

## **Supplementary Materials and Methods**

### **Plant growth and nucleic acid extraction**

Hybrids were made between *Arabidopsis thaliana* C24 and Landsberg *erecta* or Columbia by hand pollination. Parental lines, hybrids and selfed F2 seeds were sterilized and stratified for 2 days at 4°C and sown onto MS medium. At 18 days plants were transferred to soil (Debco seed raising mix supplemented with 1g/L Osmocote pellets) and grown in 16hr day cycle at 22°C/18°C (day/night). For McrBC qRT-PCR and bisulphite PCR immature inflorescence was collected in pools of plants for parents and F1 hybrids while F2 plants were individually collected. For deep sequencing immature inflorescence was collected as a pool for parental lines while immature inflorescence was collected from individuals for the F1 and F2 generations. For the analysis of developmental timing, McrBC qRT-PCR was carried out on 3Day after sowing (DAS) whole seedlings, 7 DAS aerial tissue, 15 DAS leaves, 21 DAS leaves, 28 DAS four largest leaves, and immature inflorescence. Material was frozen in liquid nitrogen and stored at -80°C for processing.

### **McrBC qRT-PCR**

Frozen immature inflorescence were ground in a mortar and pestle. DNA was extracted using the Plant DNeasy Minikit (Qiagen) following manufactures instructions. 50ng of genomic DNA was used for a negative control and a positive control. The positive control was digested with 1U McrBC enzyme, with both the negative control and positive control incubated at 37°C for 16 hours followed by heat inactivation at 65°C for 15 minutes. 10µl of a 1/20 dilution of the DNA was then used in a real-time reaction using the Corbett RG-6000 machine. All reactions were normalized to the *ACTIN* gene and then to the undigested control. PCR conditions were: step 1: 94°C-10 min; step 2: 95°C-20 sec, 58°C-20 sec, 72°C-20 sec (45 cycles). In figures containing McrBC graphs, digestion levels have been inverted to represent <sup>m</sup>C levels. Primer sequences are provided in Table S3.

### **Genotyping of At1g11450**

At1g11450 was amplified using primers described in Table S3. PCR conditions were 95°C for 5min; 95°C for 30sec, 58°C for 30sec, 72°C for 1min, 36 cycles; 72°C for 10 min. The PCR product was then digested with dra1. The PCR product from Ler contains no restriction digest sites, while C24 contains 2 restriction sites.

### **Bisulphite PCR**

Frozen immature florescence were ground in a mortar and pestle. DNA from immature inflorescence was extracted using the Plant DNeasy Minikit (Qiagen) following manufactures instructions. 500ng to 2µg of genomic DNA was then bisulphite converted using the MethylEasy Xceed kit (Human Genetic Signatures). PCR conditions for At5g26345 were 95°C for 5min; 95°C for 1min, 55°C for 2min, 72°C for 3min, 5 cycles; 95°C for 1min, 55°C for 2min, 72°C for 1.5min, 40 cycles; 72°C for 6 min. The C24 and Col alleles for At4g24620 were amplified using two different PCR settings: 95°C for 5min; 95°C for 30sec, 55°C for 30sec, 72°C for 40sec, 40 cycles for Col alleles and 95°C for 5min; 95°C for 1min, 50°C for 2min, 72°C for 40sec, 5 cycles; 95°C for 1min, 53°C for 2min, 72°C for 40sec, 40 cycles; 72°C for 6 min for C24

alleles. Primer sequences are provided in Table S3. <sup>m</sup>C levels were determined from at least 12 clones. For the hybrids this involved obtaining at least 12 clones for each allele.

### **Whole genome bisulphite sequencing (Methyl-seq)**

Whole genome bisulphite sequencing was carried out as described (12). 5µg of genomic DNA from 6 individual F2 plants was sonicated to 200-500bp in size. The true-seq DNA sample prep kit (Illumina) was used to construct libraries which were then converted twice using the EPITECT bisulphite kit (Qiagen). Libraries underwent 12 cycles of amplification using KAPA HiFi Hot start Uracil+ Ready Mix (KAPA Biosystems). Libraries were then sequenced on a Hiseq 2500 obtaining pair-ended 100bp reads. Bisulphite conversion of libraries was checked using the chloroplast genome as a control (Table S4).

### **mRNA and sRNA sequencing**

Total RNA (including large and small) was isolated from immature inflorescence using the Direct-zol RNA Miniprep (Zymos Research). Both mRNA and sRNA libraries consisted of 2 biological replicates (pools of plants) for parental lines, with three biological replicates for the reciprocal hybrids. F1 replicates represented single plants which were the parental line for each of the F2 plants sequenced. The F2 plants sequenced were the same plants used in the methyl-seq experiment. Both mRNA and sRNA library preparation was performed at the Australian Genomic Research Facility, Melbourne Australia. mRNA libraries underwent 100bp pair-ended sequencing and sRNA libraries underwent SE 35bp sequencing on a Hiseq 2500 (Illumina; Table S5 & S6).

### **Methyl-seq analysis**

Methyl-seq libraries were mapped to the Arabidopsis Col TAIR 10 reference genome using Bismark v0.13.0 with the following parameters: bowtie 1 -q -n 2 -1 20 -k 2 --best --chunksmb 512. The parental and F1 hybrid lines were mapped using single end reads and were downloaded from the Gene expression omnibus GSE34658 and GSE35542 (Table S4). F2 individual libraries were mapped using pair ended alignments. <sup>m</sup>CG DMRs between parents were defined from 100bp tiled windows which had at least an 80% difference in <sup>m</sup>C, at least 10 total reads over the tile from each sample and contained at least two CG sites. <sup>m</sup>CHG DMRs between parents were defined from 100bp tiled windows which had at least a 50% difference in <sup>m</sup>C, at least 10 total reads over the tile from each sample and contained at least two CHG sites. Due to the high variability in <sup>m</sup>CHH, DMRs were defined when there was at least 10% difference in <sup>m</sup>CHH in immature inflorescence and at least 5% difference in <sup>m</sup>C within 15DAS samples (same methylated parent). The number of CHH DMRs was further filtered by removing any DMR in which the low methylated parental value was above 10%. Any 100bp DMRS which were adjacent to each other were merged into a single DMR. Windows which had an average read coverage below 5 in an F2 sample was excluded from the analysis.

Due to technical differences between old and new data sets, a scaling (d/a) normalization method was used to define regions with altered <sup>m</sup>C states in the F2 with the methylated parent normalized to 1, the low/un methylated parent normalized to -1 and the MPV equalling 0. The following equation was used for normalization:  $met_{da} = \frac{x - avg(y,z)}{\max(y,z - avg(y,z))}$

where  $x$  = initial <sup>m</sup>C value from sample,  $y$  = C24 and  $z$  = *Ler*. Regions which were homozygous for one parent were used to define the inheritance of altered <sup>m</sup>C states. TCM regions were defined when the F2 individual was homozygous for the low methylated parent and had a d/a normalization value of > 0.5. TCdM regions were defined when the F2 individual was homozygous for the high methylated parent and had a d/a normalization value <-0.5. Bedgraph files of <sup>m</sup>C level and coverage were loaded onto the CoGe genome browser. Whisker box plots were created in R using the ggplot2 package. Outlying values larger than normalized values 2 or -2 were excluded from the whisker box plots and statistical test. An ANOVA test in R was used to look for statistical significance between F1 15DAS <sup>m</sup>C normalized values and F1 Bud <sup>m</sup>C normalized values.

