SI Appendix

SI Materials and Methods

Plant material, growth conditions and chemical treatments

All the stocks used in this study were in Col-0 genetic background. *Arabidopsis thaliana* accession Col-0, *ago4-2* [1], *dcl2,3,4* [2], *rdr2-1* (CS879934), *rdr6-12* (CS24286) and *Helianthus annuus* L. cv HA89 genotype were grown in a growth chamber at 22–24°C, under long-day photoperiods. *A. thaliana* plants were grown in Petri dishes containing 0.8% agar–Murashige and Skoog (MS) medium or in soil pots (837 cm3); sunflowers were grown in soil pots until reaching the desired developmental stage. For 5-Azacytidine (5-AZA) treatments, sunflower seeds were surface sterilized and placed into Petri dishes on filter paper saturated with a 50 µM 5-Aza-2'-deoxycytidine (Sigma-Aldrich), or DMSO as mock, aqueous solution. The seeds were incubated during 3 days in dark at 22-24°C and then transferred to sand pots watered with 1x hoagland solution in the growth chamber until samples were collected seven days post germination. Leaves/cotyledons infiltrated with agrobacterium not carrying any construct were used as mock for P19 treatments. For the experiments pooled samples of P19 and 5AZA mocks were used as no difference were detected between both controls.

Constructs and plant transformation

The sequence of *HaWRKY6* EST can be found in the EMBL/GenBank under the accession number BU024714. *HaWRKY6* promoter was isolated from HA89 seedlings. Different regions of the promoter were cloned in the pCR8/GW/TOPO plasmid (Life Technologies). The *ncRNA-W6* was subcloned under a 35S promoter. The *HaWRKY6* promoter constructs were fused to the reporter gene GUS into a pGREEN-IIS backbone vector [3]. All the promoter versions were cloned in sense and antisense with respect to the reporter gene. Constructs generated or used in this work are listed in Table S1. Primers used for cloning are listed in Table S2. *A. thaliana* plants were transformed by floral dip procedure [4]. Transient transformation of sunflower leaf and cotyledons was carried out as previously described [5, 6].

GUS staining

Fifteen homozygous independent lines for each construct were used histochemically detect GUS activity as described next. Plant tissue (seedlings, leaves, flowers, inflorescences and siliques) were collected and incubated with staining solution (2 mM Potassium-Ferrocyanide buffer, 2 mM Potassium-Ferricyanide buffer, 0,2% Triton X-100, 50 mM NaPO4 pH7.2, 2 mM X-gluc). Samples were vacuum infiltrated during 5 minutes and incubated in darkness at 37 °C overnight. De-staining was done by sequential ethanol washes (20%, 30%, 50%) 30 minutes each at room temperature with low agitation. Finally, samples were incubated in 70% ethanol until the chlorophyll was removed completely.

RNA Analysis

Total RNA was extracted using TRIZOL reagent (Thermo Fisher Scientific). RNA blots were performed as previously described [7]. Briefly, 5 µg of total RNA were resolved in 17% (v/v) polyacrylamide gels under denaturing conditions (7 M urea) and then transferred to HyBond-N+ charged nylon membranes (Amersham) by semidry electroblotting. RNA was covalently fixed to membranes in an UV Crosslinker. Membranes were hybridized over night with DNA oligonucleotide probes labeled with second generation DIG Oligonucleotide 3'-End Labeling Kit (Roche), several probes recognizing the ncRNA-W6 region were used to maximize the detection of derived small RNAs; signal was detected using CSPD ready-to-use solution, by exposure to Amersham hyperfilm ECL (GE Healthcare Life Sciences). Probe sequences are listed in Table S2.

For quantitative RT-PCR, 1 µg of total RNA was used for reverse transcription reactions using RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific). Oligonucleotides are specified in Supporting Information, Table S2. Three biological replicates, tested by duplicate, were used to calculate the standard error of the mean. Standard error of the mean (SEM) was calculated using propagation of error of the ΔCt values and expressed in figures as two times the SEM. Statistical significance was tested using a two tailed, unpaired t-Test and considered significant with p-values of less than 0.05. p-values lower than 0.05 are marked in the figures as ** while lower than 0,01 as *.

DNA Methylation Analyses

Chop qPCR assays were performed using 100 ng of restriction enzyme-digested ("chopped") genomic DNA using HpaII (methylation-sensitive) and MspI (Methylation-insensitive) as in [8]. Bisulfite sequencing was by the procedure of [9]. In brief, PCR fragments amplified from bisulfite-treated DNA EZ DNA Methylation-Lightning™ Kit (Cat. No. D5030) were cloned into pGEM-T-Easy and sequenced using a specific primer annealing to the amplified fragment. The percentage of methylated/un-methylated cytosines in all contexts was calculated by using Kismeth open software [10], as well as the dot-plots included in Figure S2B. Chop-qPCR and bisulfite sequencing primers are listed in Table S2.

