Enhanced RNAi stability through imperfect inverted repeats: nucleotide mismatches prevent intrinsic self-silencing of hpRNA transgenes in plants

Daai Zhang
Commonwealth Scientific and Industrial Research Organisation
Chengcheng Zhong
Commonwealth Scientific and Industrial Research Organisation
Neil Smith
CSIRO Plant Industry
Rob Defeyter
Commonwealth Scientific and Industrial Research Organisation
Ian Greaves
Commonwealth Scientific and Industrial Research Organisation
Steve Swain
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0002-6118-745X
Ren Zhang
University of Wollongong
Ming-Bo Wang (ming-bo.wang@csiro.au)
Commonwealth Scientific and Industrial Research Organisation

Article

Keywords: Hairpin RNA, biogenesis, plants, molecular biology

Posted Date: June 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-598237/v1

Version of Record: A version of this preprint was published at Nature Communications on July 7th, 2022.
See the published version at https://doi.org/10.1038/s41467-022-31641-5.

Abstract

Hairpin RNA (hpRNA) transgenes, with a perfect inverted-repeat (IR) DNA, have been the most successful RNA interference (RNAi) method in plants. Here we show that hpRNA transgenes were invariably methylated in the IR DNA and the adjacent promoter, causing transcriptional self-silencing and preventing the full potential of RNAi. Nucleotide substitutions in the sense sequence, which disrupts the perfect IR DNA structure, were sufficient to prevent the intrinsic DNA methylation resulting in more uniform and persistent RNAi. Substituting all cytosine (C) with thymine (T) nucleotides, in a G:U hpRNA design, prevented DNA methylation and self-silencing but still allowed for the formation of perfect hpRNA due to G:U wobble base-pairing. The G:U design induces effective RNAi in 90–96% of transgenic lines, compared to 57–65% for the traditional hpRNA design. Furthermore, while a traditional hpRNA transgene showed increasing DNA methylation and self-silencing from cotyledons to true leaves, the G:U transgenes avoided this developmental progression of self-silencing and induced RNAi throughout plant growth. The G:U and traditional hpRNA transgenes generated small interfering RNA (siRNA) with different 5’ phosphorylation, which resembled the endogenous tasiRNA and miRNA, respectively. Furthermore, our results suggest that siRNAs from the two transgene designs function differently to induce target DNA methylation, one (from traditional hpRNA) through the canonical RdDM pathway and the other (G:U hpRNA) a non-canonical pathway. Our study not only revealed a methylation-resistant RNAi transgene design but also provided new mechanistic insights into small RNA biogenesis and function in plants.

Introduction

RNA silencing is an evolutionarily conserved gene silencing mechanism in eukaryotes, where long dsRNA is processed by Dicer or Dicer-like (DCL) proteins into 20-30 nucleotide (nt) small RNA (sRNA) that induces RNA degradation via sequence complementarity[1-3]. In plants, multiple RNA silencing pathways exist, including microRNA (miRNA), trans-acting small interfering RNA (tasiRNA), repeat-associated siRNA (rasiRNA) and exogenic (virus and transgene) siRNA (exosiRNA) pathways[4]. miRNAs are 20-24 nt sRNAs processed in the nucleus by DCL1 from short self-folding RNAs transcribed from *MIR* genes[2]. tasiRNAs are 21 nt secondary siRNAs derived from DCL4 processing of dsRNA synthesized by RNA-dependent RNA polymerase 6 (RDR6) from miRNA-cleaved *TAS* RNA fragment[4]. The 24-nt rasiRNAs are generated from repetitive DNA in the genome by the combined function of DNA-dependent RNA polymerase IV (Pol IV), RDR2 and DCL3[5]. The exosiRNA pathway overlaps with the tasiRNA and rasiRNA pathways and both DCL4 and DCL3 are involved in exosiRNA processing. In addition to DCL1, DCL3 and DCL4, plant genomes encode DCL2 or equivalent, which generates 22-nt siRNAs including 22 nt exosiRNAs, and plays a key role in systemic and transitive gene silencing in plants[6]. All of these plant sRNAs are methylated at the 2’-hydroxyl group of the 3’ terminal nucleotide by HUA Enhancer 1 (HEN1), which is thought to stabilize the sRNAs[7]. miRNAs, tasiRNAs and exosiRNAs are functionally similar to sRNAs in animals, and involved in post-transcriptional RNA degradation. The rasiRNAs, however, are unique to plants and function to direct *de novo* cytosine methylation at the cognate DNA, a transcriptional gene silencing mechanism known as RNA-directed DNA methylation (RdDM)[5]. The post-transcriptional RNA silencing

mechanism has been extensively exploited as a gene knockdown technology in various eukaryotic systems, generally referred to as gene silencing or RNA interference (RNAi) technologies. In plants, the different RNA silencing pathways have led to different technical approaches, such as artificial miRNA, artificial tasiRNA and virus-induced gene silencing technologies[4]. However, long hpRNA transgenes, designed to express long hairpin-structured dsRNA, are the most widely used RNAi technology in plants, and a variety of successful applications of this technology have been demonstrated in plant biotechnology[4]. It can be anticipated that this RNAi approach will continue to be a powerful tool in many areas of crop improvements such as host-induced RNAi against pests and pathogens and metabolic engineering of novel traits through spatial and temporal gene knockdown, which is difficult to achieve using gene knockout technologies such as the CRISPR/Cas9 approach.

An hpRNA construct typically consists of a perfect inverted repeat (IR) of a target gene sequence (forming the dsRNA stem of hpRNA) separated by a spacer sequence (forming the loop). Tandem DNA repeats, particularly the IR DNA structures, are widely observed to attract strong DNA methylation causing transcriptional silencing[8,9]. Beside the IR structure, siRNAs derived from hpRNA transgenes can potentially direct DNA methylation to their own sequence via the RdDM pathway[10,11]. hpRNA transgenes therefore differ from normal transgenes and are potentially subject to self-induced transcriptional silencing. Indeed, a previous study showed that hpRNA transgene-induced RNAi in Arabidopsis was enhanced in an RdDM mutant, and that this enhanced RNAi effect correlated with reduced DNA methylation spanning from the IR DNA to the upstream promoter sequence[12]. An RNAi design that can prevent self-induced silencing would therefore be desirable for achieving durable and potent RNAi in plants.

In this study we investigated the effect of nucleotide mismatches in hpRNA-induced RNAi in plants. Introducing nucleotide mismatches to disrupt IR DNA structure resulted in uniform and persistent RNAi against a reporter gene and two endogenous genes. We discovered that the traditional hpRNA transgenes with a perfect IR structure are generally prone to self-induced methylation and transcriptional silencing (referred to as self-silencing hereafter) causing large variability in RNAi efficacy, whereas the enhanced RNAi effect of mismatched hpRNA constructs was due to the prevention of methylation in both the IR and the promoters preventing self-silencing. Additionally, we generated evidence that the IR-associated DNA methylation and self-silencing is independent of the RdDM pathway, and that siRNAs from G:U hpRNA transgenes are processed and function differently from traditional hpRNA transgenes, providing novel insights into IR-induced gene silencing and siRNA biogenesis in plants.

Results

Evenly mismatched and G:U basepaired hpRNA constructs induce uniform RNAi

We first tested three mismatched constructs in *Nicotiana tabacum* using the β-glucuronidase (*GUS*) reporter gene as the RNAi target (Figure 1A). These constructs contained the same 200 bp antisense wild-type (WT) *GUS* sequence as the traditional hpRNA construct (hpGUS[WT]) to ensure perfect sequence

complementarity between antisense siRNAs and target *GUS* mRNA. The mismatched construct hpGUS[1:4] had one nucleotide substitution in every 4 nucleotides of the 200 bp sense sequence; hpGUS[2:10] contained 2 consecutive nucleotide substitutions in every 10 nucleotides; and hpGUS[G:U] had all 52 cytosine (C) nucleotides changed to thymine (T) nucleotides (Figure 1A, S1). The C to T changes in hpGUS[G:U] disrupted the perfect IR DNA structure but did not prevent the formation of perfect hpRNA due to G:U wobble base-pairing.

The hpGUS[WT] transgenic population showed a wide range of RNAi efficiency, with 35 of the 59 independent lines analysed (59.3%) showing strong RNAi (GUS activity ≤10% of the untransformed GUS plants), 9 showing weak RNAi (GUS activity 10-30% of the untransformed), and 15 almost no silencing (Table 1 and Figure 1B), which was typical for traditional hpRNA constructs[13]. The hpGUS[2:10] construct behaved more like hpGUS[WT], inducing strong GUS RNAi in some lines (28 of 41, or 68.3%) but giving almost no GUS RNAi in the remaining 13 plants.

In clear contrast, 71 of the 74 hpGUS[G:U] lines (95.9%) tested showed strong RNAi. This uniform RNAi was not due to a uniform transgene insertion pattern across the independent lines: 16 randomly selected *GUS*-silenced lines showed a wide range of transgene copy numbers (Figure S2). All 33 hpGUS[1:4] lines showed *GUS* RNAi, although only 10 (30.3%) showed strong RNAi. Thus, relatively evenly distributed nucleotide substitutions, as in hpGUS[G:U] and hpGUS[1:4], allowed for uniform RNAi across transgenic populations. The relatively weak RNAi effect by hpGUS[1:4] coincided with its dsRNA stem having the lowest predicted thermodynamic stability (Figure 1C). Consistently, there appeared to be a good correlation between the extent of *GUS* RNAi and the predicted dsRNA stability of the four hpRNAs (Figure 1C).

As the G:U hpRNA construct induced strong and uniform RNAi against *GUS*, we tested this design against two endogenous genes in Arabidopsis, the ethylene insensitive 2 (*EIN2*) and phytoene desaturase (*PDS*) genes, silencing of which can be scored based on hypocotyl length of dark-germinated seedlings on 1-aminocyclopropane-1-carboxylic acid medium[14] and photo-bleaching[15], respectively. The hpRNA constructs (Figures 2A and 3A) were designed to target a 200 bp *EIN2* and 450 bp *PDS* mRNA sequences that contained 43 and 82 cytosines, creating 21.5% and 18.2% G:U wobble base pairs for hpEIN2[G:U] and hpPDS[G:U], respectively (Figure S3).

Analysis of 20 randomly selected hpEIN2 lines (100-200 T2 progeny for each line) showed that the hpEIN2[WT] lines had a high range of *EIN2* RNAi levels, with 7 lines (# 2, 5, 9, 10, 12, 14, 16) showing low levels of RNAi (short hypocotyl length), and the other 13 lines (65%) having moderate to strong *EIN2* RNAi (Figures 2B, C). In contrast, 18 of the 20 hpEIN2[G:U] lines (90%) displayed relatively uniform and strong *EIN2* RNAi. Individual siblings within each of the 18 lines also appeared to show less variation in hypocotyl length than those of the hpEIN2[WT] lines (Figures 2B, C), suggesting a greater sibling uniformity of RNAi.

