Supporting Information Appendix

Supplemental Results

We examined three other classes of loci from our data where, by definition, there was no association between DNA methylation and a mobile signal:

D. "Not methylated" loci correspond to mobile sRNAs loci that have no associated DNA methylation.
E. "Not remethylated" loci correspond to regions with DNA methylation in all WT grafts that overlap a mobile sRNA locus. This methylation is lost in *dcl234*/*dcl234* and is not restored by mobile sRNAs (in C24/*dcl234* grafts).
F. "DCL234-independent" loci correspond to loci methylated in all graft combinations, overlapping with sRNA loci showing similar expression in all graft combinations. It is likely that the DNA methylation at these loci is independent of DCL234-generated sRNAs.

Class D and E loci are associated with the mobile sRNA that does not control DNA methylation, whilst DNA methylation at class F loci does not appear to be associated with mobile sRNA (Fig. S1). We found relatively few such loci (Dataset S1, Table S2) and that they had distinct features from both the class A loci and one another. Class D loci are loci at which we observe mobile sRNAs but no detectable DNA methylation (Fig. S1): they harbor CG (64 loci) and CHG (124 loci), but not CHH, sequences that are not methylated (Table S2). These associations between sRNA loci and unmethylated sequences were significantly depleted relative to the number expected by chance in both CG ($Z=-29.368$, $p<0.05$) and CHG ($Z=-52.712$, $p<0.05$) contexts. However, the unmethylated CG and CHG sequences overlapped statistically significantly (Fig. S3). Class D loci were most significantly associated with TEs in unmethylated CG and CHG contexts, as well as promoters containing TEs in the unmethylated CHG context (Fig. S11). Loci of the unmethylated CG context were significantly associated with Type 1 TEs of the DNA Mariner and RC Helitron superfamilies (Fig. S12). Contrastingly, loci of the unmethylated CHG context targeted a broad distribution of TE superfamilies (Fig. S12; LINEs, RAth elements, several DNA element superfamilies, RC Helitrons). Compared to the class A loci the class D loci in the unmethylated CG context had lower levels of mobile 23-24 nt sRNAs whereas those with CHG context motifs had similar levels of mobile 23-24 nt sRNAs (Figs. S6, S7).

There were few class E loci (20 CG, 9 CHG, 17 CHH, by DNA methylation context) but the intersections between sRNA and DNA methylation loci conforming to this class definition were statistically significant in all three contexts (Table S2). CHH context class E loci overlapped statistically significantly with CG and CHG class D loci (Fig. S3). Similarly few class F loci were identified (Table S2; 7 CG, 8 CHG, 22 CHH), with the distinction that only intersections of sRNAs with CG ($Z=6.4188$, $p=6.87*10^{-11}$) and CHG ($Z=2.1556$, $p=0.015557$) context DNA methylation were statistically significant. CHH context class F loci overlapped statistically significantly with CG and CHG class F loci (Fig. S3). Both class E and F loci were associated significantly with TEs and promoters containing TEs (Fig. S11). Class E loci associated predominantly with a range of DNA elements (Type 1 TEs; Fig. S12), whilst class F loci were prevalent in association with DNA elements (Type 2 TEs). These observations contrast in both cases with the associations of class A loci, described earlier (Fig. 3). Both class E and F loci were associated with lower levels of mobile sRNAs (23, 24 nt size class) than class A loci in the CG and CHG contexts, and comparable levels in the CHH context (Figs. S6, S7).

Supplemental Materials and Methods

1. sRNA analysis

sRNA data were aligned to the TAIR9 reference genome requiring perfect matching. Reads shorter than 15 bases and longer than 36 bases were discarded, as were those that matched to more than one thousand genomic locations. Library scaling factors were calculated as the sum of the lowest 75% of expressed reads (1). sRNA loci were identified using the empirical Bayesian methods of the segmentSeq **R** package (2) using a gap parameter of 100 bases and default options otherwise.

Differentially expressed sRNA loci were identified by considering the numbers of sRNAs from all size classes within a locus, and evaluating the models described below using the baySeq **R** package (3), accounting for both library scaling factor and locus length within the models.

2. Alignment and base calling of MethylC-seq Data

We use the YAMA (Yet Another Methylome Aligner) scripts to align the sequencing data to the TAIR 10 genome (source code available at https://github.com/tjh48/YAMA). This is an unbiased (C-T conversion) aligner based on Bowtie 2 that allows reads to map to multiple locations upon the genome, allowing for greater coverage over repetitive elements. The contribution of an aligned read to the number of methylated/unmethylated cytosines observed at a given location is weighted according to the number of locations to which the read matches.

3. Locus identification from MethylC-seq Data

We use the segmentSeq **R** package to identify methylation loci and maximise the discovery of differentially methylated regions (4). The segmentSeq package classifies all potentially methylated regions as either methylated or not in a set of biological replicates by assessing whether the posterior likelihood of the locus exceeding some proportion of methylation is sufficiently high, based on a beta-binomial distribution with a non-informative Jefferys prior and a normalisation factor based on non-conversion rates. It then selects amongst those regions classified as methylated such that they contain no unmethylated region (by a similar classification) so as to maximise the number of differentially methylated regions reported. Methylation loci are identified independently for the three contexts CG, CHG and CHH. A region is classified as methylated within a replicate group if there is a greater than 99% likelihood that the proportion of methylation exceeds 20%, and not methylated if there is a greater than 99% likelihood that the proportion of methylation is lower than 20%.

4. Differential Methylation

Loci exhibiting various models of differential methylation (Fig. 1) were identified using the empirical Bayesian methods for analysis of paired data in the baySeq **R** package (4). For each sample, the number of methylated cytosines in a locus is paired with the number of unmethylated cytosines in that locus. Each model defines sets of samples under which the paired data are assumed to be distributed according to a beta-binomial distribution whose parameters are estimated empirically from the data. The posterior likelihood of each model at the locus is then calculated given the observed data.

5. Expectation and FDR of Loci

The expected number of loci conforming to any particular pattern of differential methylation/expression can readily be calculated by summing the posterior likelihoods for this pattern over all loci. These expectations define the general behaviour of the different loci without identifying specific loci that conform to a particular pattern. In order to examine overlap between loci and annotation features, we need to select specific sets of loci, which is done by controlling the estimated false discovery rates (FDRs) to be lower than 0.05.

6. Profiles of Average Methylation

Average methylation profiles for a set of genomic coordinates are calculated by sub-dividing each region into an equal number of windows and calculating the mean proportion of methylation within that window (accounting for non-conversion rates). The mean of these means then defines the average methylation profile over the coordinates.

7. mRNA-seq analysis

mRNA-seq sequenced reads were mapped to the combined transcriptome and transposable element TAIR10 reference sequences using RSEM (5). Differentially expressed coding regions and transposable elements were identified by evaluating the posterior likelihoods of the models described above at each locus using the baySeq **R** package (3, 6), accounting for both library scaling factor and gene length. mRNA-seq data were not available for the C24/C24 graft.

8. Tiling array mRNA analysis

We reanalysed the tiling array data of Laubinger and colleagues (GEO reference GSE21685) (7). We used the pre-processed gene expression values provided by Laubinger and colleagues (processed using the TAIR8 reference transcriptome) (7). We separated the data into floral/seedling tissue types, and median normalised these expression values within these tissue types. Differential expression between wild-type and mutant tissues was identified using the limma **R** package (8). Subsequently, genes were selected as differentially expressed using Benjamini & Hochberg FDR control at 0.05 threshold (9). Genes not selected as differentially expressed are reported as ND.

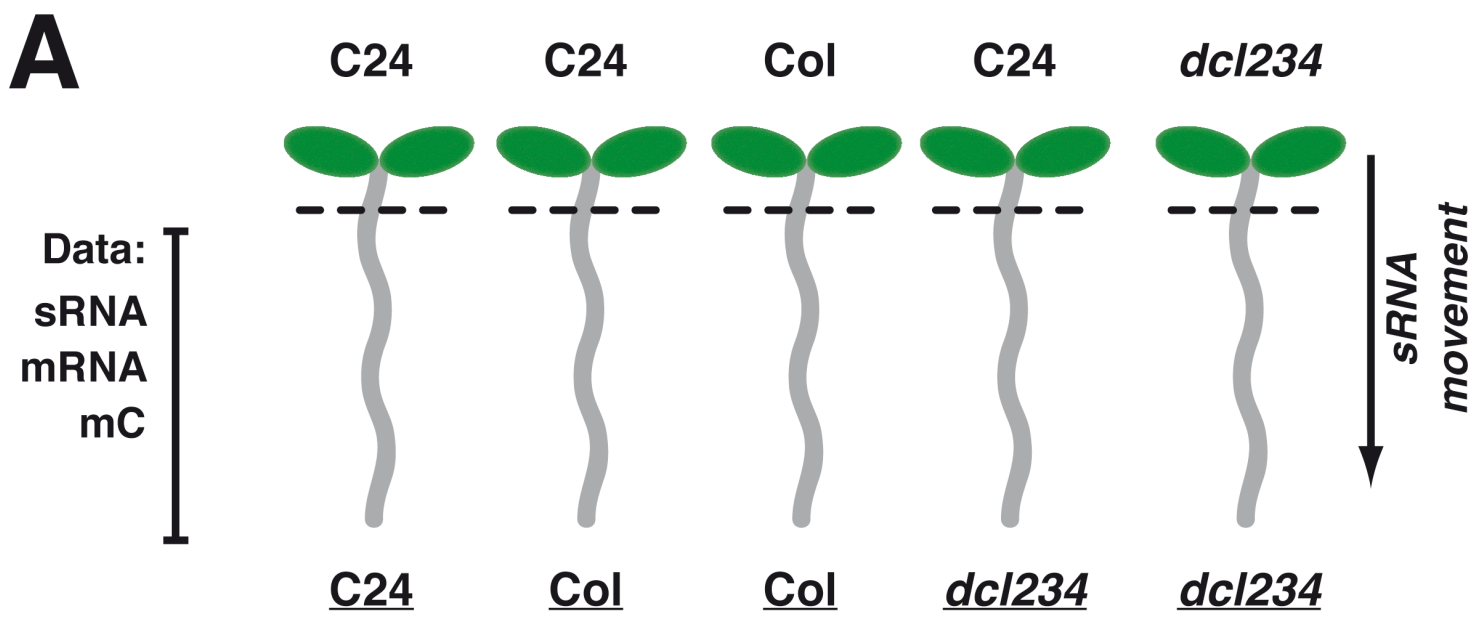
9. Association of Features

We assess association of differentially methylated regions with annotation features and differentially accumulated sRNA loci using a block-bootstrap method (10). This assesses the marginal likelihood of the observed overlap between features under the null hypothesis of no association (either positive or negative) between features, accounting for the non-uniform distribution of features upon the genome. Intersections between loci and genome features were normalized for abundance of loci and features, the sizes of loci and features and how loci and features might cluster within the genome when plotting. We did so by calculating the number of features overlapped by the DNA methylation/sRNA loci per megabase total feature length per megabase total locus length. This approach gave a visually accurate representation of significance of association. Independent measurements of statistical significance are presented beneath the bars also.