Chromosome Conformation Capture

Chromosome conformation caption (3C) was performed as described by [11]. We should adjust conditions to sunflower, starting with 12 cotyledons and 12 leaves of V4 stage plants (~15 mL). Digestions were performed overnight at 37°C with 400 U HindIII or MspI (NEB). DNA was ligated by incubation at 22° C for 5 hr in 4 mL volume using 100U of T4 DNA ligase (Thermo). After reverse crosslinking and Proteinase K treatment (Invitrogen), DNA was recovered by phenol/chloroform extraction and ethanol precipitation. Relative interaction frequency was calculated by qPCR. A region cut by neither HindIII nor MspI was used to normalize the amount of DNA. Primers used to detect and quantify loops are listed in table S2.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were performed on Sunflower V4 plants. For nuclear extraction, we used a procedure modified from [12], adapting it to sunflower tissue. Briefly, homogenates were decanted through pre-wetted Miracloth filters (100 µm), resuspended in 15 mL of Nuclei Isolation Buffer (NIB). Then, 10% Triton X-100 was added to a final concentration of 1% and gently agitated for 20 min at 4°C. The homogenate was centrifugated at 2000 x g for 10 min and the pellet resuspended gently in NIB. Then, crude preparation of nuclei was overlayed on the top of the density gradient previously assembled (5 mL Percoll 60%/5 mL Sucrose 2.5 M). The gradient was subjected to centrifugation in a swinging bucket rotor at 2000 × g for 30 min at 4°C. Then, the liquid above the gradient was diluted with 5 volumes of NIB, incubated for 10 min under gentle shaking and centrifuged at

2000 × g for 10 min. Extracted nucleus were then resuspended in Nuclei Lysis Buffer and the immunoprecipitation was performed by the procedure previously descripted [13].

Small RNA Sequencing

Small RNA libraries were prepared as indicated by the TruSeq small RNA library prep kit (Illumina). 50 ng of small RNAs purified with the ZR small-RNA PAGE Recovery Kit (Zymo Research) were used as input for the library preparation. mRNA libraries were prepared, using 1 ug of total RNA as input, as described by the TruSeq RNA sample prep V2 guide (Illumina). Small RNA and mRNA libraries size selections were performed using the BluePippin System (SAGE Science). Single-end Illumina sequencing was performed with a HiSeq3000 apparatus.

Small RNA reads were first processed to remove 3` read-through adapters using Trim_galore! with the small_rna option (version 0.4.2, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore). They were then mapped to the genomic region containing the *HaWRKY6* locus plus the 10,000 bp flanking region, 5,000 bp up- and downstream (chromosome HanXRQChr17, coordinates 18992761-18999078 on the *Helianthus annuus XRQ* r1.0 genome, [14]). This was performed with STAR (version 2.5.2b) using the following parameters: outFilterScoreMinOverLread 0 outFilterMatchNmin 16 outFilterMatchNminOverLread 0 outFilterMismatchNoverLmax 0.05 seedSearchStartLmax 30 alignIntronMax 1 alignEndsType EndToEnd. Mapped reads were additionally filtered into alignments of 21 or 24 nt long reads, the coverage profiles on Watson and Crick strands were obtained with bedtools [15] and finally coverage profiles were plotted in R (R Core Team, 2017) with the Gviz package [16]. Sequencing data are available at the European Nucleotide Archive (ENA) under the accession numbers PRJEB28614.

Identification of additional *ncRNA-W6* copies

Additional copies similar to the TE producing the *ncRNA-W6* were searched in the *Helianthus annuus XRQ* r1.0 genome with Blast. The 260 nt ncRNA-W6 sequence was used as query and hits with 90 % identity or higher and more than 200 nt aligned were

considered. A circular graph showing the genomic distribution of the hits was produced using CIRCOS (v0.69-3, [17].

Prediction of inverted repeats in the sunflower genome

In order to identify inverted repeats in promoters of sunflower genes, excluding tRNA genes, upstream regions of 1000, 2000 and 5000 bp were extracted from the *XRQ* r1.0 genome [14]. Inverted repeats were then identified using the einverted program from the EMBOSS software suite, setting a maximum repeat length of 1000 bp, a threshold of 150, a gap penalty of 8, a match score of 3 and a mismatch score of -4. Additionally, sRNA-Seq reads were mapped to the genome with bowtie (allowing 1 mismatch, the best and strata options and reporting up to five alignments) and the count of 24 nt perfect matching reads on IRs was determined with converageBED from bedtools (Quinlan et al., 2010). Reads per kilo base per million mapped reads (RPKM) were determined based on the count of mapped reads per library and the total length of the two repeats in the IR.

