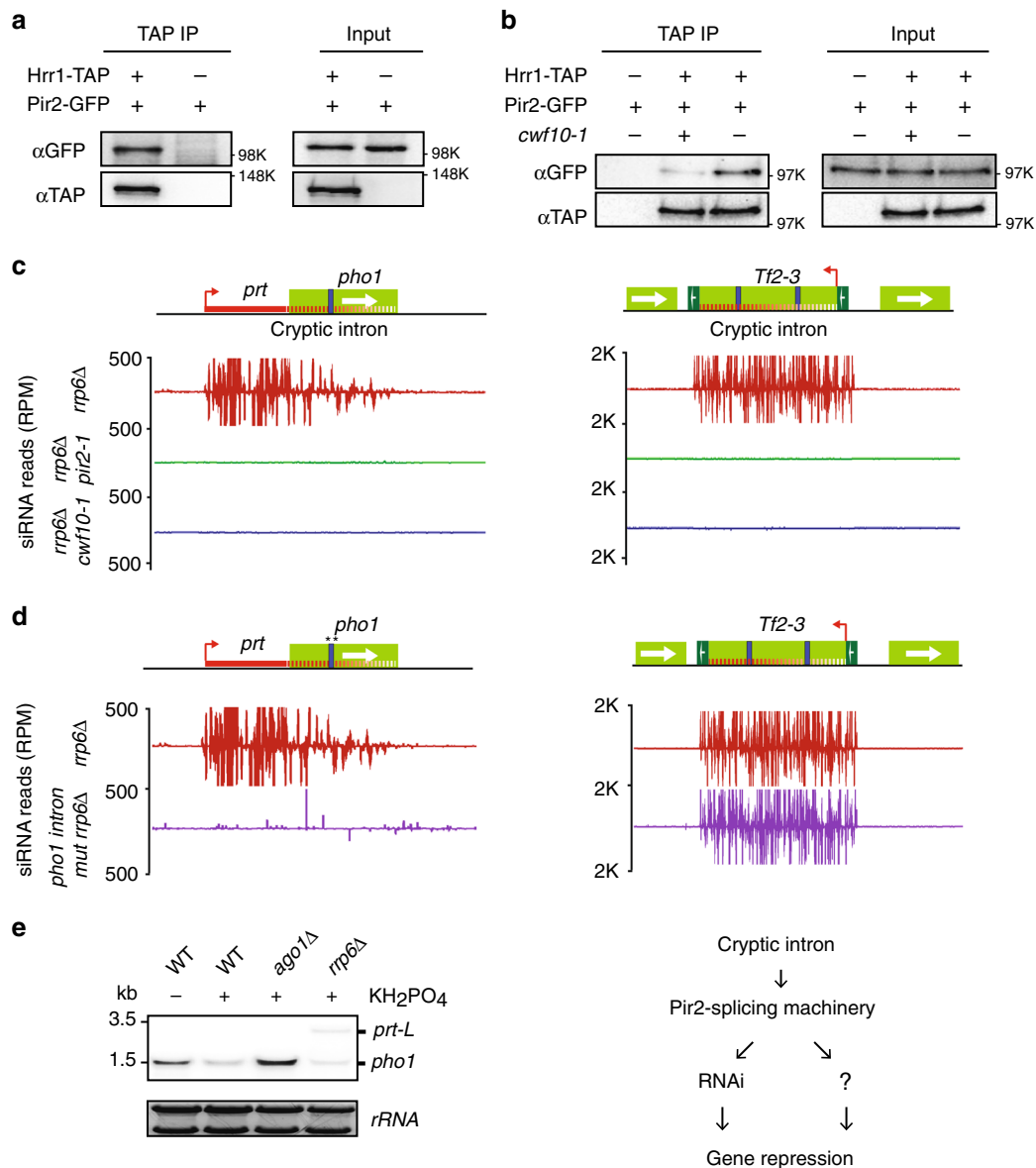
cryptic intron to promote repression of *pho1* (Fig. 3d). Together with the requirement for lncRNA to invade into the gene body, these results implicate the inclusion of the cryptic intron in the regulatory RNAs as an essential element for repression via a mechanism involving Pir2 and splicing machinery.

**Pir2 and splicing machinery collaborate genome-wide.** Cooperation between Pir2 and Cwf10 might represent a common strategy employed at other loci. Comparison of the expression profiles of *pir2-1* and *cwf10-1* revealed 435 targets repressed by both Pir2 and Cwf10, accounting for more than 50% of Pir2 target transcripts (Supplementary Fig. 3a). The targets comprised 205 mRNAs and, interestingly, 230 ncRNAs (Supplementary Data 1). Amongst the loci upregulated in *pir2-1*, we detected cryptic introns in 204 transcripts (Supplementary Fig. 3b). This is likely an underestimation due to the difficulty of detecting inefficiently spliced cryptic introns. Cryptic introns were found in mRNAs and many ncRNAs that collectively show transcript upregulation as determined by comparing RNA-seq data from *pir2-1* and *cwf10-1* cells to WT (Supplementary Fig. 3c). Notably, we detected cryptic introns in transcripts arising from retrotransposon *Tf2* elements in *pir2-1* cells (Supplementary Fig. 3d). Analysis of *pir2-1* and *cwf10-1* mutants revealed that expression of *Tf2* elements increased in both mutant strains (Supplementary Fig. 3d, e), and RIP-seq analysis showed that Pir2 binds to *Tf2* transcripts (Supplementary Fig. 3f). These results establish cryptic introns as a common feature among Pir2 targets and show that Pir2 and splicing machinery collaborate to repress genes and retrotransposons.