Supplemental References

1. Bullard JH, Purdom E, Hansen KD, & Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94.
2. Hardcastle TJ, Kelly KA, & Baulcombe DC (2012) Identifying small interfering RNA loci from high-throughput sequencing data. *Bioinformatics* 28(4):457-463.
3. Hardcastle TJ & Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11:422.
4. Hardcastle TJ (2015) Discovery of methylation loci and analyses of differential methylation from replicated high-throughput sequencing data. *bioRxiv*.
5. Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
6. Hardcastle TJ (2015) Generalised empirical Bayesian methods for discovery of differential data in high-throughput biology. *Bioinformatics* 10.1093/bioinformatics/btv569.
7. Laubinger S*, et al.* (2010) Global effects of the small RNA biogenesis machinery on the Arabidopsis thaliana transcriptome. *Proc Nat Acad Sci USA* 107(41):17466-17473.
8. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Bio* 3:Article3.
9. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289-300.
10. Bickel PJ & Freedman DA (1981) Some Asymptotic Theory for the Bootstrap. *Ann. Statist.* 9(6):1196-1217.
11. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, & Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1-2):352-364.

Supplemental Figures


A
C24
C24
Col
C24
dcl234
Data:
sRNA
mRNA
mC
sRNA movement
C24
Col
Col
dcl234
dcl234

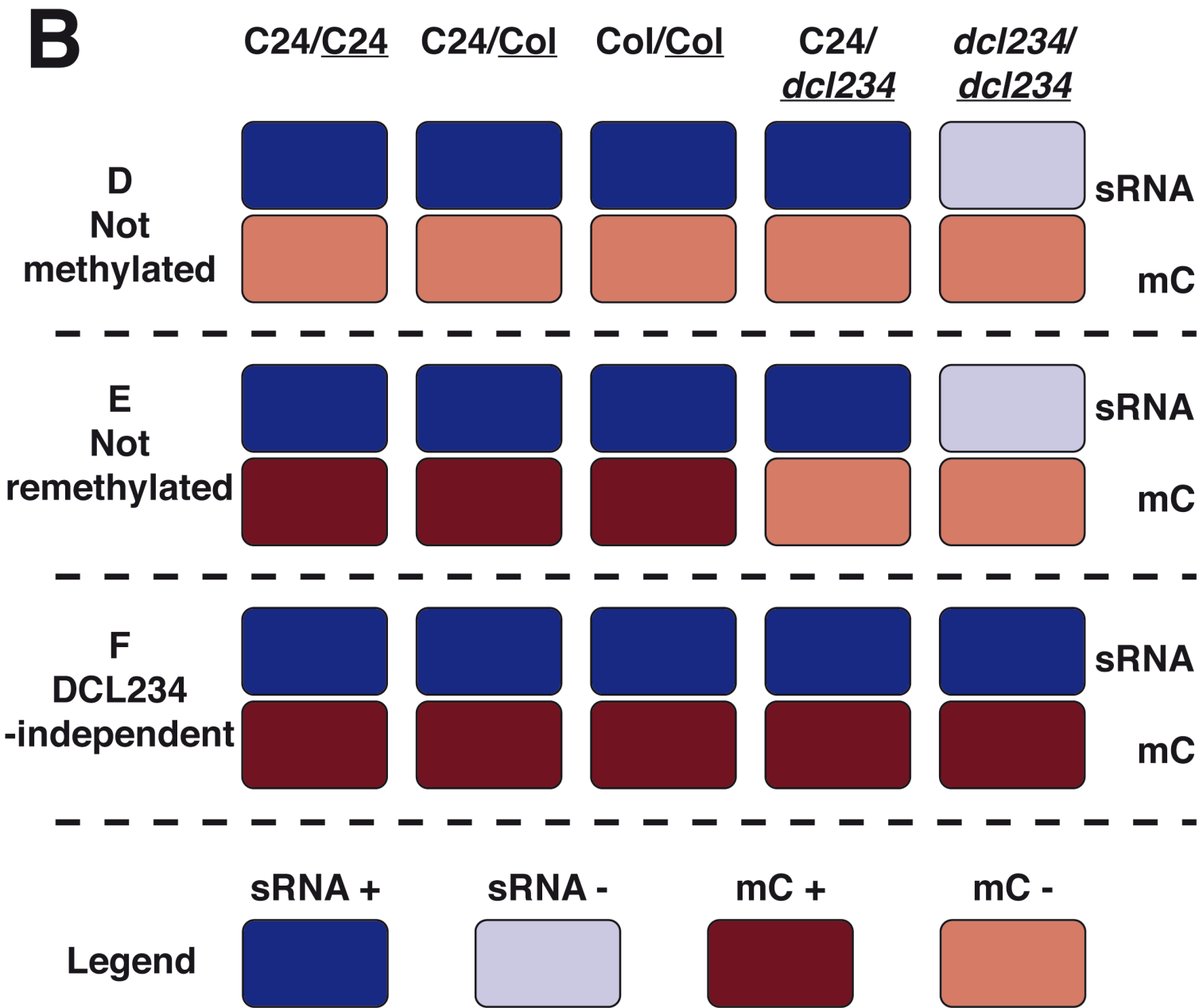

B
C24/C24
C24/Col
Col/Col
C24/ dcl234
dcl234/ dcl234
D
Not methylated
sRNA
mC
E
Not remethylated
sRNA
mC
F
DCL234 -independent
sRNA
mC
Legend
sRNA +
sRNA -
mC +
mC -

Figure S1. Grafting combinations and additional locus classes examined in this study. (A) Grafts were made between shoots and roots of various *Arabidopsis* genotypes. sRNA, mRNA or DNA methylation (mC) libraries were made from root tissue harvested five weeks after grafting. Sequenced tissue is indicated by underlined label. RNA-seq data were not available for C24/C24 grafts. (B) Locus classes D-F were defined using the same approach as for classes A-C, described in Fig. 1. Graft combinations are indicated at top, denoted shoot/root. Target loci of sRNAs and DNA methylation were identified by analyzing MethylC-seq and sRNA-seq data from roots of all graft combinations. Each locus classification was defined by a specific combination of sRNA and DNA methylation (mC) levels across the five graft combinations. "+" denotes a relatively high level of mC or sRNA, whilst "-" denotes a relatively low level. Classifications were designated D-F and are indicated on the left.