Like the hpGUS[G:U] lines, the uniform *EIN2* RNAi in the hpEIN2[G:U] lines was in general not dependent on the number of transgene insertion (as judged by the kanamycin resistant:sensitive ratio of the T2 progeny plants). The 18 hpEIN2[G:U] lines with strong RNAi had a range of insertion numbers, including high-copy number insertions (5:1 to 86:1 ratios), with only two very high-copy number lines (256:1 and 81:1) showing low levels of RNAi (Figure 2B, C). Similar to a previous study[14], the hpEIN2[WT] lines with high-copy number insertions (with 8:1, 14:1, 30:1, 51:1 ratios) tended to show low levels of RNAi.

For the *PDS* target gene, we identified and analysed 100 hpPDS[WT] and 172 hpPDS[G:U] primary transformants, all showing strong photo-bleaching in the cotyledons at the young seedlings stage (Figure 3B; 7 days). Thus, both constructs induced effective *PDS* RNAi in cotyledons. However, the two populations showed a clear difference when true leaves emerged (14 days and beyond), with a much larger number of hpPDS[WT] plants giving green leaves that indicated a loss of strong RNAi (Figure 3B). As summarised in Figure 3C, the hpPDS[G:U] population contained much higher proportions of the strongly and moderately silenced lines (63% and 30% respectively) than the hpPDS[WT] population (34% and 23%). In addition, most of the weakly silenced hpPDS[G:U] plants still showed mild mottling on true leaves, in contrast to the weakly silenced hpPDS[WT] plants that mostly had fully green leaves.

The *GUS*, *EIN2* and *PDS* RNAi results collectively confirmed that the G:U hpRNA construct induces more uniform RNAi than the traditional hpRNA construct. Importantly, the *PDS* RNAi result indicated a developmental stage variability of RNAi by the traditional hpRNA transgene, being more effective in cotyledons than leaves, and suggested that the G:U hpRNA transgenes are developmentally more stable.

DNA-mismatched hpRNA transgenes show diminished promoter methylation

The enhanced uniformity of RNAi by G:U and 1:4 mismatched hpRNA transgenes suggested a reduced level of self-silencing or an increased level of transcriptional stability of these transgenes compared to the traditional hpRNA transgenes. To investigate this, we analysed DNA methylation in the hpGUS, hpEIN2 and hpPDS transgenes, using methylation-dependent enzyme McrBC digestion-PCR and bisulfite sequencing.

hpGUS lines: Seven of the 10 hpGUS[WT] lines analysed (Figure 4A) showed a clear reduction in PCR band intensity in McrBC-digested vs undigested samples (Figure 4B), indicative of DNA methylation at the 35S-GUS junction region. These methylated lines included all the 5 lines (#2, 5, 8, 9, 10) that showed no or low levels of GUS RNAi (Figure 4A), indicating a direct link between DNA methylation and reduced RNAi efficiency. In contrast to hpGUS[WT], all the 10 hpGUS[G:U] lines showed equal or near-equal PCR amplification between McrBC-treated and untreated samples (Figure 4B), indicating no or low-levels of DNA methylation. The hpGUS[1:4] lines exhibited a generally weaker DNA methylation than the hpGUS[WT] lines, as indicated by the less frequent and extreme reduction in PCR amplification of McrBC-digested samples (Figure 4B). However 5 hpGUS[1:4] lines showed a clearly reduced PCR amplification, indicating stronger DNA methylation than the hpGUS[G:U] lines. To more accurately determine low DNA methylation levels, we performed bisulfite sequencing on two hpGUS[WT] and four hpGUS[G:U] lines that

had strong RNAi, and four hpGUS[1:4] lines with representative GUS RNAi levels (indicated by asterisks in Figure 4A). Consistent with the McrBC-digestion PCR result, all four hpGUS[G:U] lines had very low levels of DNA methylation at the 35S promoter based on bisulfite sequencing (Figure 4C). The four randomly selected hpGUS[1:4] lines all showed low to moderate levels of DNA methylation (Figure 4C), with their average promoter methylation levels correlating inversely with the extent of *GUS* RNAi (Figure S4). The two hpGUS[WT] lines, despite strong RNAi, both showed moderate levels of DNA methylation in the upstream region of the 35S promoter and high levels (60-100%) of DNA methylation near the 35S-GUS junction (Figure 4C). Thus, the hpGUS[G:U] transgene, and to a lesser degree the hpGUS[1:4] transgene, had reduced promoter methylation across the transgenic population.

hpEIN2 lines: Analysis of 12 independent hpEIN2[WT] and hpEIN2[G:U] lines each (Figure S5A) showed a clear difference in promoter methylation (Figure 5). Every hpEIN2[WT] line had some levels of DNA methylation at the 35S-EIN2 junction (Figure 5A, B). The level of DNA methylation, as judged by reduced PCR amplification of McrBC-digested DNA, showed a good inverse correlation with the extent of *EIN2* RNAi (Figure S5A). Widespread presence of promoter methylation in the hpEIN2[WT] lines was confirmed by bisulfite sequencing of three lines with the strongest *EIN2* RNAi hence least likely to be methylated (Figure S5A; #1, #11, #15). The 35S promoter region showed 20%~80% methylation for individual cytosines, with all three lines showing equally dense cytosine methylation in the sense *EIN2* sequence (Figure 5C).

The hpEIN2[G:U] lines showed clear reduction in promoter methylation, with 7 of the 12 lines showing little or no reduction in PCR amplification of McrBC-digested samples (Figure 5A; hpEIN2[G:U] #1, 2, 4, 11, 12, 13 and 14). Bisulfite sequencing analysis confirmed the low methylation levels in the three hpEIN2[G:U] lines with strong *EIN2* RNAi (Figure S5A; #1, #4, #11), all showing less than 20% methylation for all cytosines in the 35S promoter. The two hpEIN2[G:U] lines that had weak RNAi (Figure 2B; #5 and #18) showed strong DNA methylation, indicating a direct link between promoter methylation and reduced *EIN2* RNAi.

hpPDS lines: Many of the hpPDS[WT] lines showed strong *PDS* silencing in cotyledons but weak *PDS* silencing in leaves (Figure 3), suggesting increasing promoter methylation from cotyledons to leaves. The majority of the hpPDS[G:U] lines exhibited strong *PDS* silencing phenotypes in both cotyledons and true leaves, indicating low promoter methylation levels in both tissues. McrBC digestion PCR detected a clear increase in DNA methylation at the 35S promoter of the hpPDS[WT] transgene in true leaves than in cotyledons (Figure 5D, E, F), whereas promoter methylation remained similarly low in both tissues for the two hpPDS[G:U] lines analysed.

Taken together, the methylation analyses indicated that the relatively uniform RNAi of the mismatched hpRNA lines was due to diminished promoter methylation and that the traditional hpRNA transgenes are inherently prone to promoter methylation with all lines having some levels of promoter methylation. The result also suggested that promoter methylation of traditional hpRNA transgenes is developmental stage dependent.

The intrinsic methylation of traditional hpRNA transgenes is not affected in RdDM mutants

It was thought that the methylation in the IR region of a traditional hpRNA transgene is induced by hpRNA-derived siRNAs via the RdDM pathway. Consequently, it was expected that the traditional hpRNA transgenes would lose the methylation in a RdDM mutant resulting in uniform RNAi across transgenic populations. It was also expected that the traditional hpRNA transgenes would induce more effective RNAi than the G:U hpRNA transgenes in RdDM mutants due to stronger dsRNA stability. We investigated these using two Arabidopsis RdDM mutants, *nrpd1a-3* (a T-DNA insertion mutant of the upstream siRNA biogenesis factor Pol IV) and *ocp11* (a dominant-negative mutant of the downstream effector AGO4).

The traditional hpRNA constructs, targeting *PDS* or *EIN2*, indeed induced uniform RNAi in the two RdDM mutants, with over 84~100% of transgenic lines showing RNAi (Table 2). The white cotyledon-to-green leaf-type of *PDS* RNAi phenotype of the Col-0 background (Figure 3) also largely disappeared in the RdDM mutants, with most of the hpPDS[WT] plants showing relatively uniform photo-bleaching from cotyledons to leaves (Figure S6A; Figure S7A). However, to our surprise, the traditional hpRNA transgenes induced weaker RNAi than the G:U transgenes in both mutant backgrounds (Figure S6; Figure S7A). In particular, the hpPDS[G:U] construct induced extreme photo-bleaching in 100% of the transgenic lines in *ocp11*, compared with moderate to high levels of photo-bleaching in 95% of the hpPDS[WT] lines (Figure S6A; Figure S7A). The RNAi of *ocp11*/hpEIN2[G:U] lines could not be properly assayed using hypocotyl length because many lines showed poor seed germination in the dark on ACC medium (likely due to strong *EIN2* RNAi based on our previous observation). Nevertheless, under light, the T2 plants of hpEIN2[G:U] lines showed more vigorous root and foliage growth, particularly in the *ocp11* mutant (Figure S6B), indicating stronger RNAi than the hpEIN2[WT] lines. It is worth noting that in the *nrpd1a* mutant, no hpPDS[WT] lines developed strong photo-bleaching, indicating that the loss of Pol IV reduced the extent of *PDS* RNAi despite the increased uniformity of RNAi across transgenic populations.

The hpEIN2[WT] transgenes developed similarly strong DNA methylation in the IR and promoter regions in the RdDM mutants as in the wild-type Col-0 background, even in the lines with strong *EIN2* RNAi (Figure 6). In contrast, DNA methylation was largely absent in the hpEIN2[G:U] transgenes in all backgrounds (except for *nrpd1a*/hp[G:U]-13). Similarly, strong methylation in the IR region (the sense *PDS* sequence) was also detected in the hpPDS[WT] lines in the RdDM mutants (Figure S7B). Thus, strong DNA methylation inside the perfect inverted-repeat DNA, as well as its spread to the upstream promoter, was not affected in the RdDM mutants.

The *EIN2* and *PDS* genomic targets also showed DNA methylation but at a much lower level than the IR region of the hpEIN2[WT] and hpPDS[WT] transgenes, particularly at the CHG and CHH sites (Figure 6, S7B, S8). Unlike the hpRNA transgenes, target gene methylation was clearly reduced in the *nrpd1a* mutant for the hpRNA[WT] lines, indicating that Pol IV, the siRNA biogenesis component of the canonical RdDM pathway, is involved in hpRNA-induced target gene methylation. Remarkably, the G:U hpRNA transgenes also induced similar levels of non-CG methylation at the target gene loci in the WT Col-0 plants, indicative

of active RdDM, but this methylation was not reduced in the *nrpd1a* mutant (Figure 6, S7B, S8). Thus, G:U hpRNA transgenes induced RdDM independently of Pol IV.