**Pir2-splicing machinery recruit RNAi proteins for repression.** How might Pir2 trigger repression by lncRNA containing cryptic introns? Although lncRNA production is important for loading

silencing factors<sup>7,22,37,38</sup>, the exact mechanism has remained unclear. Since Pir2 homologs are involved in RNAi<sup>29,31,39,40</sup> and splicing machinery is implicated in siRNA production<sup>41–43</sup>, we examined the association of Pir2 with RNAi machinery. Co-IP analysis showed Pir2 associates with the Hrr1 subunit of RDRC (Fig. 4a). Interestingly, this interaction was impaired in the *cwf10-1* mutant, indicating that splicing factors are required for association of Pir2 with Hrr1 (Fig. 4b). We then analyzed the role of Pir2 in siRNA production in cells lacking Rrp6, which show accumulation of lncRNAs and robust repression of their target loci. We found that siRNAs, which ranged in size from 20–24 nt and mapped to lncRNAs targeting *pho1* and *byr2*, were abolished in both *pir2-1* and *cwf10-1* mutant backgrounds (Fig. 4c and Supplementary Fig. 4a, b). The remaining reads in the mutants displayed a broad length distribution consistent with degradation products (Supplementary Fig. 4b). Pir2 was also required for siRNA production at *Tf2* elements, pericentromeric repeats, and other loci (Fig. 4c, Supplementary Fig. 5 and Supplementary Data 2). Moreover, introns detected in pericentromeric repeat transcripts in RNAi mutants<sup>6,44</sup> could also be observed in *pir2-1* (Supplementary Fig. 5). Based on these results, we conclude that Pir2 targets RNAi machinery to lncRNAs and other transcripts containing cryptic introns.

**Cryptic intron is required for siRNA production at lncRNA.** We next wondered whether cryptic introns are required for Pir2-dependent generation of siRNAs. Mutations of the *pho1* cryptic intron splice sites in *rrp6Δ* cells abolished the production of siRNAs mapping to the entire *prt* lncRNA, including the region upstream of *pho1* (Fig. 4d). This result suggests that the cryptic intron acts as part of the *prt* lncRNA to engage RNAi machinery. Importantly, siRNAs mapping to other loci were not affected (Fig. 4d and Supplementary Fig. 5), indicating that the observed

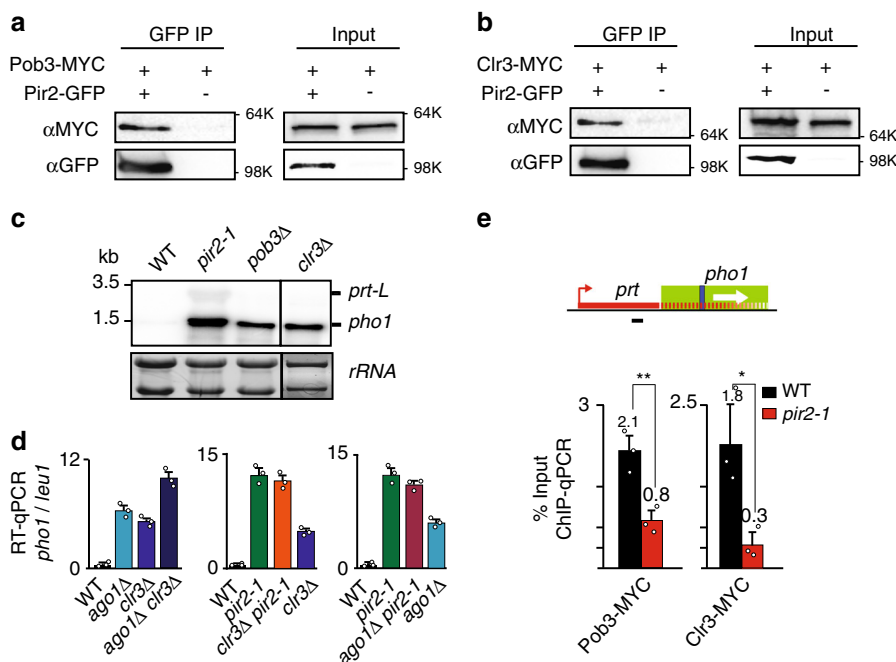


**Fig. 4 Pir2 interacts with RNAi machinery required for lncRNA-mediated gene repression.** **a, b** Co-IP analysis of Hrr1-TAP and Pir2-GFP proteins in the indicated strains. Source data are provided as a Source data file. **c** The normalized number of siRNA reads are plotted in alignment with *pho1* and *Tf2-3* loci. The signal above and below the line represent siRNAs that map to the top and bottom strands, respectively. TopHat splice junctions of introns emerging in the strains are indicated by blue arcs. **d** Normalized number of siRNAs mapping to *pho1* and *Tf2-3* loci in *rrp6Δ* (red) and *rrp6Δ* cells carrying the splice site mutation in the cryptic intron in *prt-pho1* (purple). The cryptic intron is marked by a blue box and stars indicate mutation sites. **e** Northern blot analysis of the *pho1* gene. Cells were grown in EMM medium with or without phosphate. The radioactive probe is described in Fig. 1a. Source data are provided as a Source data file.

effect was specific to *prt-pho1*. This finding is consistent with our result showing that *pho1* is upregulated in cells carrying cryptic intron splice site mutations (Fig. 3c) and led us to examine the effects of RNAi factors on lncRNA-mediated gene repression. Cells lacking Ago1 showed a considerable increase in *pho1* transcript levels as determined by northern blot analysis (Fig. 4e), but the observed effect was weaker than in *prt2-1* or *cwf10-1*, suggesting that additional factors likely cooperate with Pir2-splicing machinery.