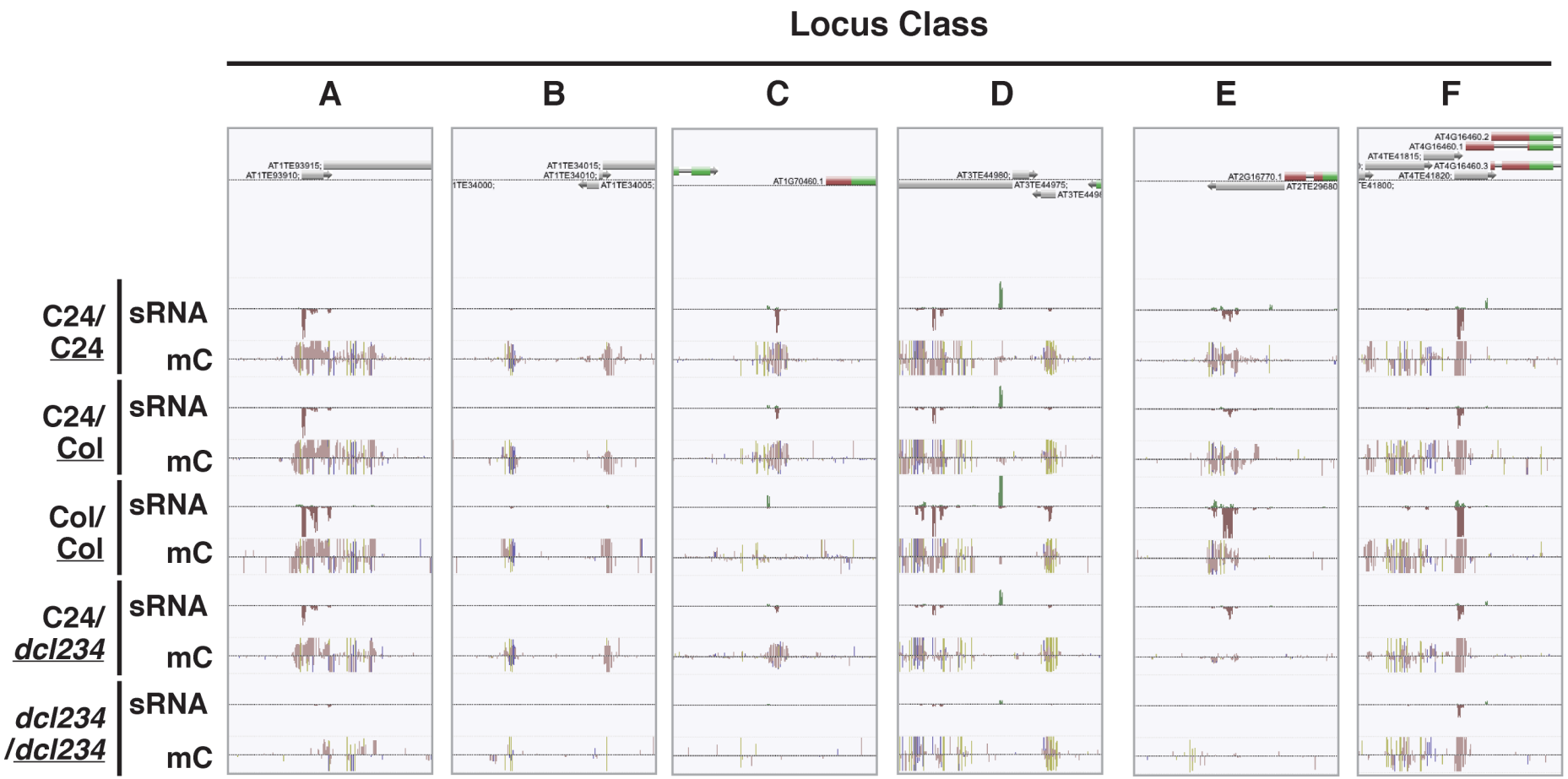

Locus Class
A
B
C
D
E
F
C24/ C24
sRNA
mC
C24/ Col
sRNA
mC
Col/ Col
sRNA
mC
C24/ dcl234
sRNA
mC
dcl234 /dcl234
sRNA
mC

Figure S2. Illustrative genome browser screenshots for each locus class. Tracks are grouped by graft combination, showing sRNA-seq (sRNA) and MethylC-seq (mC) data for each graft combination. Grafts indicated by notation shoot/root, with underline indicating sequenced tissue. Note that the same example of a class C locus is amongst those presented in Fig. 5.

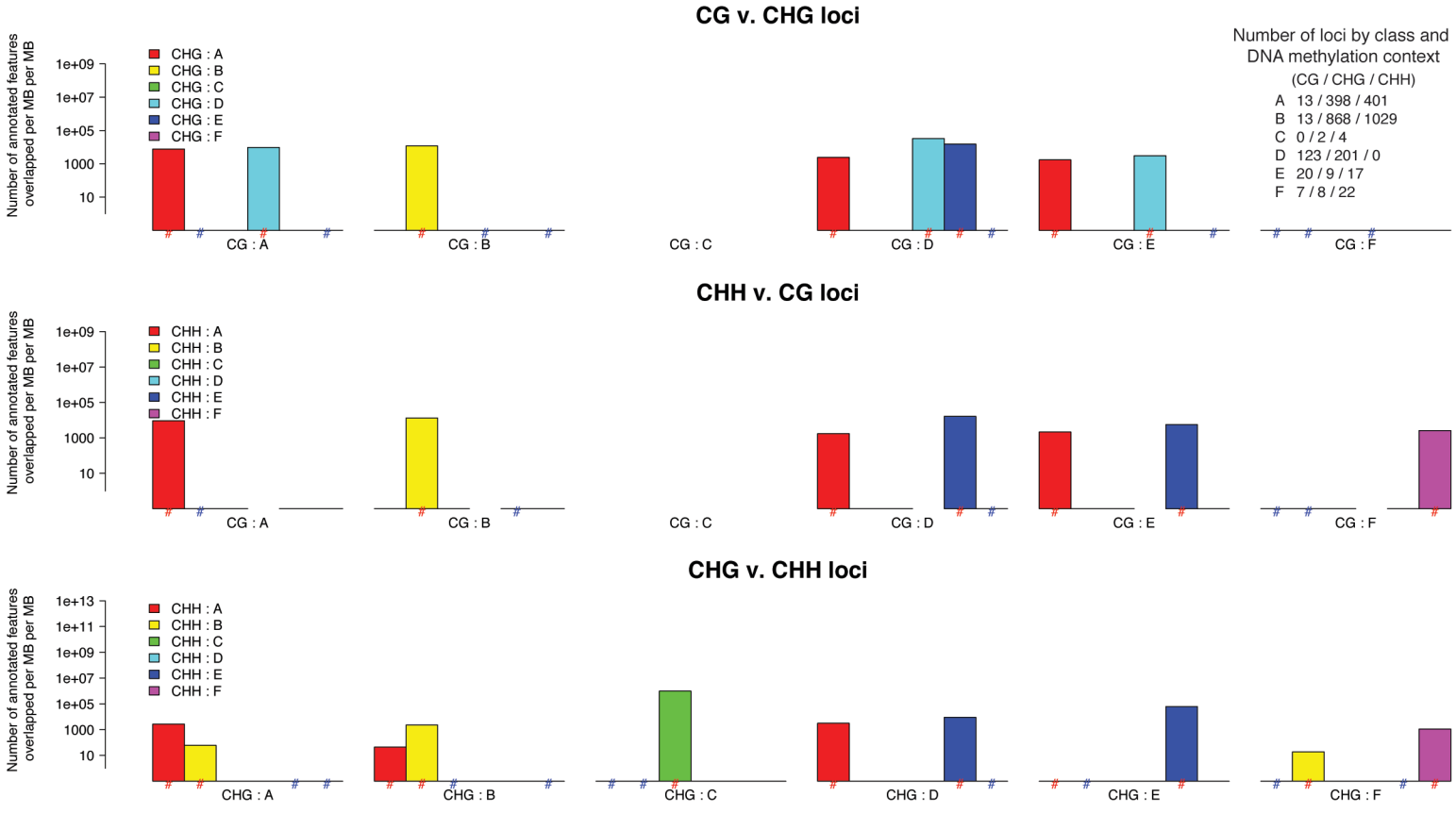

CG v. CHG loci
Number of annotated features overlapped per MB per MB
1e+09
1e+07
1e+05
1000
10
CHG : A
CHG : B
CHG : C
CHG : D
CHG : E
CHG : F
CG : A
CG : B
CG : C
CG : D
CG : E
CG : F
Number of loci by class and DNA methylation context
(CG / CHG / CHH)
A 13 / 398 / 401
B 13 / 868 / 1029
C 0 / 2 / 4
D 123 / 201 / 0
E 20 / 9 / 17
F 7 / 8 / 22
CHH v. CG loci
Number of annotated features overlapped per MB per MB
1e+09
1e+07
1e+05
1000
10
CHH : A
CHH : B
CHH : C
CHH : D
CHH : E
CHH : F
CG : A
CG : B
CG : C
CG : D
CG : E
CG : F
CHG v. CHH loci
Number of annotated features overlapped per MB per MB
1e+13
1e+11
1e+09
1e+07
1e+05
1000
10
CHH : A
CHH : B
CHH : C
CHH : D
CHH : E
CHH : F
CHG : A
CHG : B
CHG : C
CHG : D
CHG : E
CHG : F

Figure S3. Significant overlaps occur between loci of the same class but different DNA methylation contexts. For example, class A loci with CHG context DNA methylation overlap statistically significantly with class A loci containing CHH context DNA methylation (see bottom left plot, red bar, CHG:A *versus* CHH:A). Each individual plot compares the overlap between sets of loci in two different DNA methylation contexts. The x-axis label indicates one set of loci compared and their DNA methylation context (e.g. CHG:A indicates A loci with CHG context methylation, bottom left plot). The plot then represents the overlap of these loci with loci of all classes in a second DNA methylation context, indicated by the color legend of the bars. Y-axis units normalize the feature/locus overlap by both sum of feature size and sum of methylation locus size, which permits comparison between columns. These units are number of annotated features per total megabase (MB) of named feature per total MB of methylation in locus class. Significance level; # = $0 < p < 10^{-5}$, blue = under-represented, red = over-represented (relative to background). Total numbers of loci are given in the top right corner of the figure by class and context. Missing bars/plots indicate no loci exist for the comparison indicated.

