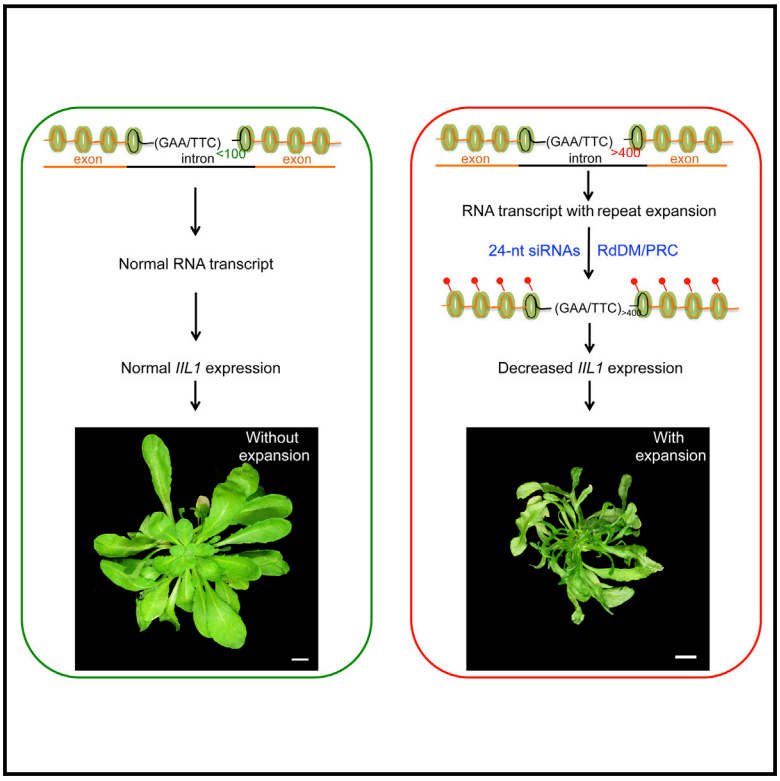
Article

Cell

RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions

Graphical Abstract



Highlights

- Triplet repeat expansions can trigger local production of 24-nt small RNAs
- Expansion-induced siRNAs correlate with the downregulation of gene expression
- Intronic repeat expansions cause gene silencing through the RdDM pathway

Authors

Hannes Eimer, Sridevi Sureshkumar, Avilash Singh Yadav, ..., Stephanie Frances Gordon, Bernard J. Carroll, Sureshkumar Balasubramanian

Correspondence

mb.suresh@monash.edu

In Brief

Triplet expansions within an *Arabidopsis* gene leads to local siRNA biogenesis and consequent transcriptional downregulation, suggesting how intronic repeat expansions may affect gene expression in other systems.

Eimer et al., 2018, Cell *174*, 1–11
August 23, 2018
https://doi.org/10.1016/j.cell.2018.06.044

CellPress

Please cite this article in press as: Eimer et al., RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.044

Cell

RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions

Hannes Eimer,[1,3] Sridevi Sureshkumar,[1,3] Avilash Singh Yadav,[1] Calvin Kraupner-Taylor,[1] Champa Bandaranayake,[1] Andrei Seleznev,[1] Tamblyn Thomason,[1] Stephen J. Fletcher,[2] Stephanie Frances Gordon,[1] Bernard J. Carroll,[2] and Sureshkumar Balasubramanian[1,4,*]

[1]School of Biological Sciences, Monash University, Clayton Campus, VIC 3800, Australia
[2]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia QLD 4072, Australia
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: mb.suresh@monash.edu
https://doi.org/10.1016/j.cell.2018.06.044

SUMMARY

Transcriptional downregulation caused by intronic triplet repeat expansions underlies diseases such as Friedreich's ataxia. This downregulation of gene expression is coupled with epigenetic changes, but the underlying mechanisms are unknown. Here, we show that an intronic GAA/TTC triplet expansion within the *IIL1* gene of *Arabidopsis thaliana* results in accumulation of 24-nt short interfering RNAs (siRNAs) and repressive histone marks at the *IIL1* locus, which in turn causes its transcriptional downregulation and an associated phenotype. Knocking down *DICER LIKE-3* (*DCL3*), which produces 24-nt siRNAs, suppressed transcriptional downregulation of *IIL1* and the triplet expansion-associated phenotype. Furthermore, knocking down additional components of the RNA-dependent DNA methylation (RdDM) pathway also suppressed both transcriptional downregulation of *IIL1* and the repeat expansion-associated phenotype. Thus, our results show that triplet repeat expansions can lead to local siRNA biogenesis, which in turn downregulates transcription through an RdDM-dependent epigenetic modification.

INTRODUCTION

Trinucleotide repeat expansions have been shown to underlie several neurogenetic diseases (Gatchel and Zoghbi, 2005). Disease-associated repeat expansions can be found both in coding and non-coding regions of the affected genes (Gatchel and Zoghbi, 2005; Pearson et al., 2005). Repeat expansions in coding regions, such as those seen in Huntington's disease are relatively small in size (~40 repeats) and mostly result in poly-glutamine stretches in the encoded protein, which can be toxic and pathogenic (Gatchel and Zoghbi, 2005; Pearson et al., 2005). By contrast, expansions in non-coding regions, as seen in Friedreich's ataxia or fragile X syndrome, can instead be massive (~up to 2,000 repeats) (Gatchel and Zoghbi, 2005; McMurray, 2010; Pearson et al., 2005). The mechanisms through which expansions in non-coding regions affect gene function appear to vary between different diseases (Gatchel and Zoghbi, 2005; Pearson et al., 2005).

Friedreich's ataxia (FRDA) is the most commonly inherited genetic ataxia, and it is caused by the expansion of GAA/TTC repeats in the first intron of the *FRATAXIN* (*FXN*) gene (Campuzano et al., 1996). This expansion is associated with a reduction in *FXN* expression, which is the primary defect and the underlying cause for FRDA (Bidichandani et al., 1998). Consequently, there are efforts to increase *FXN* expression as a potential therapeutic option for managing FRDA (Di Prospero and Fischbeck, 2005; Erwin et al., 2017; Evans-Galea et al., 2014; Li et al., 2018; Sandi et al., 2014).

Several models have been proposed to explain how intronic repeat expansions could lead to the observed transcriptional downregulation of the affected gene. It has been suggested that GAA/TTC expansions can form unusual DNA structures, which could potentially interfere with transcription (Bidichandani et al., 1998). *In vitro* transcription assays on DNA templates that contain repeat expansions indicated that GAA/TTC expansions could interfere with transcription in a length- and orientation-dependent manner (Ohshima et al., 1998; Sakamoto et al., 1999). The GAA/TTC repeat expansion in *FXN* is also thought to impede transcriptional elongation *in vivo*, as the epigenetic mark $H3K36me^3$ associated with transcriptional elongation was reduced in the *FXN* gene of FRDA cells (Punga and Bühler, 2010). Furthermore, histone deacetylase inhibitors have been shown to reverse transcriptional downregulation of *FXN* in the primary lymphocytes from individuals affected with FRDA, which suggests that the expansion of the intronic repeat induces repressive epigenetic modifications at the *FXN* locus (Herman et al., 2006; Nageshwaran and Festenstein, 2015). Moreover, the inclusion of ~200 GAA/TTC repeats in a transgene has been shown to cause silencing of the transgene in a mouse reporter cell line (Saveliev et al., 2003). It has also been shown that RNA/DNA hybrids (R-loops) could be formed due

to repeat expansions, which in turn are correlated with repressive H3K9me^2 chromatin marks (Groh et al., 2014). In addition, the epigenetic silencing of the *FXN* locus has been shown to be associated with antisense *FXN* transcription (De Biase et al., 2009). Simultaneous sense and antisense transcription of a gene could in principle lead to production of small RNAs that trigger RNA-induced transcriptional gene silencing (Borges and Martienssen, 2015; Verdel et al., 2004). However, efforts to identify small RNAs that map to *FXN* have not been fruitful (Punga and Bühler, 2010). While these findings suggest that GAA/TTC repeat expansions lead to transcriptional downregulation of the affected gene possibly via epigenetic gene silencing, it remains unclear how this occurs.

We have previously described an intronic GAA/TTC repeat expansion at the *ISOPROPYL MALATE ISOMERASE LARGE SUBUNIT 1* (*IIL1*) gene in the Bur-0 accession of *Arabidopsis thaliana* (Sureshkumar et al., 2009). This triplet repeat expansion, which is still present in natural Irish populations of *Arabidopsis thaliana* (henceforth referred to as *Arabidopsis*), causes a growth defect referred to as "*irregularly impaired leaves*" (*iil*) that is seen only at elevated temperatures (Sureshkumar et al., 2009; Tabib et al., 2016). This intronic repeat expansion is also associated with transcriptional downregulation of *IIL1*, and increasing *IIL1* expression through transgenic manipulation resulted in the suppression of the *iil* phenotype (Sureshkumar et al., 2009). Spontaneous natural revertants, which have lost the repeat expansion, show increased *IIL1* expression at high temperatures and a suppression of the *iil* phenotype confirming that the repeat expansion-mediated transcriptional downregulation underlies the *iil* phenotype (Sureshkumar et al., 2009). However, how the repeat expansion leads to transcriptional downregulation of the *IIL1* locus remains unanswered.

The intronic GAA/TTC repeat expansion at *IIL1* in *Arabidopsis* shares some parallels with the repeat expansion at *FXN* in humans that results in FRDA. In both instances, the intronic repeat expansion leads to a reduction in gene expression, and this reduction appears to be the underlying cause for the observed phenotypes (Bidichandani et al., 1998; Sureshkumar et al., 2009). In addition, in both situations, the repeat expansion displays somatic variability and genetic instability, suggesting that at least some of the underlying mechanisms may be conserved across kingdoms (Bidichandani et al., 1999; Clark et al., 2007; De Biase et al., 2007; Sureshkumar et al., 2009).

One of the well-documented mechanisms through which transcriptional gene silencing can be achieved in plants involves small RNAs (Borges and Martienssen, 2015). In particular, 24-nt siRNAs produced by DICER-LIKE 3 (DCL3) guide transcriptional gene silencing in *Arabidopsis* (Borges and Martienssen, 2015). These 24-nt siRNAs are methylated and stabilized by HUA ENHANCER 1 (HEN1), and then associate with ARGONAUTE 4 (AGO4) to recognize complementary, non-coding RNA polymerase V (Pol V) transcripts. Formation of the AGO4-Pol V complex results in recruitment of DNA methyltransferases and histone modifying enzymes that apply repressive epigenetic marks to the local chromatin (Borges and Martienssen, 2015).

Here, we used the *IIL1* repeat expansion in *Arabidopsis* as a model to investigate the molecular mechanisms through which intronic repeat expansions lead to transcriptional downregulation of the affected gene. In common with FRDA in humans, we demonstrate that the transcriptional downregulation caused by the triplet repeat expansion is associated with an increase in the repressive epigenetic marks at the *IIL1* locus. We also demonstrate that the GAA/TTC repeat expansion at *IIL1* leads to increased accumulation of 24-nt siRNAs in a temperature-dependent manner that correlates with the *iil* phenotype. We show that *DCL3* and other components of the RNA-dependent DNA methylation (RdDM) pathway are essential for this siRNA-directed epigenetic gene silencing of *IIL1*. Our results suggest that siRNA-mediated epigenetic regulation plays a fundamental role in the silencing of genes harboring intronic GAA/TTC repeat expansions in plants. Given the earlier discovery of an antisense transcript and histone modifications at *FXN* (De Biase et al., 2009), our findings suggest that RNA-dependent transcriptional gene silencing pathways should be assessed further as a possible mechanism for epigenetic attenuation of *FXN* expression in FRDA.

RESULTS

Triplet Expansion-Induced Transcriptional Downregulation of *IIL1* Is Associated with Epigenetic Changes

We have previously hypothesized that the reduction in *IIL1* expression observed in Bur-0 could be due to epigenetic gene silencing (Sureshkumar et al., 2009). To assess whether the downregulation of *IIL1* is associated with repressive epigenetic marks, we examined the abundance of H3K27me^3, which is known to be associated with a transcriptionally repressed chromatin state (Gan et al., 2015), at the *IIL1* locus. We compared the abundance of H3K27me^3 in the Bur-0 accession harboring the repeat expansion (*IIL1*-Bur-0, >400 repeats in intron 3 of *IIL1*), which is associated with the *iil* phenotype, with that of a spontaneous phenotypic revertant in the Bur-0 background called Natural Suppressor 15 (NS15, *IIL1*-NS15) that partially lost the repeat expansion (with ~100 repeats), and the reference strain Col-0, which contains 23 copies of this triplet repeat in *IIL1* (*IIL1*-Col-0). We observed increased accumulation of H3K27me^3 in Bur-0 plants compared to NS15 and Col-0 controls, which suggests that the *IIL1*-Bur-0 locus is epigenetically silenced in a triplet expansion-dependent manner (Figures 1A and S1). In contrast, the accumulation of the H3K36me^3 histone mark, which is known to be associated with active chromatin (Wagner and Carpenter, 2012), was reduced in *IIL1*-Bur-0 that harbors the repeat expansion (Figure 1B), when compared to that in *IIL1*-Col-0 and *IIL1*-NS15. This is consistent with the observed reduction in *IIL1* expression in Bur-0 and correlates with the *iil* phenotype. These results suggested that the repeat-expansion-associated transcriptional downregulation of *IIL1* observed in Bur-0 could be mediated through epigenetic modifications.

Triplet Repeat Expansion Is Associated with Increased Accumulation of Small RNAs

Small RNAs play a critical role in epigenetic gene silencing in plants (Borges and Martienssen, 2015). A 38-bp tandem repeat present in the promoter region of the *FWA* gene was shown to be associated with small RNA-dependent, DNA methylation-mediated gene silencing (Chan et al., 2006). Genetically

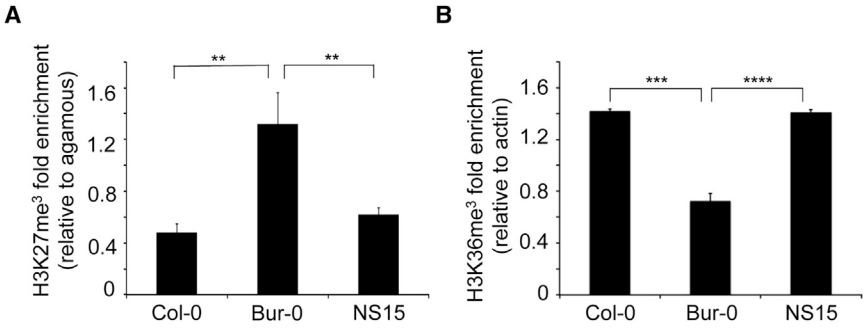


Figure 1. Triplet Expansion-Mediated Transcriptional Downregulation Is Associated with Epigenetic Changes

(A) Relative enrichment of H2K27me^3, a mark for repressive chromatin in Bur-0 compared to NS15 and Col-0 that lack the repeat expansion. The percentage of the input data are normalized as fold enrichment compared to enrichment seen in *AG*, which is used as a positive control for this repressive epigenetic mark.

(B) Depletion of H3K36me^3 in Bur-0 compared to NS15 and Col-0. The percentage of the input data are normalized as fold enrichment compared to enrichment seen in *ACTIN*, which is used as a positive control for this active epigenetic mark for transcriptional elongation. p values from one-way ANOVA are shown for the respective comparisons, $^{**}p < 0.01$, $^{***}p < 0.001$.

Error bars represent SEM. See also Figure S1.

engineered tandem repeats of similar length have also been demonstrated to generate siRNAs in *Arabidopsis* (Sasaki et al., 2014). It has been shown that the ribonuclease DICER can cleave CNG triplet repeats that form hairpins to shorter repeats that can silence specific targets (Krol et al., 2007). Furthermore, in human system, an antisense transcript, which has the potential to form double-stranded RNA, has been previously reported at the *FXN* locus, associated with its epigenetic silencing in FRDA (De Biase et al., 2009). Therefore, we considered whether small RNAs could be involved in the transcriptional downregulation of *IIL1* that underlies the *iil* phenotype.

To test whether small RNAs could contribute to the *iil* phenotype, we performed genome-wide deep sequencing of small RNAs in Bur-0, NS15, and Col-0 plants grown at 23°C and 27°C. We mapped the small RNAs to the respective genomes (Col-0 [for Col-0] or Bur-0 [for Bur-0 and NS15]), filtered, and quantified the small RNAs that map to the *IIL1* locus (Figure 2; Table 1). When compared with Col-0 or NS15, there was a significant increase in the number of small RNAs that mapped to the *IIL1* locus in Bur-0 both at 23°C and 27°C (Figure 2; Table 1, $p < 0.002$ for all comparisons, Student's t test). Consistent with the temperature-dependence of the *iil* phenotype, we also observed a significant increase in the number of small RNAs that map to the *IIL1* locus in Bur-0 plants grown at 27°C compared to 23°C (Table 1, $p = 0.0057$, Student's t test). This increase in small RNAs in Bur-0 grown at 27°C was not restricted to the triplet-derived small RNAs, but spread both upstream and downstream of the repeat (Figure 2). The majority of *IIL1* small RNAs mapped to the 5′ region of the gene up to exon 8, including the location of the repeat expansion (intron 3) and the transcriptional start site of the locus (Figure 2). There was also a significant increase in the small RNAs that perfectly matched the triplet expansion in Bur-0 grown at 27°C (Table 1, $p < 0.001$ for all comparisons, Student's t test). These findings suggested that there is a temperature-dependent and locus-specific increase in the abundance of *IIL1* small RNAs in Bur-0, correlating with the length of the triplet repeat expansion and its associated phenotype.

Knocking Down *DCL3* Suppresses the Triplet Expansion-Associated Phenotypes

In plants, different DICER molecules produce distinct species of small RNAs, which can be distinguished based on their size (Xie et al., 2004). We therefore analyzed the read length distribution of the small RNAs that map to the *IIL1* locus in our samples and found that 24-nt small RNAs represented the dominant species (Figure 2). The 24-nt siRNAs are produced by DICER LIKE 3 (DCL3) in *Arabidopsis* (Xie et al., 2004). To determine if the 24-nt siRNAs we identified have a role in the *iil* phenotype, we ascertained the effect of knocking down *DCL3* in the Bur-0 background using artificial microRNAs (*35S::amiR-DCL3*). We obtained numerous independent transgenic lines in which RNA expression analyses confirmed downregulation of *DCL3* (Figure S2B). We observed that the *iil* phenotype was indeed suppressed in *35S::amiR-DCL3* lines (Figures 3A and S2A). Of the 28 primary transgenic lines, we observed varying degrees of phenotypic suppression including complete suppression of the *iil* phenotype in 12 lines with other lines displaying varying levels of phenotypic suppression (Figures 3A and S2A). Stable phenotypic suppression as well as the presence of the repeat expansion was further confirmed up to the third generation (Figure S3A; Table S1). Suppression of the *iil* phenotype was coupled with an increase in *IIL1* expression and a decrease in the 24-nt siRNAs mapping to *IIL1* (Figures 3B, S2C, S3B, and S3C). This suggests that DCL3 activity, and in turn the 24nt-siRNAs that map to *IIL1*, are essential for the transcriptional downregulation observed in the presence of a triplet repeat expansion at the *IIL1*-Bur-0 locus.

To assess whether 24-nt siRNAs underlie the epigenetic gene silencing seen at *IIL1* in Bur-0 at 27°C, we determined the epigenetic status at the *IIL1* locus in *35S::amiR-DCL3* plants in which the *iil* phenotype is suppressed at 27°C. Interestingly, we found that H3K36me^3 levels were significantly increased in *35S::amiR-DCL3* Bur-0 lines compared to non-transgenic Bur-0 control plants (Figure 3C), despite the presence of the expanded repeat in both batches of plants. This indicates that the *DCL3*-dependent siRNAs are essential for the observed epigenetic changes caused by the expanded repeat in *IIL1* of Bur-0.

Downregulation of Pol IV and RDR2 in Bur-0 Suppresses Triplet Expansion-Associated Growth Defects

The best-known pathway for 24-nt siRNA biogenesis in *Arabidopsis* involves DCL3 processing of double-stranded RNAs synthesized by RNA Pol IV (Pol IV) and RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) (Martienssen, 2003; Onodera et al., 2005). To assess the involvement of RDR2, we knocked

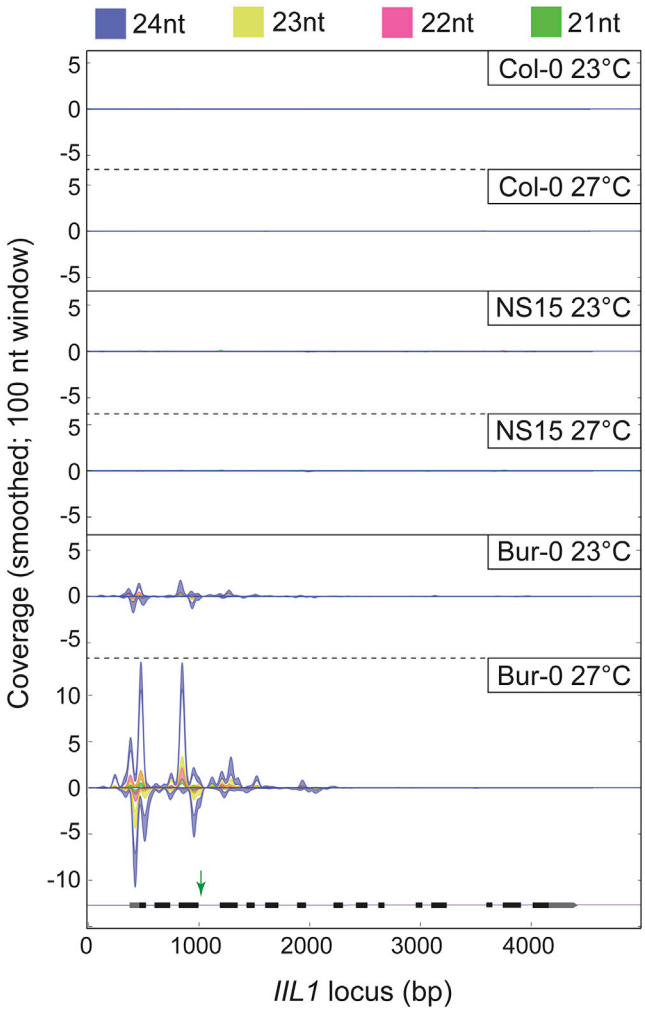


Figure 2. Abundance of siRNAs that Map to *IIL1* Locus Are Significantly Increased in Bur-0

Small RNA profiles at the *IIL1* locus generated with Small Complementary RnA Mapper (SCRAM) (Fletcher et al., 2018). The genic region is shown at the bottom with the black boxes and lines representing exons and introns, respectively. SCRAM output shows the normalized coverage along with standard error (shown as shadows) across the *IIL1* locus for the 21, 22, 23, and 24 nt small RNAs. Only the small RNAs that mapped to non-triplet repeat sequences of *IIL1* are shown in the figure. Quantification of the sense and antisense small RNAs is shown in the positive and negative dimensions, respectively, along the y axis. The green arrow indicates the position of the GAA/TTC triplet repeat in intron 3 of *IIL1*.

See also Tables S2 and S3.

down RDR2 using artificial microRNAs and analyzed the impact of this knockdown on the *iil* phenotype. Similar to the phenotypic suppression of *iil* seen in *35S::amiR-DCL3*, *35S::amiR-RDR2*-mediated knockdown of *RDR2* in the Bur-0 background resulted in transgenic plants in which the *iil* phenotype was suppressed (41 out of 49 total lines, representing 3 different amiR-constructs, Figures 4A and S4A; Table 2). Similarly, transgenic lines harboring artificial microRNAs expressed against the *NRPD1A* gene encoding the largest subunit of Pol IV (*35S::amiR-NRPD1A*) displayed suppression of the *iil* phenotype (39 out of 66 lines, Figures 4A and S4D; Table 2). Consistently, we observed an increase in *IIL1* expression in these knockdown lines (Figures 4B, S4B, S4C, and S4E), which suggests that Pol IV/RDR2 are among the components involved in the transcriptional downregulation of *IIL1* caused by the triplet repeat expansion.

HEN1, AGO4, and Pol V Are Required for the Triplet Expansion-Associated Transcriptional Downregulation of *IIL1*

Multiple loci play a role in mediating siRNA-dependent epigenetic silencing (Bologna and Voinnet, 2014). One of the initial steps in siRNA biogenesis is the methylation of small RNAs, which is mediated through *HEN1* (Yang et al., 2006). Therefore, we tested whether *HEN1* is required for the transcriptional downregulation seen in the presence of the triplet repeat expansion. Of the total 27 independent *35S::amiR-HEN1* primary transformants, 22 displayed varying levels of suppression of the *iil* phenotype at 27°C with 12 displaying complete suppression (Figures 4A and S5A; Table 2). The expression of *IIL1* was increased in these transgenic knockdown lines (Figures 4B, S5B, and S5C) along with the suppression of the *iil* phenotype, which indicates that *HEN1* is required for the observed transcriptional downregulation.

24-nt siRNAs can cause transcriptional gene silencing of a target gene through their interaction with AGO4 (Qi et al., 2006; Zilberman et al., 2003). To test whether *AGO4* is involved in the triplet expansion-induced transcriptional downregulation of *IIL1*, we knocked down *AGO4* in the Bur-0 background and analyzed its impact on the *iil* phenotype. *35S::amiR-AGO4* transgenic plants displayed suppression of the *iil* phenotype at 27°C, suggesting that AGO4 is required for the repeat expansion-mediated *iil* phenotype (Figure 4A). We observed complete phenotypic suppression in most (10 out of 11, Table 2) of the independent transgenic lines. Expression analysis in these transgenic lines revealed that the levels of the *IIL1* transcript were increased when *AGO4* levels were reduced, along with a reduction in 24-nt siRNAs mapping to *IIL1* despite the presence of the repeat expansion (Figures 4B, S3C, S5D, and S5E). This suggests that *AGO4* is essential for the transcriptional downregulation of *IIL1* caused by the repeat expansion.