Supplemental Figure Legends

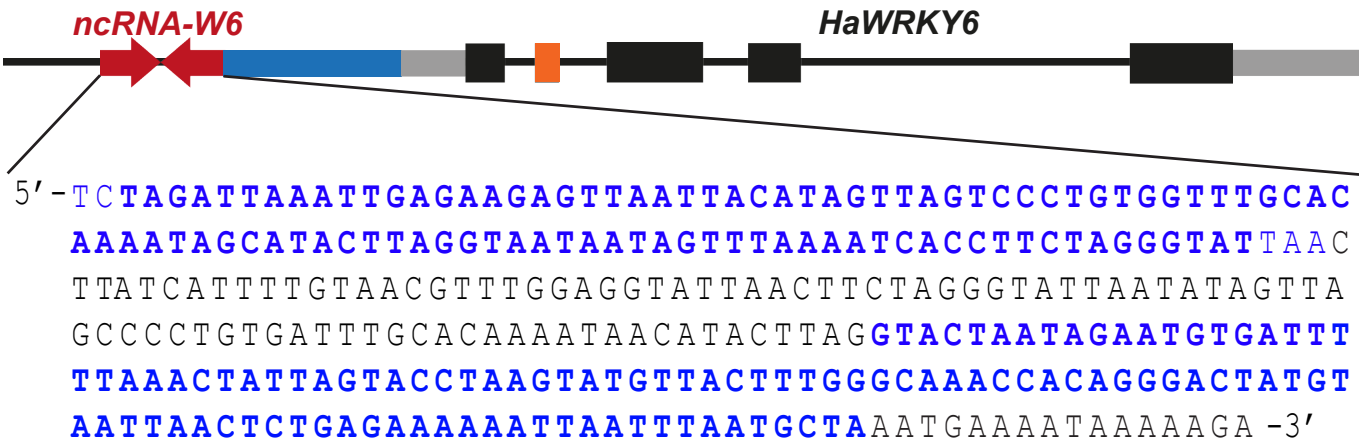

Figure S1, an inverted-repeat derived ncRNA is transcribed from the *HaWRKY6* proximal promoter. Sequence analysis of the *HaWRKY6* promoter. Bases colored in blue are the inverted repeat regions with highest complementarity. Opposed red arrows indicating the transcribed IR region, solid boxes represent exons (in grey untranslated regions, black coding sequence and in orange the alternative second exon).