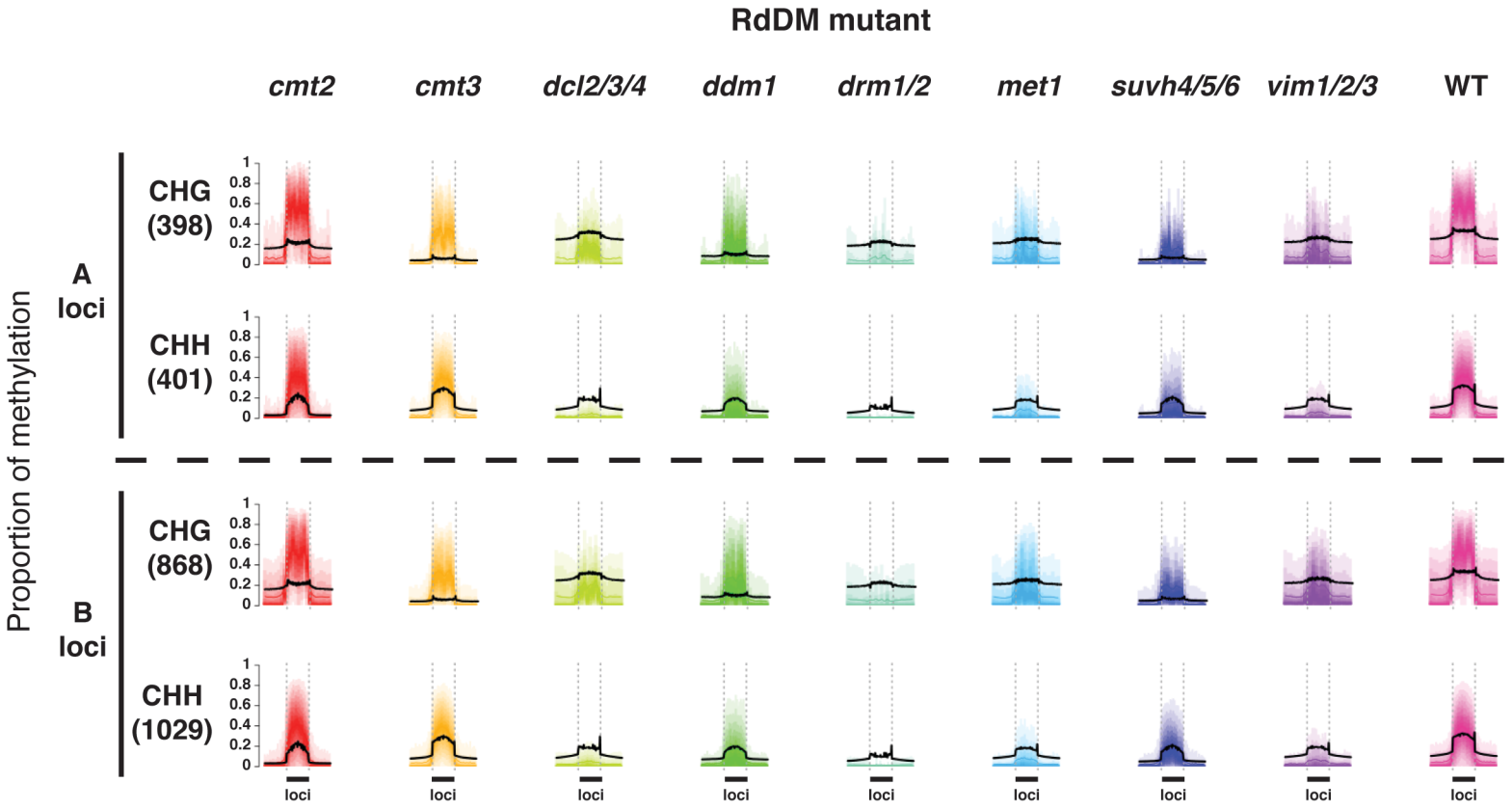
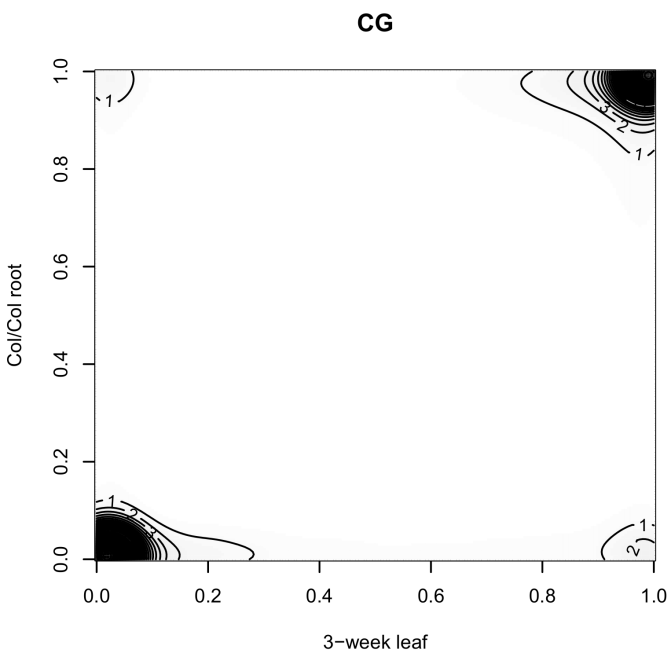
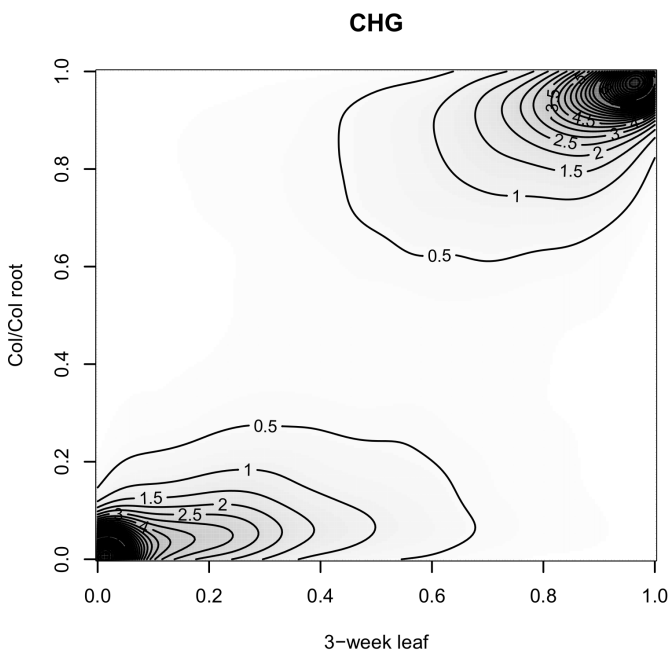

RdDM mutant
cmt2
cmt3
dcl2/3/4
ddm1
drm1/2
met1
suvh4/5/6
vim1/2/3
WT
Proportion of methylation
A loci
B loci
CHG (398)
CHH (401)
CHG (868)
CHH (1029)
1
0.8
0.6
0.4
0.2
0
loci

Figure S4. Mutation of *drm1/2* correlates with greater loss of DNA methylation at class A (direct) and class B (indirect) than the genome–wide average loss in that mutant. The plots show the mean proportion of DNA methylation across class A and B loci in mutants of RNA-directed DNA methylation (RdDM) pathway components and wild-type (WT) plants. These data complement Fig. 2. The proportion of methylation was calculated per base (between 0 and 1, unmethylated to fully methylated) for all loci within a class using published MethylC-seq data from RdDM mutants (11). Loci were then normalized to same size and the mean proportion of methylation calculated across them. The profiles of the mean proportion of DNA methylation across the size-normalized loci are plotted between the dashed vertical lines (indicated by the solid black bar labelled loci). Mean proportion of methylation in flanking DNA, 4kb upstream and downstream of the loci, are indicated to the left and right of the dashed vertical lines, respectively. The plots show ranges of DNA methylation for every fifth percentile for each mutant (*cmt2*, *cmt3*, *dcl2/3/4*, *met1*, *vim1/2/3*, *suvh4/5/6*, *drm1/2* and *ddm1*). The lightest color represents the 5^{th} to 95^{th} percentile, the next lightest represent 10^{th} to 90^{th} and so forth. Black traces overlaying the plots show mean genome-wide DNA methylation (i.e. at every locus where methylation was detected in any graft, irrespective of locus class). This reveals that the loss of DNA methylation caused by mutation of *drm1/2* and *dcl234* is greater at class A and class B loci than the genome-wide average loss. The number of loci assessed is given in parentheses below the DNA methylation context, to the left of the plots.


CG
Col/Col root
3−week leaf
0.0
0.2
0.4
0.6
0.8
1.0


CHG
Col/Col root
3−week leaf
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1
1.5
2
2.5

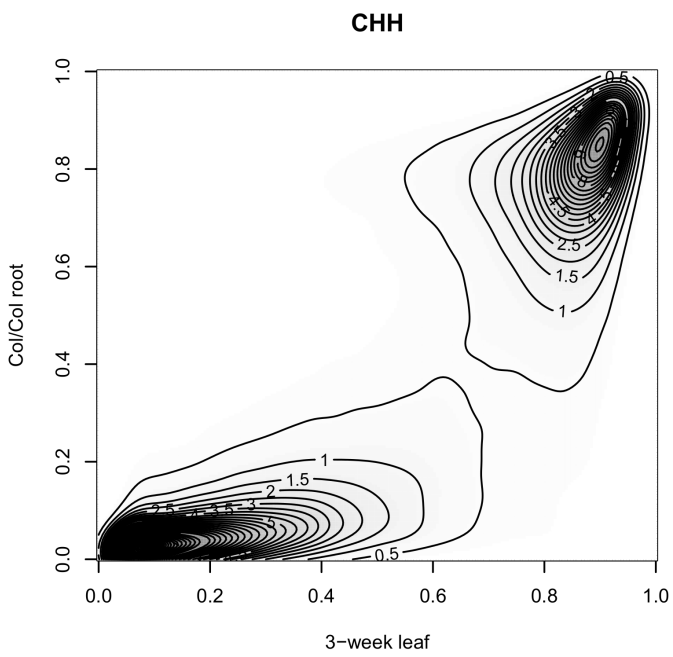

CHH
Col/Col root
3−week leaf
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1
1.5
2
2.5

Figure S5. The location of DNA methylation in root tissue of Col/Col grafts correlates with that in leaves of three week old plants. Contour maps of estimated likelihoods of DNA methylation at the CG, CHG and CHH loci identified from the Col/Col root graft methylation data for leaves of three week old plants (3-week leaf) compared with Col/Col grafted root tissue.