**Pir2 and splicing machinery target chromatin modifiers.** In addition to RNAi machinery, gene silencing by lncRNAs also requires Clr3 histone deacetylase (HDAC) and the FACT histone chaperone complex<sup>8,38</sup>, which were detected in a Pir2-purified

fraction<sup>30</sup>. Therefore, we addressed whether Pir2 also recruits Clr3 and/or FACT. Clr3 and the Pob3 subunit of FACT associated with Pir2 in our biochemical analyses (Fig. 5a, b). The loss of Clr3 or Pob3 caused an increase in *pho1* transcript levels, consistent with their involvement in repression by lncRNA<sup>8,38</sup>, but the extent of upregulation was less than in *prt2-1* (Fig. 5c). This could be due to the recruitment of multiple effectors by Pir2 to promote gene repression. To test this possibility, we compared *pho1* transcript levels in single and double mutants. Combining *pho1* with other mutants resulted in severe growth defects, precluding their analysis. However, the *ago1Δ clr3Δ* double mutant showed cumulative de-repression of *pho1* (Fig. 5d). By contrast, the *prt2-1 clr3Δ* and *prt2-1 ago1Δ* double mutant strains showed no cumulative increase in transcripts as compared to the



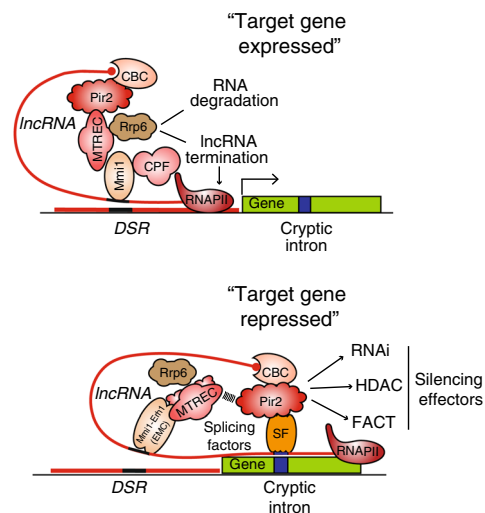
**Fig. 5 Pir2 recruits Clr3 HDAC and FACT to promote gene repression by lncRNA.** **a, b** Co-IP analysis of Pir2-GFP with Pob3-MYC and Clr3-MYC. Source data are provided as a Source data file. **c** Northern blot analysis of transcripts produced from the *pho1* locus. The radioactive probe is indicated in Fig. 1a. **d** RT-qPCR analysis of *pho1* expression, normalized to *leu1*. The amplified region is indicated by the black line. Data are presented as mean values  $\pm$  SD for  $n = 3$  biologically independent samples. Source data are provided as a Source data file. Mean data distribution is represented by white circles. **e** ChIP-qPCR of Pob3-MYC and Clr3-MYC proteins in WT (black bars) and *pir2-1* (red bars). Data are presented as mean values  $\pm$  SD for  $n = 3$  biologically independent samples. Student's *t*-test (two-tailed) was used to calculate *p*-value. Between Pob3-MYC and Pob3-MYC *pir2-1*,  $p = 0.008851$  (\*\* $p < 0.01$ ). Between Clr3-MYC and Clr3-MYC *pir2-1*,  $p = 0.008851$  (\* $p < 0.05$ ). Mean data distribution is represented by white circles.

single mutants, suggesting that Pir2 is epistatic to both RNAi and the Clr3 HDAC. Moreover, quantitative ChIP analyses showed enrichment of Clr3 and Pob3 at *pirt-pho1* in WT cells and a reduced localization in *pir2-1* cells (Fig. 5e).

Given that lncRNAs mediate repression by chromatin modifiers and RNAi, we asked whether RNA was required to mediate interactions between Pir2 and its various interacting partners. Co-IP experiments performed in the presence of Benzonase, a DNA and RNA nuclease, revealed that except for Pob3, all other interactions with Pir2 were maintained (Supplementary Fig. 6). This suggests that Pir2 interacts with components of this pathway in an RNA-independent manner. However, RNA may have a role in promoting co-transcriptional association of Pob3 with Pir2. These results suggest that lncRNA with a cryptic intron provides a docking site for the Pir2-splicing complex, which in turn recruits multiple silencing effectors to repress gene expression (Fig. 6).

**Pir2/ARS2-splicing factor connection is conserved in humans.**

The conservation of Pir2<sup>ARS2</sup> suggested that its functional interactions and role in lncRNA-mediated repression might be relevant to mammalian systems. We asked if ARS2 formed similar interactions in human cells using RIME (rapid immunoprecipitation mass spectrometry of endogenous proteins), which can detect transient chromatin associated co-transcriptional interactions. In addition to the known ARS2 interaction partners CBC, NEXT, and PAXT<sup>31,32,40</sup>, RIME analysis identified additional proteins, including factors involved in nonsense-mediated decay (NMD), pre-mRNA 3'-end processing and chromatin modifiers such as FACT and HDACs (Supplementary Fig. 7a, b and Supplementary Data 3). Notably, the most abundant associating



**Fig. 6 Pir2 and splicing machinery engage cryptic introns within lncRNA to promote repression.**

Model depicting the switch to co-transcriptional repression through the cryptic intron. For gene expression, lncRNA is terminated and degraded upstream of target loci by CPF, MTREC, and Rrp6, which are recruited by Mmi1 bound to DSR elements. Upon changes in growth conditions, the lncRNA extends into the neighboring region containing the cryptic intron, providing a scaffold for co-transcriptional recruitment of Pir2-splicing machinery, which in turn recruits silencing effectors including RNAi, HDAC, and FACT to promote gene repression. Mmi1 acts as part of the EMC to engage MTREC and other factors that together with CBC stabilize Pir2 association with lncRNAs and effector proteins.

partners of ARS2 were splicing factors (Supplementary Fig. 7b, c). Indeed, splicing factors co-eluted with ARS2 in a glycerol gradient (Supplementary Fig. 7d), analogous to our observations in *S. pombe* (Fig. 2b). Moreover, we observed association of the human counterpart of Cwf10, EFTUD2, with ARS2, suggesting that the Pir2-splicing factor connection observed in fission yeast is conserved in humans (Supplementary Fig. 7a, e). Based on these findings, we envision that ARS2 bound to splicing machinery may function similarly to bridge regulatory RNAs to gene silencing activities.