### **Mutant analysis**

<sup>m</sup>C values for Columbia and mutants in *met1*, *cmt2*, *cmt3*, *ddm1*, *ago4*, *polIV* and *polV* were downloaded from GSE39901 (21). The intersect function from BEDTools v2.24.0 with options `-wao` was used to map <sup>m</sup>C's from these files to the F2 inherited windows for each methyl context. The BEDTools function `groupby` with `-o` options `count, mean` was used to obtain coverage and average <sup>m</sup>C level for Col and the mutants in the inherited windows. Windows with no coverage in the Stroud et al data sets or were unmethylated were removed from the analysis. <sup>m</sup>CG methylated windows required 80% <sup>m</sup>C, <sup>m</sup>CHG methylated windows required at least 50% <sup>m</sup>C, while <sup>m</sup>CHH methylated windows required at least 10% <sup>m</sup>C. Enzymes important for the maintenance of <sup>m</sup>C at each window were defined when the <sup>m</sup>C in the mutants was different by at least 80% for <sup>m</sup>CG, 50% for <sup>m</sup>CHG or 10% for <sup>m</sup>CHH. Bedgraph files of <sup>m</sup>C level and coverage were loaded onto the CoGe genome browser.

### **SNP analysis**

Bis-SNP was used to detect SNPs within mapped reads from the methyl-seq reads. Output SNPs were then compared to known SNPs downloaded from the 1001 genome project (<http://www.1001genomes.org>). Only SNPs common to both lists were retained for genotyping. 10kb bins of SNP read numbers were used to define regions of the F2 as C24 homozygous, *Ler* homozygous and heterozygous. 10kb bins in which the genotype could not be defined were excluded from all analyses and the genotype designated NA (Table S1).

### **sRNA seq analysis**

Before mapping all sRNA libraries had adapters trimmed and tRNAs removed using custom perl scripts. sRNA reads were multi-mapped to the TAIR10 *Arabidopsis thaliana* Columbia genome using Biokanga v2.95 with the following options `-r5 -R500 -X` (Table S5). sRNAs were multi-mapped to avoid the possibility of calling a Methyl-seq window independent of sRNAs due to the removal of repetitive sRNA reads which would have mapped to the methyl window. The intersect function in BEDTools V2.24.0 was used to count the number of small RNA reads over each parental differentially methylated window. Only sRNAs where  $\geq 50\%$  of the read overlapped the differentially methylated windows <sup>m</sup>C window were retained. Samples were then normalized to reads per 10 million. This was done separately for <sup>m</sup>CG, <sup>m</sup>CHG and <sup>m</sup>CHH. <sup>m</sup>C windows were defined as sRNA-associated when at least one sample contained  $\geq 5$  normalized reads.

## **mRNA seq analysis**

mRNA libraries were first run through FastQC to check quality. Libraries were then mapped to the TAIR10 *Arabidopsis thaliana* Col reference genome using Biokanga v2.95 (<https://sourceforge.net/projects/biokanga/>) with the option -y45 to trim the mRNA reads (Table S6). The R package DESEQ (S1) was then used to determine differentially expressed genes using default settings. F2 expression levels were compared to the parent which matched the F2 genotype (C24, Ler or MPV). Differentially expressed genes required a p-value  $\leq 0.05$ .

## **Feature analysis**

The intersect function of BEDTools v2.24.0 with the -c option was used to count the number of Total DMRs and inherited DMRS in each feature. Each methyl context was treated separately. Windows were annotated to GeneBody/Pseudogene/Transposable Element Gene > Upstream > Downstream > Intergenic regions. These features were then split into TE associated or TE independent using the TAIR10 TE annotations. A Chi-square test was used to test for the enrichment for a particular feature in the inherited Windows. Upstream and downstream regions were  $\pm 1$ kb of the gene. If the adjacent gene was within 1kb of the gene, then the upstream or downstream region was truncated to the start/end of the next gene.

The intersect function in BEDTools v2.24.0 was used to intersect <sup>m</sup>C windows with TE positions in the genome. This was done for total DMR windows and for inherited windows. A chi-square test was used to test whether inherited windows were enriched for the presence of TEs

## **Candidate genes for expression changes associated from <sup>m</sup>C changes.**

For the scatterplots genes were identified as additive (contained either no <sup>m</sup>C window or showed no change in <sup>m</sup>C in the F2), TCM (genes which contained a TCM inherited) or TCdM (genes that contained a TCdM inherited window). Genes were further subdivided based on average expression level across all samples. Genes were divided into highly expressed (Top 25%), lowly expressed (bottom 25%) and middle expressed genes (Middle 50%).

The intersect function of BEDTools v2.24.0 with the -wa and -wb options were used to intersect inherited windows with upstream, downstream and gene bodies. Each list was then filtered for genes that contained inherited windows. Genes were further filtered by only retaining genes where an F2 individual showed a change in gene expression compared to parent ( $p \leq 0.05$ ) and the presence of an inherited window. From this list 34 genes were identified that are candidates for the altered <sup>m</sup>C states influencing gene expression patterns in the F2.

## **Northern Blots.**

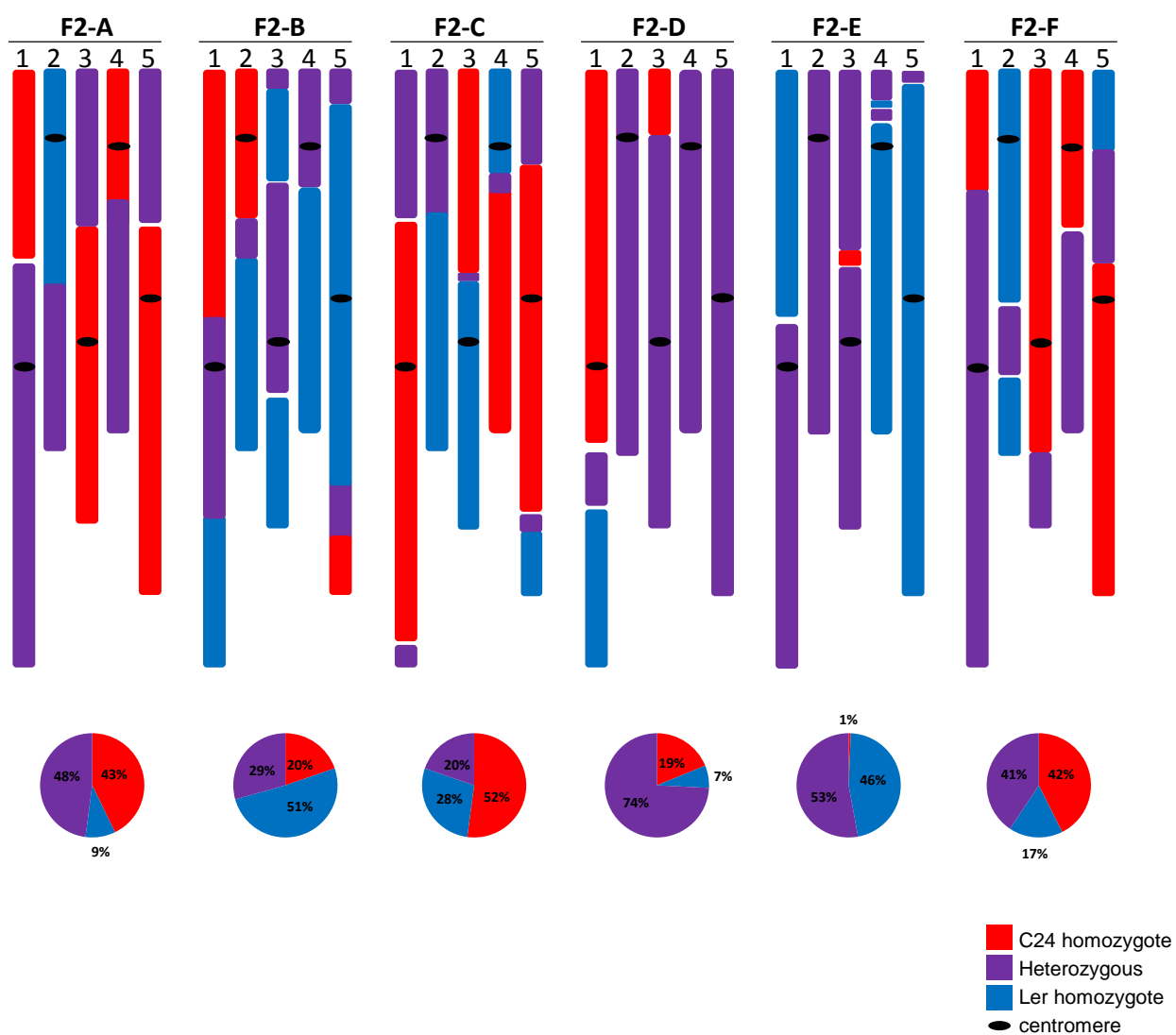
Northern Blots were carried out as previously described (S2). Primers can be found in Table S3.