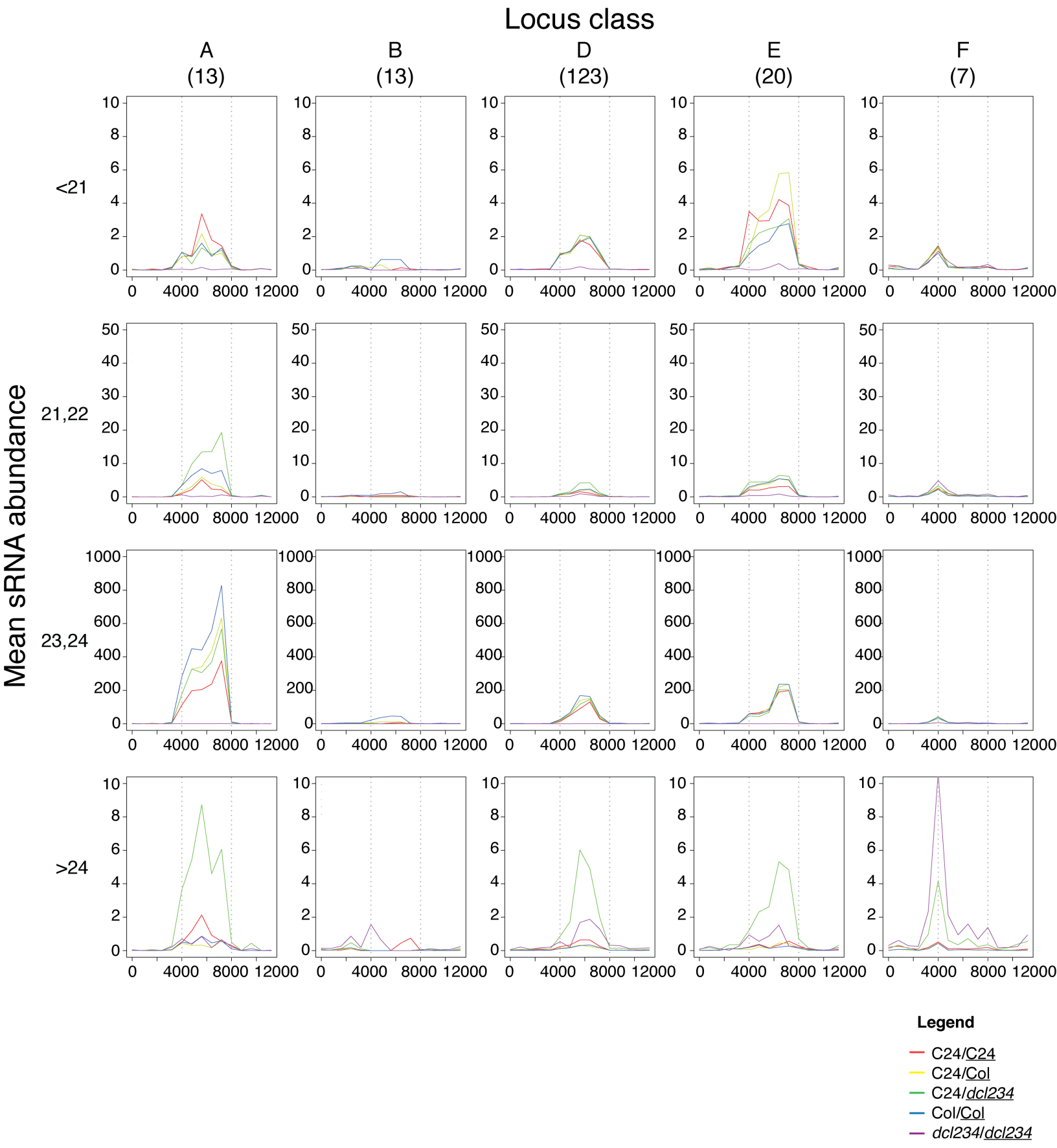

Locus class
A
(13)
B
(13)
D
(123)
E
(20)
F
(7)
Mean sRNA abundance
<21
21,22
23,24
>24
0
4000
8000
12000
Legend
C24/C24
C24/Col
C24/dcl234
Col/Col
dcl234/dcl234

Figure S6. Distribution of sRNAs targeting locus class A, B and D-F of the CG methylation context. Average sRNA abundance over loci in each graft combination are shown. Loci correspond to definitions in Figs. 1, S1 and are separated by size class of targeting sRNAs, indicated on the left. Loci were size normalized and are represented between positions 4000 and 8000 on the x axes, with average sRNA distribution over the 4000 bases upstream and downstream of loci plotted flanking them. Note that C loci are unmethylated by definition and the data consequently correspond to sRNAs targeting unmethylated CG DNA sequences. The 23,24 nt sRNAs are the canonically mobile sRNAs, but other size classes may also exhibit mobility. The number of loci assessed is given in parentheses below the locus class, above the plots.

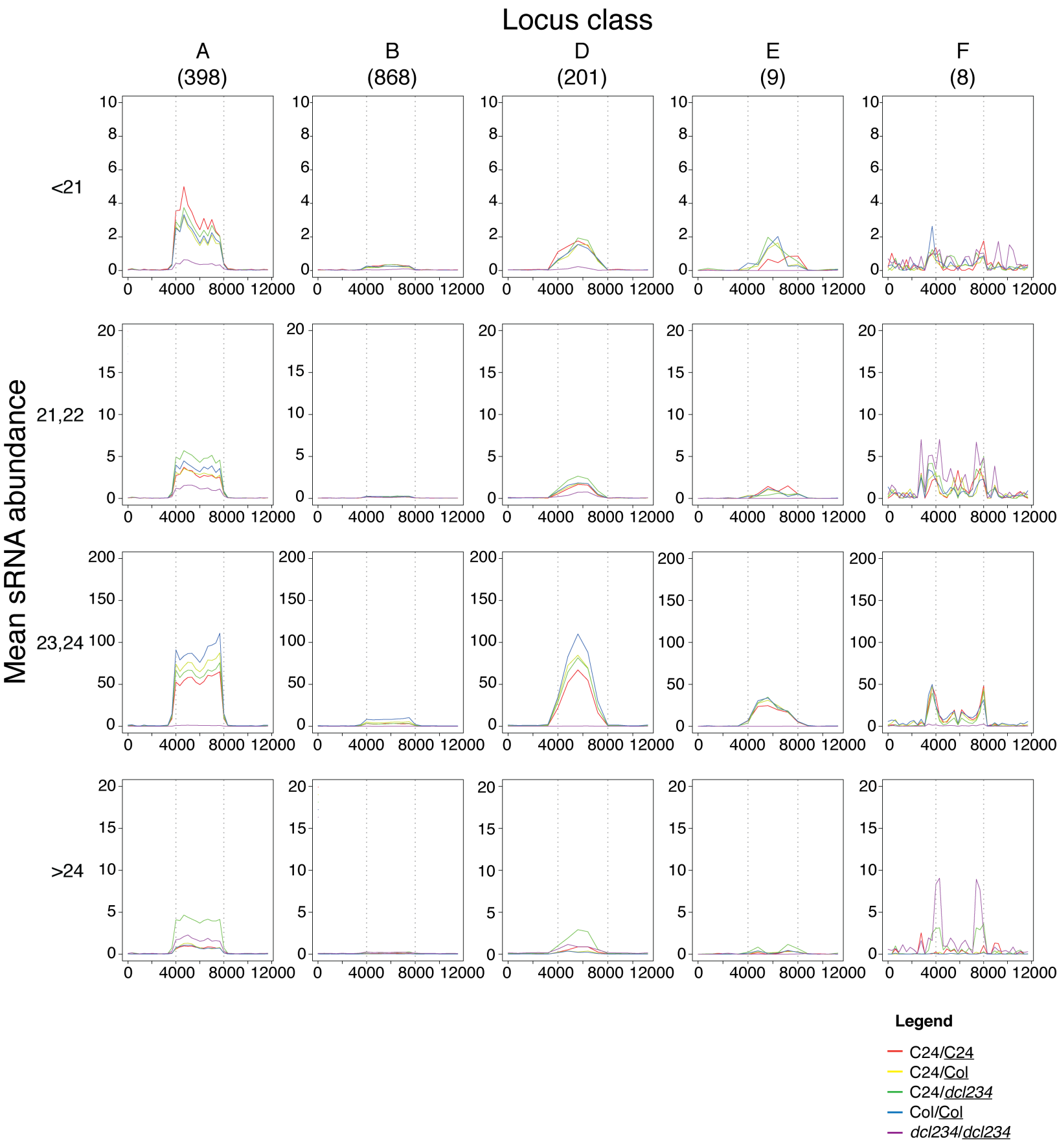

Locus class
A
(398)
B
(868)
D
(201)
E
(9)
F
(8)
<21
21,22
23,24
>24
Mean sRNA abundance
0
4000
8000
12000
Legend
C24/C24
C24/Col
C24/dcl234
Col/Col
dcl234/dcl234

Figure S7. **Distribution of sRNAs targeting locus class A, B and D-F of the CHG methylation context.** Average sRNA abundance over loci in each graft combination are shown. Loci correspond to definitions in Table 1 and are separated by size class of targeting sRNAs, indicated on the left. Loci were size normalized and are represented between positions 4000 and 8000 on the x axes, with average sRNA distribution over the 4000 bases upstream and downstream of loci plotted flanking them. Note that C loci are unmethylated by definition and the data consequently correspond to sRNAs targeting unmethylated CHG DNA sequences. The 23,24 nt sRNAs are the canonically mobile sRNAs, but other size classes may also exhibit mobility. The number of loci assessed is given in parentheses below the locus class, above the plots.