## Discussion

Our analyses reveal that a cryptic intron within the lncRNA is a crucial element for gene repression via a pathway involving Pir2<sup>ARS2</sup> and splicing factors (Fig. 6). The readthrough transcription of intergenic lncRNA incorporates a cryptic intron and creates a scaffold for the co-transcriptional recruitment of the splicing machinery–Pir2 complex, which in turn engages silencing effectors. Our findings highlight a previously unrecognized role for splicing factors in engaging Pir2<sup>ARS2</sup> to lncRNAs to dynamically control gene expression. To this end, we find that Pir2 associates with splicing factors such as Cwf21 and Cwf10 as part of a smaller complex and is not part of the larger spliceosome complex containing the NTC components. This finding suggests that the subset of splicing factors responsible for mediating gene repression is likely distinct from the active spliceosome.

Since cryptic intron-mediated silencing occurs specifically in the context of lncRNA but not the target gene transcript, additional factors bound to lncRNA are likely involved. Other factors that bind to lncRNA might help recruit and/or stabilize Pir2 with its associated silencing effectors. In this regard, we note that loss of ERH family protein, Erh1, that associates with lncRNAs and Mmi1 as part of EMC, severely affects repression of target gene loci<sup>29,45</sup>. Moreover, a mutation in Mmi1 that specifically disrupts EMC assembly without affecting its termination functions impairs lncRNA-mediated repression of neighboring genes<sup>45</sup>. EMC bound to lncRNA may directly recruit Pir2 or may act in conjunction with other factors. Consistent with the latter possibility, EMC co-purifies with MTREC<sup>29</sup>, which forms a complex with Pir2<sup>6,25,30</sup>. Therefore, MTREC, recruited by Mmi1/EMC bound to DSR elements, may act together with splicing factors engaged by the cryptic intron to promote Pir2 association with lncRNAs. In other words, the repressive effects of lncRNA require combinatorial and likely cooperative action of factors that bind to different elements embedded within regulatory RNAs (Fig. 6).

Once recruited to lncRNAs, Pir2 coordinates multiple effectors, including RNAi, to promote gene repression. In this regard, we note that Pir2 is the elusive factor that enables splicing machinery to selectively recruit RNAi to specific transcripts. RNAi processes transcripts and triggers assembly of repressive heterochromatin<sup>16–18,22</sup>. In addition, Pir2 promotes lncRNA-mediated recruitment of Clr3 HDAC and FACT, which increase nucleosome occupancy<sup>37,38,46</sup>, but may also engage additional factors to enforce gene repression. Pir2 and cryptic intron-based mechanisms repress targets throughout the genome and may be conserved in higher eukaryotes. Indeed, we note that ARS2 interacts with splicing factors in human cells, and lncRNAs such as XIST implicated in X-chromosome inactivation contain inefficiently spliced introns<sup>47,48</sup>. Considering that defects in lncRNA-mediated gene regulation contribute to human diseases including cancer and that ARS2 is commonly mutated in various types of cancer, our findings may shed light on pathways contributing to the misexpression of genes, ultimately leading to the development of specific therapeutic strategies.

## Methods

**Cell lines, strain construction, and growth conditions.** The fission yeast strains used in this study are listed in Supplementary Table 1. Strains were generated through genetic crosses or were constructed using a PCR-based method. A DNA cassette containing a selection marker with or without epitope tag was amplified using long oligonucleotides with homology to target gene loci. The PCR product was transformed and transformants were grown on appropriate selection plates. Deletions or tagged alleles were confirmed by PCR. Cells were cultured in YEA media using standard protocols unless otherwise noted in the figures. Since *pir2*, *cbc1*, and *cwf10* are essential genes, we used partial loss of function mutant alleles. *pir2-1* and *cbc1-1* were generated using an error prone PCR method<sup>6</sup>. The *pir2-1* mutant allele carries two amino acid substitutions: F165L and S316P. The *cbc1-1* mutant allele contains one amino acid substitution: L119P. *cwf10-1* is a gift from R. Allshire. For the generation of cryptic intron mutants, a strain containing a *ura4<sup>+</sup>* selectable marker inserted at the *pho1* locus was transformed with a DNA fragment containing the *pho1* ORF with mutations in the splice sites. Splice site mutations were confirmed using Sanger sequencing and production of lncRNA was confirmed by RT-PCR. *pir2-1* and *cbc1-1* mutant strains were grown at 26 °C to an OD<sub>600</sub> 0.5 prior to shifting to 33 °C for 5 h. Cells carrying the partial loss function *cwf10-1* mutant were cultured under conditions that do not affect normal splicing<sup>6</sup>. For experiments indicating phosphate or no phosphate growth conditions, Edinburgh minimal media (EMM) was prepared with or without 15.5 mM sodium phosphate and 20 mM potassium phosphate. HepG2 cells were purchased from ATCC (ATCC HB-8065). Monolayers of HepG2 cells were cultured in DMEM with 10% FBS at 37 °C and 5% CO<sub>2</sub>. According to ATCC, cells were authenticated and mycoplasma tests were done using Hoechst and direct culture method followed by microscopy techniques.