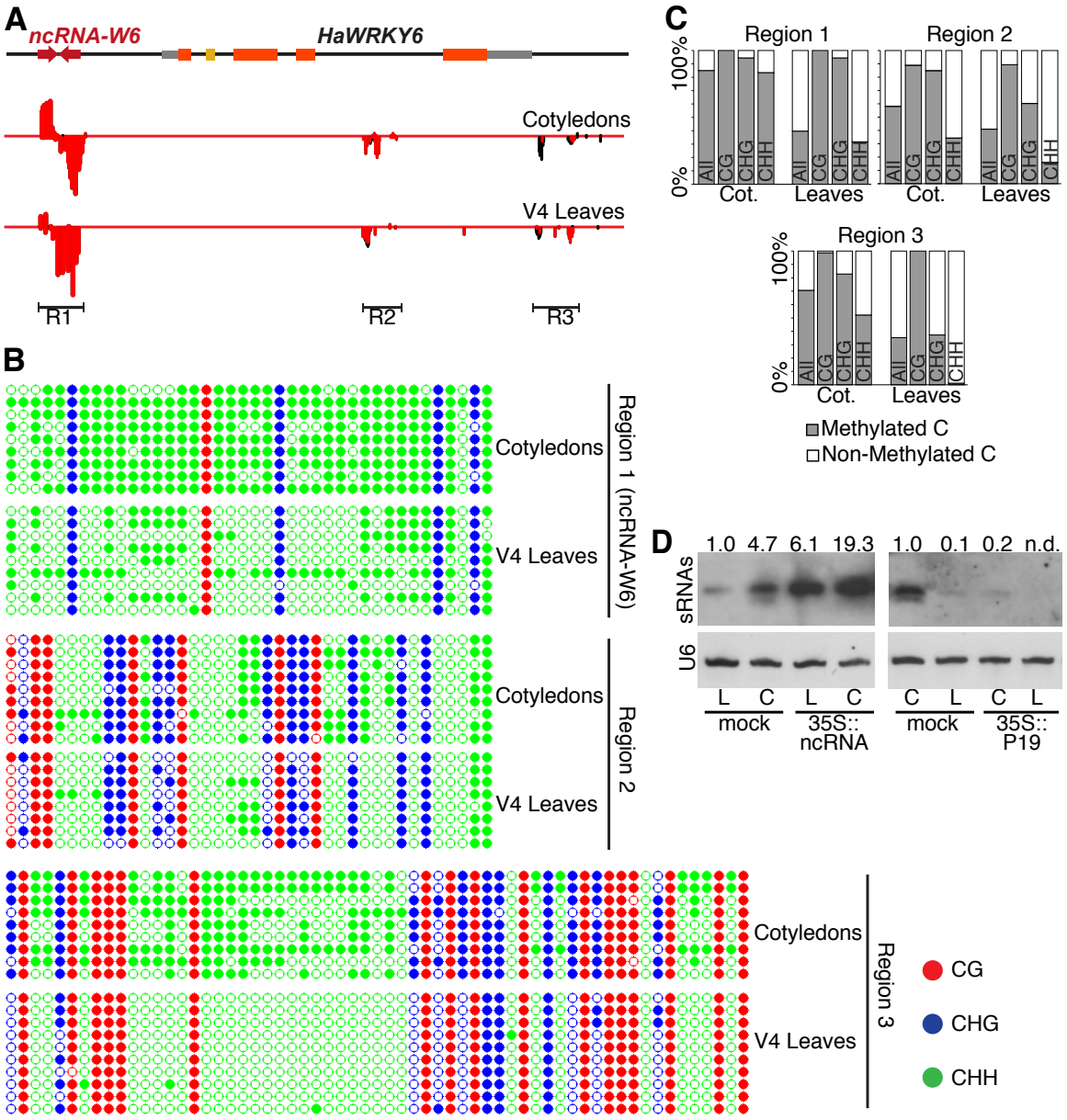

Figure S2, small RNAs are produced in *cis* and *trans* across the *HaWRKY6* locus. A. Alignment of the reads obtained by a small RNA sequencing of Sunflower Cotyledons and leaves searching for reads that uniquely map to this locus (red) or to several regions of the sunflower genome (black). Small RNAs mapping regions are labeled as R1, R2 and R3 respectively. **B.** Dot-blot showing methylated cytosines (filled dots) and un-methylated cytosines (empty dots) in all contexts (CG: red, CHG: blue, CHH: green) were generated

using Kismeth software [10]. **C.** Percentage of methylated cytosine as measured by sodium bisulphite conversion followed by amplification and cloning. **D.** RNA blots detecting *ncRNA-W6*-derived sRNAs in sunflower leaves (L) or cotyledons (C) transformed with an empty vector (mock), *35S:ncRNA-W6* or *35S:P19*. *U6* was used as a loading control and signal intensity calculated using ImageJ and expressed as relative to the first sample.

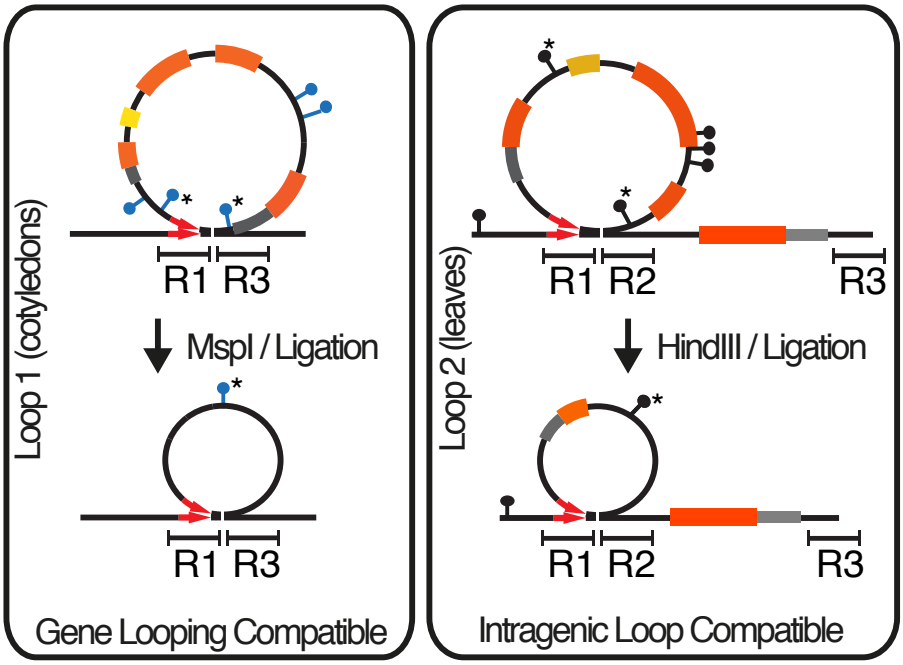

Figure S3, the *ncRNA-W6* modulates chromatin topology trough alternative loops formation at *HaWRKY6* locus. Schematic representation of the detected tissue-specific loops of the chromatin in the *HaWRKY6* locus. Ligated restriction sites are notes with a *.