Taken together, experiments with the RdDM mutants confirmed that DNA methylation of traditional hpRNA transgenes is intrinsic to the IR DNA structure and independent of the RdDM pathway. This intrinsic methylation prevented the traditional hpRNA transgenes from reaching their full RNAi efficacy, even in the RdDM mutants. However, the increased cross-line uniformity of *PDS* and *EIN2* RNAi in the RdDM mutants suggested that the RdDM pathway contributes to genomic position or copy number-dependent silencing of hpRNA transgenes.

siRNAs from traditional and G:U hpRNA are differently processed

One obvious question was whether G:U base-paired hpRNAs were efficiently processed by Dicer into siRNAs. Northern blot analysis detected abundant siRNAs from the hpEIN2[G:U] (Figure 7A) and hpGUS[G:U] (Figure S9A) plants. The amount of siRNAs looked more even across the independent G:U hpRNA lines than the traditional hpRNA lines, and showed good correlation with the extent of RNAi (Figure S5B, C). Thus, the uniform RNAi across independent G:U hpRNA lines could be attributed to relatively even amounts of siRNAs. The strong RNAi lines of hpEIN2[G:U] accumulated similar amounts of siRNAs to the strong RNAi lines of hpEIN2[WT] (except for hpEIN2[WT]-17), indicating that G:U hpRNA was efficiently processed by Dicer. The strong hpGUS[WT] lines, however, accumulated much higher amounts of siRNAs than the hpGUS[G:U] lines (Figure S9A). A possible explanation for the discrepancy is that *GUS* target mRNA, from a transgene, may serve as template for production of secondary siRNAs by RDR, whereas the *EIN2* mRNA, from an endogenous gene, did not serve as RDR template[16].

siRNAs from both traditional and G:U hpRNA form two dominant bands on the gel, consistent with hpRNA transgenes generating primarily 21 and 24 nt siRNAs[17]. However, the two siRNA bands of hpEIN2[G:U] (and hpGUS[G:U]) plants showed faster gel mobility (Figure 7A, Figure S9A), which occurred in all Arabidopsis genotypes (Figure 7B, top). Small RNA deep sequencing (sRNA-seq) detected no clear shift in siRNA size profiles between the traditional and G:U hpRNA lines, with the 21 nt siRNAs being always the dominant followed generally by the 24 nt or 22-23 nt siRNAs (Figure 8, Figure S9B). This result suggested that siRNAs from the traditional and G:U hpRNAs possess different chemical modifications at the termini.

Dicer-processed sRNAs were assumed to have 5’ monophosphate but in *C. elegans* many siRNAs are found to possess di- or tri-phosphate which increases gel mobility[18]. Alkaline phosphatase treatment homogenised the gel mobility of hpRNA[WT] and hpRNA[G:U]-derived siRNAs (Figure 7C), indicating that the two siRNA populations have a similar size profile and that the differential gel mobility of untreated siRNAs was due to different 5’ phosphorylation. The siRNA bands of hpEIN2[WT] plants aligned well with the 21 and 24 nt sRNA size markers that were monophosphorylated with radioactive ^{32}P (Figure 7A; Figure S10A), suggesting that these siRNAs are largely monophosphorylated. The G:U hpRNA-derived siRNAs, with faster mobility, were therefore likely to possess 5’ di- or multi-phosphate. This possibility was

further suggested by the under-representation of hpEIN2[G:U]-derived antisense sRNAs compared to hpEIN2[WT]-derived antisense siRNAs in the sRNA-seq data, despite the similar or even stronger northern blot bands of hpEIN2[G:U] siRNAs (Figure 8, Figure S10B, Figure S11); a standard sRNA-seq protocol involves adaptor ligation with 5' monophosphorylated sRNAs but not di or multi-phosphorylated sRNAs[18,19]. Northern blot hybridisation detected high amounts of long dsRNA species in the hpEIN2[G:U] lines but not in the hpEIN2[WT] plants (Figure 7B), suggesting that the two types of hpRNA are processed differently, which could account for the differential 5' phosphorylation of the siRNAs. The Arabidopsis microRNA miR168 resembled hpRNA[WT]-derived siRNAs in gel mobility, whereas the trans-acting siRNA tasiR255 were similar to hpRNA[G:U]-derived siRNAs (Figure 7c, Figure S10a), suggesting that plant endogenous sRNAs also possess different 5' phosphorylation.

Deep sequencing of hpGUS[1:4] lines also detected siRNAs but with a much lower abundance (Figure S9c), which was consistent with the relatively low extent of *GUS* RNAi and suggested that the low thermodynamic stability of the hpGUS[1:4] RNA reduced Dicer processing efficiency. In addition, while the bulk of siRNAs from the perfect and G:U hpRNAs were derived from the dsRNA stem of hpRNA, a large proportion of siRNAs in the hpGUS[1:4] lines came from the target GUS gene downstream of the RNAi target region (Figure S11), suggesting an involvement of transitive siRNAs in hpGUS[1:4]-induced RNAi.

Discussion

In this study we showed that the traditional hpRNA transgenes are invariably methylated at the IR DNA structure and the adjacent promoter sequences compromising RNAi efficiency. This widespread intrinsic DNA methylation and self-silencing of hpRNA transgenes were not reported before but is nevertheless unsurprising. IR DNA structures have long been reported to attract DNA methylation that can extend short distance to upstream promoters in plants, and the methylated IR locus can induce homology-dependent trans-methylation of single-copy loci in the genome[20-22]. The best studied IR DNA is the naturally occurring *PAI1-PAI4* locus in Arabidopsis ecotype Wassilewskija, which always carries dense DNA methylation independently of its transcriptional activity or RdDM factors[23]. Evidence exists that supports DNA:DNA pairing in IR-induced methylation, but dsRNA and sRNA signals are also suggested to contribute the methylation particularly at the homologous trans-methylated non-IR loci[21,24]. Our results showed that strong DNA methylation in the hpRNA transgenes was largely retained in the RdDM mutants of both the upstream siRNA biogenesis factor Pol IV and the downstream effector AGO4, which seems to support a RdDM-independent DNA:DNA pairing model in IR methylation. However, the increased cross-plant uniformity of RNAi in the RdDM mutants by the traditional as well as G:U hpRNA transgenes suggest that hpRNA transgenes, like any type of transgenes, are subject to insertion pattern or position-dependent transcriptional silencing, and that RdDM plays a key role in this type of transgene silencing.

It is interesting to note that RNAi potency was generally reduced in the Pol IV mutant compared to wild-type Col-0 and the AGO4 mutant, as indicated by the uniform but weak photo-bleaching phenotypes of hpPDS lines and the low amount of hpEIN2-derived siRNAs in the *nrpd1* background. It has been proposed previously that Pol IV may use either methylated DNA and/or dsRNA as template to generate

dsRNA and siRNAs[25]. A more direct evidence for the dsRNA-templated model came from a study showing that RNA virus-derived siRNAs, without RNAi transgenes, are strongly reduced in a Pol IV mutant[26]. The hpEIN2[G:U] plants accumulated high amounts of long dsRNA species, and the bulk of sense siRNAs had the C to U-modified sequence, indicating that siRNAs were mostly derived from direct Dicer processing of the primary G:U hpRNA transcript independent of Pol IV. For the hpEIN2[WT] transgenes, however, long dsRNA was almost undetectable and there was a strong reduction in siRNA accumulation in the *nrpd1a-3* background (Figure 7B). This raises the possibility that Pol IV may contribute specifically to siRNA production from the traditional hpRNA transgenes using the low amounts of the primary perfect dsRNA as template. Interestingly, siRNA bands of hpRNA[WT] looked more scattered on the gel blot than those of hpRNA[G:U] (Figure 7), which implies that hpRNA[WT]-derived siRNAs are a mixture of different biogenesis processes with different 5' phosphorylation hence gel mobility (e.g. direct Dicer processing of primary hpRNA plus Pol IV-mediated amplification), unlike the G:U hpRNA-derived siRNAs that are largely derived from the primary hpRNA transcript.

The key finding of this study is that C to T substitutions or around 25% nucleotide modifications in the sense DNA sequence prevented the intrinsic methylation of the hpRNA transgenes, resulting in uniform RNAi across independent transgenic lines. The C to T substitutions also prevented the cotyledon to true leaf progression of methylation and self-silencing observed for the hpPDS[WT] transgene, a phenomenon that has not been reported before but has important implications in studying developmental stage-dependent RNAi and transcriptional gene silencing. Thus, disruption of perfect IR DNA structures is sufficient to block IR methylation and self-silencing of hpRNA transgenes. The promoter methylation level in the C-to-T modified transgenes (hpGUS[G:U] and hpEIN2[G:U]) was more reduced than in the 1-in-4 mismatched transgene (hpGUS[1:4]), suggesting that depletion of cytosines, the target of DNA methylation, in the sense sequence, further reduces promoter methylation. It is interesting to note that microRNA precursors in plants usually contain mismatches or G:U base pairs in the duplex regions. Considering the results from our study, this structural feature may have evolved to disrupt IR DNA structure preventing transcriptional self-silencing of miRNA genes.

As illustrated by the different *GUS* RNAi efficacy by the four hpGUS constructs, reduced dsRNA stability due to nucleotide modifications in the sense strand reduces RNAi efficiency presumably because of inefficient Dicer processing. Weak to moderate RNAi can have specific applications, particularly when the target genes are required for plant viability. The potential drawback of reduced RNAi, however, is largely overcome with the G:U hpRNA constructs, where the C-to-T changes in the sense sequence disrupt the IR DNA structure but still allow the formation of perfect hpRNA structure due to G:U wobble base-pairing. Consequently, all three G:U hpRNA constructs tested induced strong and uniform RNAi. hpRNAs containing multiple G:U base-pairs (up to 17.5%) has been previously shown to induce RNAi in animals and confer virus resistance in plants[27,28]. In our study all cytosines in the sense sequence, constituting 18~26% of the target sequences, were substituted in the G:U hpRNA constructs. Future studies should examine the number of C-to-T substitutions that are required for reducing self-silencing while maximizing RNAi efficiency.

Our study indicated that G:U hpRNA-derived siRNAs have different 5’ phosphorylation to those of traditional hpRNA. This was unexpected, as Dicer processed small RNAs in plants were assumed to carry 5’ monophosphate and standard sRNA cloning strategies have been based on this 5’ chemistry. This finding has important implications in the interpretation of current sRNA-seq data and development of sRNA sequencing protocols, as sRNAs with different 5’ phosphorylation requires different cloning methods[18]. Methylation analysis of hpRNA transgenes in the RdDM mutants suggested that the G:U hpRNA-derived siRNAs, unlike those of the traditional hpRNA, induce RdDM through a Pol IV-independent pathway. Thus, G:U hpRNA-derived siRNAs may have distinct functional properties from the traditional hpRNA-derived siRNAs, possibly due to different biogenesis or 5’ modification. How the traditional and G:U hpRNA-derived siRNAs are differentially phosphorylated is beyond the current study. The similar 5’ phosphorylation between the two groups of siRNAs and miR168 (a nuclear processed sRNA) or tasiR255 (a likely cytoplasmic sRNA) raises the possibility for siRNA processing in the nuclei for perfect hpRNA and cytoplasms for G:U hpRNAs.

In conclusion, our study uncovered a new RNAi construct design that overcomes transcriptional self-silencing to induce more uniform and persistent RNAi than the traditional hpRNA design, and shed new light on the RNAi pathways in plants. Apart from theoretical interest, future studies should investigate if G:U-modified and other mis-matched hpRNA transgenes also have increased long-term stability inducing effective RNAi in multiple generations, which would be important for field applications of RNAi in crop improvements.