**Sporulation assay.** Cells were spotted onto EMM (Fig. 1f) or Pombe minimal glutamate (PMG) medium (Fig. 2e) plates and grown at 30 °C for 3 days prior to exposure to iodine vapor. They were subsequently mounted on a 2% agarose pad for differential interference contrast (DIC) imaging using SoftWoRx V7.0 software on a DeltaVision Elite fluorescence microscope (Applied Precision, GE Healthcare) with Olympus 100×/1.40 objective. Fiji V1.0 (ImageJ, National Institutes of Health) was used for processing the images and counting the sporulation frequencies. Sporulation efficiency was monitored in more than 1000 cells from three independent isolates for each strain.

**Co-immunoprecipitations.** For co-IP experiments in fission yeast, 1 L of *S. pombe* cells was grown overnight to OD<sub>600</sub> 0.8 and cells were harvested by vacuum filtration. Cell pellets were flash frozen in liquid nitrogen. Cells were lysed using a CryoMill (Retsch) at a setting of 30 (frequency per second) for 1 min and repeated 3 times with 30 s intervals. Lysed cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 10% glycerol, 1% NP-40, 5 mM EDTA) containing Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific). Lysates were incubated with either anti-GFP (Roche) bound to protein A magnetic beads (NEB) or IgG sepharose (GE) for 2 h at 4 °C. To remove DNA and RNA, proteins bound to beads were washed and treated with 50 U/ml Benzonase for 30 min. Antibody–protein complexes were eluted off the beads using 1× sample buffer and heated to 95 °C for 5 min. For western blot analysis, samples were run on a 10% Tris-glycine gel, transferred to PVDF (Thermo Fisher Scientific) and probed using anti-FLAG (M2 Sigma, F1804) at dilution 1/1000, anti-GFP (Roche) at dilution 1/1000, or anti-c-MYC (Covance, 9E10) antibodies at dilution 1/250.

For co-IP experiments in mammalian cells, human HepG2 cells were grown to 70–90% confluency and the cell nuclei were isolated in hypotonic buffer (20 mM HEPES pH 7.4, 10 mM KCl, 2 mM MgCl<sub>2</sub>, 1 mM EDTA, protease inhibitors) and lysed in lysis buffer (50 mM Tris pH 8, 0.3 M KCl, 0.5 mM EDTA/EGTA, 0.25% NP-40) by resuspending the nuclear pellet using a 20G needle. Lysates were incubated with anti-ARS2 antibody (Abcam, ab192999) for 2 h and washed 5 times with lysis buffer. Elution and western blotting were performed as described above. Anti-EFTUD2 (Abcam, ab72456) and anti-ARS2 (Abcam, ab192999) antibodies were used to probe the western blots.

**Glycerol gradient analysis.** To prepare yeast cell lysates, cells were lysed using a CryoMill (Retsch) as described above and the cell powder was resuspended in 10% glycerol buffer (20 mM Tris pH 7.5, 137 mM NaCl, 10% Glycerol, 0.5% NP40, 2 mM EDTA). For human HepG2 cells, nuclear extract was prepared as described above and the nuclei were lysed in 10% glycerol buffer by passing through a 20G needle 10 times. The lysate was resolved by loading 150 µg onto a linear 20–50% glycerol gradient prepared in an ultracentrifuge tube (Beckman, 347357). The gradients were spun in an Optima TLX-ultracentrifuge (Beckman) at 81,400 × g for 19 h. Fractions were collected by pipetting and were resolved on a 4–12% gradient gel (Invitrogen, NP0336BOX), followed by western blot analysis with ARS2 and EFTUD2 antibodies as described above. Western blotting to detect *S. pombe* proteins was performed with the following antibodies: GFP (Pir2-GFP detection; Roche, 11814460001) at dilution 1/1000, HA (Cdc5-HA and Spp42-HA detection; Biolegend, 901501) at dilution 1/1000 and FLAG (Cbc1 detection; M2 Sigma, F1804) at dilution 1/1000.



**Northern blot analysis.** For most experiments, cells were cultured in YEA medium that contains phosphate. In experiments comparing levels of *pho1* expression in the presence or absence of phosphate, EMM was used with or without 15.5 mM sodium phosphate and 20 mM potassium phosphate. Total RNA was isolated by incubating cells in hot phenol heated to 65 °C for 10 min followed by 3 additional extractions using phenol-chloroform. RNA was precipitated using the sodium-acetate-ethanol method. Northern blots were performed according to the published protocol<sup>6</sup>. 10 µg of RNA was resolved on a 1% formaldehyde-agarose denaturing gel and capillary transferred using NorthernMAX transfer buffer (Thermo Fisher Scientific) onto positively charged BrightStar-Plus nylon membrane (Ambion) and crosslinked using UV Stratalinker 2400 (Stratagene). The T7 *in vitro* transcription kit (Promega) was used to generate α-<sup>32</sup>P-UTP (PerkinElmer) labeled RNA probes (Supplementary Table 2) that were hybridized to the membrane overnight at 65 °C in ULTRAhyb buffer (Ambion). The membrane was exposed and scanned using a Typhoon FLA 9500 phosphor imager (GE Healthcare).

**RT-qPCR.** Total RNA was extracted as described above and treated with RQ1 DNase (Promega) followed by phenol-chloroform extraction and ethanol precipitation. Strand specific reverse transcription was performed using Revertaid Reverse Transcriptase (Thermo Fisher Scientific) with gene specific reverse primers (Supplementary Table 2) and quantified by performing qPCR with iTaq Universal SYBR Green Supermix (Bio-Rad) on the QuantStudio 3 platform (Thermo Fisher Scientific).