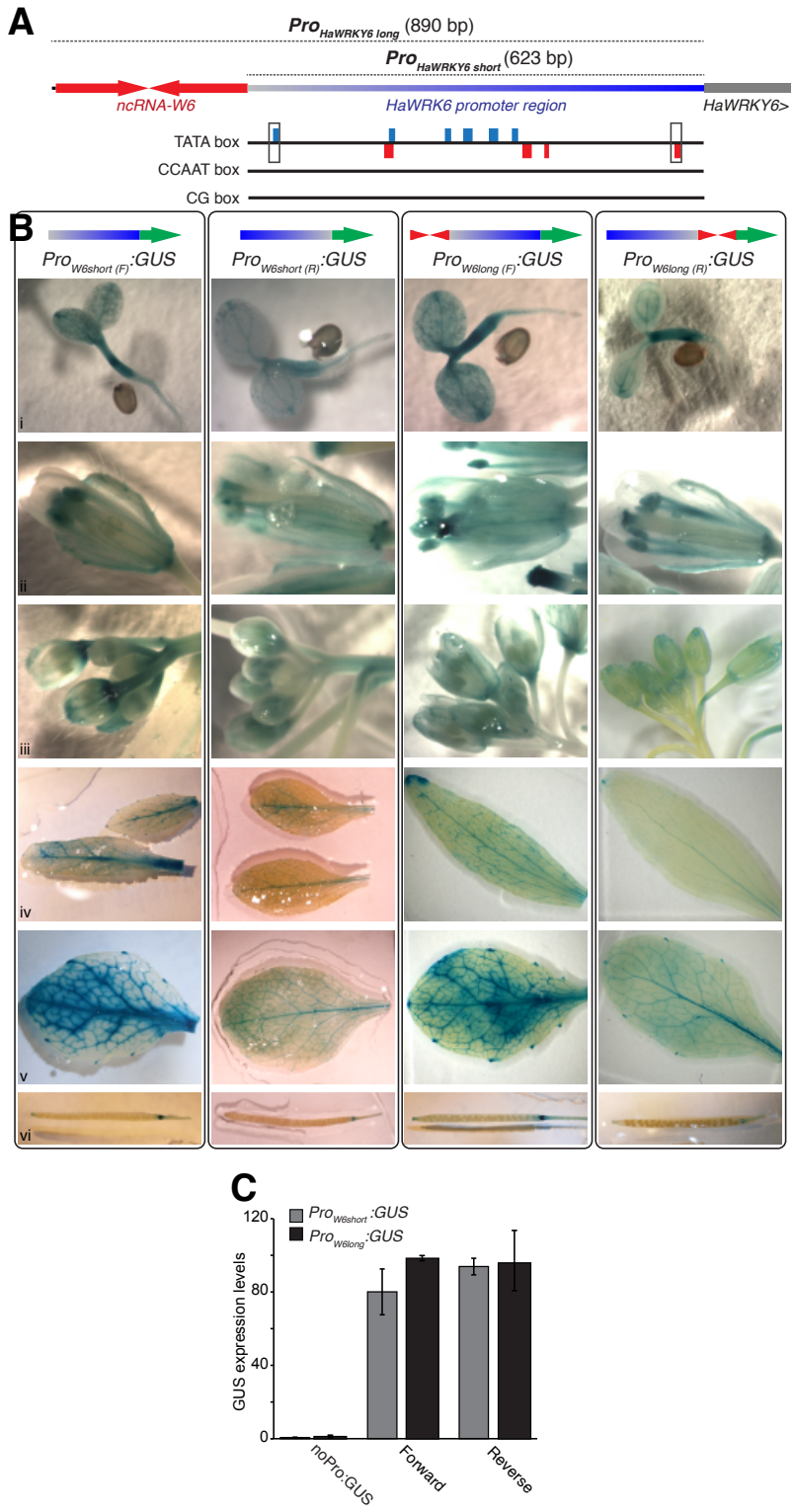

Figure S4, Chromatin looping in the *HaWRKY6* locus changes the transcription directionality of *HaWRKY6* promoter. A. Schematic representation of the *HaWRKY6*

regulatory region showing the segments cloned as “short and Long” promoters. On the bottom is shown a informatics prediction of TATA, CCAAT and CG boxes. Empty boxes highlight the -25 to -35 regions from each transcript as potential sites of active TATA boxes. **B.** Histochemical GUS staining of 3-day-old seedlings, 34-day-old caulinar and rosette leaves, siliques, flowers and inflorescences of Arabidopsis plants transformed with different promoter constructs cloned in sense (F) or antisense (R) orientation, to the reporter gene. **C.** Expression levels of the reporter gene GUS quantified by RT-qPCR in transiently transformed sunflower cotyledons. Error bars represent 2 x SEM, p-values of less than 0.05 (**) or 0,01 (*) in a two-tailed, unpaired t-test were considered significant.