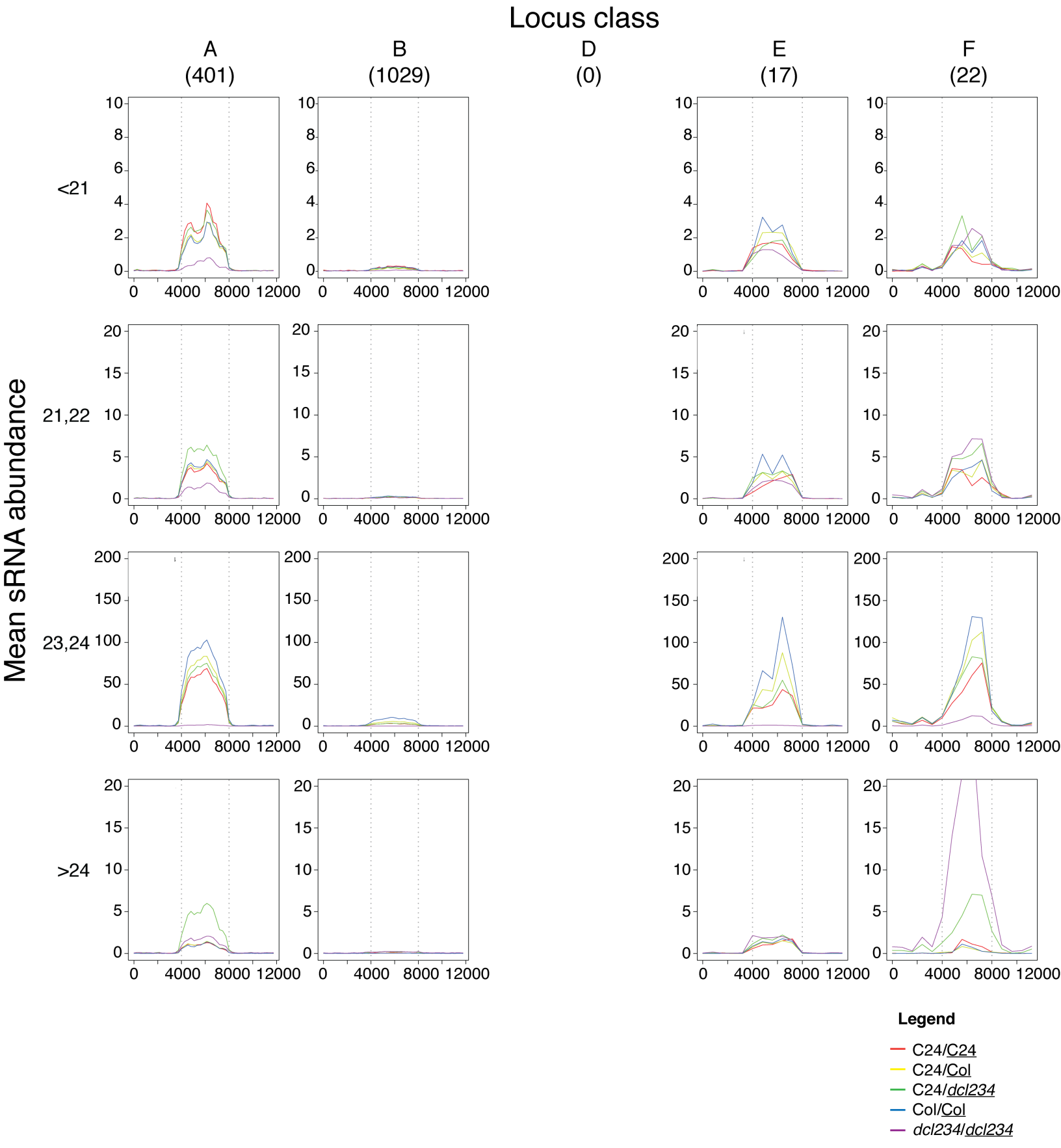

Locus class
A
(401)
B
(1029)
D
(0)
E
(17)
F
(22)
Mean sRNA abundance
<21
21,22
23,24
>24
0
4000
8000
12000
Legend
C24/C24
C24/Col
C24/dcl234
Col/Col
dcl234/dcl234

Figure S8. **Distribution of sRNAs targeting locus class A, B and D-F of the CHH methylation context.** Average sRNA abundance over loci in each graft combination are shown. Loci correspond to definitions in Table 1 and are separated by size class of targeting sRNAs, indicated on the left. Loci were size normalized and are represented between positions 4000 and 8000 on the x axes, with average sRNA distribution over the 4000 bases upstream and downstream of loci plotted flanking them. Note that no CHH context C loci were identified and this column is left blank. The 23,24 nt sRNAs are the canonically mobile sRNAs, but other size classes may also exhibit mobility. The number of loci assessed is given in parentheses below the locus class, above the plots.

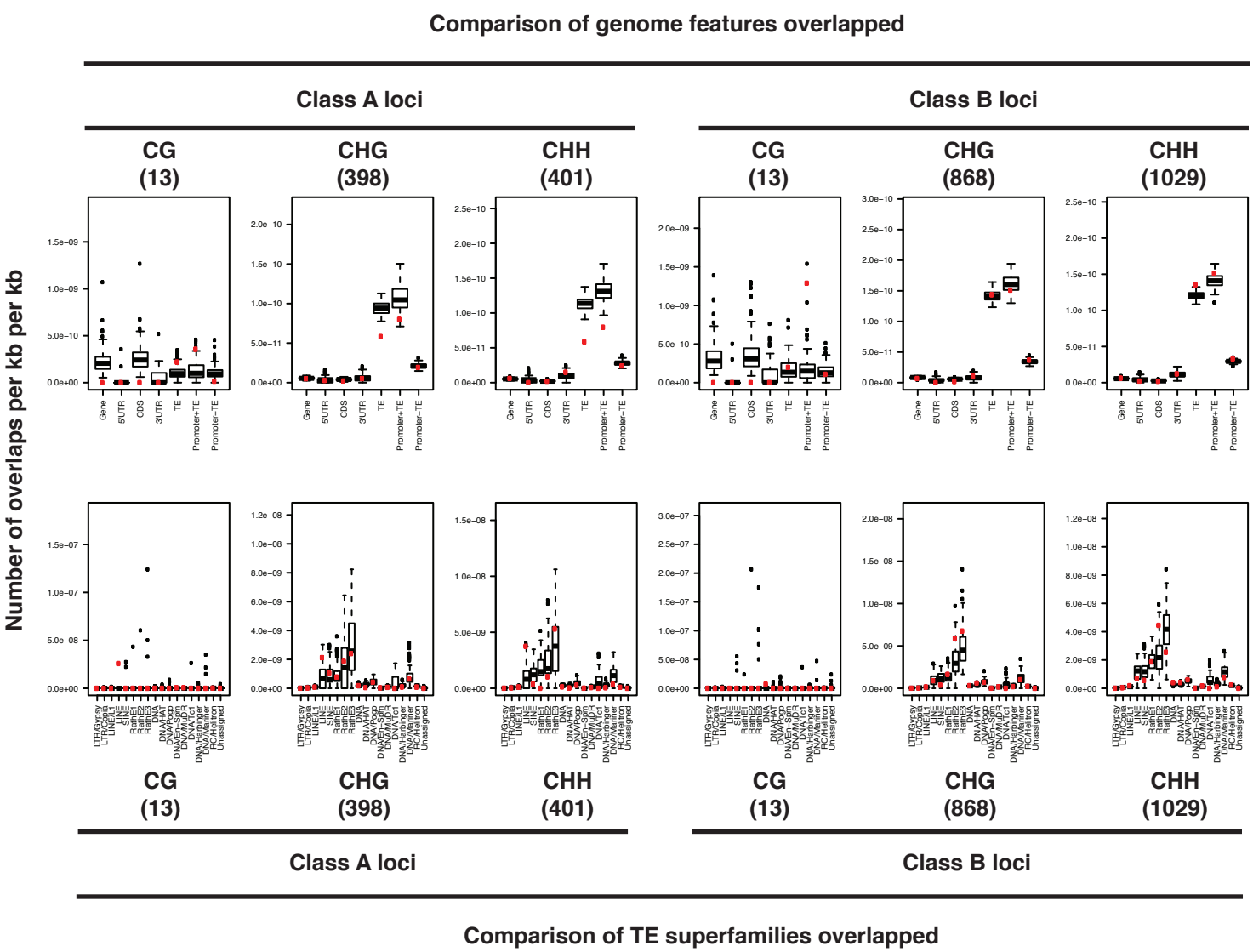

Comparison of genome features overlapped
Class A loci
Class B loci
CG (13)
CHG (398)
CHH (401)
CG (13)
CHG (868)
CHH (1029)
Number of overlaps per kb per kb
Gene
5'UTR
CDS
3'UTR
TE
Promoter+TE
Promoter−TE
LTR/Gypsy
LTR/Copia
LINE/L1
LINE/I
SINE
RathE1
RathE2
RathE3
DNA
DNA/HAT
DNA/Pogo
DNA/MuDR
DNA/Tc1
DNA/Harbinger
DNA/Mariner
RC/Helitron
Unassigned
Comparison of TE superfamilies overlapped

Figure S9. Feature association of class A and B loci is driven by the associations of *dcl234* dependent methylation loci. The genome features associated with class A and B loci were compared with genomic loci that lose DNA methylation in root tissue of *dcl234*/*dcl234* grafts. The observed number of overlaps for class A and B loci against each feature are shown in red. Overlaps per kilobase per kilobase observed in repeated block bootstrap samplings of DNA methylation loci showing a loss of methylation in *dcl234*/*dcl234* grafted root tissue are shown as black box-and-whisker plots. Data were normalized per kilobase total feature length per kilobase total locus length (per kb per kb), in the same manner described by Supplemental Method 9. Numbers of loci analysed are given in parentheses beneath DNA methylation context labels. Abbreviations: CDS – coding sequence, Promoter+TE – promoter containing transposable element, Promoter-TE – promoter not containing transposable element, UTR – untranslated region.