**ChIP and ChIP-sequencing.** ChIP experiments were performed according to the published procedure<sup>49</sup>. *S. pombe* cells at OD<sub>600</sub> 0.5 were crosslinked using 1% formaldehyde at room temperature for 20 min. Cells were spun at 2000 × g for 10 min and pellets were lysed using glass beads and lysis buffer (50 mM Hepes/KOH, 140 mM NaCl, 1 mM EDTA, 1% Triton X100, 0.1% DOC plus protease inhibitors). Cell lysates were sheared using a Bioruptor-300 (Diagenode) to an approximate size of 300–600 bp. Immunoprecipitation of the protein of interest was accomplished using 50 µl of anti-c-MYC affinity gel (Sigma, A7470) or 5 µg of anti-GFP antibody (Abcam, ab290). Protein A magnetic beads (NEB) were used to capture the GFP-protein-antibody complexes. Beads were washed twice with lysis buffer, twice with lysis buffer containing 0.5 M NaCl and once in TE buffer pH 8. Chromatin-antibody complex was eluted using TES buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) and, along with whole-cell extract (WCE) input, were de-crosslinked by heating to 65 °C overnight and purified using PCR purification columns (Qiagen). Immunoprecipitated and input DNA were assessed using iTaq-qPCR SYBR green supermix (Bio-Rad) according to the manufacturer's recommendations. Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina according to the manufacturer's protocol and analyzed using an Agilent 4200 TapeStation system (Agilent). Sequencing was performed on the NextSeq500 platform (Illumina).

**Small RNA-sequencing.** Total RNA was isolated from a total of 4 OD<sub>600</sub> units of log-phase cells using the hot-phenol method as previously described above for northern blot analysis. smRNAs between 15 and 30 nucleotides were excised from 17% Urea-PAGE gels and subsequently eluted using Corning Costar Spin-X columns (Sigma Aldrich) followed by ethanol precipitation overnight. Pellets were resuspended using DEPC treated water and libraries were constructed using the NEBNext Small RNA Library for Illumina (NEB) according to the manufacturer's protocol. Qiagen MinElute PCR-purification columns were used to purify the resulting libraries which were further resolved through a 6% PAGE gel, eluted using Corning Costar Spin-X columns and ethanol precipitated. The precipitated RNA was resuspended in 1×TE buffer and analyzed as described above. The resulting libraries were sequenced on the MiSeq platform (Illumina).

**RNA-sequencing.** Total RNA was isolated as described above. rRNA was removed using the Ribo-Zero Magnetic Gold Kit for yeast (Illumina). Libraries were made using the NEBNext Ultra Directional RNA Library Prep kit for Illumina (NEB) according to the manufacturer's instructions. Libraries were analyzed and sequenced on the MiSeq platform (Illumina) as described above.

**RNA immunoprecipitation.** *S. pombe* cells expressing either GFP-tagged Pir2 (Pir2-GFP) or untagged Pir2 were grown in 50 ml of YEA medium at 30 °C to an OD<sub>600</sub> 0.5. Cells were then crosslinked by adding formaldehyde to a final concentration of 1% for 20 min with gentle shaking. After adding glycine to a final concentration of 0.2 M to stop crosslinking, cells were then resuspended in lysis buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) supplemented with complete EDTA-free proteinase inhibitor cocktail (Roche) and RNase inhibitors (Thermo Fisher Scientific, am2694) and lysed by bead beating. The lysate was sheared by sonication using a Bioruptor-300 (Diagenode). 1.5% of the original lysate for RNA preparation was set aside as input. The rest of the lysate was pre-cleaned using 0.9 mg of pre-washed protein G Dynabeads (Invitrogen, 10004D) at 4 °C for 1 h. Anti-GFP antibody (Abcam, ab290) was added to the lysate and incubated with gentle rotation at 4 °C overnight. Antibody-protein complexes were captured using 1.2 mg of protein G Dynabeads for 2 h at 4 °C. Beads were washed once in 900 µl of lysis buffer, once in

900 µl of lysis buffer with 300 mM NaCl, once in 900 µl of LiCl buffer (50 mM HEPES pH7.5, 250 mM LiCl, 0.5% NP-40, 0.1% sodium deoxycholate), and once in 900 µl of TE pH 7. Beads were eluted twice in 75 µl of RIP elution buffer (50 mM Tris pH 8, 10 mM EDTA, 300 mM NaCl, 1% SDS) at 37 °C for 10 min. To the 50 µl input samples, 100 µl of RIP elution buffer was added to a final volume of 150 µl. Then, 20 µg of proteinase K (Thermo Fisher Scientific, am2548) was added to both IP and input samples, and the mixtures were incubated at 37 °C for 1 h and then at 65 °C for 1 h to de-crosslink. The samples were then extracted once with phenol-chloroform and once with chloroform, precipitated with ethanol, and resuspended in 80 µl of DEPC treated water. The samples were further treated with 20 units of RNase-free DNase I (Thermo Fisher Scientific, am2222) for 1 h at 37 °C, extracted with phenol-chloroform, and ethanol precipitated as above. IP and input samples were resuspended in 30 µl and 100 µl water, respectively. RNA concentrations were determined using a Qubit fluorometer (Thermo Fisher Scientific). Libraries were directly generated from the IP sample. The total input sample was first subjected to ribosomal RNA removal before library preparation. The libraries for IP and input samples were prepared and analyzed as described above for RNA-seq. The libraries were sequenced using the NextSeq 500 platform (Illumina).