Methods

Plant materials and growth conditions

Plants used in the experiments included *Arabidopsis thaliana* (ecotype Col-0), and transgenic *Nicotiana tabacum* Wisconsin 38 lines PPGH11 and PPGH24. These are two independent lines homozygous for the single-copy transgene expressing *GUS* driven by a promoter from the *Cucurbita pepo* PP2 gene[29]. The PP2-GUS plants were chosen as the testing plants because the PP2 promoter came from an endogenous gene with a different sequence to the 35S promoter used to drive the expression of the hpRNA transgenes, which therefore would prevent transcriptional silencing of the target GUS gene by promoter trans-inactivation. Plant seeds were sown either directly into soil, or placed first on MS plate for germination followed by transferring seedlings to soil. Plants were grown in a growth room (16 hours light/8 hours dark) at 22-24 °C.

Construct preparation

Preparation of GUS hpRNA constructs: The 200 bp *GUS* ORF sequence (nt. 801-1000 from the translational start codon ATG) was PCR-amplified using the oligonucleotide primer pair GUS-WT-F and GUS-WT-R (Supplementary Table 1), containing *Xho*I and *Bam*HI sites or *Hin*dIII and *Kpn*I sites, respectively. PCR fragment was inserted into pGEM-T Easy (Promega), the correct nucleotide sequence

confirmed by sequencing, and inserted as a *Bam*HI/*Hin*dIII fragment into pKannibal[30] forming the 35S-P::PDK intron::antisense GUS::Ocs-T cassette (pMBW606). This plasmid was used as the base vector for assembling the four *GUS* hpRNA constructs as follows.

hpGUS[WT]: The 200 bp *GUS* PCR fragment was excised from the pGEM-T Easy plasmid with *Xho*I and *Kpn*I, and inserted into the same sites in pMBW606 between the 35S promoter and the PDK intron. The resulting 35S::sense GUS[WT]::PDK intron::antisense GUS::OCS-T cassette was excised with *Not*1 digestion and inserted into the *Not*I site of pART27[31], forming hpGUS[WT].

hpGUS[1:4], hpGUS[2:10] and hpGUS[G:U]: The 200 bp 1 in 4 mismatched, 2 in 10 mismatched and C to T modified sequences were assembled by annealing the respective pair of overlapping oligonucleotides (GUS-4M-F + GUS-4M-R for GUS[1:4], GUS-10M-F + GUS-10M-R for GUS[2:10], and GUS-GU-F + GUS-GU-R for GUS[G:U]; Supplementary Table 1) followed by PCR extension of 3' ends using the high fidelity LongAmp Taq polymerase (NEB). Nucleotide substitutions in GUS[1:4] and GUS[2:10] followed the following rule: C is changed to G, G to C, A to T and T to A. The PCR fragments were ligated into the pGEM-T Easy vector, the correct nucleotide sequences confirmed by sequencing, and then inserted as a *Xho*I/*Kpn*I fragment into pMBW606. The resulting 35S promoter::hpRNA::OCS terminator cassette was excised with *Not*1 and inserted into the *Not*I site of pART27, forming the three mismatched constructs.

Perfect and G:U base-paired *EIN2* and *PDS* hpRNA constructs: DNA fragments spanning the 200 bp regions of the wild-type *EIN2* cDNAs were PCR-amplified from *Arabidopsis thaliana* Col-0 cDNA using the oligonucleotide primer pairs EIN2wt-F and EIN2wt-R (Supplementary Table 1) and cloned into pGEM-T Easy. The 200 bp C to T modified sense sequence (EIN2[G:U]) was assembled by annealing the overlapping oligonucleotides EIN2-GU-F and EIN2-GU-R (Supplementary Table 1), followed by PCR extension of 3' ends using LongAmp Taq polymerase, and also cloned into pGEM-T Easy and sequenced. DNA fragments of 450 bp wild-type and C-to-T modified sequences of *PDS* cDNA (Supplementary Table 2) were synthesized by GeneArt™.

The 35S-P::sense fragment::PDK intron::antisense fragment::OCS-T cassettes were prepared in the same way as for the hpGUS constructs. Essentially, the wild-type sequences were excised from the respective pGEM-T Easy plasmids by digestion with *Hin*dIII and *Bam*HI, and inserted into pKannibal between the *Bam*HI and *Hin*dIII sites so they would be in the antisense orientation relative to the 35S promoter. The wild-type or C to T modified fragments were then excised from the respective plasmids using *Xho*I and *Kpn*I and inserted into the same sites of the respective antisense-containing clone. All of the cassettes in the pKannibal vector were then excised with *Not*I and inserted into pART27 to form the final binary vectors for plant transformation.\

Stable transformation and identification of transgenic lines

All four *GUS* hpRNA constructs were transformed into the GUS-expressing tobacco lines PPGH11 and PPGH22 using the *Agrobacterium*-mediated leaf-disk method[32]. *EIN2* and *PDS* hpRNA constructs were

transformed into *A. thaliana* by the "floral dipping" method[33]. To select for transgenic Arabidopsis lines, mature seeds were sterilized[34] and spread on MS plates containing 50 µg/mL kanamycin plus 150 µg/mL timentin to inhibit Agrobacterium growth. The phenotype of *PDS* silencing was recorded for the primary (T1) transformants. The surviving T1 lines of *PDS* hpRNA constructs, and those of *EIN2* hpRNA construct, were transferred to soil, self-fertilised and grown to maturity. Seed collected from these plants (T2 seed) was used to establish T2 plants that were used for further gene silencing, DNA methylation, and transgene segregation analyses.

Analysis of *GUS* and *EIN2* silencing

GUS activity was quantitatively determined using fluorimetric 4-methylumbelliferyl-β-D-glucuronide (MUG) assay[34]. The relative GUS activity represents slope value per 5 mg of protein. For T0 plants (the primary transformants), protein used for the MUG assay was extracted from 3 leaves of an individual plant, while for the second generation, protein was extracted from a pool of multiple (20-50) transgenic plants.

For *EIN2* silencing assay, Arabidopsis seeds were sterilized[34] and plated on half-strength MS salt medium (without organics) containing 5mg/L 1-aminocyclopropane-1-carboxylic acid (ACC). The plates were imbibed for 3 days at 4 °C in the dark, transferred to 22 °C under lights for 10 hours to improve germination, and then incubated for 4 days in the dark. Around 10-12 seedlings from each transgenic line, representing the overall hypocotyl length distribution, were selected from the half-strength MS salt medium and positioned horizontally onto agar plates containing blue stain to visualize hypocotyl length. The hypocotyl length of the seedlings was photographed using a digital camera and measured using ImageJ (http://rsb.info.nih.gov/ij).

DNA and RNA analysis

DNA, small RNA and large RNA from all transgenic tobacco lines were prepared following the phenol extraction method as previously described[10]. Total RNA from the T2 transgenic Arabidopsis lines was extracted using TRIzol® Reagent (Ambion® USA) according to the manufacturer's instructions. The genomic DNA from the T2 transgenic plants was isolated from plant leaves using a Cetyltrimethyl Ammonium Bromide (CTAB) method[35].

Southern blot hybridization was performed as described[10] using a full length octopine synthase (OCS) terminator sequence as probe, which was excised from pART7[31] with *Bam*HI and *Not*I digestion, gel purified and radioactively labelled with [α-^{32}P] dCTP using the DecaLabel DNA Labeling Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The labelled DNA probe was purified using G-25 columns (GE Healthcare).

For detection of *GUS* and *EIN2* siRNAs, 20 µg of total RNA samples was separated in 17% denaturing acrylamide gel, electroblotted and UV crosslinked to HyBond-N$^+$ membrane (GE Healthcare), and hybridized with 200-nt *GUS* and *EIN2* sense RNA probe following the procedure in Wang, et al.[10].

Analysis of 5’ phosphorylation of sRNAs

Alkaline phosphotase treatment to remove 5’ phosphate from sRNAs was performed according to Pak and Fire[18]. Twenty µg of total RNA was incubated for 2 hours at 37°C in 100 µl reaction containing 1× CutSmart Buffer (New England Biolabs) and 50 units of Calf Intestinal alkaline phosphotase (CIP) (New England BioLabs). After incubation, RNA was purified with phenol/chloroform extraction, precipitated with 10 µl 3M NaOAc and 250 µl of ethanol at -20°C overnight, and dissolved in 5 µl H_2O. sRNA northern hybridization analysis of the CIP-treated RNA and untreated samples was performed the same way as described above.

DNA methylation analysis using McrBC digestion PCR and bisulphite PCR

Plant genomic DNA (~500 ng) was digested with 30 units of McrBC (NEB) in a 50 µL reaction volume at 37 °C overnight. For McrBC-minus controls, the same amount of DNA was incubated overnight at 37 °C in 50 µL reaction volumes containing the same buffer, but without the McrBC enzyme. 1 µL (50 ng) of digested and undigested DNA of each sample was used to set up PCR reactions using Taq DNA polymerase along with ThermoPol buffer (NEB). The PCR product was electrophoresed on a 2% agarose gel, stained with ethidium bromide, and visualized by UV illumination.

Bisulfite conversion and purification were performed using the EpiTect Bisulfite kit (QIAGEN) following the procedures recommended by the manufacturer. Bisulfite PCR was performed as a nested PCR (two PCR reactions). The primers used in the first and second round PCR was listed in the Supplementary Table 3. The PCR cycles was the same as described previously[36]. The PCR products from the second PCR were purified using an UltraClean DNA Purification Kit (MO BIO) following the manufacturer’s instructions. Approximately 50-200ng of purified bisulfite PCR product was sequenced with BigDye Terminator V3.1 premix (Applied Biosystems) using one of the nested primers. Cytosine methylation levels were determined from the sequencing trace files using the same procedure as described in Le et al.[37].

Declarations

Acknowledgements

We thank Carl Davies for photography and Craig Wood for supporting sRNA sequencing of RdDM mutant samples. D.Z. was supported by a China Scholarship Council (CSC) scholarship and University of Wollongong PhD tuition waiver, and C.Z. was supported by a CSC scholarship.

Author contributions

D.Z. performed the majority of experiments; C.Z. conducted bioinformatic analysis of sRNA deep sequencing data; N.S. performed bisulfite sequencing and data analysis; R.D. contributed to data analysis and writing; I.G. contributed to the RdDM mutant experiment; S.S had input in project

development and writing; R.Z. contributed to project initiation, supervision of D.Z. and writing; M.B.W. initiated and designed the work and did most of the writing.

References

1. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356-363 (2004).

2. Eamens, A., Wang, M.-B., Smith, N.A. & Waterhouse, P.M. RNA Silencing in Plants: Yesterday, Today, and Tomorrow. *Plant Physiology* **147**, 456 (2008).

3. Ghildiyal, M. & Zamore, P.D. Small silencing RNAs: an expanding universe. *Nature Reviews Genetics* **10**, 94-108 (2009).

4. Guo, Q., Liu, Q., Smith, N.A., Liang, G. & Wang, M.-B. RNA Silencing in Plants: Mechanisms, Technologies and Applications in Horticultural Crops. *Current genomics* **17**, 476-489 (2016).

5. Matzke, M.A., Kanno, T. & Matzke, A.J.M. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annual Review of Plant Biology* **66**, 243-267 (2015).

6. Taochy, C. *et al.* A Genetic Screen for Impaired Systemic RNAi Highlights the Crucial Role of DICER-LIKE 2. *Plant Physiology* **175**, 1424-1437 (2017).