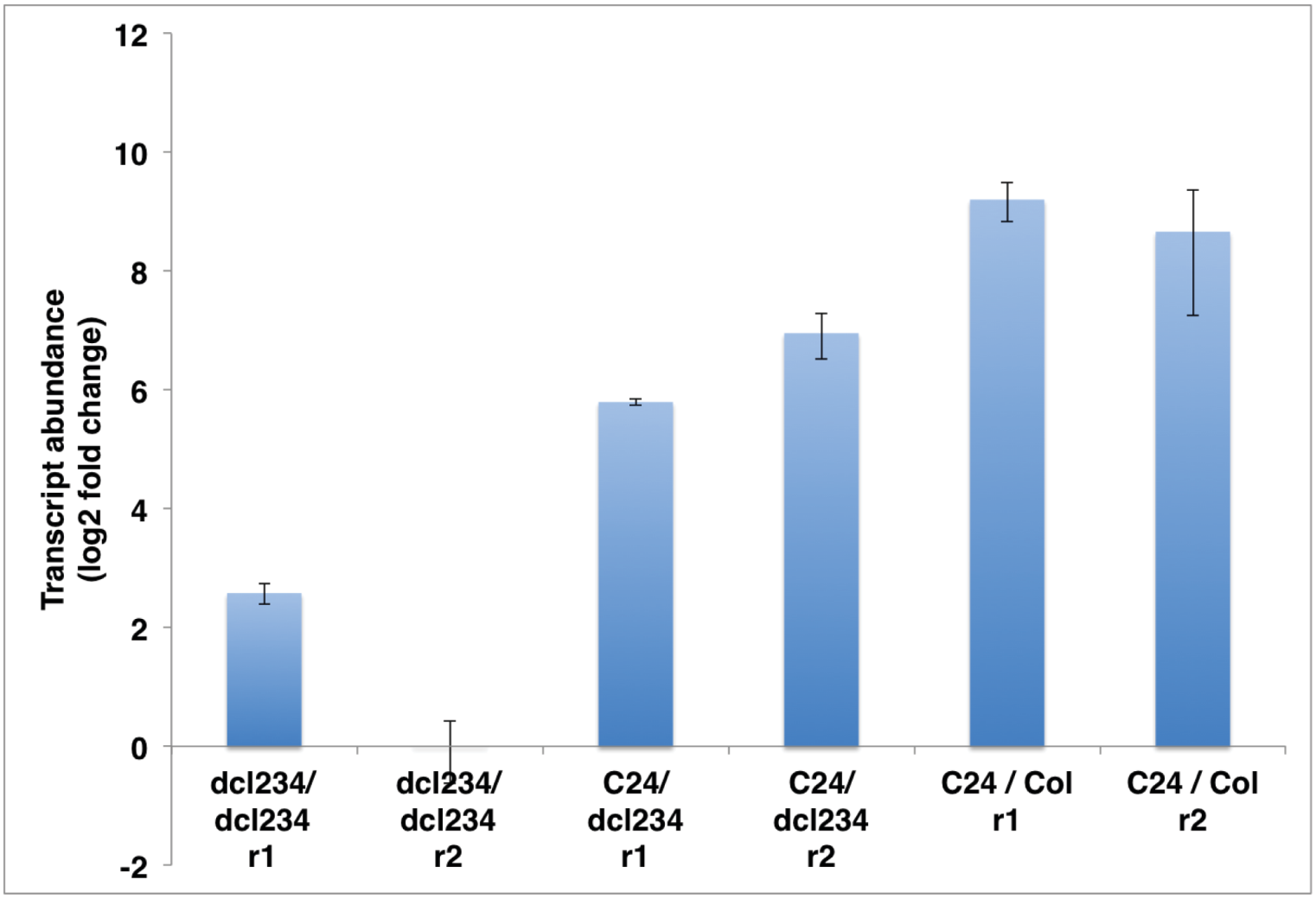

Transcript abundance
(log2 fold change)
12
10
8
6
4
2
0
-2
dcl234/
dcl234
r1
dcl234/
dcl234
r2
C24/
dcl234
r1
C24/
dcl234
r2
C24 / Col
r1
C24 / Col
r2

Figure S10. Quantitative real time PCR confirms RNA-seq measurements of *REPRESSOR OF SILENCING 1* transcript abundance. Lower transcript abundance was detected in roots of *dcl234*/*dcl234* grafts compared with roots of C24/*dcl234* and C24/Col grafts. This agrees with RNA-seq analyses (Dataset S3). Measurements were taken from two independent grafted plants for each combination, denoted r1 and r2. Error bars show 95% confidence intervals of fold change.

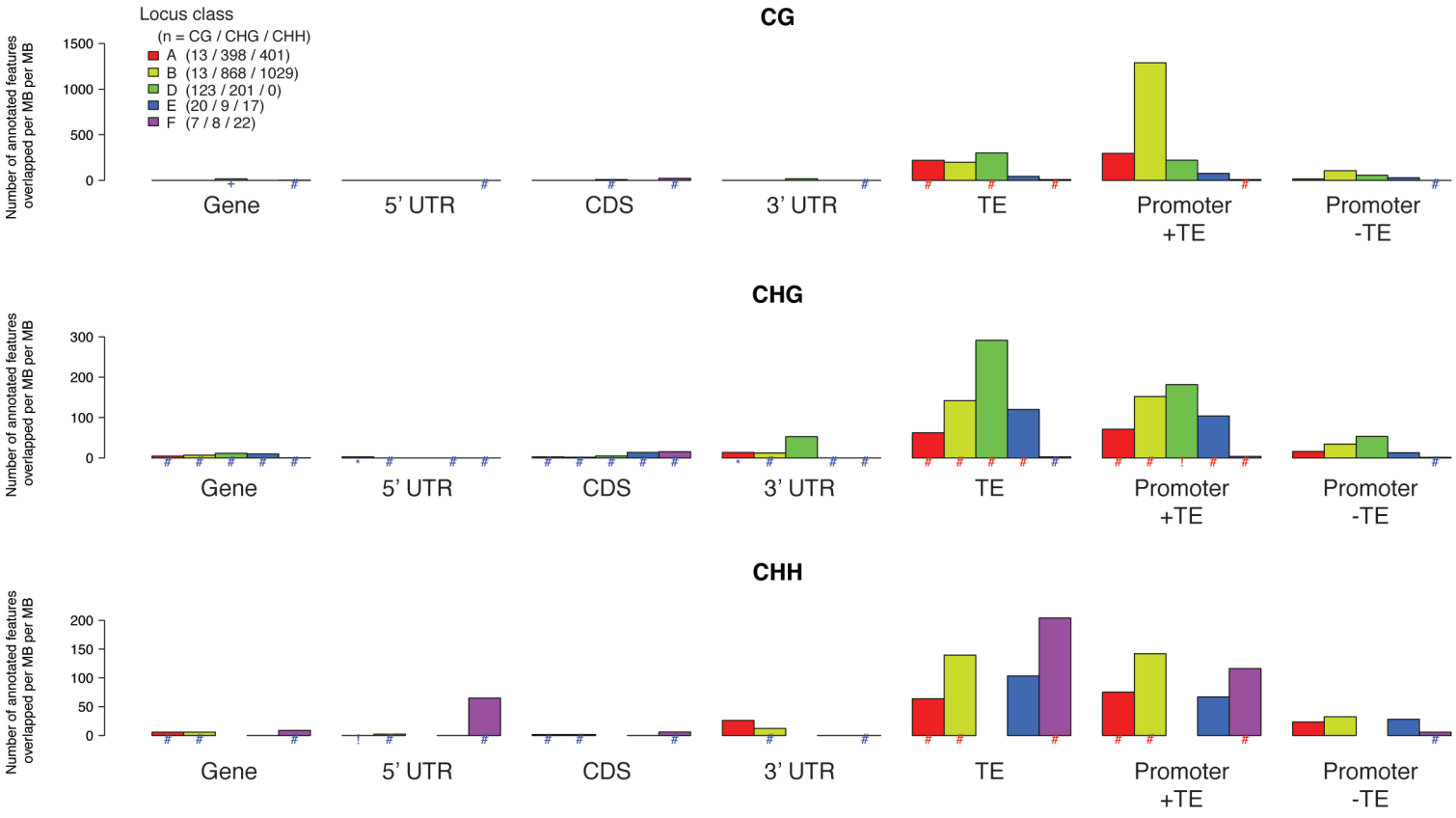

Locus class
(n = CG / CHG / CHH)
A (13 / 398 / 401)
B (13 / 868 / 1029)
D (123 / 201 / 0)
E (20 / 9 / 17)
F (7 / 8 / 22)
CG
CHG
CHH
Number of annotated features overlapped per MB per MB
1500
1000
500
0
300
200
100
50
150
Gene
5' UTR
CDS
3' UTR
TE
Promoter +TE
Promoter -TE

Figure S11. Locus classes A, B and D-F target distinct sets of genome features. Association of each locus class with defined genome features, separated by DNA methylation context. Locus classes correspond to Figs. 1, S1. Y-axis units normalize the feature/locus overlap by both sum of feature size and sum of methylation locus size, which permits comparison between columns. These units are number of annotated features per total megabase (MB) of named feature per total MB of methylation in locus class. Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * = $10^{-3} < p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to background). CDS – coding sequence, TE – transposable element, UTR – untranslated region. Gaps within x axes denote that no loci were identified of that class/methylation context combination were identified overlapping the indicated feature. Legend shows number of loci (n) in parentheses for each class by DNA methylation context.