**RIME.** Rapid immunoprecipitation mass spectrometry of endogenous proteins technique was performed according to the published procedure<sup>50</sup>. Monolayers of 1E8 HepG2 cells were crosslinked with 0.8% formaldehyde for 8 min at room temperature. Crosslinking was quenched with 0.1 M (final) glycine and cell nuclei were isolated using buffer LB1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40/Igepal CA-630 and 0.25% Triton X-100). Nuclei were washed with buffer LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) and lysed with buffer LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (wt/vol) sodium deoxycholate and 0.5% (vol/vol) *N*-lauroylsarcosine). A Bioruptor-300 (Diagenode) was used to sonicate the lysate to yield DNA fragments of size range 200–500 bp followed by full speed centrifugation at 4 °C for 10 min. The resulting supernatant was incubated with 10 µg of anti-ARS2 antibody (Abcam, ab192999) or 10 µg of rabbit IgG mock (Abcam, ab37415) bound to protein A magnetic beads (NEB) for 4 h at 4 °C with slow rotation. Beads were washed 10 times with RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7% (wt/vol) sodium deoxycholate, 1% NP-40 and 0.5 M LiCl) and twice with 100 mM ammonium hydrogen carbonate buffer. The protein-antibody complex was trypsinized overnight while on the beads and the peptides were extracted using Pierce C18 Spin Columns (Thermo Fisher Scientific). Extracted peptides were dried using a Vacufuge (Eppendorf).

For mass spectrometry analysis, tryptic peptides were trapped on a trapping column and separated on a 75 µm × 15 cm, 2 µm Acclaim PepMap reverse phase column (Thermo Scientific) using an UltiMate 3000 RSLCnano HPLC (Thermo Scientific). Peptides were separated at a flow rate of 300 nL/min followed by online analysis by tandem mass spectrometry using a Thermo Orbitrap Fusion mass spectrometer. Peptides were eluted into the mass spectrometer using a linear gradient from 96% mobile phase A (0.1% formic acid in water) to 55% mobile phase B (0.1% formic acid in acetonitrile) over 210 min. Parent full-scan mass spectra were collected in the Orbitrap mass analyzer set to acquire data at 120,000 FWHM resolution; ions were then isolated in the quadrupole mass filter, fragmented within the HCD cell (HCD normalized energy 32%, stepped ± 3%), and the product ions analyzed in the ion trap. Proteome Discoverer 2.2 (Thermo) was used to search the data against human proteins from the UniProt database (downloaded January 2017) using SequestHT v1.17. The search was limited to tryptic peptides, with maximally two missed cleavages allowed. Cysteine carbamidomethylation was set as a fixed modification and methionine oxidation set as a variable modification. The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 0.6 Da. The Percolator node was used to score and rank peptide matches using a 1% false discovery rate. Protein false discovery rate was set at 1% and a minimum of 1 unique peptide required for protein identification. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018373 and 10.6019/PXD018373.

**Bioinformatic analyses.** ChIP sequences were aligned to the genome using Bowtie2<sup>51</sup>. RIP and total RNA sequencing were aligned to the genome using TopHat2 and for RNA-seq, FPKM was calculated using cufflinks<sup>52</sup>. Normalized (RPKM) bedgraph files for aligned BAMs were generated using DeepTools<sup>53</sup>. Introns were assessed using TopHat2 junctions function and mutant cryptic introns were derived by subtracting annotated and WT introns. Unlike annotated introns that are spliced in majority of reads, cryptic introns are inefficiently spliced in less than 5% of the total reads spanning the region containing the intron. Cryptic intron percentage was calculated by taking the number of reads that contained a split read, indicating a spliced event, and dividing that number by the total number of reads that overlap the intron junction. For RIP-seq, the signal obtained from the untagged strain was subtracted from Pir2-GFP signal. siRNA sequencing reads were aligned using Novoalign V2 (Novocraft) and aligned reads were processed to retain 21–24 nucleotide reads and generate SGR files using a python script. The resulting file was normalized to million mapped reads. The *S. pombe* genome version used in this study is ASM294v2. For RIME (Supplementary Fig. 7), *p*-values for peptide detection were estimated from the observed distribution of log2

(#PSMs-Ars2/#PSMs-Ig-Control). This work employed the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

**Statistics and reproducibility.** Experiments for Figs. 1a, b, e, f; 2a–c, e; 4a, b, e and Supplementary Figs. 1b, e; 2; 6; 7d, e were successfully repeated independently with identical results at minimum two times. The statistical significance for qPCR results was assessed using two-tailed *t*-test.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Datasets are available on NCBI Gene Expression Omnibus (Accession no. GSE135161). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018373 and 10.6019/PXD018373. The source data underlying Figs. 1a, b, e, 2a–d, 3c, d, 4a, b, e, and 5a–e and Supplementary Figs. 1b, e, 2, 3e, 6, 7d, e are provided as a Source data file. All data is available from the corresponding author upon reasonable request.

### Code availability

All codes used in this study are publicly available.

Received: 3 September 2019; Accepted: 21 April 2020;

Published online: 15 May 2020

### References

- Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
- Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
- Rinn, J. L. lncRNAs: linking RNA to chromatin. *Cold Spring Harb. Perspect. Biol.* **6**, a018614(2014).
- Kim, T. K. & Shiekhattar, R. Diverse regulatory interactions of long noncoding RNAs. *Curr. Opin. Genet. Dev.* **36**, 73–82 (2016).
- Hirota, K. et al. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130–134 (2008).
- Lee, N. N. et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* **155**, 1061–1074 (2013).
- Shah, S., Wittmann, S., Kilchert, C. & Vasiljeva, L. lncRNA recruits RNAi and the exosome to dynamically regulate *pho1* expression in response to phosphate levels in fission yeast. *Genes Dev.* **28**, 231–244 (2014).
- Ard, R., Tong, P. & Allshire, R. C. Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nat. Commun.* **5**, 5576 (2014).
- Atkinson, S. R. et al. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* **24**, 1195–1213 (2018).
- Fauquenoy, S. et al. Repression of cell differentiation by a cis-Acting lincRNA in fission yeast. *Curr. Biol.* **28**, 383–391 e383 (2018).
- Touat-Todeschini, L. et al. Selective termination of lincRNA transcription promotes heterochromatin silencing and cell differentiation. *EMBO J.* **36**, 2626–2641 (2017).
- Chatterjee, D., Sanchez, A. M., Goldgur, Y., Shuman, S. & Schwer, B. Transcription of lincRNA prt, clustered prt RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of *pho1* gene expression under phosphate-replete conditions in fission yeast. *RNA* **22**, 1011–1025 (2016).
- Vo, T. V. et al. CPF recruitment to non-canonical transcription termination sites triggers heterochromatin assembly and gene silencing. *Cell Rep.* **28**, 267–281 e265 (2019).
- Sanchez, A. M., Shuman, S. & Schwer, B. RNA polymerase II CTD interactome with 3' processing and termination factors in fission yeast and its impact on phosphate homeostasis. *Proc. Natl Acad. Sci. U. S. A.* **115**, E10652–E10661 (2018).
- Sanchez, A. M., Garg, A., Shuman, S. & Schwer, B. Inositol pyrophosphates impact phosphate homeostasis via modulation of RNA 3' processing and transcription termination. *Nucleic Acids Res.* **47**, 8452–8469 (2019).
- Reyes-Turcu, F. E. & Grewal, S. I. Different means, same end-heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Curr. Opin. Genet. Dev.* **22**, 156–163 (2012).
- Hall, I. M. et al. Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
- Volpe, T. A. et al. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Motamedi, M. R. et al. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
- Noma, K. et al. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat. Genet.* **36**, 1174–1180 (2004).
- Verdel, A. et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Yamanaka, S. et al. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature* **493**, 557–560 (2013).
- Harigaya, Y. et al. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45–50 (2006).
- Sugiyama, T. & Sugiyoka-Sugiyama, R. Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *EMBO J.* **30**, 1027–1039 (2011).
- Zhou, Y. et al. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat. Commun.* **6**, 7050 (2015).
- Meola, N. et al. Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell* **64**, 520–533 (2016).
- Yamashita, A. et al. Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open Biol.* **2**, 120014 (2012).
- Kilchert, C. et al. Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep.* **13**, 2504–2515 (2015).
- Sugiyama, T. et al. Enhancer of rudimentary cooperates with conserved RNA-processing factors to promote meiotic mRNA decay and facultative heterochromatin assembly. *Mol. Cell* **61**, 747–759 (2016).
- Egan, E. D., Braun, C. R., Gygi, S. P. & Moazed, D. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* **20**, 867–881 (2014).
- Gruber, J. J. et al. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell* **138**, 328–339 (2009).
- Andersen, P. R. et al. The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* **20**, 1367–1376 (2013).
- Kuhn, A. N. & Kaufer, N. F. Pre-mRNA splicing in *Schizosaccharomyces pombe*: regulatory role of a kinase conserved from fission yeast to mammals. *Curr. Genet.* **42**, 241–251 (2003).
- Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Luhrmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846–850 (2008).
- Nguyen, T. H. D. et al. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
- Yan, C., Wan, R. & Shi, Y. Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* **11**, a032409 (2019).
- Ard, R. & Allshire, R. C. Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res.* **44**, 10619–10630 (2016).
- Watts, B. R. et al. Histone deacetylation promotes transcriptional silencing at facultative heterochromatin. *Nucleic Acids Res.* **46**, 5426–5440 (2018).
- Montgomery, T. A. & Carrington, J. C. Splicing and dicing with a SERRATED edge. *Proc. Natl Acad. Sci. USA* **105**, 8489–8490 (2008).
- Sabin, L. R. et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in *Drosophila*. *Cell* **138**, 340–351 (2009).
- Bayne, E. H. et al. Splicing factors facilitate RNAi-directed silencing in fission yeast. *Science* **322**, 602–606 (2008).
- Dumesic, P. A. et al. Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* **152**, 957–968 (2013).
- Newman, M. A. et al. The surveillance of pre-mRNA splicing is an early step in *C. elegans* RNAi of endogenous genes. *Genes Dev.* **32**, 670–681 (2018).
- Mutazono, M. et al. The intron in centromeric noncoding RNA facilitates RNAi-mediated formation of heterochromatin. *PLoS Genet.* **13**, e1006606 (2017).
- Xie, G. et al. A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. *Nat. Commun.* **10**, 251 (2019).
- Aygun, O., Mehta, S. & Grewal, S. I. HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin. *Nat. Struct. Mol. Biol.* **20**, 547–554 (2013).
- Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lincRNAs. *Genome Res.* **22**, 1616–1625 (2012).
- Krchnakova, Z. et al. Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res.* **47**, 911–928 (2019).
- Cam, H. P. et al. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* **37**, 809–819 (2005).

50. Mohammed, H. et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
52. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
53. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

### Acknowledgements

We thank R. Allshire for the *cvf10-1* mutant strain, members of the Grewal laboratory for valuable assistance and J. Barrowman for editing the paper. We also thank M. Zofall for suggestions and protocols, and V. Balachandran for strain constructions. This work was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute. Bioinformatics analysis in this study used the Helix and Biowulf Linux cluster at the National Institutes of Health.

### Author contributions

S.I.S.G. and G.T. designed the study. G.T., H.X., S.H., and J.D. performed experiments. H.X. performed RIP-seq and ultracentrifugation experiments. S.H. performed sporulation frequency experiment. J.D. generated the *cbc-1* mutant. D.W. performed statistical analysis of RIME and provided bioinformatics support. L.M.J. performed mass spectrometry analysis; G.T. performed all other experiments and bioinformatics analyses. S.I.S.G. and G.T. prepared data figures. S.I.S.G. and G.T. wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-16280-y>.

**Correspondence** and requests for materials should be addressed to S.I.S.G.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020