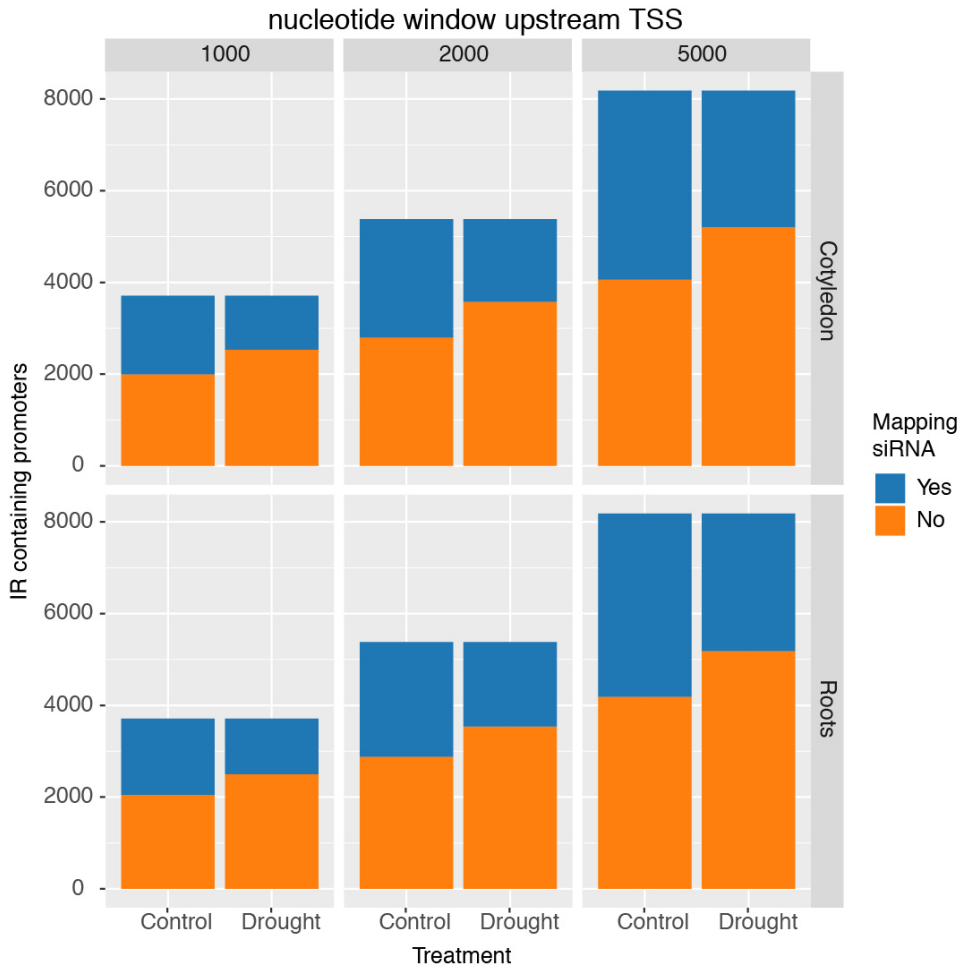

Figure S5, IR prediction and siRNA mapping in Sunflower genes promoter regions. IR prediction was performed using the sequences 1000, 2000 and 5000 upstream the

transcription start sites (TSS) of all Sunflower annotated genes. Small RNAs population in control of drought treated plants (roots and cotyledons) was mapped against the predicted IR regions. Promoter containing IRs with mapped siRNAs are noted in blue while IRs without mapped siRNAs are in orange.

Supplemental Tables

Table S1. Related to STAR Methods. Transgenes. Constructs are based on pGREEN and confer either Basta or kanamycin resistance in plants. Additional information available upon request.

Transgene	**Name**	**Description**	**Reference**
35S::ncRNA W6 F	pDG007	35S promoter (Pro_{35S}) driving *ncRNA-W6*.	This work
35S::ncRNA W6 R	pDG008	Pro_{35S} driving antisense *ncRNA-W6*.	This work
$Pro_{HaWRKY6\ short\ (F)}$:GUS	pDG009	Short version of the *HaWRKY6* promoter, excluding the *ncRNA-W6* region ($Pro_{HaWRKY6\ short}$) driving *Beta-Glucuronidase (GUS)* reporter gene.	This work
$Pro_{HaWRKY6\ short\ (R)}$:GUS	pDG010	$Pro_{HaWRKY6\ short}$ antisense driving GUS.	This work
$Pro_{HaWRKY6\ (F)}$:GUS	pDG011	*HaWRKY6* promoter driving GUS.	This work
$Pro_{HaWRKY6\ (R)}$:GUS	pDG012	*HaWRKY6* promoter antisense driving GUS.	This work
pUBQ::ncRNA W6 R	pDG014	UBIQUITIN-10 promoter driving antisense *ncRNA-W6*.	This work
gHaWRKY6	pDG019	Genomic fragment of the *HaWRKY6* locus from the 5' end of the *ncRNA-W6* region to 93 bp downstream region 3.	This work
*gHaWRKY6(*ΔIR)	pDG020	Genomic fragment of the *HaWRKY6* locus excluding the *ncRNA-W6* IR region to 93 bp downstream region 3.	This work

Table S2. Related to STAR Methods. DNA oligonucleotide primers and probes.