7. Yu, B. *et al.* Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science* **307**, 932 (2005).

8. Yamasaki, T., Miyasaka, H. & Ohama, T. Unstable RNAi effects through epigenetic silencing of an inverted repeat transgene in Chlamydomonas reinhardtii. *Genetics* **180**, 1927-1944 (2008).

9. Stam, M. *et al.* Distinct features of post-transcriptional gene silencing by antisense transgenes in single copy and inverted T-DNA repeat loci. *The Plant Journal* **21**, 27-42 (2000).

10. Wang, M.-B. *et al.* Hairpin RNAs derived from RNA polymerase II and polymerase III promoter-directed transgenes are processed differently in plants. *RNA (New York, N.Y.)* **14**, 903-913 (2008).

11. Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A. & Matzke, A.J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *The EMBO journal* **19**, 5194-5201 (2000).

12. Dong, L. *et al.* DRD1-Pol V-dependent self-silencing of an exogenous silencer restricts the non-cell autonomous silencing of an endogenous target gene. *The Plant journal : for cell and molecular biology* **68**, 633-645 (2011).

13. Smith, N.A. *et al.* Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319-320 (2000).

14. Finn, T.E. *et al.* Transgene expression and transgene-induced silencing in diploid and autotetraploid Arabidopsis. *Genetics* **187**, 409-423 (2011).

15. Qin, G. *et al.* Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Research* **17**, 471-482 (2007).

16. Vaistij, F.E., Jones, L. & Baulcombe, D.C. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *The Plant Cell* **14**, 857-867 (2002).

17. Fusaro, A.F. *et al.* RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO reports* **7**, 1168-1175 (2006).

18. Pak, J. & Fire, A. Distinct Populations of Primary and Secondary Effectors During RNAi in *C. elegans*. *Science* **315**, 241 (2007).

19. Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. *Science* **294**, 858 (2001).

20. Luff, B., Pawlowski, L. & Bender, J. An Inverted Repeat Triggers Cytosine Methylation of Identical Sequences in Arabidopsis. *Molecular Cell* **3**, 505-511 (1999).

21. Muskens, M.W.M., Vissers, A.P.A., Mol, J.N.M. & Kooter, J.M. Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. *Plant Molecular Biology* **43**, 243-260 (2000).

22. Matzke, M.A., Aufsatz, W., Kanno, T., Mette, M.F. & Matzke, A.J.M. 8 - Homology-Dependent Gene Silencing and Host Defense in Plants. in *Advances in Genetics*, Vol. 46 (eds. Dunlap, J.C. & Wu, C.t.) 235-275 (Academic Press, 2002).

23. Melquist, S., Luff, B. & Bender, J. Arabidopsis PAI gene arrangements, cytosine methylation and expression. *Genetics* **153**, 401-413 (1999).

24. Bender, J. Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. *Trends in Biochemical Sciences* **23**, 252-256 (1998).

25. Onodera, Y. *et al.* Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell* **120**, 613-622 (2005).

26. Eamens, A., Vaistij, F.E. & Jones, L. NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. *The Plant Journal* **55**, 596-606 (2008).

27. Taylor, S.H. *et al.* Construction of effective inverted repeat silencing constructs using sodium bisulfite treatment coupled with strand-specific PCR. *BioTechniques* **52**, 254-262 (2012).

28. Akashi, H. *et al.* Escape from the interferon response associated with RNA interference using vectors that encode long modified hairpin-RNA. *Mol Biosyst* **1**, 382-90 (2005).

29. Wang, M.B., Boulter, D. & Gatehouse, J.A. Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of Cucurbita pepo. *Plant molecular biology* **24**, 159-170 (1994).

30. Helliwell, C.A. & Waterhouse, P.M. Constructs and Methods for Hairpin RNA-Mediated Gene Silencing in Plants. in *Methods in Enzymology*, Vol. 392 24-35 (Academic Press, 2005).

31. Gleave, A.P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology* **20**, 1203-1207 (1992).

32. Ellis, J.G., Llewellyn, D.J., Dennis, E.S. & Peacock, W.J. Maize Adh-1 promoter sequences control anaerobic regulation: addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. *The EMBO Journal* **6**, 11-16 (1987).

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

34. Chen, S. *et al.* A novel T-DNA vector design for selection of transgenic lines with simple transgene integration and stable transgene expression. *Functional plant biology : FPB.* **32**, 671-681 (2005).

35. Vaillancourt, B. & Buell, C.R. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv*, 783159 (2019).

36. Wang, M.B., Wesley, S.V., Finnegan, E.J., Smith, N.A. & Waterhouse, P.M. Replicating satellite RNA induces sequence-specific DNA methylation and truncated transcripts in plants. *RNA (New York, N.Y.)* **7**, 16-28 (2001).

37. Le, T.-N. *et al.* DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. in *Genome biology* Vol. 15 458 (2014).

Tables

Table 1. Summary of *GUS* gene silencing by the four hpRNA constructs based on MUG assay data.

Constructs	Total No. transgenic lines	Strong silencing (<10% of control)	Weak silencing (10%-30% of control)	Almost no silencing (>30% of control)
hpGUS[WT]	59	35 (59.3%)	9 (15.3%)	15 (25.4%)
hpGUS[G:U]	74	71 (95.9%)	1 (1.4%)	2 (2.7%)
hpGUS[1:4]	33	10 (30.3%)	23 (69.7%)	0
hpGUS[2:10]	41	28 (68.3%)	0	13 (31.7%)

Table 2. Summary of *PDS* and EIN2 RNAi in RdDM mutant backgrounds.

Constructs	Total No. transgenic lines scored	Silenced lines	Unsilenced/weakly silenced lines
Col/hpPDS[WT]	100	57 (57%)	43
Col/hpPDS[G:U]	172	160 (93%)	12
nrpd1/hpPDS[WT]	52	46 (88.5%)	6
nrpd1/hpPDS[G:U]	40	40 (100%)	0
ocp11/hpPDS[WT]	59	56 (94.9%)	3
ocp11/hpPDS[G:U]	100	100 (100%)	0
Col/hpEIN2[WT]	20	13 (65%)	7
Col/hpEIN2[G:U]	20	18 (90%)	2
nrpd1a/hpEIN2[WT]	20	20 (100%)	0
nrpd1a/hpEIN2[G:U]	20	20 (100%)	0
ocp11/hpEIN2[WT]	19	16 (84.2%)	3
ocp11/hpEIN2[G:U]*	11	10 (90.9%)	1

*A number of lines had poor germination on ACC medium in the darkness, possibly due to extreme *EIN2* RNAi, so the plant number is too small for proper comparison with the other transgenic populations.

Figures

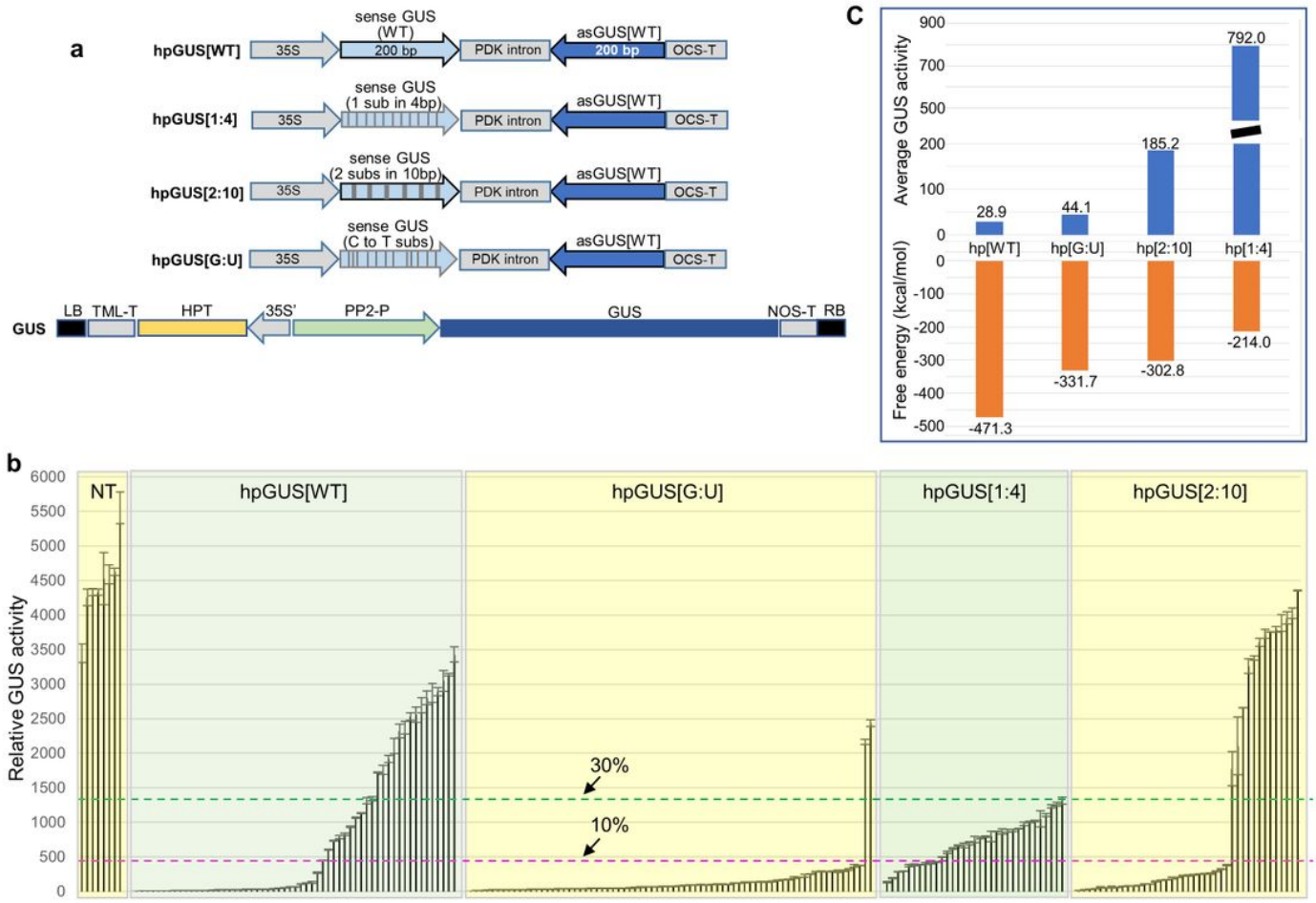

Figure 1

G:U and 1:4 mismatched GUS hpRNA constructs induce uniform GUS RNAi in tobacco. a. Schematic diagram of the four GUS hpRNA constructs and the target GUS gene in PPGH11 and PPGH24 plants. 35S, the longer version (~1.3 kb) of the cauliflower mosaic virus 35S promoter; 35S', a shorter version (337 bp from nt. -285 to +50) of the 35S promoter from pTRA151 vector; OCS-T, NOS-T, TML-T, the transcriptional terminator sequences of the Agrobacterium tumefaciens octopine synthase, nopaline synthase, and tumour morphology large genes, respectively. HPT, hygromycin phosphotransferase gene; PP2-P, the Cucurbit pepo PP2 protein gene promoter. b. GUS activity in independent T0 plants transformed with the conventional and modified hpRNA constructs. Each bar represents an independent T0 transgenic plant. The dashed green and pink lines indicate the 30% and 10% GUS activity levels of the untransformed PPGH11 and PPGH24 plants (NT). c. The average levels of GUS silencing in the strongly silenced lines of hpGUS[WT], hpGUS[G:U] and hpGUS[2:10] and all the silenced plants of hpGUS[1:4] show a good correlation with the thermodynamic stability of predicted hpRNA structures.