AGO4-dependent transcriptional downregulation involves RNA polymerase V (Pol V), a plant-specific RNA polymerase (Herr et al., 2005; Lahmy et al., 2009). Multiple independent transgenic plants (12 out of 12) in which RNA Pol V was knocked down using artificial microRNAs (*35S::amiR-NRPE1)* in the Bur-0 background appeared normal at 27°C (Figure 4A; Table 2). Consistent with a reduction in *NRPE1* levels, *IIL1* expression was increased in these lines at 27°C despite the presence of the repeat expansion along with a reduction in siRNAs mapping to *IIL1* (Figures 4B, S3C, S5D, and S5F). These results suggest that Pol V is essential for the observed transcriptional downregulation of *IIL1* caused by the repeat expansion.

Perturbation of DNA Methylation Suppresses Triplet Expansion-Associated Transcriptional Downregulation

Canonical models for small RNA-mediated transcriptional silencing invoke a role for DNA methylation (Cuerda-Gil and Slotkin, 2016; Matzke and Mosher, 2014), mediated through the *DOMAINS REARRANGED METHYL TRANSFERASE 2* (*DRM2*),

Table 1. Abundance of siRNAs that Map to the *IIL1* Locus Is Increased in Bur-0 Grown at Higher Temperatures

			23°C			27°C		
Genotype	Region	Type	Rep 1	Rep 2	Ave ± SD	Rep 1	Rep 2	Ave ± SD
Col-0	repeat	sense	0	0.04	0.02 ± 0.02	0	0.06	0.03 ± 0.04
		antisense	1.31	2.52	1.92 ± 0.86	3.25	9.24	6.25 ± 0.86
	non-repeat	sense	1.31	0.74	1.03 ± 0.4	0.33	1.02	0.68 ± 0.49
		antisense	0	0	0	0.07	0.11	0.09 ± 0.03
		total	2.62	3.3	2.96 ± 0.48	3.65	10.43	7.04 ± 4.79
Bur-0	repeat	sense	0	0.09	0.05 ± 0.06	0.18	0.49	0.34 ± 0.22
		antisense	3.43	7.87	5.65 ± 3.14	97.35	103.34	100.35 ± 4.24
	non-repeat	sense	17.52	18.32	17.92 ± 0.57	135.27	124.47	129.87 ± 7.64
		antisense	9.62	15.64	12.63 ± 4.26	82.8	77.22	80.01 ± 3.94
		total	30.57	41.92	36.25 ± 8.03	315.6	305.52	310.56 ± 7.13
NS15	repeat	sense	0	0.17	0.09 ± 0.12	0.19	0.06	0.13 ± 0.09
		antisense	13.63	29.64	21.64 ± 11.32	16.17	18.05	17.11 ± 1.33
	non-repeat	sense	1.86	1.7	1.78 ± 0.11	1.36	1.41	1.39 ± 0.04
		antisense	0.29	0.4	0.35 ± 0.08	0.49	0.35	0.42 ± 0.10
		total	15.78	31.91	23.85 ± 11.4	18.21	19.87	19.04 ± 1.17

Normalized sequence reads (by library size, in reads per million) of siRNAs that mapped to the *IIL1* locus in sense and antisense direction are shown. Datasets of each library include normalized reads perfectly matching to the *IIL1* locus (from −400 upstream of transcriptional start site and +200 from transcriptional stop site, shown as non-repeat) and the reads that map perfectly to the TTC/GAA repeats (shown as repeat) for Bur-0, a natural suppressor NS15 (lacks the repeat expansion in the Bur-0 background), and Col-0 grown at 23°C and 27°C in short days. SD, standard deviation of the two biological replicates. The siRNA counts in Bur-0 at 27°C is significantly different from all others ($p < 0.002$ for all comparisons, Student's t test).

which carries out *de novo* methylation of the DNA. We compared the methylation profile of *IIL1*-Bur-0 with *IIL1*-Col-0, using the epigenome data from the 1001 Genomes Project (Kawakatsu et al., 2016). *IIL1*-Bur-0 appears to be hypermethylated in the 5′ region of the gene up to exon 8 including the repeat expansion, particularly in CHG and CHH sequence contexts, when compared with the reference strain Col-0 (Figure S6A). Furthermore, analysis of all the epigenome data from the 1001 Genome Project revealed that the methylation profile of *IIL1*-Bur-0 was unique, correlating with the presence of the repeat expansion (Figure S6B). To assess whether *DRM2* is required for the transcriptional downregulation caused by expanded repeats, we knocked down *DRM2* using artificial microRNAs in the Bur-0 background. The *35S::amiR-DRM2* lines, in which *DRM2* levels were reduced, displayed suppression of the *iil* phenotype. Of the total 48 independent T1 lines harboring two distinct *amiR-DRM2* constructs, 41 showed phenotypic suppression of *iil*, coupled with an increase in the *IIL1* expression (Figures 4A and S6C–S6E; Table 2). While DRM2 is required for *de novo* RNA-directed DNA methylation in all sequence contexts, maintenance of methylation at CHG and CG require methyltransferases CHROMOMETHYLASE 3 (CMT3) and DNA METHYLTRANSFERASE 1 (MET1), respectively (Cao and Jacobsen, 2002; Chan et al., 2005; Law and Jacobsen, 2010). In addition, loss of MET1 has also been shown to modulate histone methylation mediated by polycomb group (PcG) proteins (Deleris et al., 2012). To assess the role of *CMT3* and *MET1* in the *iil* phenotype, we generated transgenic lines harboring knockdown constructs for *CMT3* and *MET1* in the Bur-0 background (Figures 4A and S6F–S6K; Table 2). We obtained 58 and 33 independent primary transgenic lines for *35S::amiR-CMT* and *35S::amiR-MET1,* respectively, in the Bur-0 background. We observed a statistically significant increase in the phenotypic suppression of *iil* coupled with an increase in *IIL1* expression with both constructs; ~50% of T1 lines for both *35S::amiR-CMT3* (27/58) and *35S::amiR-MET1* (16/33), (Figures 4B and S6F–S6K; Table 2). Taken together, these results confirm that DNA methylation is an essential component of the triplet expansion-induced transcriptional downregulation of *IIL1*.

Histone Methyltransferases Are Essential for the Transcriptional Downregulation Caused by Triplet Repeat Expansions

Because our results have shown the involvement of the RdDM pathway in the transcriptional downregulation of *IIL1* induced by the repeat expansion and because there was enrichment for the H3K27me^3 in the presence of repeat expansion (Figure 1A), we asked whether perturbing histone methyltransferases could suppress the *iil* phenotype. To assess this, we generated artificial microRNA constructs against the polycomb group protein LIKE HETEROCHROMATIN 1 (LHP1) and the histone N-lysine methyltransferase, CURLY LEAF, a component of the polycomb repressive complex (PRC) (Mozgova and Hennig, 2015). The *35S::amiR-LHP1* (20 primary transgenic lines) and *35S::amiR-CLF* (38 independent primary transformants) plants displayed very strong phenotypes similar to *lhp* and *clf* mutants (Figure S7A), which prevented us from assessing whether the *iil* phenotype was suppressed in these lines. However, *IIL1* expression analysis in few *35S::amiR-CLF* lines suggested a potential increase in *IIL1* expression when CLF was perturbed (Figure S7B). To assess this possibility further, we generated transgenic lines that expressed artificial microRNAs against *CLF* under the control of an estradiol-inducible promoter (*XVE::amiR-CLF).* We

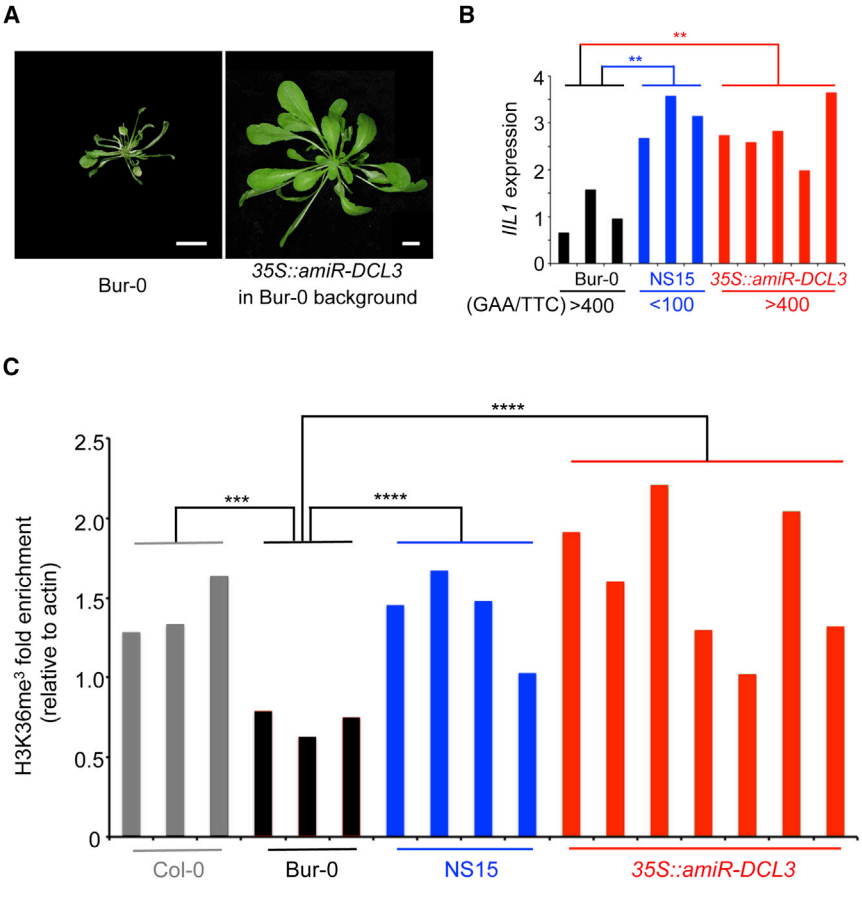


Figure 3. *DCL3* Is Required for Transcriptional Downregulation of *IIL1* Induced by an Intronic GAA/TTC Triplet Repeat Expansion

(A) Knocking down *DCL3* suppresses the *irregularly impaired leaves* (*iil*) phenotype in the *Arabidopsis* accession Bur-0. Scale bar, 1 cm.

(B) Expression levels of *IIL1* in different genotypes harboring varied intronic GAA/TTC triplet repeat lengths. All genotypes are in the Bur-0 genetic background. Expression levels from individual plants are shown thus lacking the error bars. ANOVA revealed that the expression differences are significant ($p < 0.0001$).

(C) Relative enrichment of $H3K36me^3$ in chromatin immunoprecipitation (ChIP) experiments across various genotypes measured through qPCR. The data shown in (B) and (C) are for individual plants and the p values from one-way ANOVA are shown above the comparisons. **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

See also Figures S2 and S3 and Table S1.

induced the *amiR-CLF,* when the primary transgenic plants were at the 7-leaf stage, which is when the Bur-0 plants typically start to display the *iil* phenotype when grown at 27°C (Sureshkumar et al., 2009). Among the transgenic lines, we selected lines in which the *clf* phenotype (leaf curling) was observed and asked how many of these lines also displayed the suppression of the *iil* phenotype. Among the 36 independent primary transgenic lines that showed leaf curling, we detected 21 to be suppressed for the *iil* phenotype (Figures 4A and S7C–S7E; Table 2), which indicates that *CLF* is required for the transcriptional downregulation of *IIL1* caused by triplet repeat expansion in Bur-0.

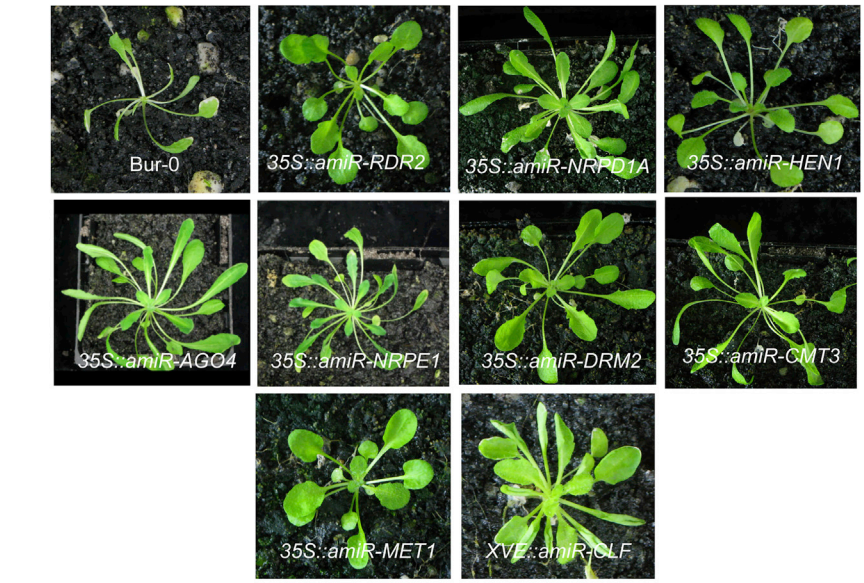
Taken together, these results indicate that GAA/TTC triplet repeat expansions can lead to an increase in siRNA accumulation and that these siRNAs direct transcriptional downregulation of the affected gene via epigenetic silencing. The siRNAs appear to modulate epigenetic silencing of the loci harboring the repeat expansion through the RdDM pathway, eventually resulting in repressive epigenetic marks being deposited on chromatin by the PRC.