Gene/small RNA	**Sequence**	**Purpose**
HaWRKY6	TTTGCCCAAAGTAACATACTT	RNA blot

HaWRKY6 new	AAGTATGTTACTTTGGGCAAA	RNA blot
U6	GCTAATCTTCTCTGTATCGTTCC	RNA blot
ACTIN	F: GGTAACATTGTGCTCAGTGGTGG R: GGAGATCCACATCTGCTGGAATG	qPCR
Beta-glucuronidase	F: TCCGATCACCTGCGTCAATG R: ACAGCACATCAAAGAGATCGC	qPCR
Basta	F: CTCCCGCAGACGGACGAGG R: AAGCACGGTCAACTTCCGTA	qPCR
Kanamycin	F: CTATCAGGACATAGCGTTGGCTAC R: GATACCGTAAAGCACGAGGAAG	qPCR
HaWRKY6	F: ATACGCTTTAGCAAATGAGAGAGAG R: AGCAAGTTGTGAGGTGAAATTTGGG	qPCR
ncRNA W6	F: ACTTTTCATTTTGTAACGTTTGGAGG R: GTCCCTGTGGTTTGCACAAAG	qPCR
Bisulfite reg 1	F: GAAAGYTTTGTGTTTGGATTATTAAAAAAG R: TTCATTTARCATTAAATTAATTTTTTCTCA	Bisulphite-PCR-sequencing
Bisulfite reg 2	F: AGATAAYGATGTYTGGTGGTATAA R: CCACATAAACCTARAAARACCAAAA	Bisulphite-PCR-sequencing
Bisulfite reg 3	F: TGTGGATTGGTGATYYAYAAAAGATTA R: AAAACAATCCRCTCAATTCAAATTCAAT	Bisulphite-PCR-sequencing
Mitochondrial ATPase	F: TGAAYGAGATTYAAGYTGGGGAAATGGT R: CCCTCTTCCATCAATARRTACTCCCA	Bisulphite-PCR-sequencing
Region 1	F: AAATTTGTAAAAAGTGAGAAGAYATGAAAA R: CATRRTTTTCAATAAAATTTATTCCAAAAT	Chop-qPCR
Region 2	F: GTCCCTAGATGATTCTGTAGTCCCTG R: TATTCATGTCACCCAGGTACAGTG	Chop-qPCR
Region 3	F: TCGGATCGGATCGGCTCATGAAC R: CAAAAGATTACAGAGATTCGGCTC	Chop-qPCR

Loop 1	F: ATTAATGAAAGCTTTGTGTTTGG R: GTGTCTTCTTATAAAACCGAGCG	3C-PCR
Loop 1	F: TCGGATCGGATCGGCTCATGAAC R: GGGACTATGTAATTAACTCTGAG	3C-qPCR
Loop 2	F: GCCACATCAGTTACCCTTGATC R: GTGGAAACCTAACAATAGAG	3C-PCR
Loop 2	F: GATCATCATGACACCCCAGTAA R: GGCCAAATGACTAAGGGTTTG	3C-qPCR
Loop 1 from transgene	F: ACAAGTTTGTACAAAAAAGCAGGCT R: GTGGAAACCTAACAATAGAG	3C-PCR
Loop 1 from transgene	F: ACAAGTTTGTACAAAAAAGCAGGCT R: GGCCAAATGACTAAGGGTTTG	3C-qPCR
Loop 2 from transgene	F: GGGACTATGTAATTAACTCTGAG R: AACTCAGTAGGATTCTGGTGTGTGC	3C-PCR
Loop 2 from transgene	F: GGGACTATGTAATTAACTCTGAG R: TCGGATCGGATCGGCTCATGAAC	3C-qPCR
RNAPII occupancy assay P1	F: ACTTTTCATTTTGTAACGTTTGGAGG R: GTCCCTGTGGTTTGCACAAAG	ChIP-qPCR
RNAPII occupancy assay P2	F: AAACAGGTGTGCCGCAATCTAAC R: GGGACTATGTAATTAACTCTGAG	ChIP-qPCR
RNAPII occupancy assay P3	F: GATCATCATGACACCCCAGTAA R: TCGTCGTCATCTATATAAGCGACTTC	ChIP-qPCR
RNAPII occupancy assay P4	F: GGGACAAGTATAACTCTCTAACAAC R: GACGGGTCTACTTTTACATAAACTC	ChIP-qPCR
RNAPII occupancy assay P5	F: CAAAAATCAATGTCCCTAGATGATTC R: TATTCATGTCACCCAGGTACAGTG	ChIP-qPCR
RNAPII occupancy	F: ATACGCTTTAGCAAATGAGAGAGAG	ChIP-qPCR