## **Supplementary references.**

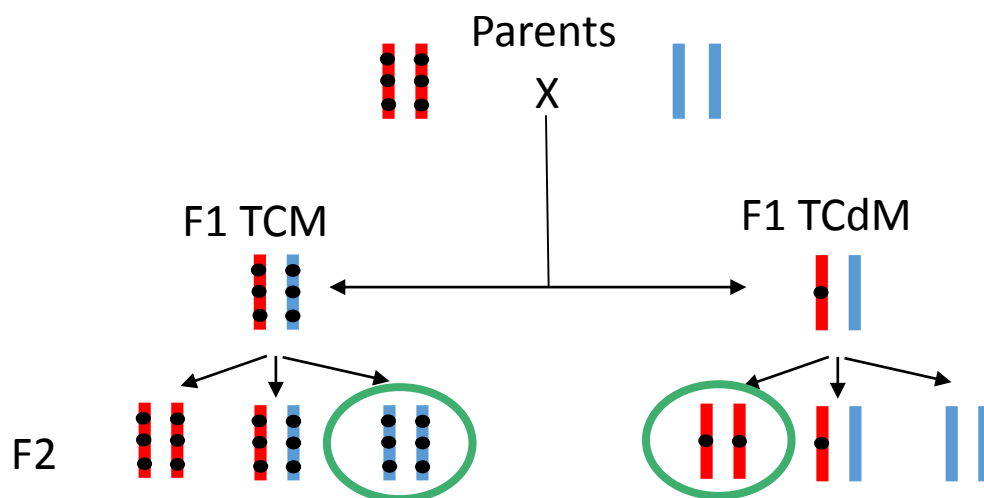
- S1. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* 11(10)
- S2. Smith NA, Eamens AL, & Wang MB (2010) The presence of high-molecular-weight viral RNAs interferes with the detection of viral small RNAs. *Rna-a Publication of the Rna Society* 16(5):1062-1067.

# Supplementary figure 1

A.

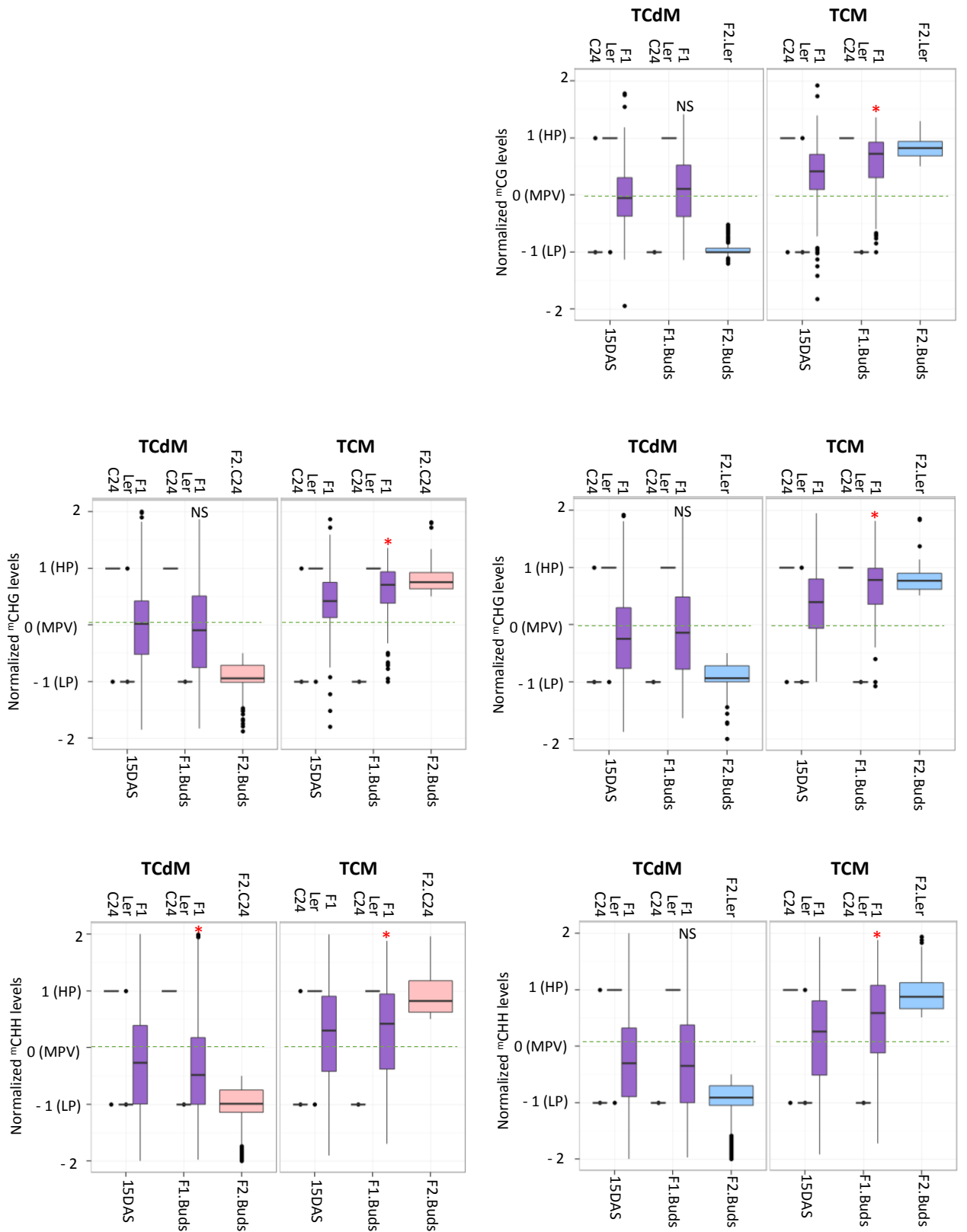


B.



**Figure S1: Genotypes of F2 individuals.** A) Genotypes for each F2 individual. Pie charts give overall genotype frequency in each individual. Red represent C24 homozygous, blue represents *Ler* homozygous and purple represents heterozygous genotypes. B) Experimental design. F2 plants were analyzed for changes in  $^{13}\text{C}$  in regions homozygous for one of the parents. TCdM inherited windows were identified at homozygous regions in the F2 with  $^{13}\text{C}$  levels lower than that of the parental genotype. TCM inherited windows were identified at homozygous regions in the F2 individuals which demonstrated a  $^{13}\text{C}$  level above the parental genotype. The green circles highlight the allele combinations required to define inherited windows.

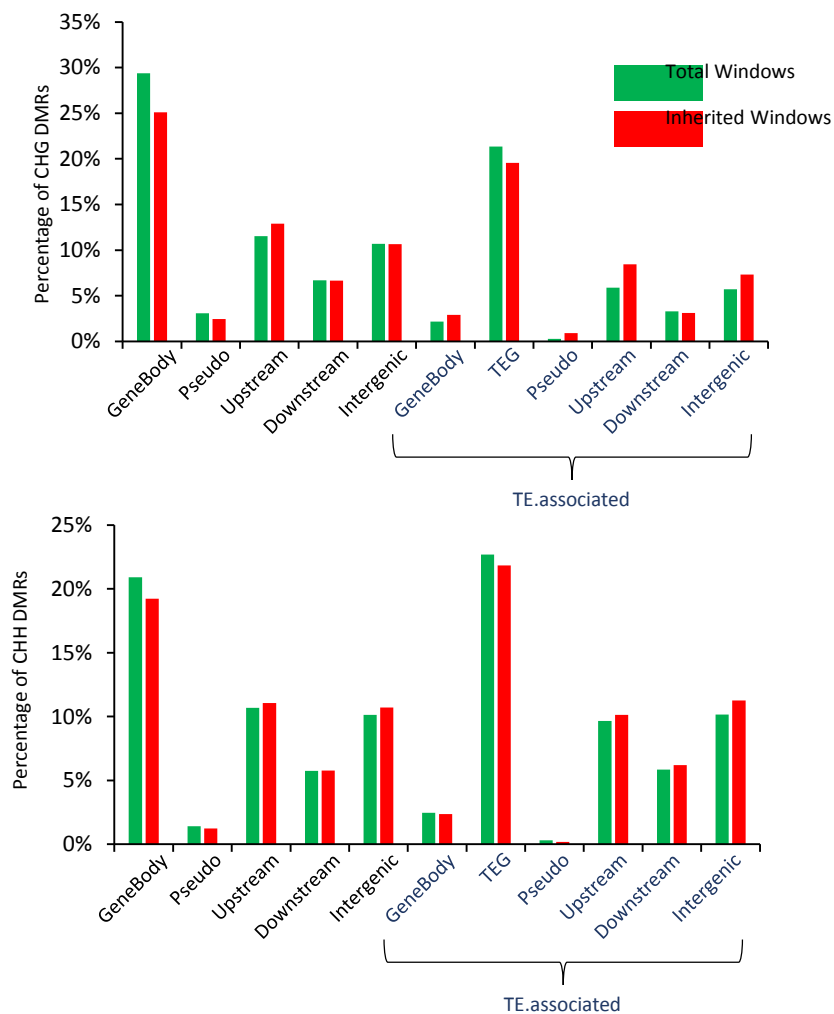
# Supplementary figure 2





**Figure S2: Inherited windows in the F2 show TCM in two F1 time points.** Whisker box plots of <sup>m</sup>C in F1 15 das seedlings (15DAS), F1 immature inflorescence (F1.Buds) and F2 immature inflorescence (F2.Buds). F1 <sup>m</sup>C levels at the two points are represented in purple. F1 <sup>m</sup>C levels represent windows in the F1 which were homozygous for C24 or homozygous for Ler in the F2. C24.F2 represents inherited windows which were homozygous for C24 (light red). *Ler*.F2 represents inherited windows homozygous for *Ler* (light blue). Y axis represents context specific normalized <sup>m</sup>C levels with C24 or *Ler* as the methylated parent (HP; 1) or low methylated parent (LP; -1). MPV (0) is represented as a dotted green line. The \* above F1.Buds represents a statistical difference in normalized <sup>m</sup>C levels between F1 15DAS and F1 Buds ( $p \leq 0.01$ ; ANOVA test). NS = normalized <sup>m</sup>C levels between F1 15DAS and F1 Buds which are not statistically significant.

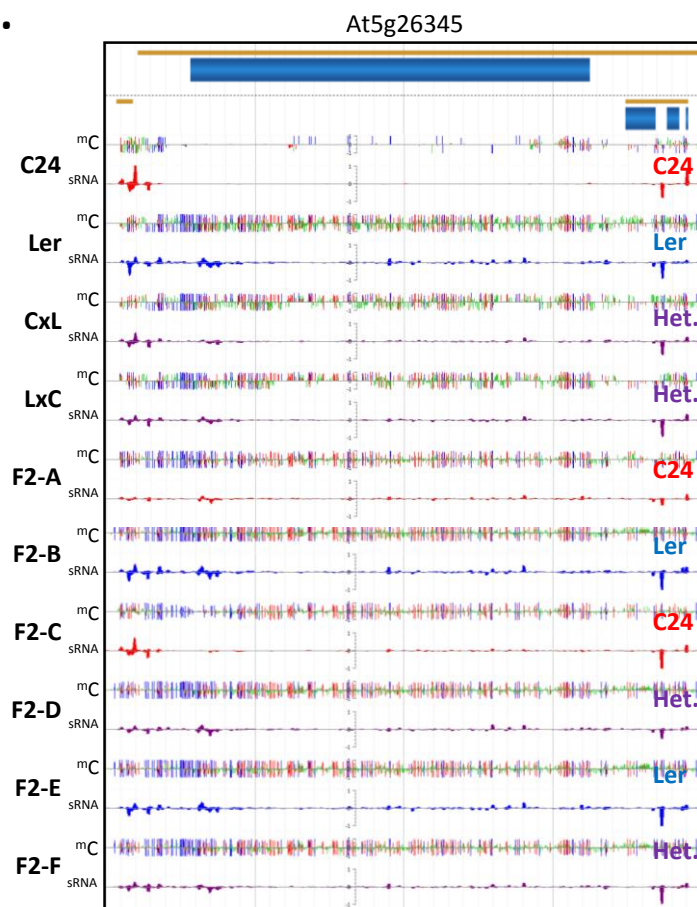
## Supplementary figure 3



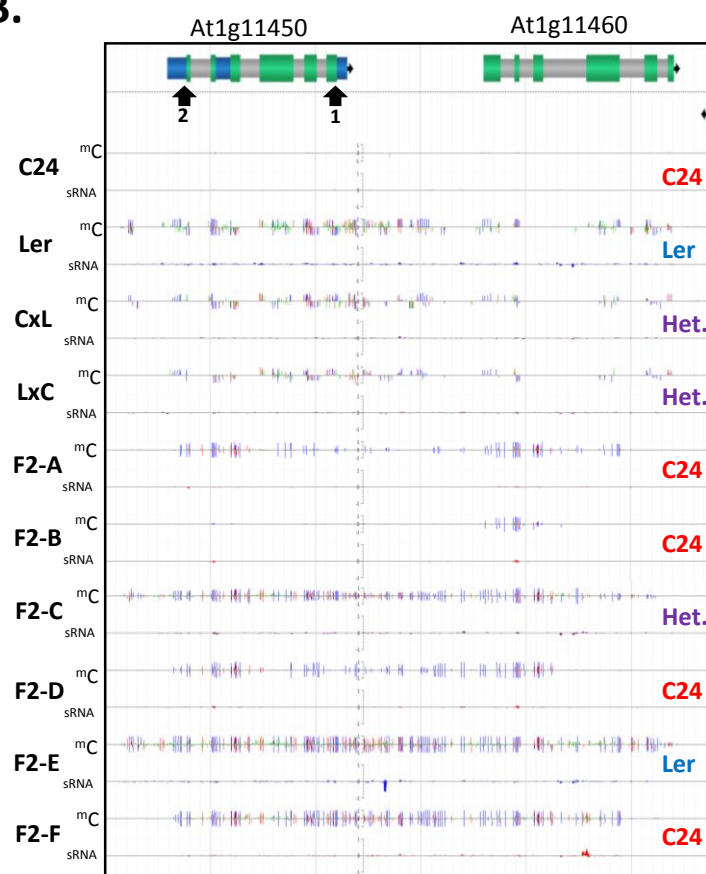
**Figure S3: Genome locations of inherited windows for  $mCHG$  and  $mCHH$ .** Features were split into those overlapping TEs (TE.associated) and features not associated with TEs. Features mapped to include intragenic (GeneBody), Transposable Element Genes (TEG), Pseudogenes (Pseudo), 1kb upstream (Upstream), 1kb Downstream (Downstream), Intergenic

# Supplementary figure 4

A.

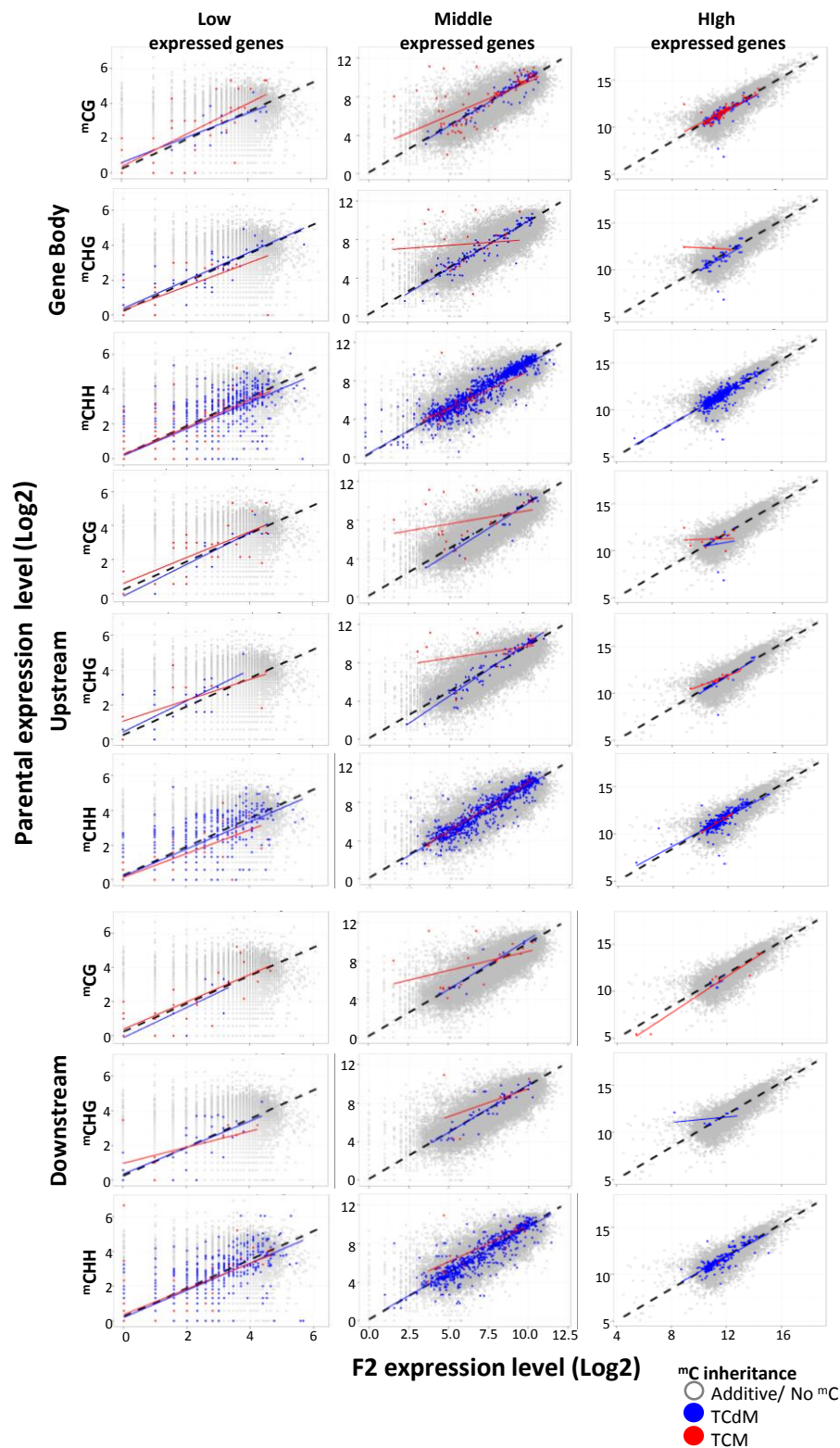


B.



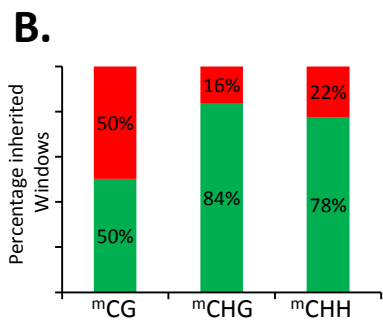
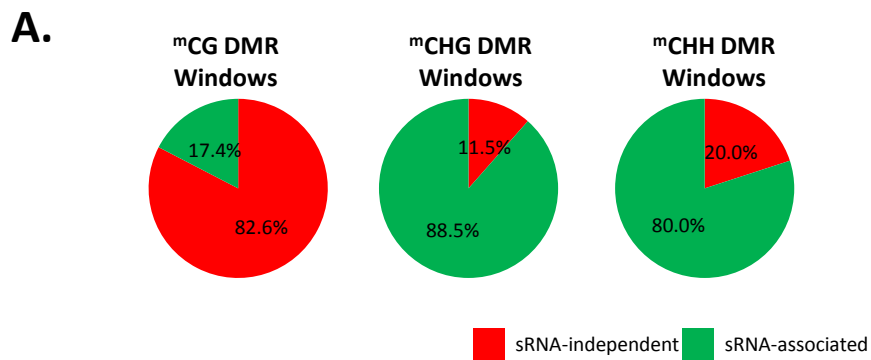
**Figure S4: Examples of stable (A) and variable inheritance (B).** Genome browser shot of <sup>m</sup>C and sRNA levels in parents, C24 x *Ler* F1 hybrid (CxL), *Ler* x C24 F1 hybrid (LxC) and six individual F2 plants. In sRNA tracks genotypes are on the right of each track with sRNA track colour representing C24 homozygous (red), *Ler* homozygous (blue) and heterozygous (purple). For <sup>m</sup>C tracks blue represents <sup>m</sup>CG, red represents <sup>m</sup>CHG and green represents <sup>m</sup>CHH. In (B) arrows represent primer sets used in Figure 3A.

## Supplementary figure 5

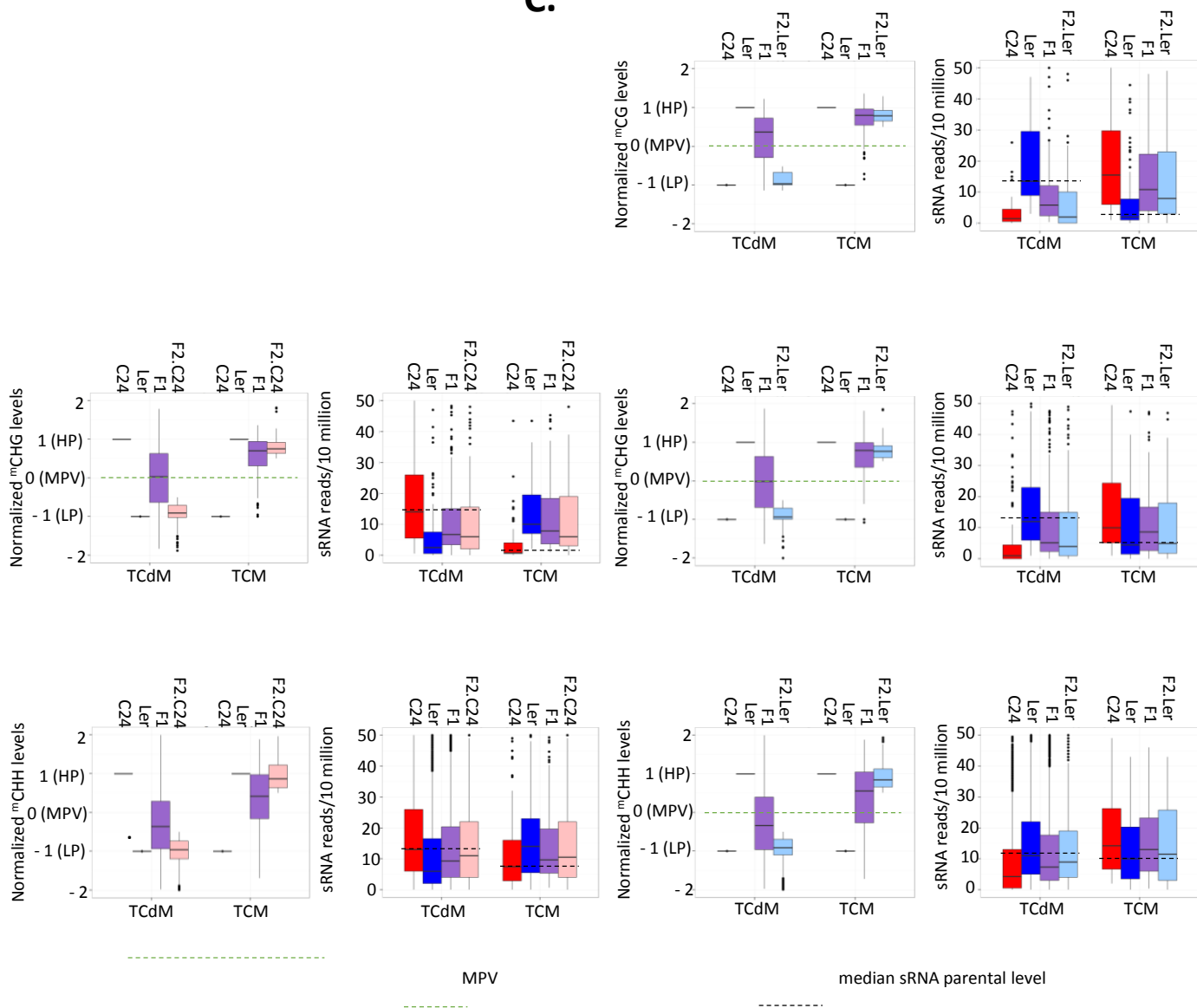


**Figure S5: mC inherited windows impact on gene expression.** mC scatterplots of gene expression levels in homozygous regions of the F2 (X axis) and expression level in the parent of origin (Y axis). Genes with additive mC or no mC are in grey. Genes overlapping a TCM inherited window are in red and Genes overlapping a TCdM window are in blue. Regression lines for additive, TCM and TCdM are drawn in dotted black, red and blue respectively.

# Supplementary figure 6

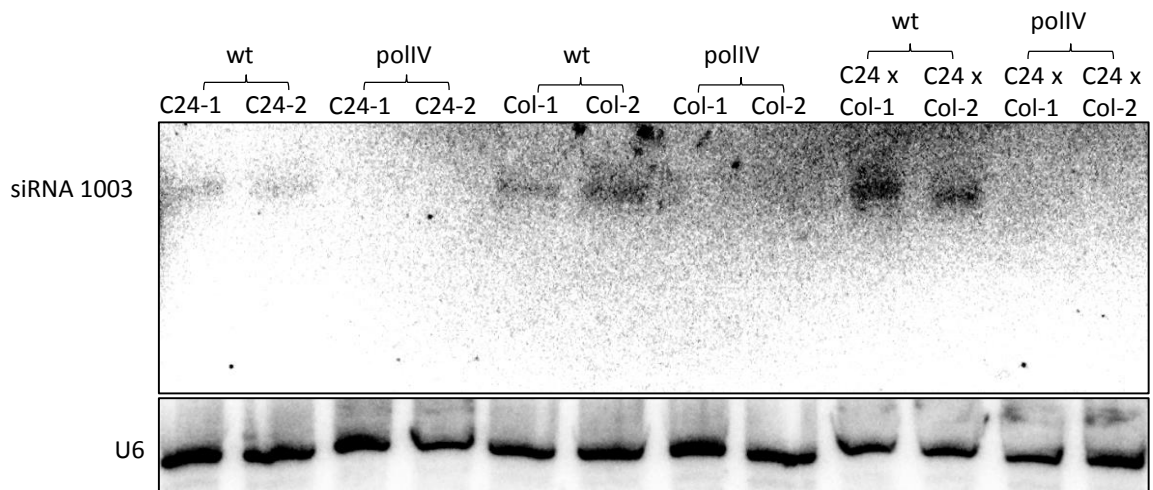


**C.**



**Figure S6: Relationship between inherited windows and sRNAs.** A) Proportion of DMR windows which were sRNA-associated or sRNA-independent. B) Proportion of inherited windows which were sRNA-associated or sRNA-independent. C) Whisker box plots of <sup>m</sup>C levels in sRNA-associated inherited windows and their corresponding levels of sRNAs in parental lines, F1 hybrids (immature inflorescence) and the F2 generation. C24 is represented as red with F2 homozygous for C24 (C24.F2) represented as light red. *Ler* is represented as blue with F2 homozygous for *Ler* (*Ler*.F2) represented as light blue. In the <sup>m</sup>C whisker box plots Y axis represents context specific normalized <sup>m</sup>C levels with C24 or *Ler* as the methylated parent (HP; 1) or low methylated parent (LP; -1). TCM defines inherited windows with <sup>m</sup>C levels above that of the parental genotype. TCdM defines inherited windows with <sup>m</sup>C levels below that of the parental genotype.

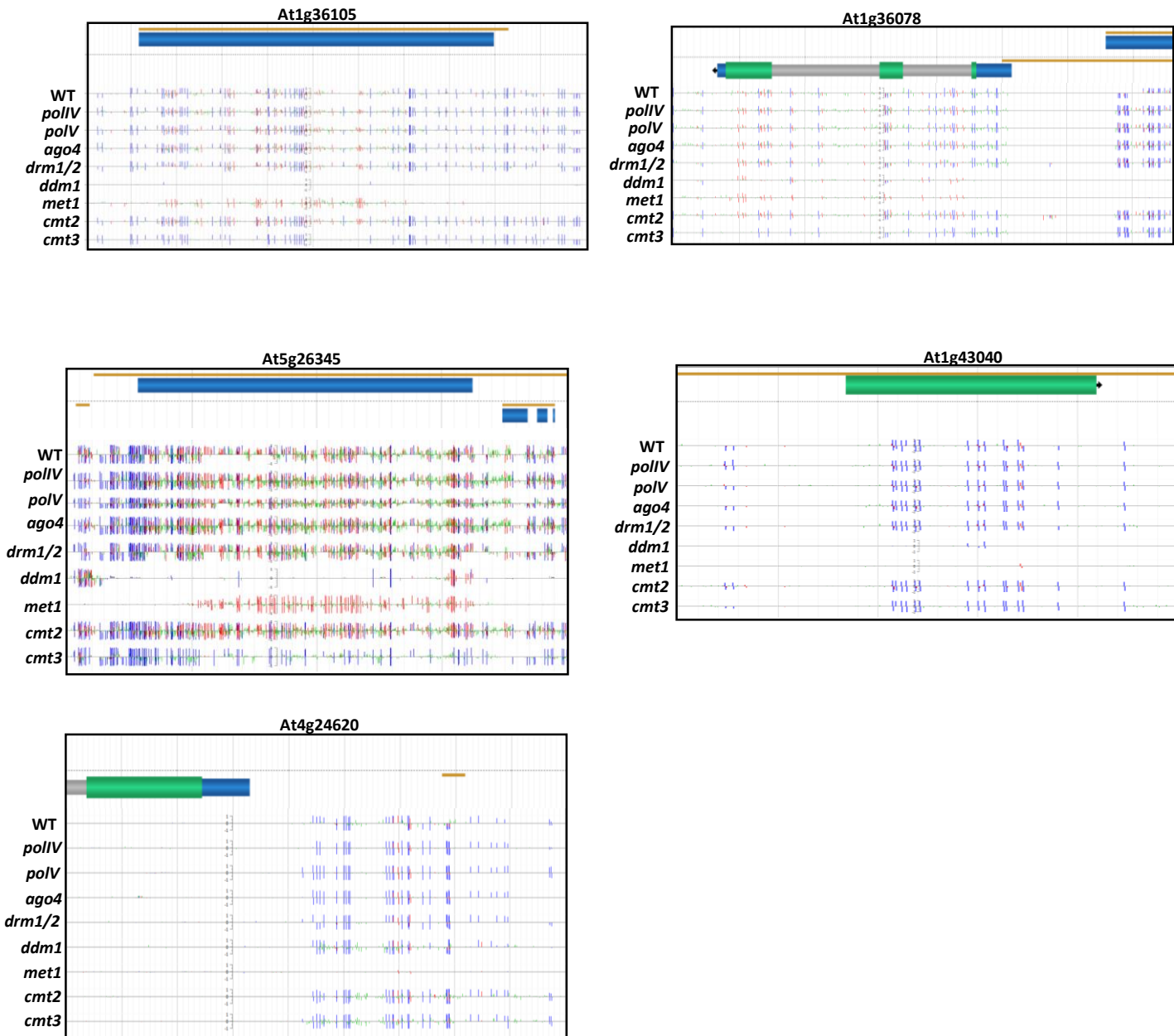
## Supplementary figure 7



**Figure S7: Importance of sRNAs for the initiation of TCM events.** *polIV* hybrids lack the 24nt siRNA 1003. There are two biological replicates for each sample. U6 acts as a loading control for the northern.



## Supplementary figure 8



**Figure S8: mC patterns at regions which undergo TCM in mutants for mC maintenance enzymes.** A browser shot of Col wt and Col mutant lines at five regions known to undergo TCM in C24/Ler hybrids. For mC tracks blue represents mCG, red represents mCHG and green represents mCHH.

# Supplementary Table 1

F2-A				F2-B				F2-C				F2-D				F2-E				F2-F			
Chr	Start	Stop	Genotype	Chr	Start	Stop	Genotype	Chr	Start	Stop	Genotype	Chr	Start	Stop	Genotype	Chr	Start	Stop	Genotype	Chr	Start	Stop	Genotype
1	0	9870000	C24	1	0	12730000	C24	1	0	7780000	Both	1	0	19040000	C24	1	0	12750000	Ler	1	0	6140000	C24
1	9870001	9919999	NA	1	12730001	12759999	NA	1	7780001	7829999	NA	1	19040001	19419999	NA	1	12750001	12909999	NA	1	6140001	6179999	NA
1	9920000	30427671	Both	1	12760000	22900000	Both	1	7830000	29270000	C24	1	19420000	22140000	Both	1	12910000	30427671	Both	1	6180000	30427671	Both
2	0	11040000	Ler	1	22900001	22929999	NA	1	29270001	29399999	NA	1	22140001	22319999	NA	2	0	19698289	Both	2	0	11890000	Ler
2	11039999	11099999	NA	1	22930000	30427671	Ler	1	29400000	30427671	Both	1	22320000	30427671	Ler	3	0	9260000	Both	2	11890001	11939999	NA
2	11100000	19698289	Both	2	0	7640000	C24	2	0	7490000	Both	2	0	19698289	Both	3	9260001	9279999	NA	2	11940000	15570000	Both
3	0	8290000	Both	2	7640001	7659999	NA	2	7490001	7499999	NA	3	0	3300000	C24	3	9280000	10020000	C24	2	15570001	15609999	NA
3	8290000	23459830	C24	2	7660000	9720000	Both	2	7500000	19698289	Ler	3	3300001	3379999	NA	3	10020001	10039999	NA	2	15610000	19698289	Ler
4	0	6990000	C24	2	9720001	9769999	NA	3	0	10570000	C24	3	3380000	23459830	Both	3	10040000	23459830	Both	3	0	19410000	C24
4	6989999	6999999	NA	2	9770000	19698289	Ler	3	10570001	10579999	NA	4	0	18585056	Both	4	0	1610000	Both	3	19410001	19499999	NA
4	7000000	18585056	Both	3	0	1100000	Both	3	10580000	10940000	Both	5	0	26975502	Both	4	1610001	1619999	NA	3	19500000	23459830	Both
5	0	7980000	Both	3	1100000	5790000	Ler	3	10940001	10999999	NA	4	1620000	2050000	Ler	4	0	8080000	C24	4	0	8080000	C24
5	7979999	8009999	NA	3	5790001	5859999	NA	3	11000000	23459830	Ler	4	2050001	2099999	NA	4	8080001	8139999	NA	4	8080001	8139999	NA
5	8010000	26975502	C24	3	5860000	16620000	Both	4	0	5490000	Ler	4	2100000	2780000	Both	4	8140000	18585056	Both	4	8140000	18585056	Both
				3	16620001	16779999	NA	4	5490001	5499999	NA	4	2780001	2789999	NA	5	0	4090000	Ler	5	0	4090000	Ler
				3	16780000	23459830	Ler	4	5500000	6350000	Both	4	2790000	18585056	Ler	5	4090001	4099999	NA	5	4090001	4099999	NA
				4	0	6100000	Both	4	6350001	6359999	NA	5	0	750000	Both	5	4100000	9990000	Both	5	4100000	9990000	Both
				4	6100000	18600000	Ler	4	6360000	18585056	C24	5	750001	759999	NA	5	9990000	26975502	C24	5	9990000	26975502	C24
				5	0	1900000	Both	5	0	4950000	Both	5	760000	26975502	Ler								
				5	1900001	1909999	NA	5	4950001	4969999	NA												
				5	1910000	21280000	Ler	5	4970000	22620000	C24												
				5	21280001	21289999	NA	5	22620001	22759999	NA												
				5	21290000	23820000	Both	5	22760000	23640000	Both												
				5	23820001	23829999	NA	5	23640001	23669999	NA												
				5	23830000	26975502	C24	5	23670000	26975502	Ler												

**Table S1: F2 genotypes.** C24 means C24 homozygous, Ler means Ler homozygous and Both means heterozygous regions. Regions where the genotype could not be defined were labelled as NA.

# Supplementary Table 2

Gene.ID	Gene family	Name	Short description
AT1G11450	protein_coding	nodulin MtN21 /EamA-like transporter family protein	nodulin MtN21 /EamA-like transporter family protein
AT1G11460	protein_coding	nodulin MtN21 /EamA-like transporter family protein	nodulin MtN21 /EamA-like transporter family protein
AT1G33970	protein_coding	P-loop containing nucleoside triphosphate hydrolases superfamily protein	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G36078	protein_coding	unknown protein	.
AT1G42852	transposable_element_gene		transposable element gene
AT1G47210	protein_coding	cyclin-dependent protein kinase 3	cyclin-dependent protein kinase 3;2
AT1G70890	protein_coding	MLP-like protein 43 (MLP43)	MLP-like protein 43
AT1G78160	protein_coding	pumilio 7 (PUM7)	pumilio 7
AT2G15160	transposable_element_gene		transposable element gene
AT2G16140	transposable_element_gene	similar to DNA binding [Arabidopsis thaliana] (TAIR:AT3G47680.1)	transposable element gene
AT2G25370	protein_coding	RING/U-box superfamily protein	RING/U-box superfamily protein
AT2G29350	protein_coding	senescence-associated gene 13 (SAG13)	senescence-associated gene 13
AT2G43160	protein_coding	ENTH/VHS family protein	ENTH/VHS family protein
AT3G32023	transposable_element_gene		transposable element gene
AT3G43350	transposable_element_gene	similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G52960.1)	transposable element gene
AT3G44205	transposable_element_gene		transposable element gene
AT3G50625	transposable_element_gene		transposable element gene
AT4G04030	protein_coding	ovate family protein 9 (OFP9)	ovate family protein 9
AT4G09640	protein_coding	Protein of unknown function (DUF803)	Protein of unknown function (DUF803)
AT4G10201	pseudogene		.
AT4G10220	protein_coding	Protein of Unknown Function (DUF239)	Protein of Unknown Function (DUF239)
AT4G16162	protein_coding	Leucine-rich repeat (LRR) family protein	Leucine-rich repeat (LRR) family protein
AT4G20480	protein_coding	Putative endonuclease or glycosyl hydrolase	Putative endonuclease or glycosyl hydrolase
AT4G26140	protein_coding	beta-galactosidase 12 (BGAL12)	beta-galactosidase 12
AT5G07770	protein_coding	Actin-binding FH2 protein	Actin-binding FH2 protein
AT5G20750	transposable_element_gene	similar to Ulp1 protease family protein [Arabidopsis thaliana] (TAIR:AT3G42690.1)	transposable element gene
AT5G26345	transposable_element_gene		transposable element gene
AT5G27345	transposable_element_gene		transposable element gene
AT5G47530	protein_coding	Auxin-responsive family protein	Auxin-responsive family protein
AT5G52067	pseudogene		.
AT5G54585	protein_coding	unknown protein	.
AT5G59630	pseudogene		.
AT5G60430	protein_coding	drug transmembrane transporters	drug transmembrane transporters;antiporters
AT5G66380	protein_coding	folate transporter 1 (FOLT1)	folate transporter 1

**Table S2: Candidate genes.** Genes which may have their expression patterns influenced by the inheritance of altered <sup>m</sup>C states into the F2 generation.

# Supplementary Table 3

## Primers for McrBC qRT-PCR

At3g43340 forward	GGTCACGCTCACAGTTCTCTC
At3g43340 reverse	TCTTTTAGACTCCACCGCTCA
At5g26345 forward	AGAGACAGCAATTGCGAAG
At5g26345 reverse	TGCAAGAATCTCACCTTCA
At4g24620 forward	TGATACGATGGTAGCTTTTCG
At4g24620 reverse	AACTTTAACAGGATCCGACG
At1g36078 forward	ACTGGGGACAAATACTTACTGC
At1g36078 reverse	AACATTCTGCTGCTCGTAGT
At1g11450-1 forward	GCTATACAGTGTGATTGGATCAG
At1g11450-1 reverse	ACTAATTTACGACAATGAAGCCA
At1g11450-2 forward	TTCAAGAAAGAGAGAACGAGGG
At1g11450-2 reverse	AAATAGGCGATAGGAACCAGA
ACTIN forward	CTAAGCTCTCAAGATCAAAGGC
ACTIN reverse	AACATTGCAAAGAGTTTCAAGG

## genotype primers for At1g11450

At1g11450 geno forward	GCTATACAGTGTGATTGGATCAG
At1g11450 geno reverse	TCTCCCAATATGGGTATCGAA

## Primers for Bisulphite PCR

At4g24620 forward	AATGTTGATTGGGAGTAGGGTTGTT
At4g24620 reverse	ACRTAATTACATTTTTATCTTTTTRTTT
At5g26345 forward	ATTAAGGATGAGATAATGATGAGGTATA
At5g26345 reverse	ATCTCCATACACCTTCTTCAATTA

## Primers for Northern

siRNA 1003	ATGCCAAGTTTGGCCTCACGGTCT
U6	AGGGGCCATGCTAATCTTCTC

**Table S3: Primer sets.** For bisulphite PCR primers Y = C or T, while R = A or G

# Supplementary Table 4

Sample	sequence_type	# raw_reads	# ft_reads	%reads_pass_ft	mapping_eff	unique_map	no_align	multi_align	prop_CG	prop_CHG	prop_CHH	conversion_rate
C24 15das	se	151580154	142371568	0.939	41.7	59421294	37083269	45867005	27.7	9.8	4.7	99.49
Ler 15 das	se	136470326	129093613	0.946	42.5	54886741	27808512	46398360	28.8	10.1	4	99.66
CxL 15 das	se	133946590	129209005	0.965	46.2	59727776	19989126	49492103	26.9	8.9	3.2	99.85
LxC 15 das	se	129567407	123082281	0.95	45.1	55549484	22197618	45335179	29.3	10.7	4.2	99.81
C24 Buds	se	95538869	94200750	0.986	52.6	49563153	10304029	34333568	31.7	14.7	4.7	99.4
Ler Buds	se	96575435	96136492	0.995	57.3	55119854	8771838	32244800	30.5	14	4.5	99.51
CxL Buds	se	89633151	89444102	0.998	53.7	48014033	11234834	30195235	39.2	20.8	6.7	99.15
LxC Buds	se	85381984	85153743	0.997	55.5	47278791	9976515	27898437	35.3	17	4.8	99.93
F2-A	pe	227558376	211987838	0.932	56.4	59812488	37988263	8193168	18.3	5	0.9	99.93
F2-B	pe	252962014	234368164	0.926	53.3	62426780	44664048	10093254	23	7.1	1.2	99.93
F2-C	pe	214304058	205718350	0.96	56	57614035	37723678	7521462	20.3	5.4	1	99.95
F2-D	pe	220498502	212091928	0.962	57.7	61233647	35689575	9122742	19.7	5.2	0.9	99.95
F2-E	pe	233844754	228271582	0.976	59.2	67581050	37324805	9229936	20.8	6.1	1.1	99.95
F2-F	pe	201910086	197259618	0.977	58.1	57347740	32768753	8513316	19.4	5.3	0.9	99.95

**Table S4: Mapping of methyl-seq libraries.**

pe = paired end, se = single end

# Supplementary Table 5

	Starting read Numbers	Filtered read numbers (ex. tRNA)	Unique	Multi	No Alignment	loci aligned to	Mapped	Normalization factor to 10,000,000
C24_RepA	38967013	16103555	7645238	6879909	1578401	59121426	14525147	0.688461
C24_RepB	34867470	11018145	5325708	4810210	882226	40672269	10135918	0.986590
Ler_RepA	37121173	11035432	5212509	5070941	751977	40840458	10283450	0.972436
Ler_RepB	40900325	13480295	6982759	5511203	986329	47919445	12493962	0.800387
C24xLer1	16396764	8960089	4714434	3534131	711524	33273727	8248565	1.212332
C24xLer6	16285609	8807191	4716323	3383991	706877	32000136	8100314	1.234520
C24xLer7	18335163	10154555	5295011	4066651	792893	38205803	9361662	1.068186
LerxC245	19203437	11915249	6084895	4910124	920230	45690108	10995019	0.909503
LerxC247	16785334	10362400	5540810	3977353	844237	38585364	9518163	1.050623
LerxC2410	19054826	12673778	6600511	5109443	963824	47638910	11709954	0.853974
F2-A	16150068	9758759	4960628	4018883	779248	38245718	8979511	1.113646
F2-B	19774107	11879938	6699094	4226499	954345	43211657	10925593	0.915282
F2-C	15329531	8607839	4941626	2974428	691785	27616815	7916054	1.263256
F2-D	16076577	10152266	5291239	4063192	797835	39645496	9354431	1.069012
F2-E	16237841	10938496	5790901	4274461	873134	41927212	10065362	0.993506
F2-F	16272803	9682653	5175494	3733943	773216	36451672	8909437	1.122405

**Table S5: Mapping of sRNA libraries.** sRNA libraries were multi-mapped using Biokanga v 2.95.0. For the analysis reads were normalized to reads per 10 million

# Supplementary Table 6

	Starting read Numbers	Aligned	Multi loci	No Alignment
C24_RepA	91458684	55415989	10753486	25047510
C24_RepB	54115674	38286641	5410875	10233292
Ler_RepA	93704896	59255991	10580653	23622095
Ler_RepB	50998016	19195733	25273254	6419668
C24xLer1	21629786	17840289	1042566	2623703
C24xLer6	72572934	60378078	3505332	8286232
C24xLer7	20662110	17051049	755179	2729948
LerxC245	39527950	32916322	5333576	1058505
LerxC247	51728084	39206533	1306651	11039807
LerxC2410	20696852	17502041	625418	2477332
F2-A	32068534	25977421	2093945	3862064
F2-B	61614816	50626864	2029247	8734583
F2-C	153384582	113897228	23732328	15236983
F2-D	74235640	63734278	1591789	8520717
F2-E	49835074	35608591	6532984	7438692
F2-F	87648612	61543354	15318188	10364256

**Table S6: Mapping of mRNA libraries.** mRNA libraries were uniquely mapped using Biokanga v 2.95.0.