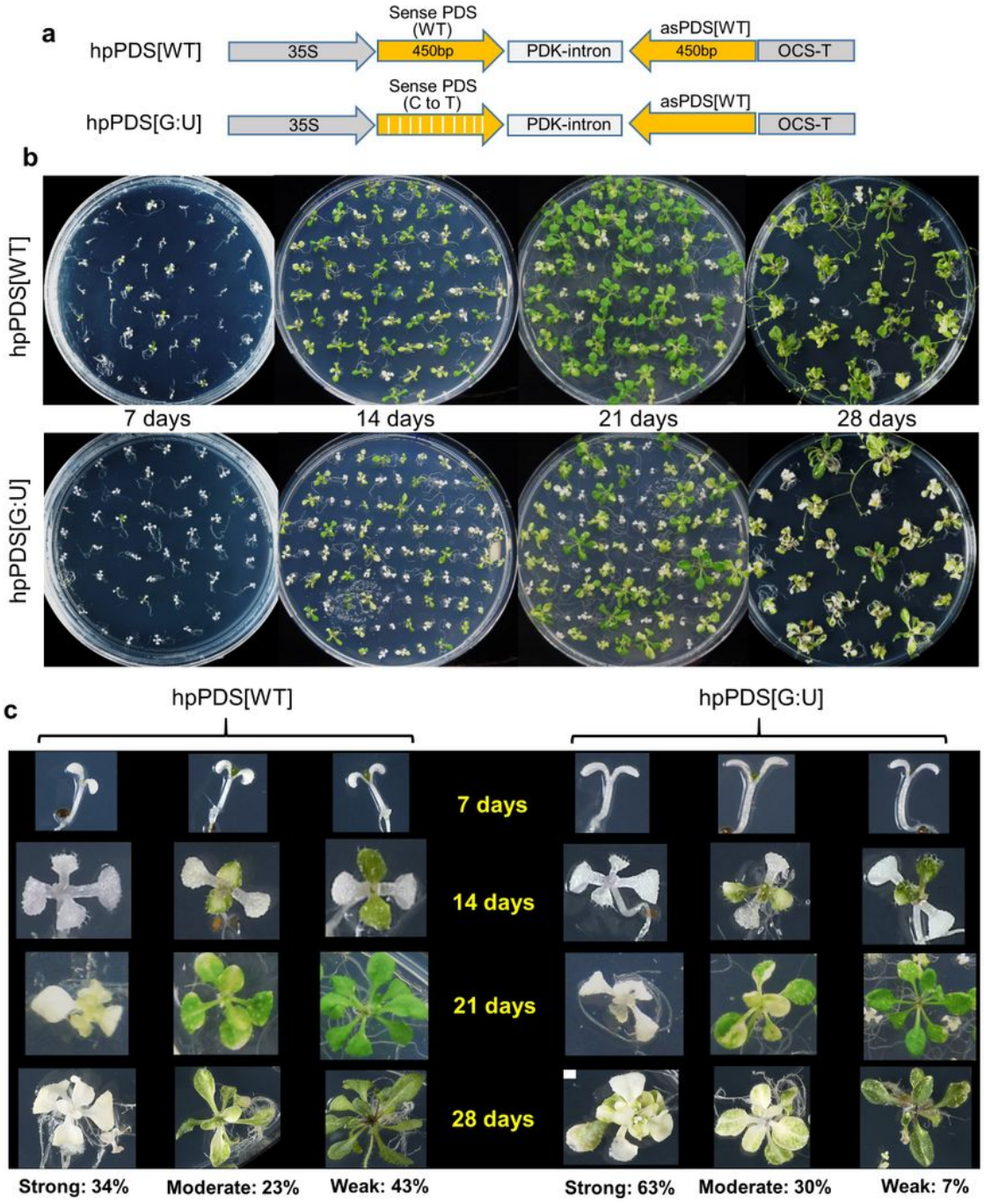

Figure 2

G:U modified hpRNA construct induces uniform RNAi of the endogenous EIN2 gene in Arabidopsis. a. Schematic of the traditional (hpEIN2[WT]) and the G:U modified (hpEIN2[G:U] constructs. b. Hypocotyl length phenotypes of 20 independent lines each of the two constructs, with longer hypocotyls indicating stronger EIN2 RNAi. Approximately 12 T2 siblings with representative hypocotyl lengths were placed side

by side and photographed. Untransformed (WT) Arabidopsis Col-0 plants were used as control. c. Measured lengths of the plants shown in (b).

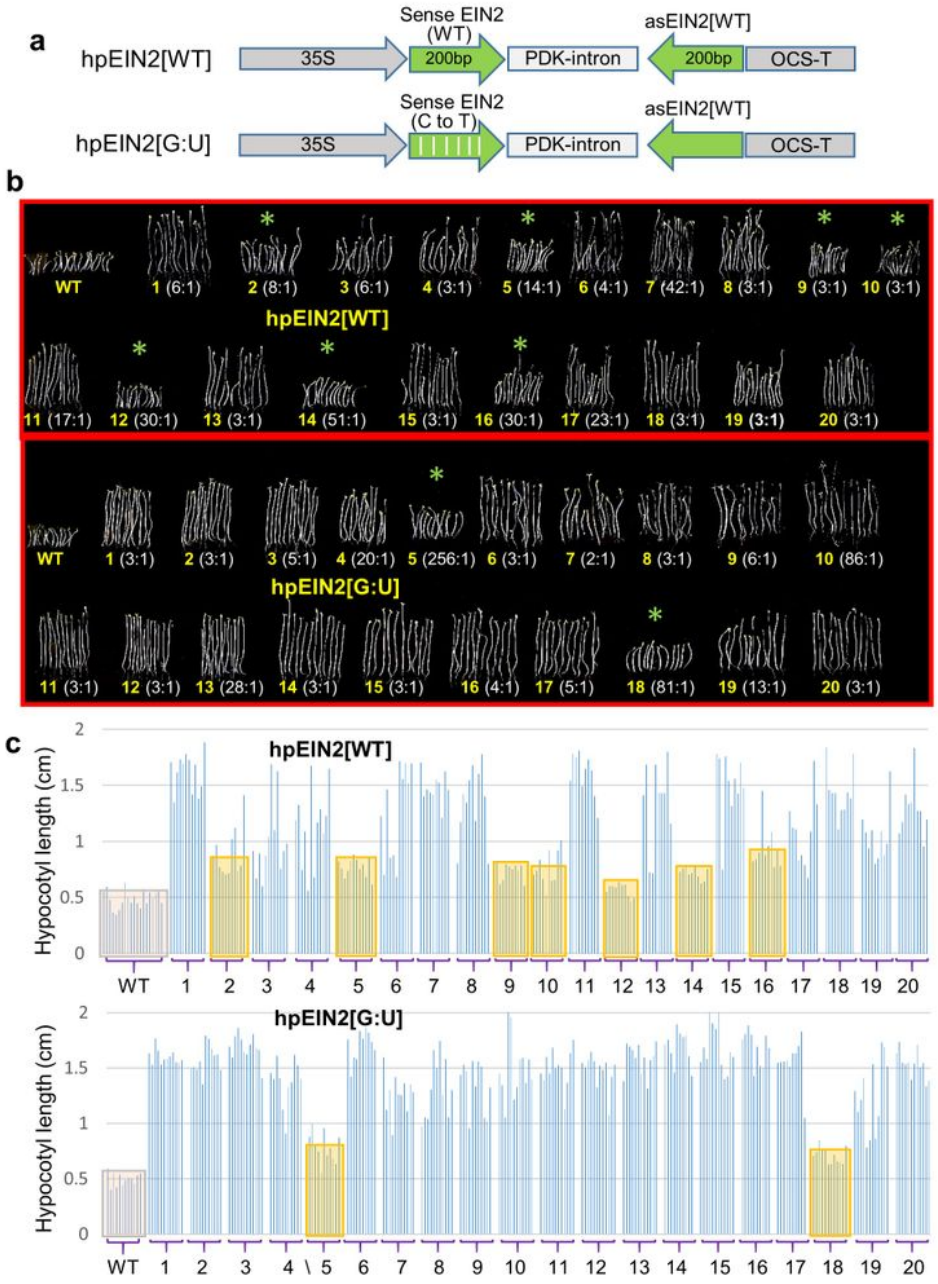

Figure 3

G:U modified hpRNA construct induces uniform and persistent RNAi of the endogenous PDS gene in Arabidopsis. a. Schematic of the traditional (hpPDS[WT]) and the G:U modified (hpPDS[G:U]) constructs. b. Phenotypes of primary independent transformants with the hpPDS[WT] and hpPDS[G:U] constructs.

Note that all lines have photo-bleached cotyledons, indicating strong PDS RNAi in cotyledons, but some lines developed green true leaves indicating loss of PDS silencing. c. Summary of PDS RNAi in the primary transformants, showing a higher frequency of persistent PDS RNAi in the hpPDS[G:U] population. The plants were classified into three groups based on strong (strong photo-bleaching in the whole plant), moderate (pale green or mottled leaves) and weak (fully green or weakly mottled leaves) PDS RNAi.