assay P6	R: AGCAAGTTGTGAGGTGAAATTTGGG	
ncRNA-W6 Forward	F: CACCTTCTAGATTAAATTGAGAAGAGTTAATTAC R: AGAGTTAATTACATAGTCCCTGTGG	Cloning
ncRNA-W6 Reverse	F: AGAGTTAATTACATAGTCCCTGTGG R: CACCTTCTAGATTAAATTGAGAAGAGTTAATTAC	Cloning
HaWRKY6 promoter short Forward	F: CACCTAATTAACTCTGAGAAAAAATTAATTTAATGC R: TGTTAAGTTAGTTGATGGTTTGAAG	Cloning
HaWRKY6 promoter short Reverse	F: TGTTAAGTTAGTTGATGGTTTGAAG R: CACCTAATTAACTCTGAGAAAAAATTAATTTAATGC	Cloning
HaWRKY6 promoter long Forward	F: CACCTTCTAGATTAAATTGAGAAGAGTTAATTAC R: TGTTAAGTTAGTTGATGGTTTGAAG	Cloning
HaWRKY6 promoter long Reverse	F: TGTTAAGTTAGTTGATGGTTTGAAG R: CACCTTCTAGATTAAATTGAGAAGAGTTAATTAC	Cloning
HaWRKY6 genomic	F: CACCTTCTAGATTAAATTGAGAAGAGTTAATTAC R: GTGTCTTCTTATAAAACCGAGCG	Cloning
HaWRKY6 genomic ΔIR	F: CACCTAATTAACTCTGAGAAAAAATTAATTTAATGC R: GTGTCTTCTTATAAAACCGAGCG	Cloning

1. Agorio, A. and P. Vera, *ARGONAUTE4 is required for resistance to Pseudomonas syringae in Arabidopsis.* Plant Cell, 2007. **19**(11): p. 3778-90.
2. Deleris, A., et al., *Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense.* Science, 2006. **313**(5783): p. 68-71.
3. Hellens, R.P., et al., *pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation.* Plant Mol Biol, 2000. **42**(6): p. 819-32.
4. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.* Plant J, 1998. **16**(6): p. 735-43.
5. Manavella, P.A. and R.L. Chan, *Transient transformation of sunflower leaf discs via an Agrobacterium-mediated method: applications for gene expression and silencing studies.* Nat Protoc, 2009. **4**(11): p. 1699-707.
6. de Felippes, F.F. and D. Weigel, *Transient assays for the analysis of miRNA processing and function.* Methods in molecular biology, 2010. **592**: p. 255-64.
7. Tomassi, A.H., et al., *Nonradioactive Detection of Small RNAs Using Digoxigenin-Labeled Probes.* Methods Mol Biol, 2017. **1640**: p. 199-210.
8. Zhang, H., et al., *Protocol: a beginner's guide to the analysis of RNA-directed DNA methylation in plants.* Plant Methods, 2014. **10**: p. 18.
9. Foerster, A.M. and O. Mittelsten Scheid, *Analysis of DNA methylation in plants by bisulfite sequencing.* Methods Mol Biol, 2010. **631**: p. 1-11.

10. Gruntman, E., et al., *Kismeth: Analyzer of plant methylation states through bisulfite sequencing.* BMC Bioinformatics, 2008. **9**:371.
11. Louwers, M., et al., *Tissue- and expression level-specific chromatin looping at maize b1 epialleles.* Plant Cell, 2009. **21**(3): p. 832-42.
12. Sikorskaite, S., et al., *Protocol: Optimised methodology for isolation of nuclei from leaves of species in the Solanaceae and Rosaceae families.* Plant Methods, 2013. **9**: p. 31.
13. Ariel, F., et al., *Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop.* Mol Cell, 2014. **55**(3): p. 383-96.
14. Badouin, H., et al., *The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution.* Nature, 2017. **546**(7656): p. 148-152.
15. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
16. Hahne, F. and R. Ivanek, *Visualizing Genomic Data Using Gviz and Bioconductor.* Methods Mol Biol, 2016. **1418**: p. 335-51.
17. Krzywinski, M., et al., *Circos: an information aesthetic for comparative genomics.* Genome Res, 2009. **19**(9): p. 1639-45.