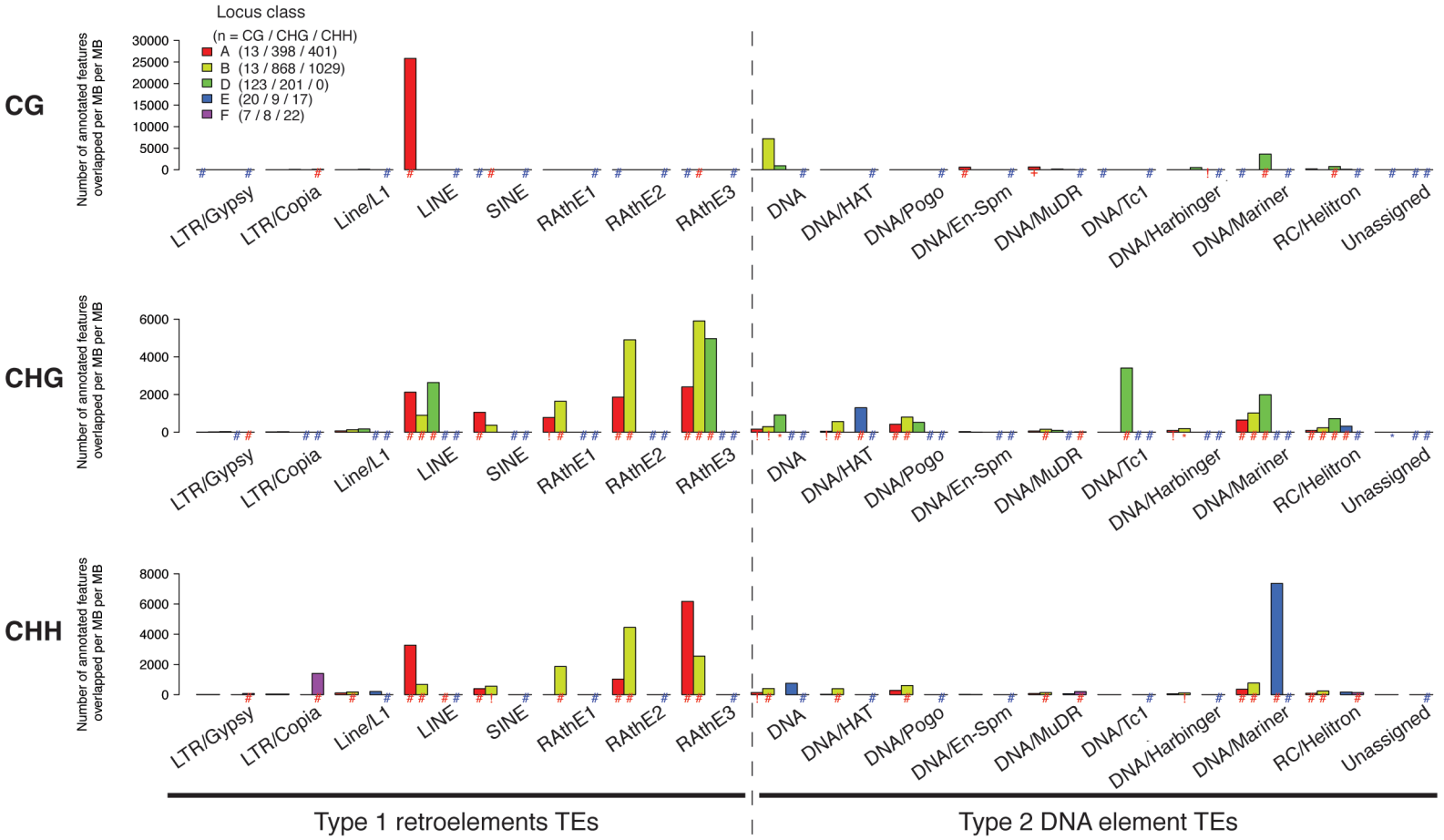

CG
CHG
CHH
Number of annotated features overlapped per MB per MB
Locus class
(n = CG / CHG / CHH)
A (13 / 398 / 401)
B (13 / 868 / 1029)
D (123 / 201 / 0)
E (20 / 9 / 17)
F (7 / 8 / 22)
LTR/Gypsy
LTR/Copia
Line/L1
LINE
SINE
RAthE1
RAthE2
RAthE3
DNA
DNA/HAT
DNA/Pogo
DNA/En-Spm
DNA/MuDR
DNA/Tc1
DNA/Harbinger
DNA/Mariner
RC/Helitron
Unassigned
Type 1 retroelements TEs
Type 2 DNA element TEs

Figure S12. Class A and B loci are associated with similar TE classes. Locus classes correspond to Figs. 1, S1. Data are plotted according to DNA methylation context of the loci. Note that D loci are unmethylated by definition and the data consequently are plotted according to unmethylated CG/CHG/CHH DNA sequences. Y-axis units normalize the feature/locus overlap by both sum of feature size and sum of methylation locus size, which permits comparison between columns. These units are number of annotated features per total megabase (MB) of named feature per total MB of methylation in locus class. Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * = $10^{-3} < p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to background).

Supplemental Tables

Table S1. Summary of sizes of transposable element classes. Units of measurement – nucleotides.

Class	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
RathE3_cons	20	101.8	142	135	172.2	244
RathE1_cons	28	94	144	135.4	169	287
RathE2_cons	21	84.25	130	163.1	243.2	371
DNA/Tc1	26	176.5	243	215.8	258	288
DNA/Mariner	33	213	250	231.9	267.5	404
SINE	22	103.5	197	244.1	335	1423
LINE	23	137	257	327.1	361	1247
DNA	20	64	189	375.7	460	9068
DNA/Pogo	20	105	367.5	389.3	649	2114
DNA/HAT	20	126	242	445.5	561.5	8407
RC/Helitron	20	98	361	583.1	784	18770
DNA/Harbinger	21	156	328	655.4	885.5	6427
DNA/MuDR	11	106	260.5	765.7	566.8	23000
LINE/L1	10	134.2	310.5	914.3	897.5	14250
LTR/Copia	12	105	253	1077	966	14970
DNA/En-Spm	20	117	259	1251	1218	18470
LTR/Gypsy	14	135	587	1668	1895	31020
Unassigned	123	886	1267	1723	2250	5724

Table S2. Numbers of class D, E and F loci in each DNA methylation (mC) context. Columns "mC loci" and "sRNA loci" give the total numbers of mC and sRNA loci identified corresponding to each locus class D-F, as described in Fig. S1B. The mC loci were counted individually by mC sequence context (CG, CHG, CHH). sRNA loci are not context dependent, so only one set exists for each class D-F. The sRNA loci were intersected with the mC loci from the three sequence contexts within each class D-F. Comparisons were made in two directions; the number of mC loci overlapping sRNA loci (column "mC overlaps to sRNA") and the number of sRNA loci overlapping mC loci (column "sRNA overlaps to mC"). The median lengths of the loci in nucleotides (nt) are given in parentheses. Reversing the direction of comparison yielded different numbers of overlaps between sRNA and mC loci. This is possible because the sRNA and mC loci are of different sizes. Hence, one sRNA locus may overlap multiple mC loci. * indicates which overlaps were carried forward for downstream analyses of that class. Positive Z-score (Z) indicates an enriched association between the sRNA and methylation loci, whilst a negative Z-score indicates depleted association, relative to the overlap expected by chance. Column "p" gives the p values associated with the overlaps. Note that certain locus classes require low levels of mC or sRNAs in specific graft combinations (see Fig. S1B). For example, the number of overlaps column for class D gives the number of genomic loci where sRNA abundance was high in roots of all graft combinations, except *dcl234/dcl234*, and where mC levels were low in all graft combinations. Consequently, for this model the "mC loci" column gives the number of loci where mC levels were low in all graft combinations.

Class	**mC context**	**mC loci (median length)**	**sRNA loci (median length)**	**mC overlaps to sRNA**	**sRNA overlaps to mC**	**Z**	**p**
D	CG	132313 (346)	925 (147)	123	64*	-29.368	$6.98*10^{-190}$
	CHG	24981 (2483)		201	124*	-52.712	0
	CHH	1 (660)		0	0*	NA	NA
E	CG	1633 (10)	925 (147)	20*	17	7.3741	$8.27*10^{-14}$
	CHG	115 (145)		9*	7	22.878	0
	CHH	203 (132)		17*	13	23.302	0
F	CG	321 (272)	235 (247)	7*	5	6.4188	$6.87*10^{-11}$
	CHG	434 (1656)		8*	5	2.1556	0.015557
	CHH	3287 (1)		22*	20	1.3893	0.082375

Table S3. MethylC-seq library statistics.

Graft combination	Bioreplicate number	Non-conversion (%)	Total number of reads	Reads with at least one alignment	Reads with no alignments	Total number of alignments
C24/C24	1	0.39	97282476	53468497 (54.96%)	43813979 (45.04%)	108031755
C24/C24	2	0.56	77257652	52736588 (68.26%)	24521064 (31.74%)	251482201
C24/C24	3	0.54	232282436	67258086 (28.96%)	165024350 (71.04%)	445365903
C24/C24	4	0.36	46525685	32199788 (69.21%)	14325897 (30.79%)	174468042
C24/Col	1	4.31	44828712	38046442 (84.87%)	6782270 (15.13%)	73303467
C24/Col	2	3.26	82763841	66266543 (80.07%)	16497298 (19.93%)	311422104
C24/Col	3	5.07	145755159	52506196 (36.02%)	93248963 (63.98%)	277195396
C24/Col	4	1.72	42018007	38345661 (91.26%)	3672346 (8.74%)	164201615
C24/*dcl234*	1	5.55	81666429	68968313 (84.45%)	12698116 (15.55%)	129484256
C24/*dcl234*	2	2.29	34574555	31315085 (90.57%)	3259470 (9.43%)	132264182
Col/Col	1	5.08	58355688	49234769 (84.37%)	9120919 (15.63%)	96047853
Col/Col	2	6.40	45805338	30916686 (67.50%)	14888652 (32.50%)	167109490
Col/Col	3	2.55	48736570	34541290 (70.87%)	14195280 (29.13%)	147993082
dcl234/*dcl234*	1	7.12	89887953	76095824 (84.66%)	13792129 (15.34%)	145038964
dcl234/*dcl234*	3	2.29	30862638	20117034 (65.18%)	10745604 (34.82%)	85527551
dcl234/*dcl234*	2	3.54	55146553	32291691 (58.56%)	22854862 (41.44%)	151630559

Supplemental Datasets S1-S5

See attached files.