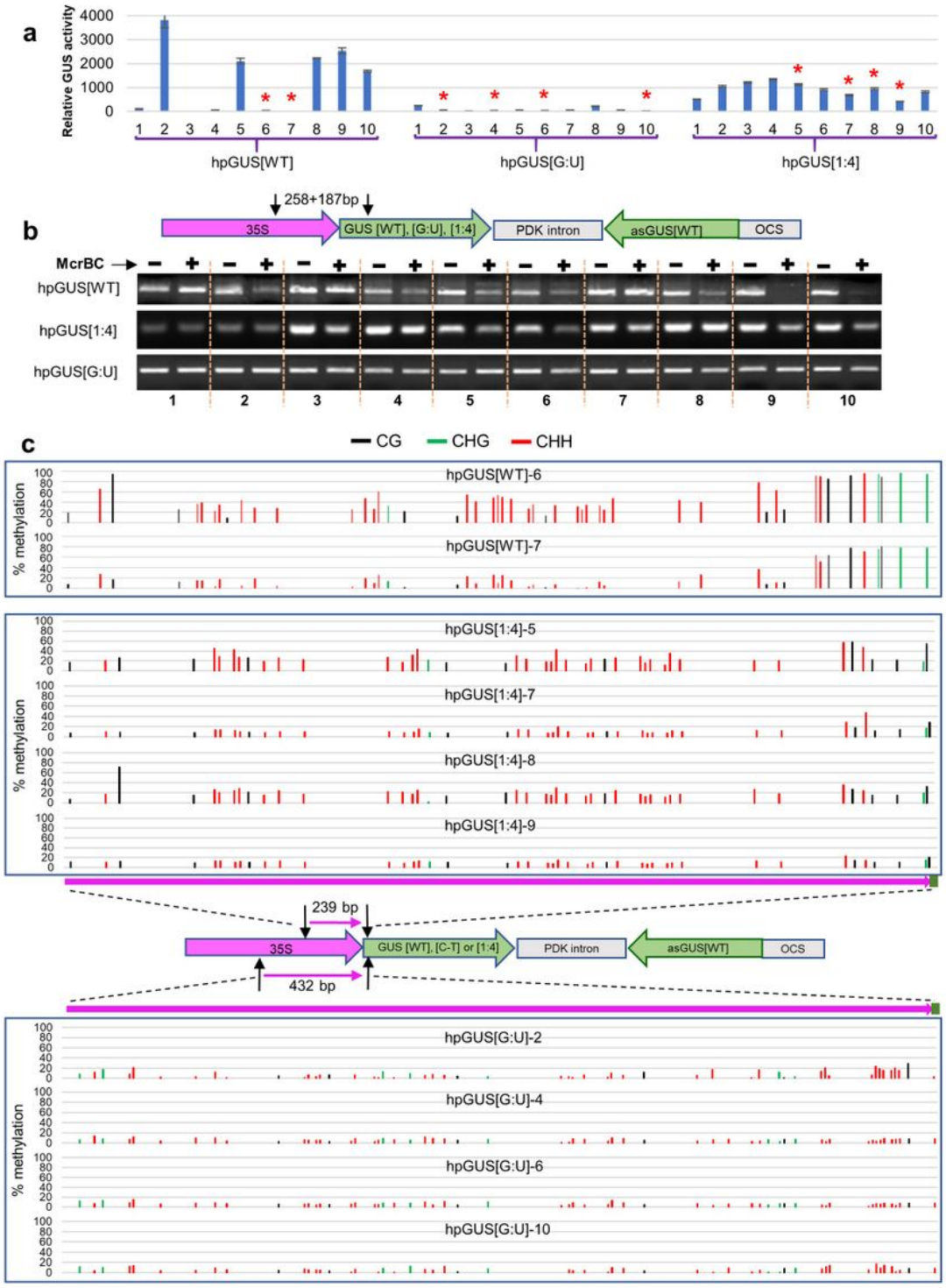

Figure 4

a. GUS expression patterns of the independent transgenic lines analysed by McrBC-digestion PCR. The red asterisks indicate the lines analysed by bisulfite sequencing in (c). b. McrBC-digestion PCR analysis of the 35S promoter-sense junction regions in the hpGUS transgenes as defined by the arrows. McrBC is an endonuclease which only cleaves DNA containing methylated cytosine bases. Comparing PCR amplification of McrBC-treated with untreated DNA therefore provide an estimation of DNA methylation levels. DNA samples were either untreated (-) or treated (+) with McrBC before PCR amplification. c. Bisulfite sequencing of the 35S promoter regions as indicated by the arrows. Bisulfite treatment of genomic DNA converts unmethylated cytosine bases to uracil (U) (shown as thymine in PCR product) but methylated cytosines are not affected. PCR amplification of bisulfite-treated DNA followed by sequencing therefore detects methylated cytosines at a single-nucleotide resolution. PCR primers were designed to specifically amplify only the 35S promoter sequences of the hpGUS transgenes but not the one driving HPT expression in the target GUS gene (Figure 1A; 35S').

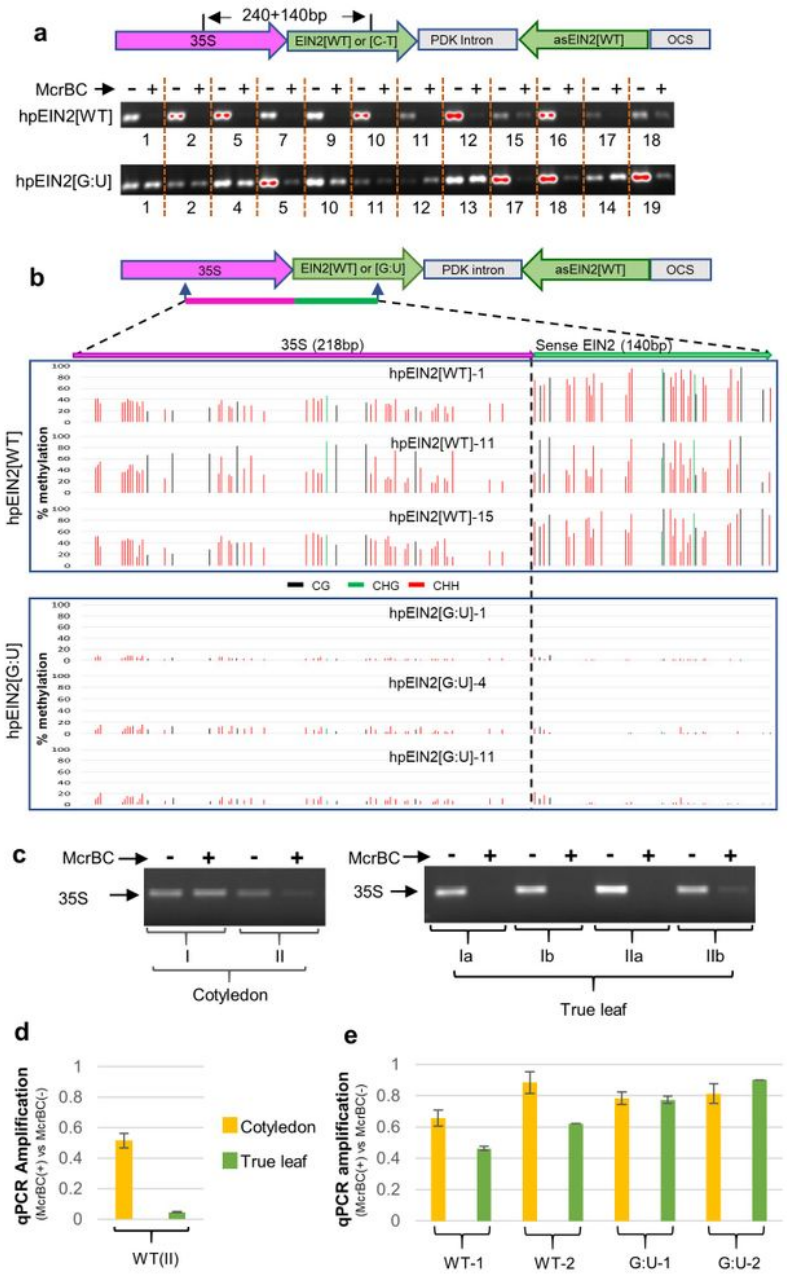

Figure 5

The G:U modified EIN2 hpRNA transgene shows greatly diminished DNA methylation at the promoter region. a. McrBC-digestion PCR analysis of the 35S-EIN2 junctions regions in the hpEIN2[WT] and hpEIN2[G:U] transgenes. EIN2 silencing phenotypes of these lines are shown Figure S5A. b. Bisulfite sequencing analysis of the 35S promoter-EIN2 junction region. c, d, e. The hpPDS[WT] transgene shows stronger 35S promoter methylation in true leaves than in cotyledons. c and d. McrBC-digestion PCR or

qPCR of primary T1 transgenic lines. Primary T1 transgenic lines were randomly divided into two pools (I and II) and photo-bleached cotyledons from multiple T1 transgenic plants within each pool were collected and combined to generate two DNA samples (photo-bleached cotyledons were very small and multiple plants were needed to obtain sufficient DNA). Young leaf tissues were also harvested from the same two groups of hpPDS[WT] plants but they were divided into four pools (two pools for each of the two cotyledon pools) resulting in four DNA samples (Ia, Ib and IIa, IIb). e. McrBC-digestion qPCR of T2 transgenic plants. As T1 plants with strong PDS RNAi did not grow to seed, only two lines each with moderate PDS RNAi were analysed. For each of the four lines, photobleached cotyledons and the first true leaves (younger than the true leaves in (d) and (e)) that had just emerged were harvested from ~25 T2 progeny, and used for DNA extraction and McrBC-digestion qPCR analysis.

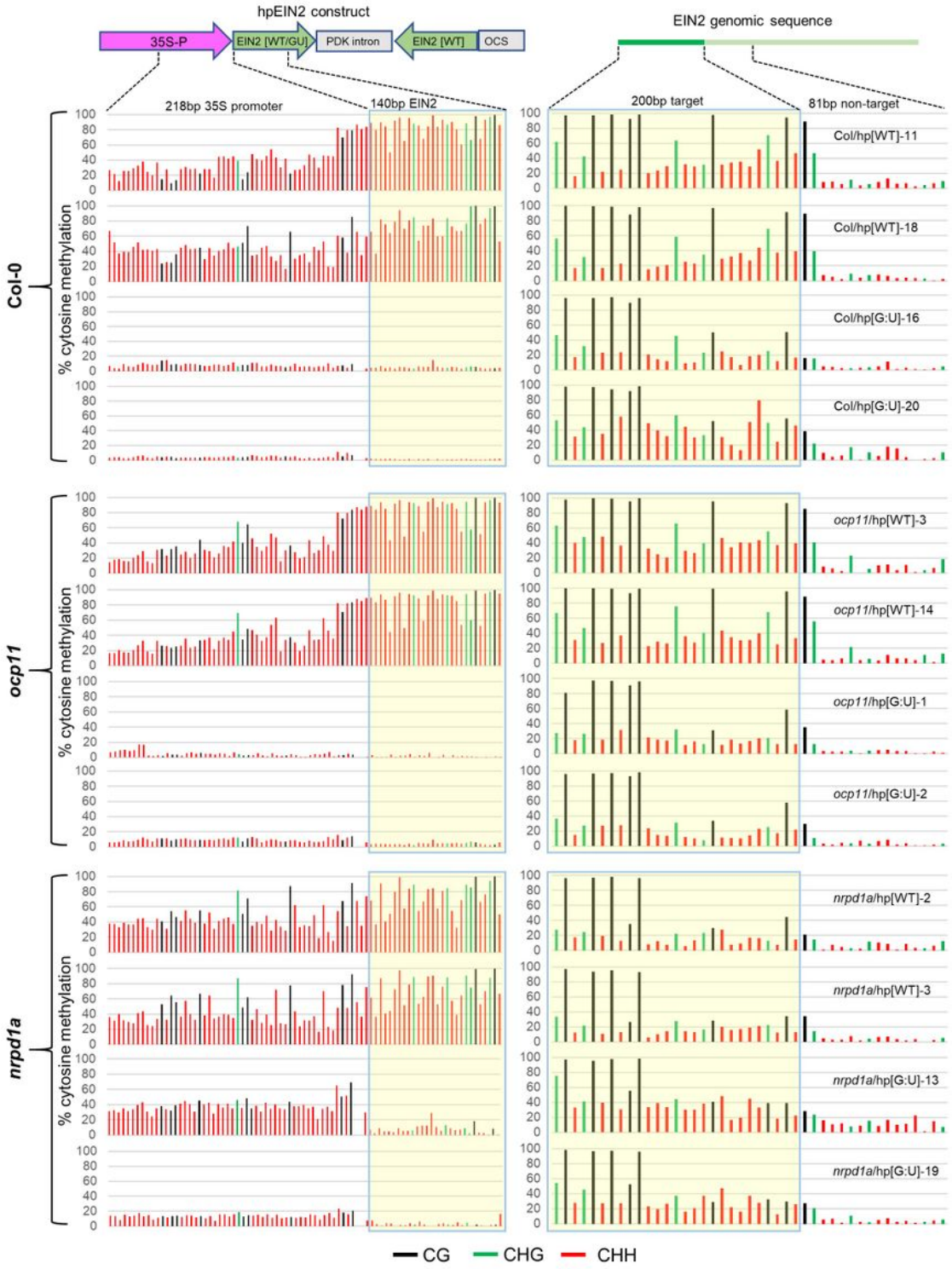

Figure 6

Heavy DNA methylation in the traditional hpEIN2[WT] transgene is not affected in RdDM mutants. yellow-highlighted areas represent the IR (left) and target EIN2 genomic (right) regions. Left, both promoter and IR DNA methylation of hpEIN2[WT] is retained in the RdDM mutants, with hpEIN2[G:U] showing almost no methylation except for the nrpd1a-3/hp[G:U]-3 line. Right, methylation levels of the hpRNA-targeted EIN2

genomic sequence, showing much lower levels of CHG and CHH methylation than the hpEIN2[WT] sequence on the left, particularly in the nrpd1a/hpEIN2[WT] lines.