DISCUSSION

Triplet Repeat Expansions in Transcribed Regions Can Trigger Production of Small RNAs

Using *Arabidopsis* as a model, we have demonstrated that an intronic GAA/TTC triplet repeat expansion can trigger the local production of small RNAs in its vicinity. We have also shown that this triplet expansion-associated increase in small RNAs is dependent on the length of the triplet expansion, and the abundance of these small RNAs is enhanced at elevated temperature. Furthermore, we have demonstrated that these small RNAs can then target and silence the genes harboring the repeat expansion. Triplet expansion derived small RNAs have been previously reported in a *Drosophila* genetic model for Myotonic dystrophy (Yu et al., 2011). In that study involving transgenic *Drosophila* lines, a CAG/CTG repeat expansion was shown to produce 21-nt small RNAs (Yu et al., 2011). Although our study indicates accumulation of 24-nt small RNAs, it appears that the production of siRNAs derived from triplet repeats is not restricted to plants.

We have shown that the number of triplet-derived siRNAs is significantly higher in plants harboring the repeat expansion (Table 1). It is unlikely that the production of the repeat-specific siRNAs came from a GAA/TTC triplet repeat elsewhere in the genome, as the accumulation of these siRNAs was dependent on the length of the triplet repeats, since the repeat-specific siRNA levels are much lower in NS15 plants than in Bur-0 plants. In the Col-0 accession GAA/TTC triplet-derived siRNAs can be barely detected. There are a few GAA/TTC repeat stretches present in intergenic regions of the Col-0 genome, the longest being a 489bp (163 repeats) stretch between At4g06597 and At4g06635. While this is significantly lower than the number of repeats present in Bur-0 (> 400), it is somewhat similar to the number of repeats seen in the NS15 suppressor (in the Bur-0 background) that harbors around 100 repeats. However, in contrast to Col-0, we could detect a significant number of triplet-derived siRNAs in NS15 ($p < 0.01$, Student's t test). While genetic differences between Col-0 and Bur-0 could account for



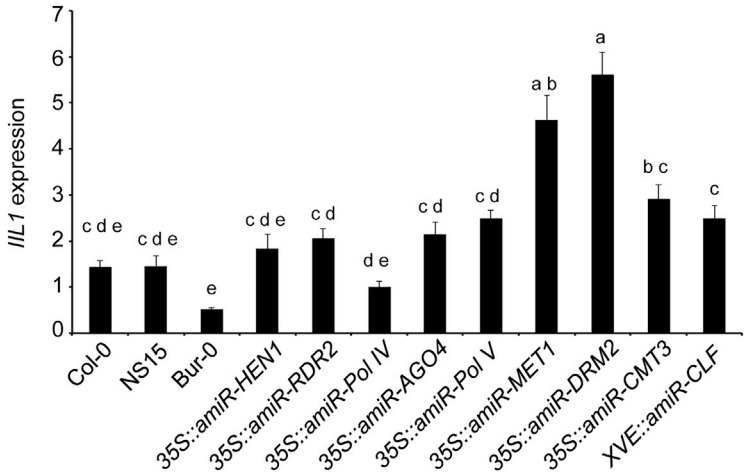


Figure 4. Multiple Components of the RdDM Pathway Are Required for Triplet Expansion-Associated Transcriptional Downregulation in Plants

(A) Phenotypes of Bur-0, *35S::amiR-RDR2*, -*35S::amiR-NRPD1A* (Pol IV), *35S::amiR-HEN1*, *35S::amiR-AGO4*, and *35S::amiR-NRPE1* (Pol V), *35S::amiR-DRM2*, *35S::amiR-CMT3*, *35S::amiR-MET1*, and *XVE::amiR-CLF* plants in the Bur-0 genetic background.

(B) The expression levels of *IIL1* in Col-0, NS15, Bur-0, and across various primary transgenic lines harboring artificial microRNAs against the individual components of the RdDM pathway in Arabidopsis. For transgenic lines, average *IIL1* expression levels (±SE) across multiple independent transgenic lines are shown. Please see Figures S4, S5, S6, and S7 for the data on individual primary transformants. p values are based on Tukey's post hoc test, and the lines that are not harboring the same letters are significantly different from each other ($p < 0.05$). Error bars represent SEM.

this difference, it is also conceivable that triplet repeats in the transcribed regions are more likely to lead to siRNAs as opposed to the repeats that are present in the intergenic regions of the genome. This would suggest that a high level of transcription and/or the engagement of transcriptional machinery may be essential for the production of siRNAs from triplet repeat expansions.

Triplet-Derived siRNAs Are Increased at Higher Temperatures

We have previously shown that the *iil* phenotype in Bur-0 is temperature-dependent and it is correlated with the temperature-dependent decrease in *IIL1* expression (Sureshkumar et al., 2009; Tabib et al., 2016). Our results now demonstrate a correlation with a temperature-dependent increase in the 24-nt siRNAs that map to *IIL1*. It is currently unclear why the accumulation of the triplet-derived siRNAs is higher at elevated temperatures. Previous studies have suggested that viral as well as transgene-induced silencing are dependent on temperature, with higher temperatures resulting in increased levels of siRNAs that lead to efficient silencing and lower temperatures making the plants more susceptible to viral attacks (Szittya et al., 2003). Therefore, one possibility is that the siRNA biogenesis itself is dependent on the ambient growth temperature. However, a comparison of the normalized counts of siRNAs that are highly expressed and are common between the two different temperatures failed to reveal a general increase in siRNAs at 27°C across the genome within the Bur-0 accession (Table S2). We also did not observe a general increase in siRNAs in Bur-0, compared to NS15, which lacks the repeat expansion (Table S3). Therefore, the temperature-dependent increase seems to be specific to the production of siRNAs that map to the *IIL1* locus.

While it is clear that siRNAs at the *IIL1* locus mediate the epigenetic gene silencing and transcriptional downregulation, it is not clear how these siRNAs originate from the repeat expansion. In the case of FRDA, an antisense RNA transcript has been previously shown to be associated with epigenetic silencing of the *FXN* locus (De Biase et al., 2009). However, so far our efforts have failed to detect long non-coding antisense RNA at the *IIL1*-Bur-0 locus. We cannot, however, rule out the possibility of an unstable antisense transcript that is quickly processed or degraded, thus escaping our detection. It is also feasible that there might be secondary structures in transcripts with the repeat expansion that act as substrates

Table 2. Phenotypic Summary of Transgenic Lines that Express Artificial MicroRNAs against Components of the RdDM Pathway

Genotype	Construct	Total	Phenotype		% of Suppression	p Value (Chi-Squared Test)
			iil1	Sup		
Bur-0	Expt1	32	28	4	13	
	Expt2	30	26	4	13	
	Expt3	12	8	4	33	
Bur-0 total		74	62	12	16	
35S::amiR-RDR2	G54	8	3	5	8	
35S::amiR-RDR2	pSS1	15	2	13	87	
35S::amiR-RDR2	pSS2	26	3	23	88	
35S::amiR-RDR2 total		49	8	41	84	****
35S::amiR-HEN1	miR1	11	1	10	90	
35S::amiR-HEN1	miR2	16	4	12	75	
35S::amiR-HEN1 total		27	5	22	81	****
35S::amiR-NRPD1	G65	6	5	1	17	
35S::amiR-Pol IV (*NRPD*)	pSS3	24	12	12	50	
35S::amiR-Pol IV (*NRPD*)	pSS4	36	10	26	72	
35S::amiR-Pol IV total		66	27	39	59	****
35S::amiR-AGO4 total	G67	11	1	10	90	****
35S::amiR-POL V total	G74	12	0	12	100	****
35S::amiR-DRM2	pSS7	15	3	12	80	
35S::amiR-DRM2	pSS8	33	4	29	88	
35S::amiR-DRM2 total		48	7	41	85	****
35S::amiR-CMT3	pSS11	45	23	22	49	
35S::amiR-CMT3	pSS12	13	8	5	38	
35S::amiR-CMT3 total		58	31	27	47	****
35S::amiR-MET1	pSS9	24	13	11	46	
35S::amiR-MET1	pSS10	9	4	5	56	
35S::amiR-MET1 total		33	17	16	48	****
35S::amiR-LHP1 total	pCB1	20			strong *lhp* phenotype	
35S::amiR-CLF	pSS5	14			strong *clf* phenotype	
35S::amiR-CLF	pSS6	24			strong *clf* phenotype	
35S::amiR-CLF total		38				
XVE:amiR-CLF	*pGREAT-SS5*	27	11	16	59	
XVE::amiR-CLF	*pGREAT-SS6*	9	4	5	56	
XVE::amiR-CLF total		36	15	21	58	****

The last column represents whether the phenotypic suppression observed in the transgenic lines is statistically significant based on a chi-squared test.

to trigger small RNA production. It should also be noted that most of the siRNAs map to the 5′ region of the locus, while they are barely detected in the 3′ region. This suggests that these siRNAs may be associated with the regular Pol II-mediated transcription of *IIL1* and either R-loops or varied RNA structures might trigger siRNA production from the *IIL1* locus. Once siRNA biogenesis is initiated, tandem repeats have been hypothesized to sustain RNA-dependent RNA polymerase activity, as the siRNAs that are generated can act both upstream and downstream of the site of initiation, unlike single-copy repeats where they can act only downstream (Martienssen, 2003). As a result, tandem repeats have been proposed to be better substrates for gene silencing than single copy sequences (Martienssen, 2003).

siRNA-Mediated Epigenetic Silencing Underlies the Triplet Expansion-Associated Transcriptional Downregulation of the Affected Gene

Although it has been known for some time that the intronic GAA/TTC triplet repeat expansion seen in FRDA is associated with epigenetic changes (Chutake et al., 2014; Herman et al., 2006; Kim et al., 2011; Kumari et al., 2011; Punga and Bühler, 2010), how the repeat expansion leads to epigenetic modifications remains unknown. We have shown that the transcriptional downregulation of the *IIL1* locus caused by the intronic GAA/TTC repeat expansion is associated with epigenetic changes that are dependent on the 24-nt siRNAs. Our findings reveal that the GAA/TTC repeat expansion in the *IIL1* intron lead to an increase in 24-nt siRNAs, which leads to epigenetic silencing of the *IIL1*

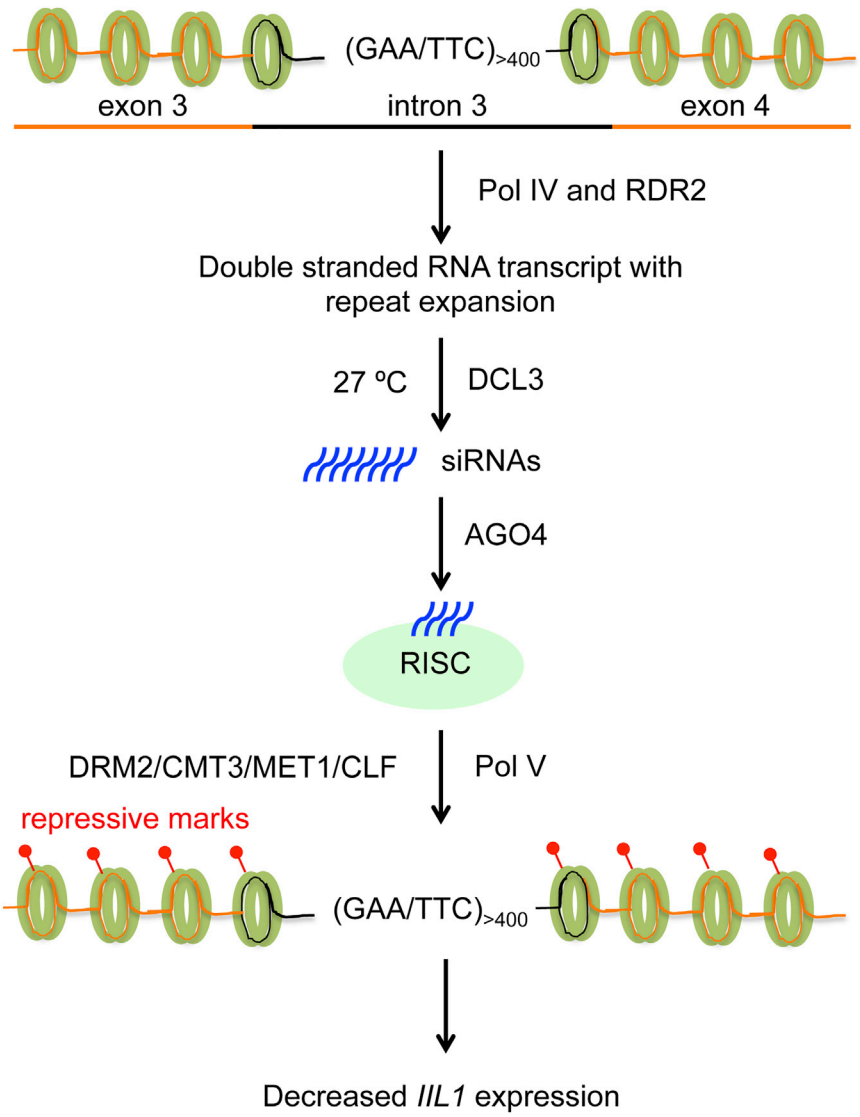


Figure 5. A Model for Intronic Triplet Repeat Expansion-Induced Transcriptional Downregulation in Plants

The *IIL1* locus across the 3rd intron harboring the expansions with the flanking exons is color-coded (orange, exon; black, intron) and shown around the nucleosomes (green). The different genes that act at varied levels are shown in the side. Triplet repeats lead to 24-nt siRNAs through *DCL3*, which gets loaded on to the RISC complex that modulates the epigenetic status of the locus through a pathway involving RdDM components. Subsequent epigenetic changes result in the transcriptional downregulation of *IIL1*.

locus through the RdDM pathway and the polycomb repressive complex. We have thus defined the pathway through which intronic triplet expansions can lead to epigenetic silencing in *Arabidopsis* (Figure 5).

It has been previously shown in the case of FRDA that the GAA/TTC repeat expansion can lead to epigenetic changes, which in turn results in blocking of transcriptional elongation (Chutake et al., 2014; Herman et al., 2006; Kim et al., 2011; Kumari et al., 2011; Punga and Bühler, 2010). However, the involvement of siRNAs in this process is currently unclear. A previous study that profiled small RNA in one of the FRDA cell lines did not find an increase in small RNAs mapping to the *FXN* locus in FRDA cell lines compared with control cell lines derived from unaffected siblings (Punga and Bühler, 2010). However, as described earlier, it has been suggested that the epigenetic silencing in FRDA is associated with an antisense transcript (De Biase et al., 2009). While a pathway identical to RdDM has not been reported in mammalian system, several components of the pathway (e.g., *DICERs*, *ARGONAUTEs*, and *DNMTs)* and the PRC complexes exist in mammals. Based on our results, it is not inconceivable that sense and antisense transcription at the *FXN* locus could potentially lead to siRNAs, which could eventually result in epigenetic changes at the *FXN* locus. The potential role of an antisense *FXN* transcript and the potential role for siRNAs in downregulation of the *FXN* gene in FRDA require further investigation.

In conclusion, we have demonstrated that triplet expansions in transcribed regions of the genome have the potential to generate siRNAs, which in turn can target the locus harboring the repeat expansion for epigenetic gene silencing. Epigenetic changes have been implicated in several triplet expansion disorders (Nageshwaran and Festenstein, 2015). It has also been suggested that the repeats that undergo expansion have a distinct association with epigenetic features (Essebier et al., 2016). Our findings reinforce the importance of epigenetic changes in establishing the disease state caused by triplet repeat expansions. It would be interesting to assess whether siRNA-mediated epigenetic silencing is of significance in triplet expansion diseases such as FRDA in the human system. Future studies should explore additional components of this pathway involving chromatin modifications that result from trinucleotide repeat expansions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plants

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and four tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.06.044.

ACKNOWLEDGMENTS

We thank David Powell, Shilpi Singh, and Kirill Tsyganov for suggestions on the bioinformatic analysis and Traude Beilharz, Richard Burke, Craig Dent, David Smyth, and Marco Todesco for discussions and comments on this manuscript. This work is supported by an NHMRC project grant (APP1004112 to S.B.), Australian Research Council Discovery Projects (DP1095325 to S.B., DP120103966 to S.S. and B.J.C., and DP150104048 to B.J.C.), an ARC Future Fellowship (FT100100377 to S.B.), and funds from the Larkins Fellowship program from Monash University (to S.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, S.B.; Methodology, H.E., S.S., B.J.C., and S.B.; Software, A.S.Y., A.S., B.J.C., S.J.F., and S.B.; Formal Analysis, S.S., A.S.Y., C.K.-T., A.S., S.J.F., and S.B.; Investigation, H.E., S.S., C.K.-T., T.T., S.F.G., and C.B.; Writing—Original Draft, H.E. and S.B.; Writing—Review & Editing, H.E., S.S., B.J.C., and S.B.; Visualization, H.E., S.S., S.J.F., and S.B.; Supervision, S.S., B.J.C., and S.B.; Project Administration, S.S. and S.B.; Funding Acquisition: B.J.C., S.S., and S.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 30, 2017
Revised: April 20, 2018
Accepted: June 22, 2018
Published: July 26, 2018

REFERENCES

Bidichandani, S.I., Ashizawa, T., and Patel, P.I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am. J. Hum. Genet. *62*, 111–121.

Bidichandani, S.I., Purandare, S.M., Taylor, E.E., Gumin, G., Machkhas, H., Harati, Y., Gibbs, R.A., Ashizawa, T., and Patel, P.I. (1999). Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. Hum. Mol. Genet. *8*, 2425–2436.

Bologna, N.G., and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. Annu. Rev. Plant Biol. *65*, 473–503.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423–1427.

Cao, X., and Jacobsen, S.E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr. Biol. *12*, 1138–1144.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nat. Rev. Genet. *6*, 351–360.

Chan, S.W., Zhang, X., Bernatavichute, Y.V., and Jacobsen, S.E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biol. *4*, e363.

Chutake, Y.K., Lam, C., Costello, W.N., Anderson, M., and Bidichandani, S.I. (2014). Epigenetic promoter silencing in Friedreich ataxia is dependent on repeat length. Ann. Neurol. *76*, 522–528.

Clark, R.M., De Biase, I., Malykhina, A.P., Al-Mahdawi, S., Pook, M., and Bidichandani, S.I. (2007). The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Hum. Genet. *120*, 633–640.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. *16*, 735–743.

Cuerda-Gil, D., and Slotkin, R.K. (2016). Non-canonical RNA-directed DNA methylation. Nat. Plants *2*, 16163.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. *139*, 5–17.

De Biase, I., Chutake, Y.K., Rindler, P.M., and Bidichandani, S.I. (2009). Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. PLoS ONE *4*, e7914.

De Biase, I., Rasmussen, A., Monticelli, A., Al-Mahdawi, S., Pook, M., Cocozza, S., and Bidichandani, S.I. (2007). Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics *90*, 1–5.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. PLoS Genet. *8*, e1003062.

Di Prospero, N.A., and Fischbeck, K.H. (2005). Therapeutics development for triplet repeat expansion diseases. Nat. Rev. Genet. *6*, 756–765.

Erwin, G.S., Grieshop, M.P., Ali, A., Qi, J., Lawlor, M., Kumar, D., Ahmad, I., McNally, A., Teider, N., Worringer, K., et al. (2017). Synthetic transcription elongation factors license transcription across repressive chromatin. Science *358*, 1617–1622.

Essebier, A., Vera Wolf, P., Cao, M.D., Carroll, B.J., Balasubramanian, S., and Bodén, M. (2016). Statistical enrichment of epigenetic states around triplet repeats that can undergo expansions. Front. Neurosci. *10*, 92.

Evans-Galea, M.V., Pébay, A., Dottori, M., Corben, L.A., Ong, S.H., Lockhart, P.J., and Delatycki, M.B. (2014). Cell and gene therapy for Friedreich ataxia: progress to date. Hum. Gene Ther. *25*, 684–693.

Fletcher, S.J., Boden, M., Mitter, N., and Carroll, B.J. (2018). SCRAM: a pipeline for fast index-free small RNA read alignment and visualization. Bioinformatics. Published online March 18, 2018. https://doi.org/10.1093/bioinformatics/bty161.

Gan, E.S., Xu, Y., and Ito, T. (2015). Dynamics of H3K27me3 methylation and demethylation in plant development. Plant Signal. Behav. *10*, e1027851.

Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. Nat. Rev. Genet. *6*, 743–755.

Groh, M., Lufino, M.M., Wade-Martins, R., and Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. *10*, e1004318.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nat. Chem. Biol. *2*, 551–558.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118–120.

Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., et al.; 1001 Genomes Consortium (2016). Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. Cell *166*, 492–505.

Kim, E., Napierala, M., and Dent, S.Y. (2011). Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. Nucleic Acids Res. *39*, 8366–8377.

Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M., and Krzyzosiak, W.J. (2007). Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. Mol. Cell *25*, 575–586.

Kumari, D., Biacsi, R.E., and Usdin, K. (2011). Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. J. Biol. Chem. *286*, 4209–4215.

Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. (2009). PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. Proc. Natl. Acad. Sci. USA *106*, 941–946.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Li, L., Shen, X., Liu, Z., Norrbom, M., Prakash, T.P., O'Reilly, D., Sharma, V.K., Damha, M.J., Watts, J.K., Rigo, F., and Corey, D.R. (2018). Activation of frataxin protein expression by antisense oligonucleotides targeting the mutant expanded repeat. Nucleic Acid Ther. *28*, 23–33.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics *25*, 1966–1967.

Martienssen, R.A. (2003). Maintenance of heterochromatin by RNA interference of tandem repeats. Nat. Genet. *35*, 213–214.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394–408.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. *11*, 786–799.

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

Nageshwaran, S., and Festenstein, R. (2015). Epigenetics and triplet-repeat neurological diseases. Front. Neurol. *6*, 262.

Ohshima, K., Montermini, L., Wells, R.D., and Pandolfo, M. (1998). Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J. Biol. Chem. *273*, 14588–14595.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613–622.

Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet. *6*, 729–742.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.

Punga, T., and Bühler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol. Med. *2*, 120–129.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature *443*, 1008–1012.

Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D., and Wells, R.D. (1999). Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol. Cell *3*, 465–475.

Sandi, C., Sandi, M., Anjomani Virmouni, S., Al-Mahdawi, S., and Pook, M.A. (2014). Epigenetic-based therapies for Friedreich ataxia. Front. Genet. *5*, 165.

Sasaki, T., Lee, T.F., Liao, W.W., Naumann, U., Liao, J.L., Eun, C., Huang, Y.Y., Fu, J.L., Chen, P.Y., Meyers, B.C., et al. (2014). Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive transsilencer locus in *Arabidopsis thaliana*. Plant J. *79*, 127–138.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature *422*, 909–913.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell *18*, 1121–1133.

Sureshkumar, S., Todesco, M., Schneeberger, K., Harilal, R., Balasubramanian, S., and Weigel, D. (2009). A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*. Science *323*, 1060–1063.

Szittya, G., Silhavy, D., Molnár, A., Havelda, Z., Lovas, A., Lakatos, L., Bánfalvi, Z., and Burgyán, J. (2003). Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. EMBO J. *22*, 633–640.

Tabib, A., Vishwanathan, S., Seleznev, A., McKeown, P.C., Downing, T., Dent, C., Sanchez-Bermejo, E., Colling, L., Spillane, C., and Balasubramanian, S. (2016). A Polynucleotide Repeat Expansion Causing Temperature-Sensitivity Persists in Wild Irish Accessions of *Arabidopsis thaliana*. Front. Plant Sci. *7*, 1311.

Tasset, C., Singh Yadav, A., Sureshkumar, S., Singh, R., van der Woude, L., Nekrasov, M., Tremethick, D., van Zanten, M., and Balasubramanian, S. (2018). POWERDRESS-mediated histone deacetylation is essential for thermomorphogenesis in *Arabidopsis thaliana*. PLoS Genet. *14*, e1007280.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672–676.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. *13*, 115–126.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. *2*, E104.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. Nucleic Acids Res. *34*, 667–675.

Yu, Z., Teng, X., and Bonini, N.M. (2011). Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a Drosophila model for myotonic dystrophy. PLoS Genet. *7*, e1001340.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

Zuo, J., Niu, Q.W., and Chua, N.H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. *24*, 265–273.

Please cite this article in press as: Eimer et al., RNA-Dependent Epigenetic Silencing Directs Transcriptional Downregulation Caused by Intronic Repeat Expansions, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.044

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-H3K36me3 antibody	AbCam	Cat # Ab9050; RRID: AB_306966
Anti-H3K27me3 antibody	Millipore	Cat # 07-449; RRID: AB_310624
Chemicals, Peptides, and Recombinant Proteins		
TRIzol Reagent	Invitrogen	Cat # 15596026
Basta Non-Selective Herbicide	Bayer Cropscience	Cat # 84442615
pDONR207	Invitrogen	Cat # 12213-013
Gateway BP Clonase II Enzyme mix	Invitrogen	Cat # 11789020
Gateway LR Clonase II Enzyme Mix	Invitrogen	Cat # 11791020
Critical Commercial Assays		
Transcriptor First Strand cDNA Synthesis Kit	Roche	Cat # 04379012001
LightCycler 480 SYBR Green I Master	Roche	Cat # 04707516001
Deposited Data		
Whole genome small RNA sequence from Bur-0 at 23°C – Replicate 1	NCBI	GSM3123722
Whole genome small RNA sequence from Bur-0 at 23°C – Replicate 2	NCBI	GSM3123723
Whole genome small RNA sequence from Col-0 at 23°C – Replicate 1	NCBI	GSM3123718
Whole genome small RNA sequence from Col-0 at 23°C – Replicate 2	NCBI	GSM3123719
Whole genome small RNA sequence from NS15 at 23°C – Replicate 1	NCBI	GSM3123726
Whole genome small RNA sequence from NS15 at 23°C – Replicate 2	NCBI	GSM3123727
Whole genome small RNA sequence from Bur-0 at 27°C – Replicate 1	NCBI	GSM3123724
Whole genome small RNA sequence from Bur-0 at 27°C – Replicate 2	NCBI	GSM3123725
Whole genome small RNA sequence from Col-0 at 27°C – Replicate 1	NCBI	GSM3123720
Whole genome small RNA sequence from Col-0 at 27°C – Replicate 2	NCBI	GSM3123721
Whole genome small RNA sequence from NS15 at 27°C – Replicate 1	NCBI	GSM3123728
Whole genome small RNA sequence from NS15 at 27°C – Replicate 2	NCBI	GSM3123729
Experimental Models: Organisms/Strains		
Arabidopsis thaliana accession Bur-0	N/A	N/A
Arabidopsis thaliana accession Col-0	N/A	N/A
Arabidopsis thaliana accession NS15 in the Bur-0 background	N/A	N/A
A. tumefaciens GV3103 and GV3101	N/A	N/A
E. coli TOP10	Invitrogen	Cat # C404010
Recombinant DNA		
35S::amiR-DCL3	N/A	N/A
35S::amiR-RDR2	N/A	N/A
35S::amiR-AGO4	N/A	N/A
35S::amiR-NRPE1	N/A	N/A
35S::amiR-HEN1 mir1	N/A	N/A
35S::amiR-HEN1 mir2	N/A	N/A
35S::amiR-NRPD1	N/A	N/A
35S::amiR-RDR2 [pSS1]	N/A	N/A
35S::amiR-RDR2 [pSS2]	N/A	N/A
35S::amiR-NRPD1 [pSS3]	N/A	N/A
35S::amiR-NRPD1 [pSS4]	N/A	N/A
35S::amiR-CLF [pSS5]	N/A	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
35S::amiR-CLF [pSS6]	N/A	N/A
35S::amiR-DRM2 [pSS7]	N/A	N/A
35S::amiR-DRM2 [pSS8]	N/A	N/A
35S::amiR-MET1 [pSS9]	N/A	N/A
35S::amiR-MET1 [pSS10]	N/A	N/A
35S::amiR-CMT3 [pSS11]	N/A	N/A
35S::amiR-CMT3 [pSS12]	N/A	N/A
35S::amiR-LHP1	N/A	N/A
DEX::amiR-CLF [pSS5i]	N/A	N/A
DEX::amiR-CLF [pSS6i]	N/A	N/A
Software and Algorithms		
Bowtie	N/A	N/A
SOAPAligner	N/A	N/A
Custom scripts in Python	N/A	N/A
SCRAM	N/A	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further requests for reagents and resources should be directed to and will be fulfilled by the Lead Contact, Sureshkumar Balasubramanian (mb.suresh@monash.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plants

Two *A. thaliana* accessions Col-0 (Stock No.: CS1092) and Bur-0 (Stock No.: CS1028) along with a Natural Suppressor 15 (NS15), which is in the Bur-0 background and lost the repeat expansion, were used for comparative analysis of the small RNAs. All the genotypes have been previously described (Sureshkumar et al., 2009; Tabib et al., 2016). Plants were grown at 23°C and 27°C under short day (SD, 8 hours light, 16 hours darkness) conditions for small RNA profiles.

METHOD DETAILS

Small RNA sequencing

For small RNA sequencing Col-0, Bur-0 and NS15 plants were grown for 45 days at 23°C or 27°C in short days and the total shoot tissue was harvested 5 hours after the beginning of the day (light exposure). Total RNA was extracted using TRIzol reagent (Ambion). Small RNA sequencing consisting of two biological replicates for each of the genotypes was performed by BGI-Shenzhen on the Illumina HiSeq 2000 platform. Quality controlled reads devoid of adaptor, low quality tags, contaminants and shorter reads (< 18 bp) were mapped to the TAIR10 *Arabidopsis* genome using SOAPAligner (Li et al., 2009), Bowtie (Langmead and Salzberg, 2012) or SCRAM (Fletcher et al., 2018) and reads that aligned perfectly to *IIL1* were extracted. This analysis does not typically result in the identification of the siRNAs that are perfectly matching the repeats, as they do not necessarily map to the *IIL1* repeat. Parsing through the collapsed FASTA (clean.txt) file of each of the siRNA libraries independently identified these siRNAs. Counts of these small RNAs were normalized by library size. To assess whether small RNAs against *IIL1* was indeed reduced in transgenic lines, we sequenced small RNAs from pooled transgenic lines along with Bur-0 controls grown at 27°C for 45 days. Small RNA sequencing for two biological replicates were carried out at the Monash Medical Genomics Facility and the sequences were analyzed in house using SCRAM by aligning the reads against the *IIL1*-Bur-0 locus.

Constructs and transgenics

The online tool WMD3 (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi) was used to design artificial miRNAs against *DCL3, RDR2, NRPD1A, HEN1, AGO4, NRPE1, DRM2, MET1, CMT3, LHP1, CLF*. The primers A and B (Schwab et al., 2006) were modified to contain attB sites (Table S4). In a series of overlapping PCR reactions gene specific amiRNA precursors were amplified by site-directed mutagenesis using the plasmid pRS300 as a backbone, as previously described (Schwab et al., 2006). The newly introduced attB sites facilitated cloning of the isolated PCR fragments into pDONR207 (Invitrogen) by conducting BP reactions according to the manufacturer's protocol (Gateway® BP Clonase® II Enzyme mix, Invitrogen). For *RDR2, NRPD1A, DRM2, CLF, MET1* and *CMT3* the

amiRNA constructs were commercially synthesized in plasmids (IDT, USA) and then sub-cloned into an entry vector. All artificial miRNA precursors were sub cloned into the plasmid pFK210 by conducting LR reactions according to the manufacturer's protocol (Gateway® LR Clonase® II Enzyme mix, Invitrogen). To facilitate inducible expression at a desired developmental stage, we modified the XVE-system (Zuo et al., 2000) for Gateway-compatibility and generated the vector pGREAT. We amplified the gateway cassette from pFK210 and the PCR fragment was digested with KpnI and BamHI and ligated into the pINCHIE-CLAW vector. The 35S/XVE-pea rcbS E9 terminator fragment was excised from pINCHIE grabs vector with SpeI and AscI and ligated downstream of the OlexA/TATA-attR1-Cm^r-ccdB-attr2-pea 3A terminator fragment in pINCHIE claw. The OlexA and XVE fragments were excised together with NotI and ligated into pMLBART. The resulting vector was named as pGREAT. Two artificial microRNAs (pSS5 & pSS6), were cloned in pGREAT. All sequence verified constructs were transformed into *A. tumefaciens* GV3103 cells for plant transformation.

Arabidopsis Col-0, Bur-0, and NS15 plants were transformed using the floral dip method (Clough and Bent, 1998). Primary transformants were grown at 27°C SD and continuously watered with nutrient water containing 120 mg/L BASTA (glufosinate ammonium, BAYER) for selection. For the inducible system, plants were sprayed with 50μM Estradiol with 0.1% Silwet L-77, every two days after the emergence of the 7th leaf. We obtained multiple independent primary transgenic lines and most of the transgenic lines displayed phenotypic suppression. The number of primary independent transformants that displayed complete phenotypic suppression for each of the transgenes is listed below (*35S::amiR-DCL3 – 10; 35S::amiR-RDR2 – 41; 35S::amiR-NRPD1 - 39; 35S::amiR-HEN1 –22; 35S::amiR-AGO4 – 10; 35S::amiR-NRPE1 – 12; 35S::amiR-DRM2 - 41; 35S::amiR-CMT3 - 27; 35S::amiR-MET1 - 16; 35S::XVE::amiR-CLF - 21).* Subsequent analysis with second and third generation showed 65% of plants harboring the *35S:amiR-DCL3* display phenotypic suppression (Table S2). The presence of each transgene was confirmed by PCR using primers listed in Table S4 and expression levels of each artificial miRNA's target gene were determined by qRT-PCR. Expanded GAA/TTC repeat tracts at the *IIL1* locus were confirmed by PCR (primer pair: oSKB_608 and oSKB_609, Table S4). After phenotypic analysis, primary transformants were transferred to 23°C LD conditions for seed collection.

Expression analysis

For qRT-PCR, total RNA was extracted using TRIzol reagent (Ambion). cDNA was synthesized with Anchored-oligo $(dT)_{18}$ primers and the Transcriptor First Strand cDNA Synthesis Kit (Roche). After synthesis, cDNA was diluted five fold with water and 2 μL of the diluted cDNA served as template for each qRT-PCR. Each amplification reaction was set up as a 20 μL reaction containing each primer at 250 nM and 1x SYBR Green (Roche). Real-time PCR was performed with a thermal profile of 95°C for 10 min and 40 amplification cycles of 95°C, 10min; 60°C, 30sec and 72°C, 45sec. *AT1G13320* (primer pair: oSKB_3019 and oSKB_3020) and *AT5G15710* (primer pair: oSKB_3023 and oSKB_3024) were used as endogenous standards to quantify the target gene expression levels (Czechowski et al., 2005).

ChIP Assays

Arabidopsis Col-0, Bur-0, NS15, and two Bur-0 *35S:amiR-DCL3* second generation transgenic lines were continuously grown for 48 days at 27°C SD. Chromatin crosslinking and chromatin preparation was conducted as previously described (Tasset et al., 2018). Input DNA and purified DNA were used to determine enrichment by PCR. PCR was performed with a thermal profile of 94°C for 2 minutes and 35 amplification cycles of 94°C (20 s), 56°C (30 s), 72°C (45 s) followed by a final incubation step of 10 minutes at 72°C. To analyze enrichment, we systematically analyzed the *IIL1* locus with a set of 4 different primer pairs (Table S4) to determine a region appropriate for the analysis and subsequently this region was analyzed in all samples. To account for differences in the immuno precipitation efficiency across biological replicates, the data is presented after normalizing against a positive control. For H3K36 trimethylation, *ACTIN 7* (*AT5G09810*) was used as the positive control and for H3K27 trimethylation, *AGAMOUS* was used as a positive control and the enrichment was calculated relative to the positive controls. The data across the biological replicates is averaged and presented.

QUANTIFICATION AND STATISTICAL ANALYSIS

Variation in gene expression was analyzed through quantitative real-time PCR analysis using the $2^{-\Delta\Delta cT}$ method (Pfaffl, 2001). The statistical significance of the difference in gene expression between specific samples is analyzed through Student's t test or ANOVA. For ChIP experiments, the data is expressed after normalizing with a positive control and the statistical significance was analyzed through a Student's t test. The error bars represent standard errors of mean.

DATA AND SOFTWARE AVAILABILITY

The accession number for the small RNA data reported in this paper is NCBI SRA: GSE113923.

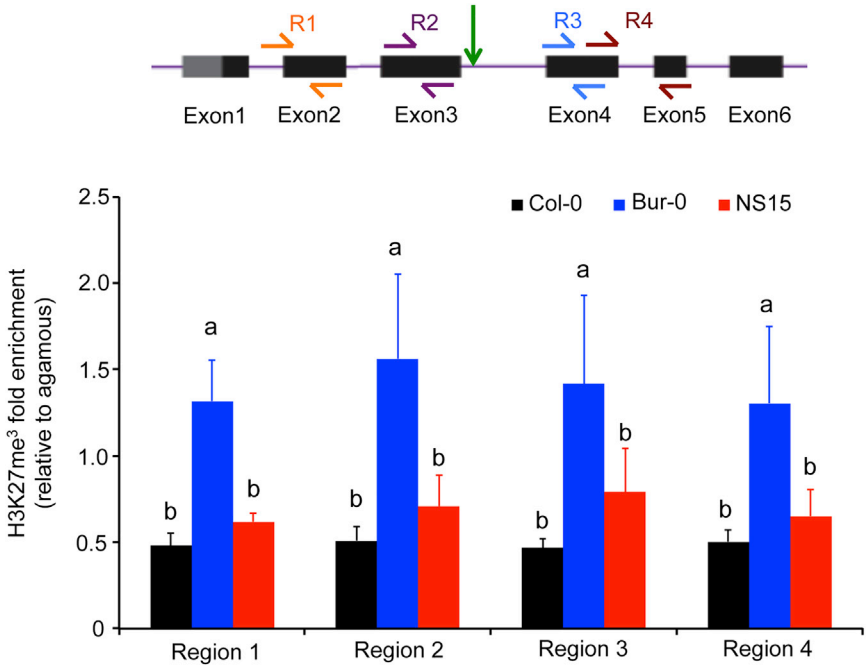


Figure S1. *IIL1* Locus in Bur-0 Is Epigenetically Silenced, Related to Figure 1

Accumulation of the H3K27me^3 marker in Bur-0 compared to Col-0 and NS15. Data are from the averages of four biological replicates. The first 6 exons of the *IIL1* locus are shown above with the transcriptional start site and the four different regions that were assessed for H3K27me^3 enrichment are shown above. Green arrow points to the position of the repeat expansion. The letters above the bar represent statistical grouping defined through an ANOVA followed by Tukey's Post hoc tests. Samples that are significantly different are shown with different letters. Error bars represent SEM.

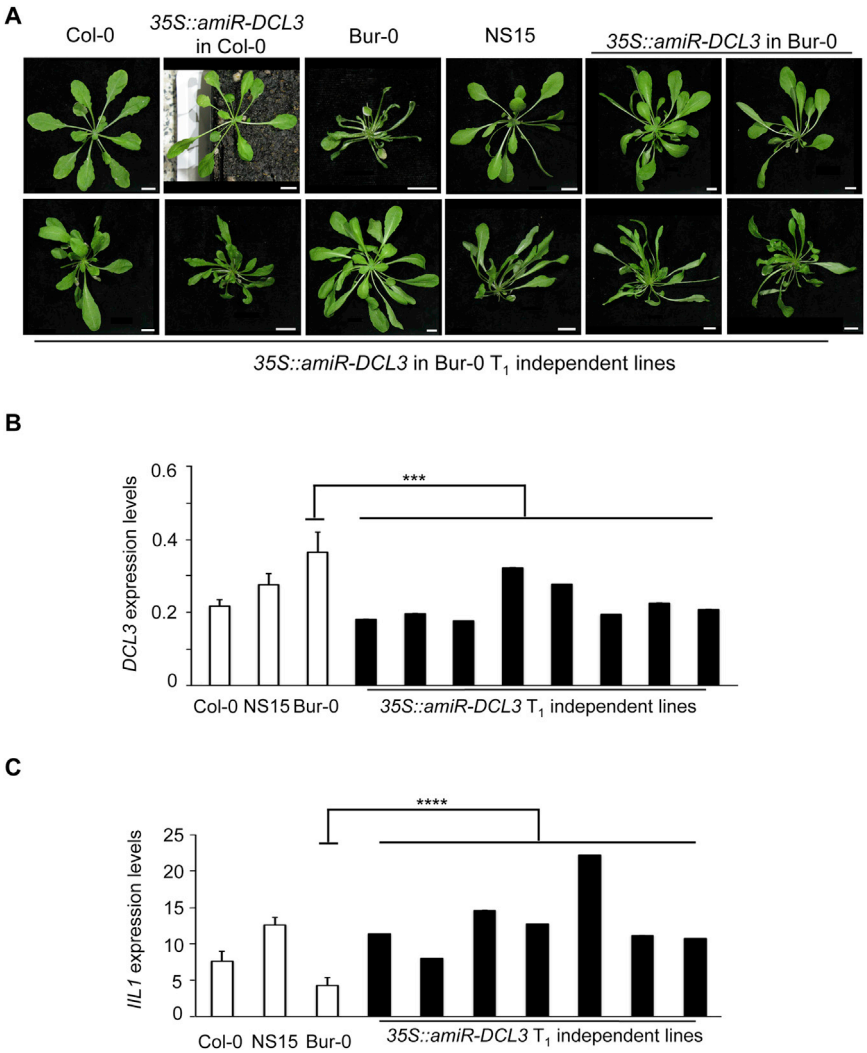


Figure S2. Triplet Expansion-Associated *iil* Phenotype Is Suppressed with an Increase in *IIL1* Expression in *DCL3* Knockdown Transgenic Lines, Related to Figure 3

(A) Independent primary transgenic lines displaying the phenotypic suppression of the *iil* phenotypes are shown. Scale bar = 1cm

(B) *IIL1* expression levels in independent *35S::amiR-DCL3* primary transgenic plants compared to Bur-0, NS15 and Col-0.

(C) Expression levels of *DCL3* in the same *35S::amiR-DCL3* transgenic lines. The data for transgenic lines is from individual plants. *p-value* from one-way ANOVA is shown above the comparisons. *** < 0.001, **** < 0.0001. Error bars represent SEM.

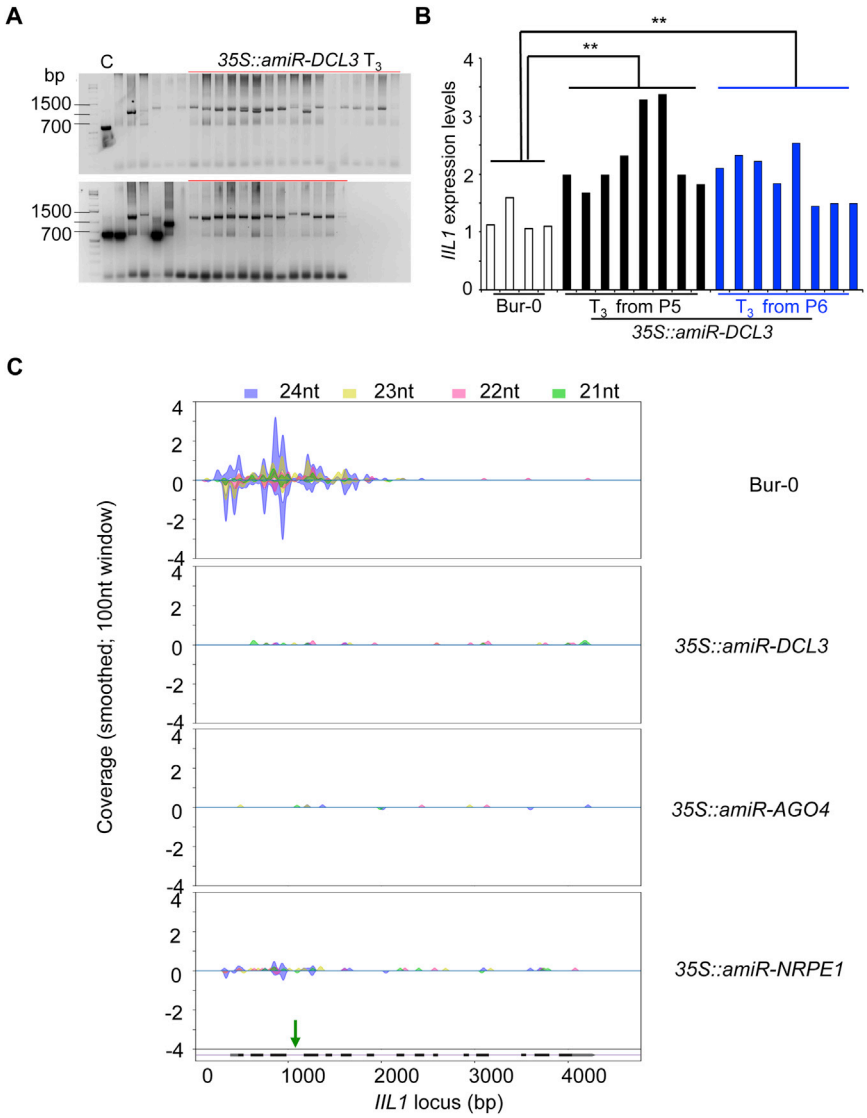


Figure S3. Plants that Harbor an Artificial MicroRNA to Knock Down *DCL3* Still Retain the Repeat Expansion and Have Elevated *IIL1* Expression with a Reduction in Small RNAs that Map to *IIL1*, Related to Figure 3

(A) Agarose gel electrophoresis of the GAA/TTC repeat expansion at the *IIL1* locus in different genotypes analyzed through PCR. *35S::amiR-DCL3* third generation plants contain the expanded repeat at the *IIL1* locus. The plants were grown for 2-months at 27 °C short days (top) or 23 °C short days (bottom).

(B) *IIL1* expression in individual third generation plants derived from two second generation parental lines are shown along with the expression in Bur-0 controls. Plants were grown for 4 weeks at 27° C in short days. The data is for individual plants. *p-values* from one-way ANOVA are shown above the comparisons. ** < 0.01

(C) siRNAs that map to *IIL1* locus are mostly absent in transgenic lines in which *DCL3, AGO4 or NRPE1 (Pol V)* is downregulated. Mapping of small RNAs at the *IIL1* locus using Small Complementary RNA Mapper (SCRAM2). The genic region is shown at the bottom with the black boxes representing exons and lines as introns. SCRAM2 output showing the normalized read counts along with standard error (shown as shadows) across the *IIL1* locus for the 21, 22, 23 and 24nt small RNAs is shown. Only the small RNAs that are non-repeat associated are shown in the figure. Quantification of the sense and antisense small RNAs is shown in the positive and negative dimensions, respectively, along the y axis.

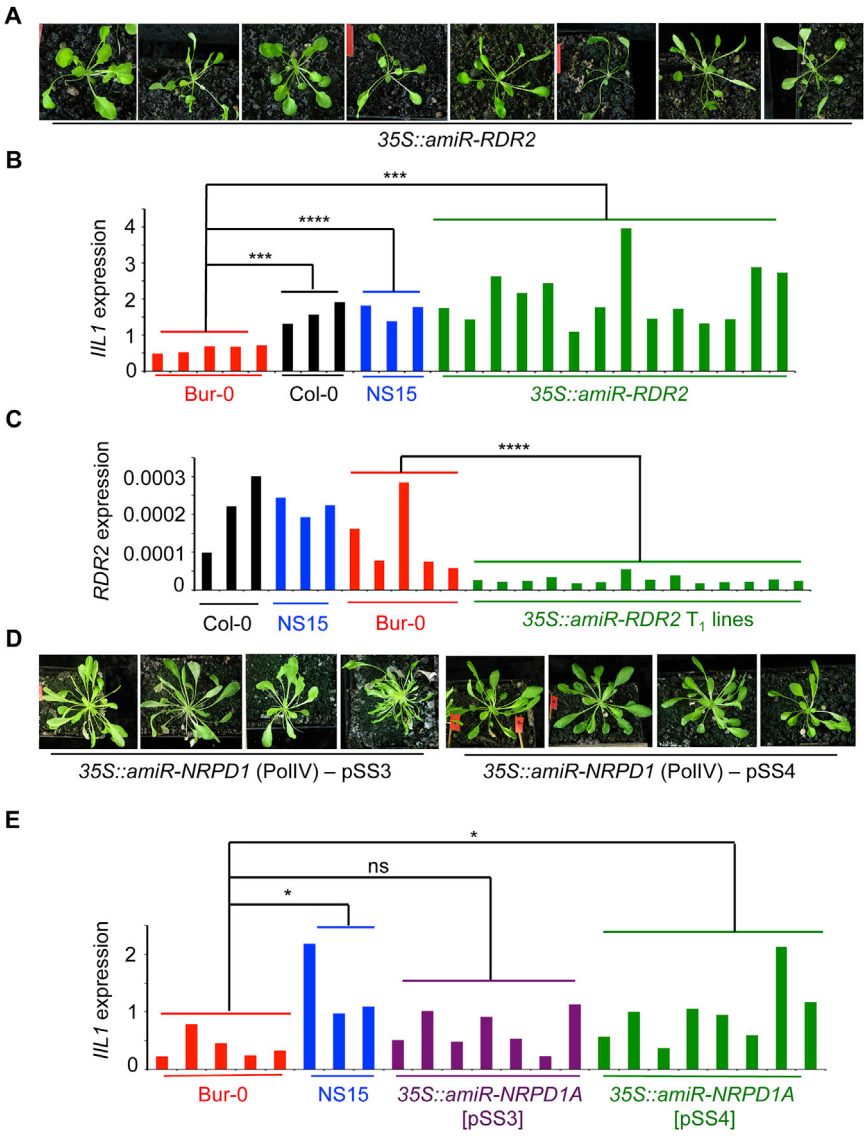


Figure S4. Knocking Down *RDR2* and *NRPD1* Suppresses the *iil* Phenotype and Restores *IIL1* Expression, Related to Figure 4

(A) Few independent primary transgenic lines that show the varying phenotypes are shown above.

(B) *IIL1* expression levels in T1 transgenic lines for *35S::amiR-RDR2.* The data is for individual plants and thus lack standard deviation.

(C) *35S::amiR-RDR2* results in knocking down *RDR2* expression.

(D) Phenotypes seen in independent T1 lines for *35S::amiR-NRPD1* with two distinct artificial microRNAs. pSS4 shows a stronger suppression than pSS3.

(E) Increase in *IIL1* expression in *35S::amiR-NRPD1* primary transgenic lines. The data is for individual plants. *p-values* from one-way ANOVA are shown above the comparisons. .* < 0.05, *** < 0.001, **** < 0.0001 ns-not significant.

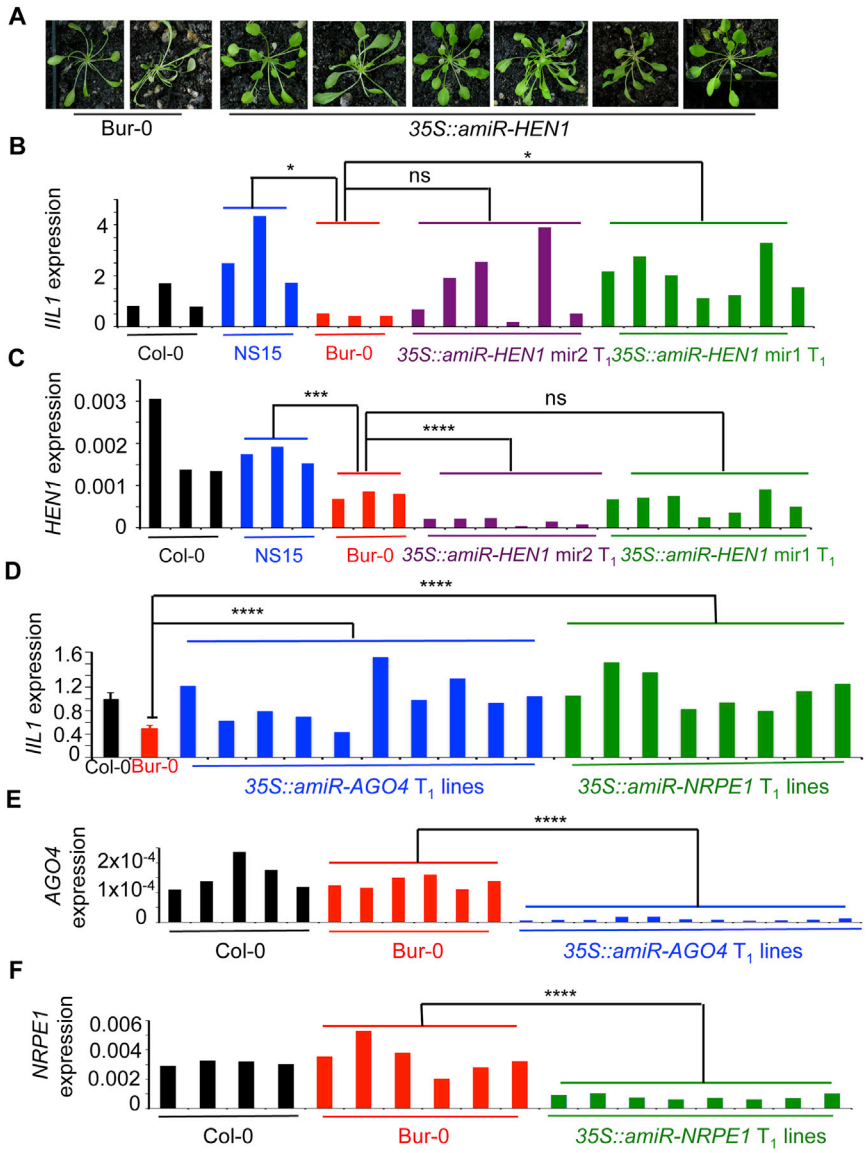


Figure S5. Knocking Down *HEN1*, *AGO4*, and *NRPE1* Is Associated with an Increase in *IIL1* Expression Coupled with a Suppression of the *iil* Phenotype, Related to Figure 4

(A) Few independent primary transgenic lines that show the varying phenotypes are shown above.

(B) *IIL1* expression levels in T1 transgenic lines for *35S::amiR-RDR2*.

(C) Downregulation of *HEN1* expression in *35S::amiR-HEN1* knockdown lines.

(D) Expression levels of *IIL1* in independent primary transgenic lines for *35S::amiR-AGO4* or *35S::amiR-NRPE1* are shown.

(E) *AGO4* expression in multiple independent *35S::amiR-AGO4* primary transgenic lines.

(F) *NRPE1* expression in multiple independent *35S::amiR-NRPE1* primary transgenic lines. The data is for individual plants. *p-value*s from one-way ANOVA are shown above the comparisons. * < 0.05, *** < 0.001, **** < 0.0001, ns- not significant.

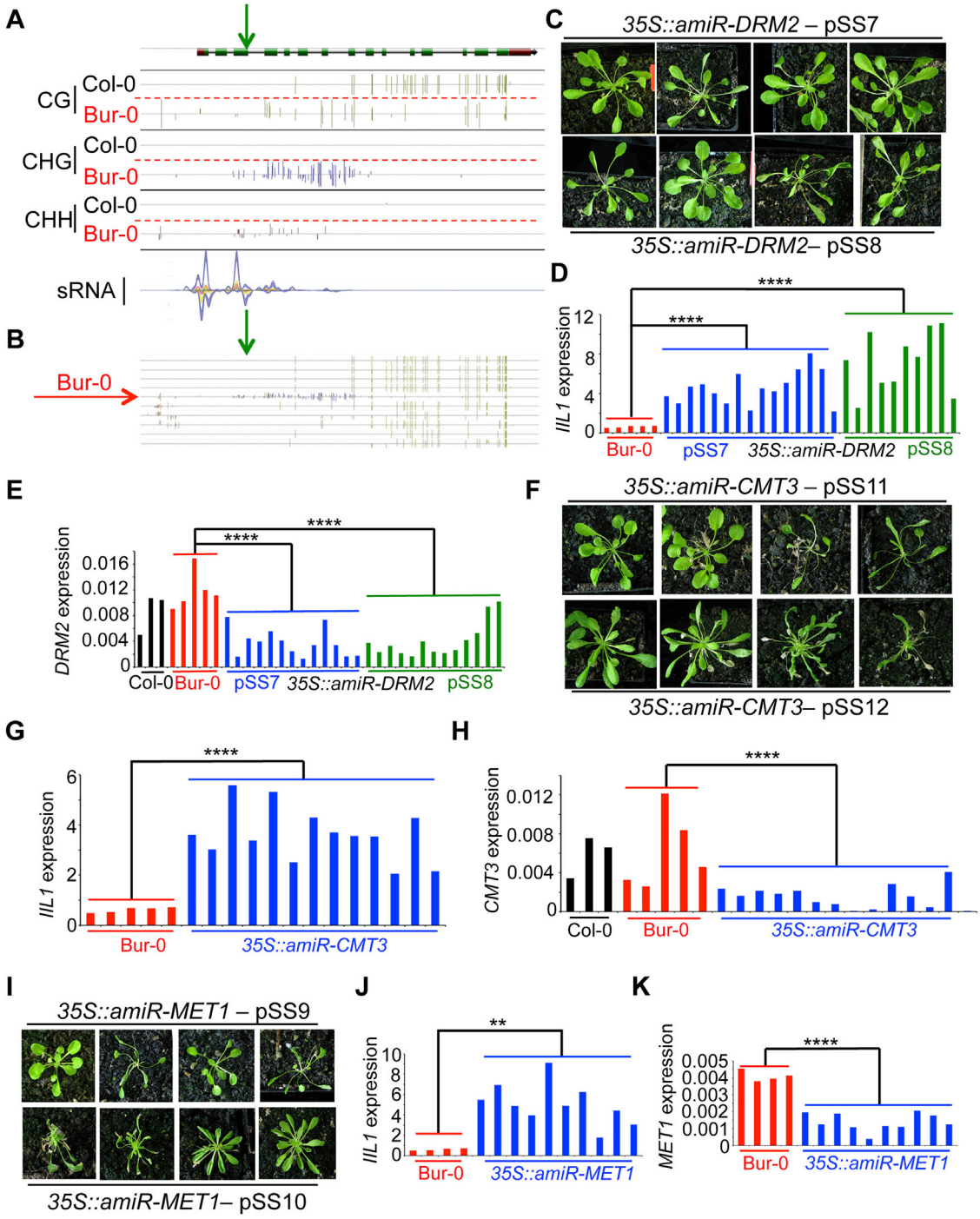


Figure S6. *IIL1*-Bur-0 Harbors a Distinct Methylation Profile and Knocking Down DNA Methyl Transferases Suppresses the *iil* Phenotype, Related to Figure 4

(A) Comparison of the methylation profile of *IIL1*-Bur-0 and *IIL1*- Col-0 in CG, CHG and CHH sequence contexts. The data are from the 1001 epigenomes dataset (Kawakatsu et al., 2016). For comparison small RNA profile from of Bur-0 from Figure 2 is included here as well.

(B) Comparison of the methylation of profiles at *IIL1* locus across a random set of strains including Bur-0. The entire 1001 epigenome reads at *IIL1* locus was visually inspected and a random representative screenshot is shown above. The green arrow indicates the position of the intronic repeat expansion in *IIL1*-Bur-0. Please note the epigenome data from Kawakatsu et al. is from plants grown at different temperatures in long days (10°C, 16°C and 23°C for various lines), while the smallRNA profile shown is for Bur-0 plants grown at 27°C in short days.

(C) Phenotypes seen in independent primary transgenic lines for *35S::amiR-DRM2* with two distinct artificial microRNAs.

(D) Increase in *IIL1* expression in *35S::amiR-DRM2* primary transgenic lines.

(legend continued on next page)

(E) *DRM2* expression in multiple independent *35S::amiR-DRM2* primary transgenic lines.
(F) Phenotypes seen in independent primary transgenic lines for *35S::amiR-CMT3* with two distinct artificial microRNAs.
(G) Increase in *IIL1* expression in *35S::amiR-CMT3* primary transgenic lines.
(H) *CMT3* expression in multiple independent *35S::amiR-CMT3* primary transgenic lines.
(I) Phenotypes seen in independent primary transgenic lines for *35S::amiR-MET1* with two distinct artificial microRNAs.
(J) Increase in *IIL1* expression in *35S::amiR-MET1* primary transgenic lines.
(K) *MET1* expression in multiple independent *35S::amiR-MET1* primary transgenic lines. The data is for individual plants. *p-value*s from one-way ANOVA are shown above the comparisons. ** < 0.01, **** < 0.0001.

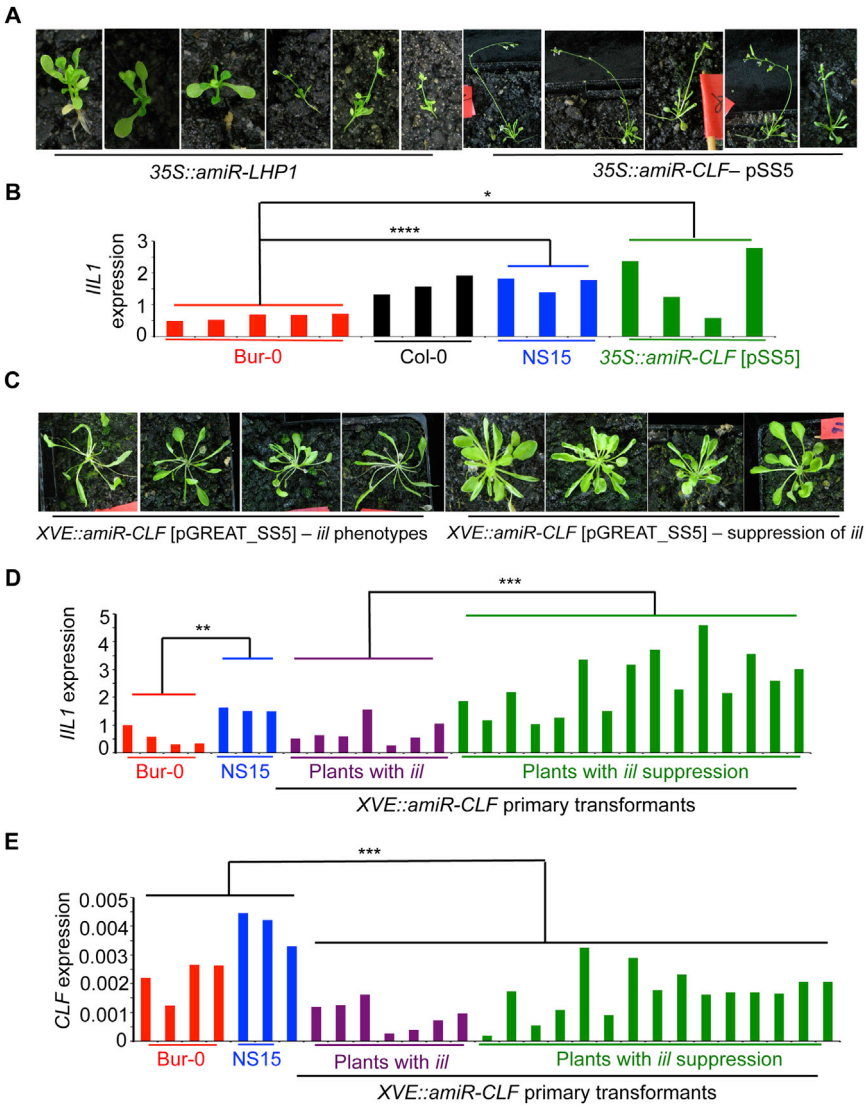


Figure S7. Knocking Down *LHP1* and *CLF* Results in Severe Developmental Phenotypes and Inducible Knockdown of *CLF* Suppresses the *iil* Phenotype, Related to Figure 4

(A) Phenotypes seen in independent primary transgenic lines for *35S::amiR-LHP1* and *35S::amiR-CLF* artificial microRNAs.

(B) Increase in *IIL1* expression in *35S::amiR-CLF* primary transgenic lines. The data is for individual plants.

(C) Phenotypes seen in independent primary transgenic lines for *DEX::amiR-CLF*.

(D) Increase in *IIL1* expression in *DEX::amiR-CLF* primary transgenic lines.

(E) Comparison of *CLF* expression between non-transgenic plants (Bur-0 & NS15) and transgenic plants that displayed *clf* phenotype (*XVE::amiR-CLF*). The variability in the data is due to both differences in developmental heterogeneity as well as variation in the estradiol-induction efficiency among transgenic lines. *p-value*s from one-way ANOVA are shown above the comparisons.* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001.