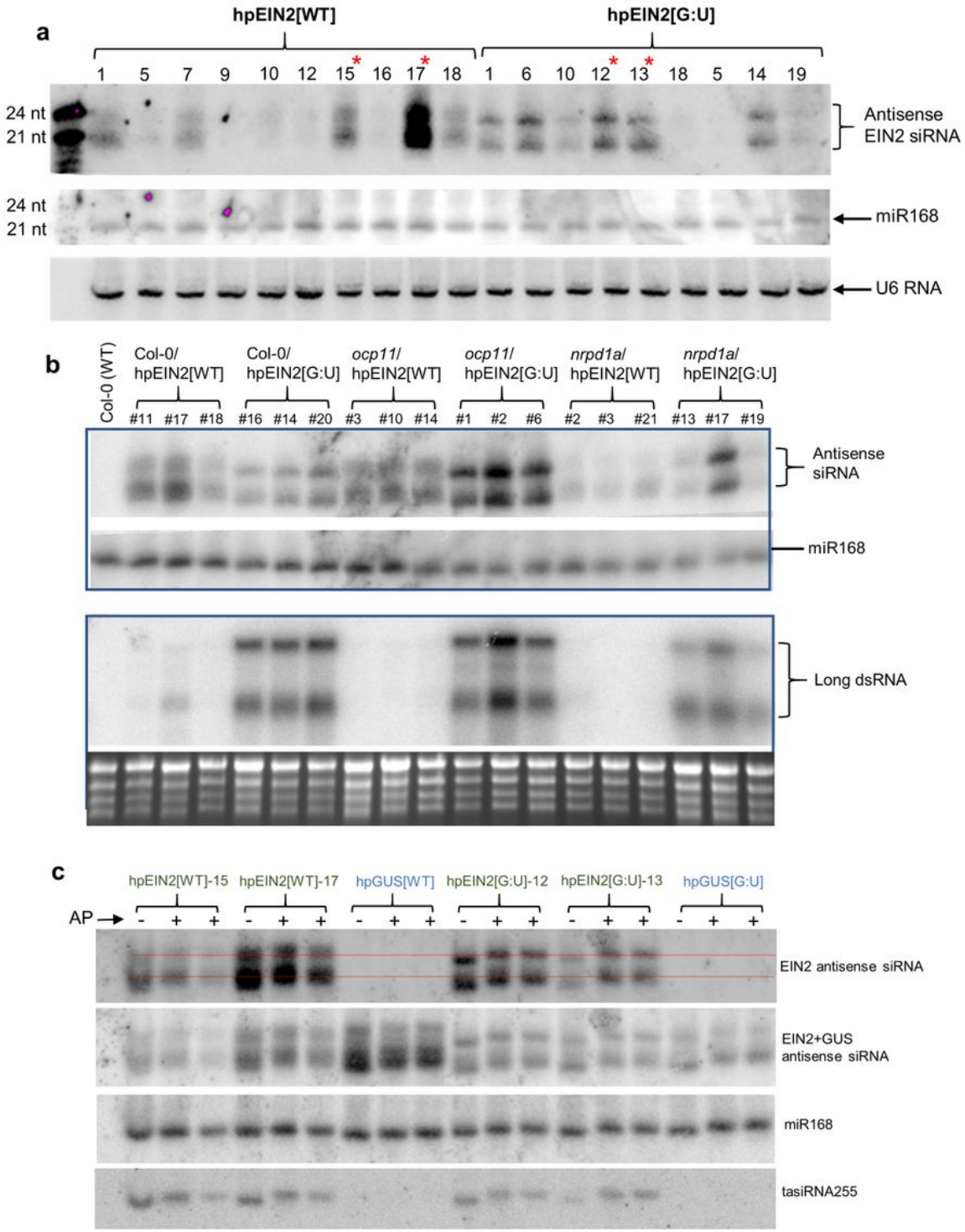

Figure 7

G:U hpRNAs are processed into siRNAs that show different 5’ phosphorylation to traditional hpRNA-derived siRNAs. a. Northern blot hybridisation to detect antisense sRNAs from T2 transgenic hpEIN2[WT] and hpEIN2[G:U] Arabidopsis plants using sense RNA transcripts of the 200 bp EIN2 target sequence as

probe. The EIN2 RNAi phenotypes of these lines are shown in Figure 2 and Figure S5B, and the line numbers correspond to those of Figure 2. The asterisks indicate the samples used for sRNA deep sequencing. The blots were hybridized with U6 RNA probe for loading control. b. Northern blot hybridisation to detect antisense sRNAs (upper panel) and long dsRNA (lower panel) from T2 transgenic hpEIN2[WT] and hpEIN2[G:U] plants in Col-0, ocp11 and nrpd1a-3 backgrounds, using the same sense EIN2 probe as in (a). c. Alkaline phosphatase (AP) treatment assay of hpRNA transgene-derived and endogenous sRNAs. Top panel: sRNAs from traditional and G:U modified hpRNA transgenes were treated (+) or untreated (-) with AP, separated in 17% denaturing polyacrylamide gel, and hybridized with the same sense EIN2 probe. The two red colored lines indicate the homogenized position of the 21 and 24 nt siRNA bands after phosphatase treatment. Second panel: The same gel blot shown above was hybridized with sense GUS RNA showing both the EIN2 and GUS antisense siRNAs. Third panel: The same gel blot was stripped and re-hybridized with antisense miR168 oligonucleotide probe. The bottom panel: The same blot was stripped and re-hybridized with antisense tasiR255 oligonucleotide probe. Note that the AP-caused gel mobility shift of miR168 and tasiR255 bands resembles that of traditional hpRNA-derived siRNAs and G:U hpRNA-derived siRNAs, respectively, which is more clearly shown in Figure S10A.

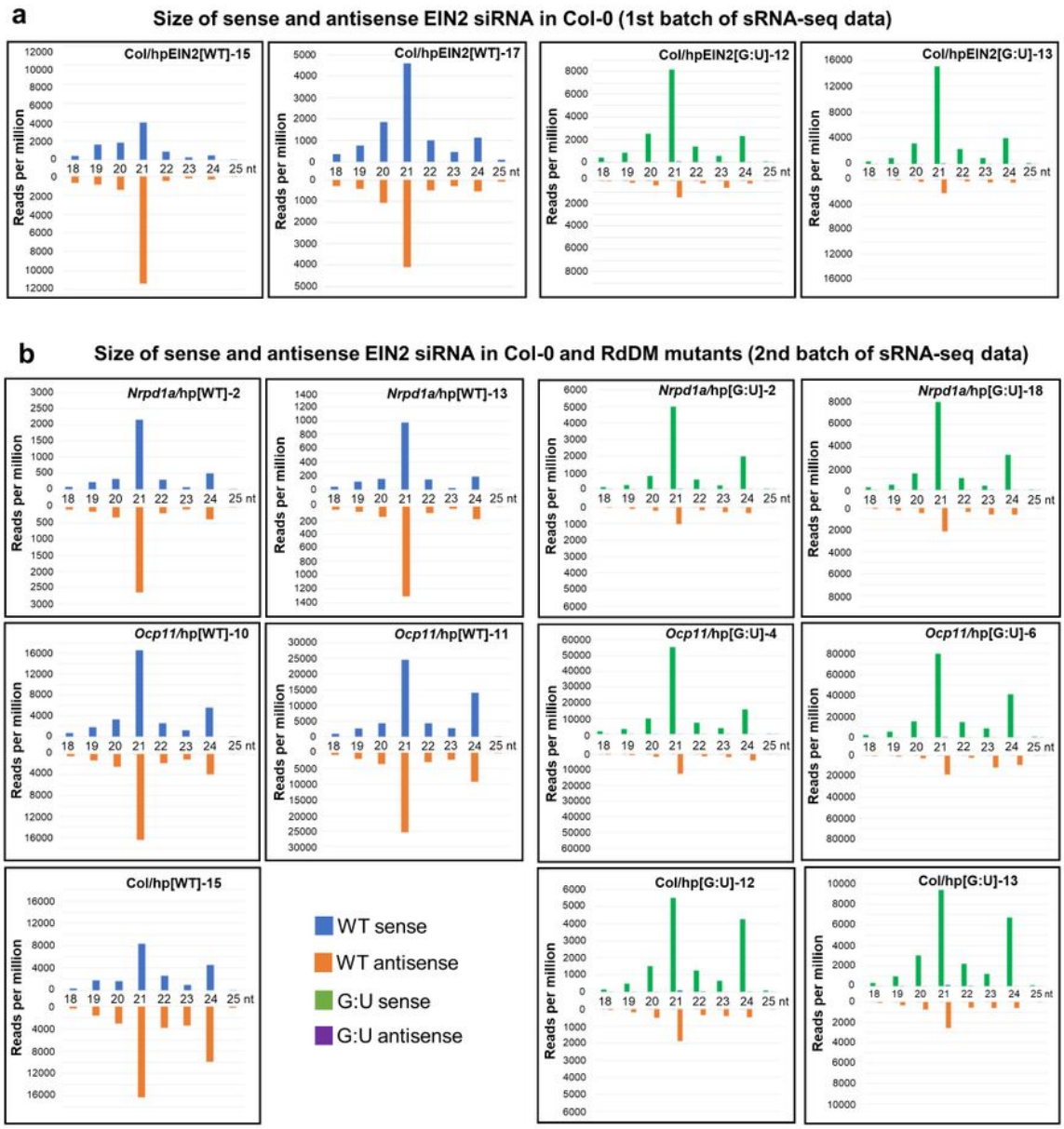

Figure 8

Summary of sRNA deep sequencing data showing size distribution of sense and antisense siRNAs derived from the dsRNA stem of the hpEIN2[WT] and hpEIN2[G:U] in Col-0, nrpd1a-3 and ocp11 backgrounds. Note that for the G:U hpRNA lines, the bulk of the sense siRNAs have the C to U converted sequence of the sense strand, indicating no RDR-synthesized secondary siRNAs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ZhangetalSupplementarymaterials2021.docx