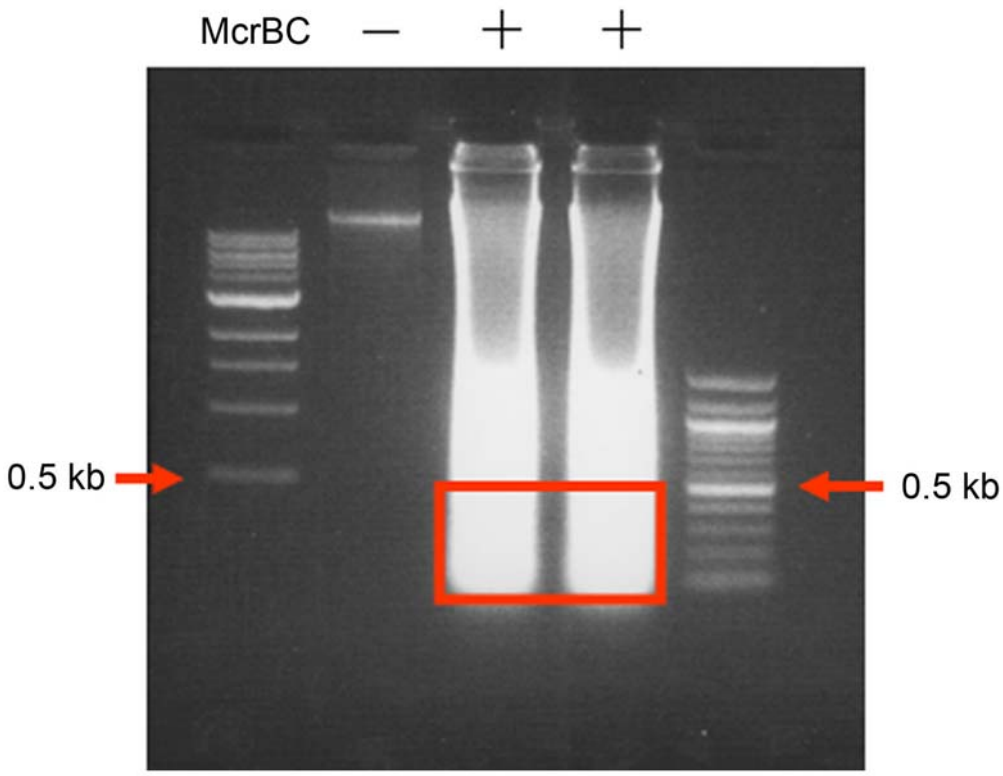
Supplemental Data

For

High resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation and gene expression

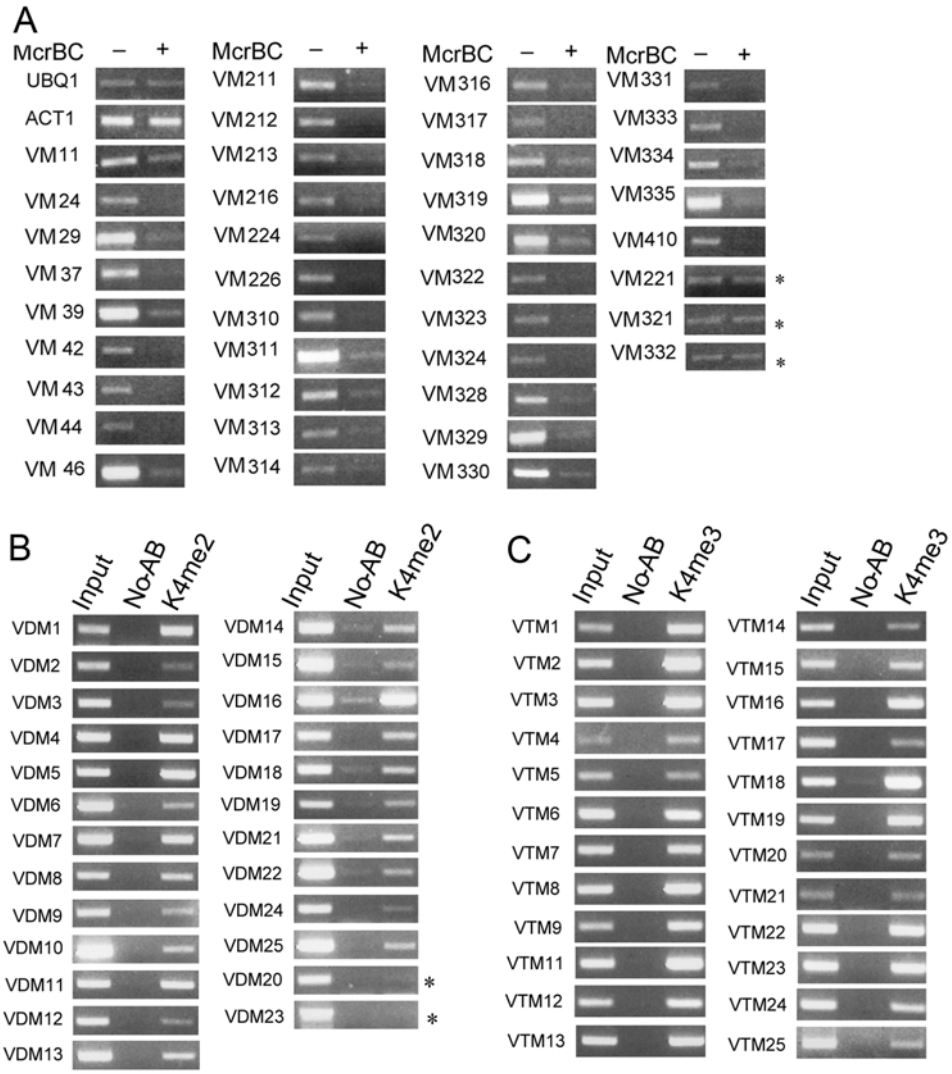
Xueyong Li[1, 2]*, Xiangfeng Wang[1, 2, 3]*, Kun He[2, 3], Yeqin Ma[2], Ning Su[2], Hang He[1], Viktor Stolc[4], Waraporn Tongprasit[4], Weiwei Jin[5], Jiming Jiang[5], William Terzaghi[2], Songgang Li[3] and Xing Wang Deng[1, 2, 3]†

1National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China. 2Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA. 3Peking-Yale Joint Research Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China. 4Genome Research Facility, NASA Ames Research Center, MS 239-11, Moffett Field, CA 94035, USA. 5Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706



Supplemental Figure 1. Isolation of methylated DNA by the McrBC digestion method.

20 µg of genomic DNA were digested overnight with 200 units of McrBC (New England Biolabs), which only digests methylated DNA at [A/G]mC (N40-3000) [A/G]mC. Highly methylated regions were cut into small fragments. McrBC digests were separated on a 1.0% agarose gel. DNA fragments smaller than 0.5 kb, which are mainly composed of methylated DNA, were collected and purified using Qiagen Qiaquick gel extraction kits.



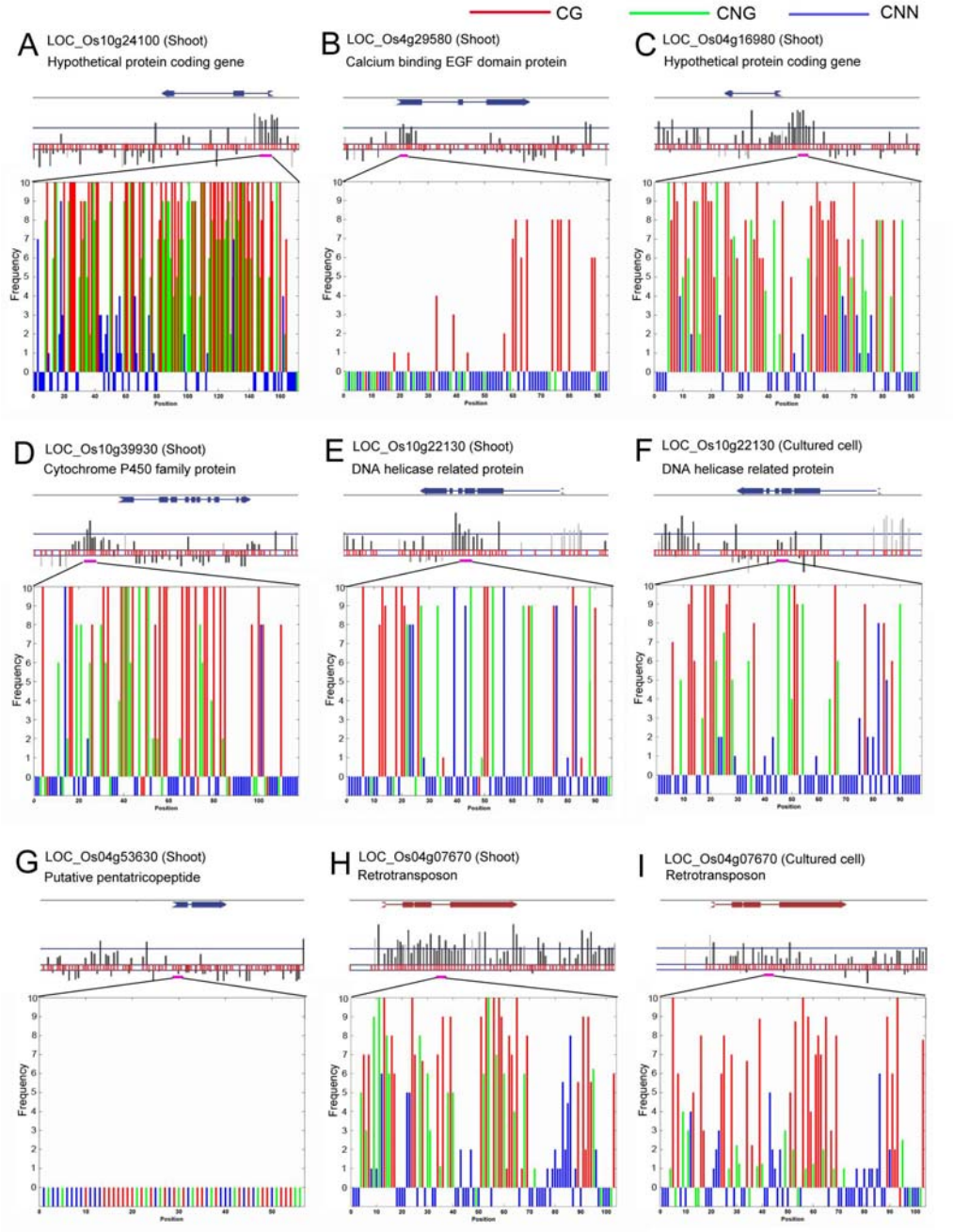
Supplemental Figure 2. Experimental confirmation of DNA methylation, H3K4me3 and H3K4me2 regions.

To assess the reliability of the threshold ($P<0.05$), experimental validations of selected modified regions identified by the Wilcoxon test were performed.

(A) McrBC-PCR validation of DNA methylation. Methylation of the indicated sequences was evaluated by the McrBC-PCR method as described in "methods." "+" = samples treated with McrBC, "-" = samples treated with heat-inactivated McrBC. "*" = unvalidated sample.

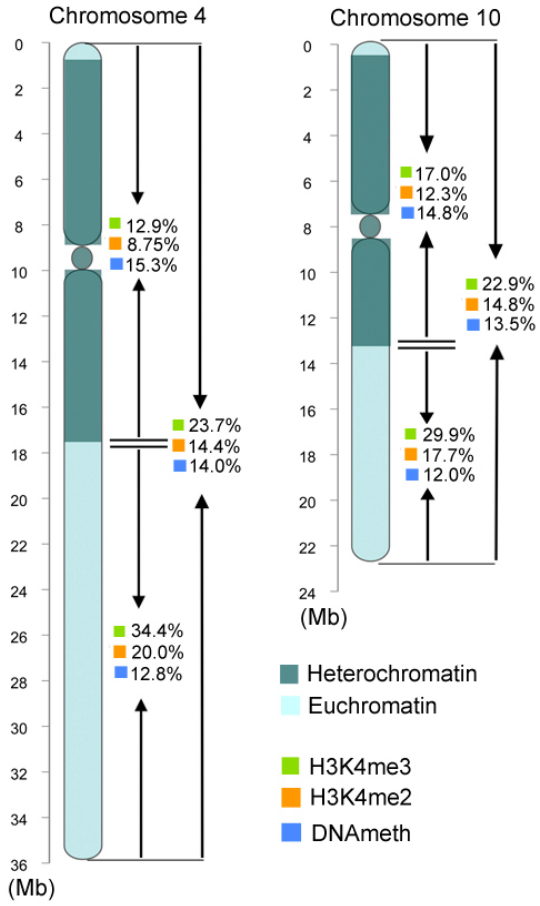
(B) Confirmation of H3K4me2 regions by ChIP-PCR. Modification of the indicated sequences by H3K4Mme2 was tested by ChIP-PCR as described in "methods." "Input"= whole cell extract, No-AB = samples treated for ChIP except antibody was omitted, K4me2 = samples immunoprecipitated with anti-H3K4me2. "*" = unvalidated sample.

(C) Confirmation of H3K4me3 regions by ChIP-PCR. Modification of the indicated sequences by H3K4me3 was tested by ChIP-PCR as described in "methods." "Input"= whole cell extract, No-AB = samples treated for ChIP except antibody was omitted, K4me3 = samples immunoprecipitated with anti-H3K4me2.



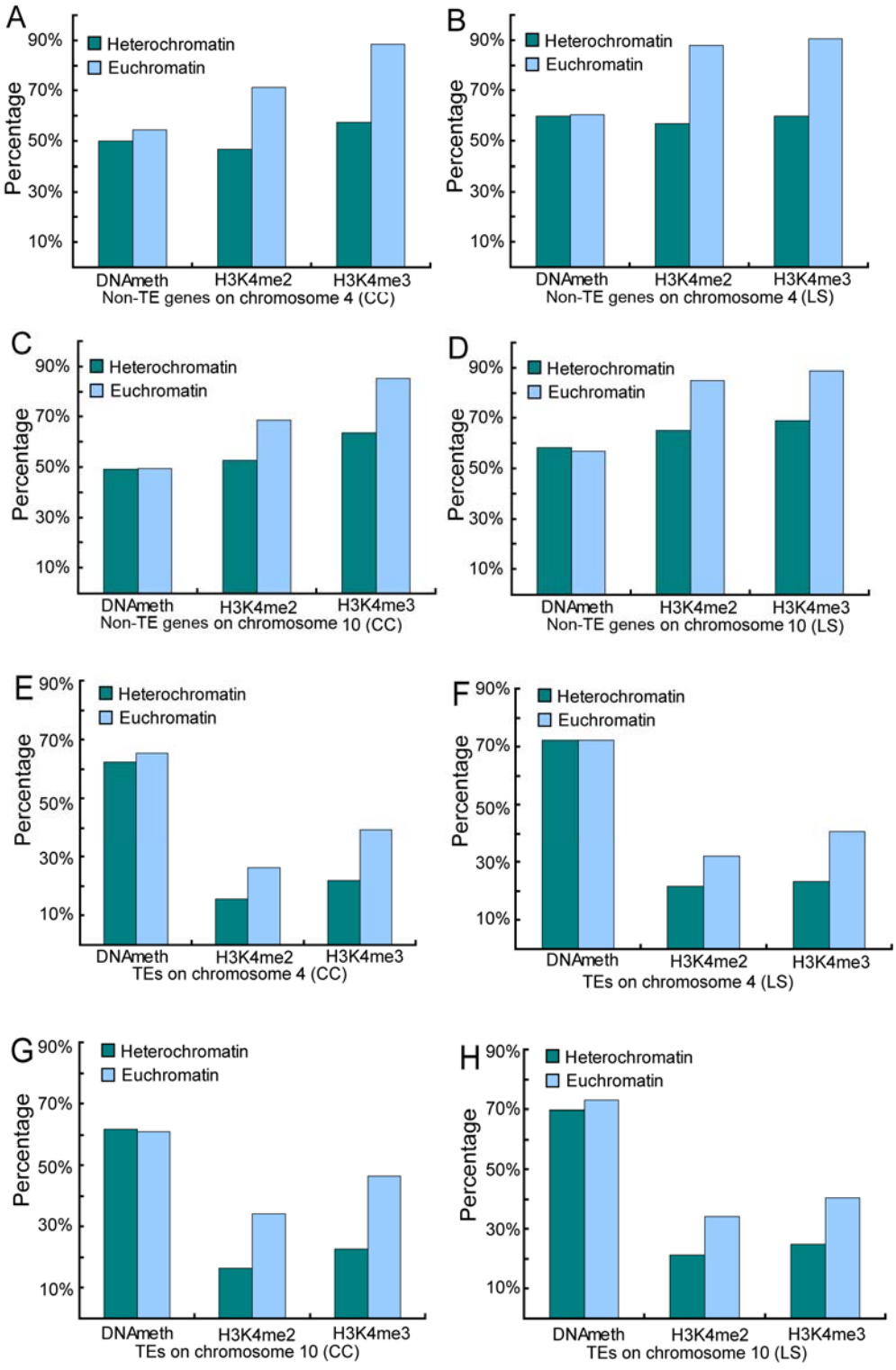
Supplemental Figure 3. Experimental validation of DNA methylation regions by genomic bisulfite sequencing.

Methylated cytosines were detected in the indicated genes from the indicated tissues by bisulfite sequencing as described in "methods." CG, CNG, and CNN = methylcytosines detected in the indicated contexts.

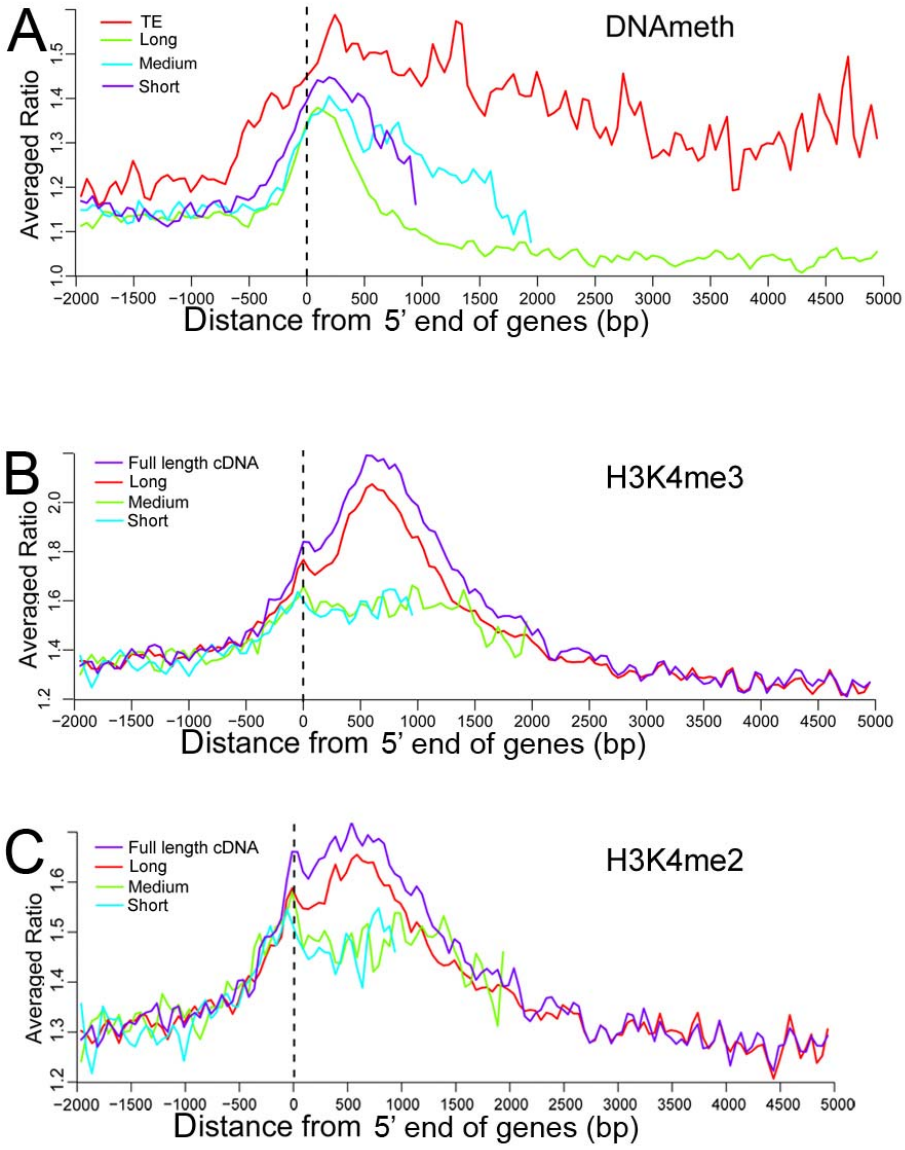


Supplemental Figure 4. Coverage of rice chromosomes 4 and 10 by three types of epigenetic modifications in suspension-cultured cells.

Coverage is calculated as the sum of the specified modified regions divided by the length of the indicated region for heterochromatin, euchromatin and the entire chromosome, respectively.



Supplemental Figure 5. Proportions of epigenetically-modified protein-coding and TE genes in heterochromatin and euchromatin.

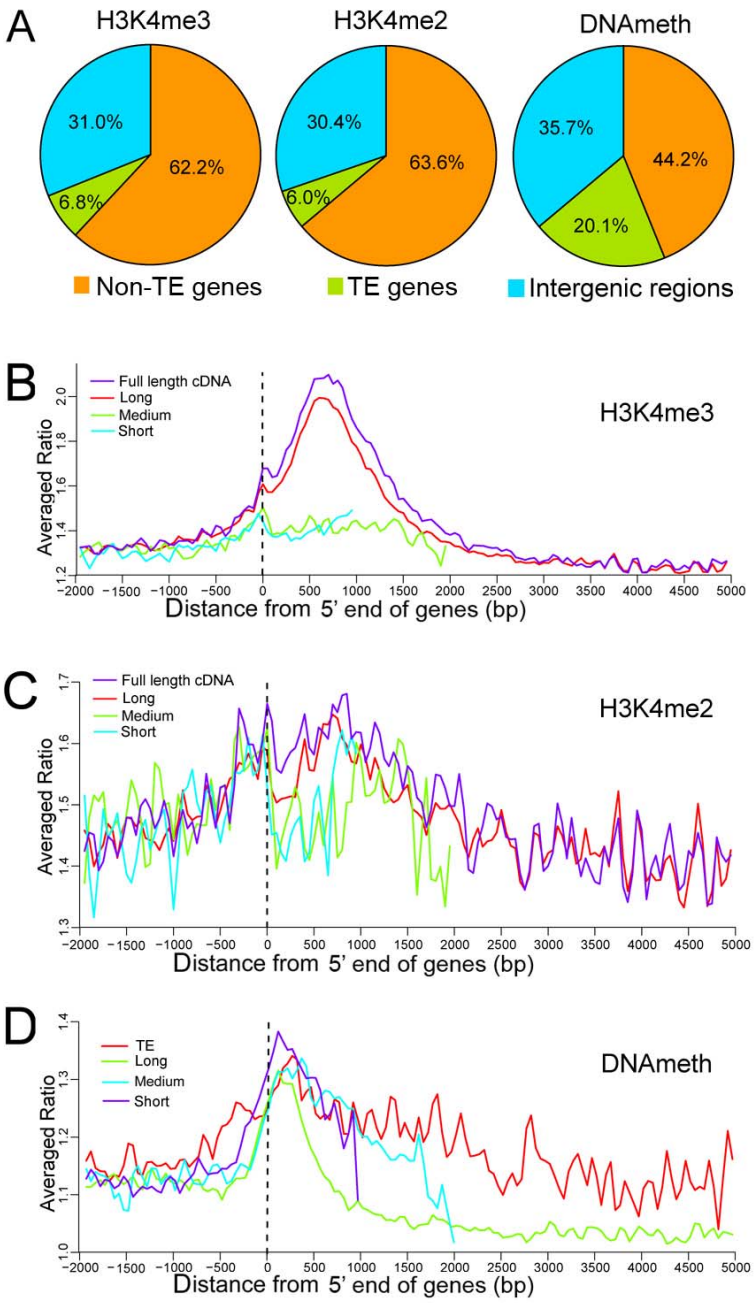


Supplemental Figure 6. Distribution of H3K4me3, H3K4me2 and DNA methylation within aligned genes in light-grown rice shoots.

(A) Distribution of H3K4me3 along aligned long, medium and short rice genes. Full length cDNAs were used as reference.

(B) Distribution of H3K4me2 along aligned long, medium and short rice genes. Full length cDNAs were used as reference.

(C) Distribution of DNA methylation along aligned long, medium and short rice genes. TEs were used as reference.



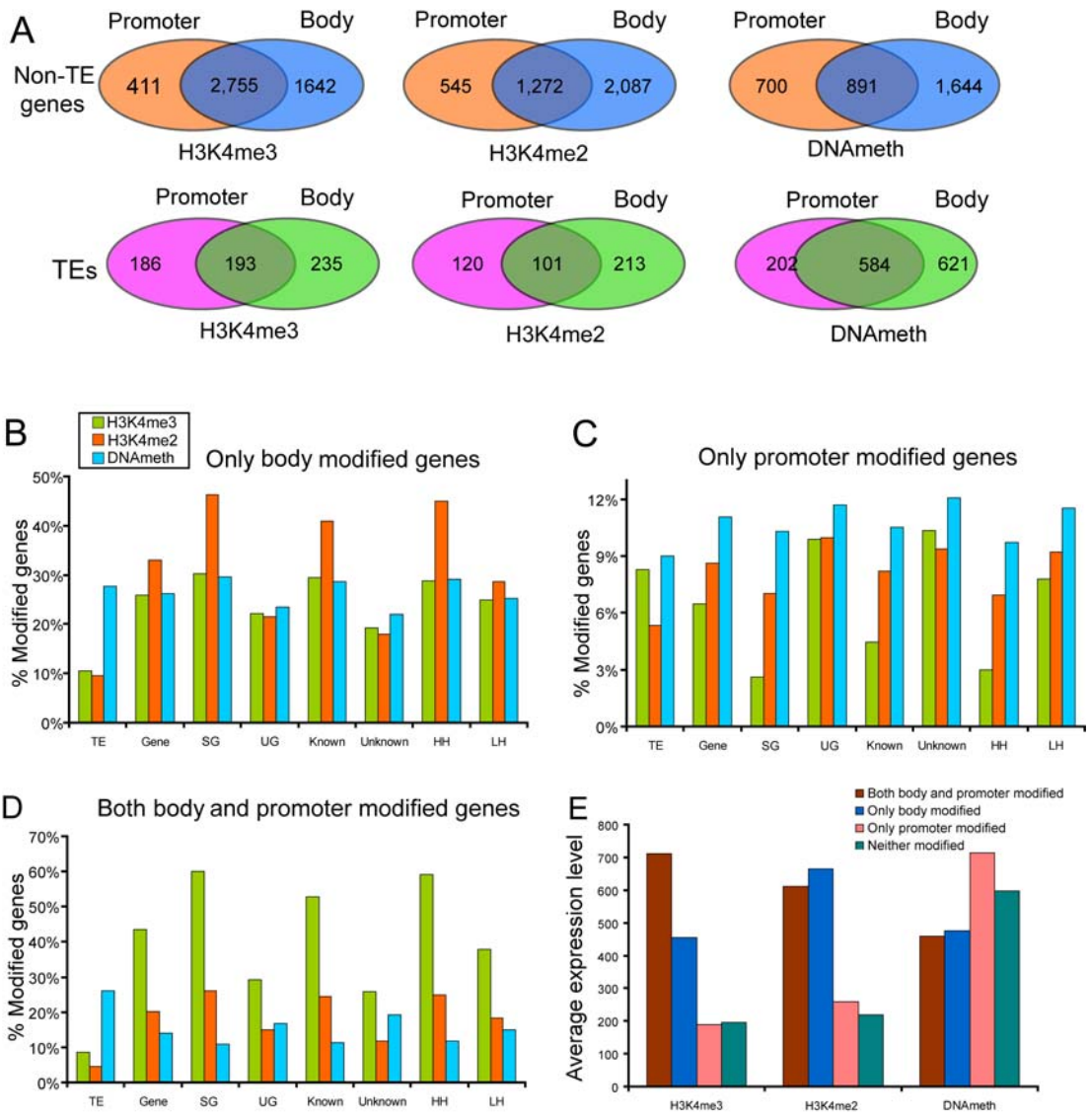
Supplemental Figure 7. Distributions of H3K4me3, H3K4me2 and DNA methylation within aligned genes in cultured rice cells.

(A) Frequencies of H3K4me3, H3K4me2 and DNA methylation occurring on gene, TE and intergenic regions.

(B) Distribution of H3K4me3 along aligned long, medium and short rice genes. Full length cDNAs were used as reference.

(C) Distribution of H3K4me2 along aligned long, medium and short rice genes. Full length cDNA were used as reference.

(D) Distribution of DNA methylation along aligned long, medium and short rice genes. TEs were used as reference.



Supplemental Figure 8. Occurence of DNA methylation, H3K4me3 and H3K4me2 in promoter or body regions of genes in cultured rice cells.

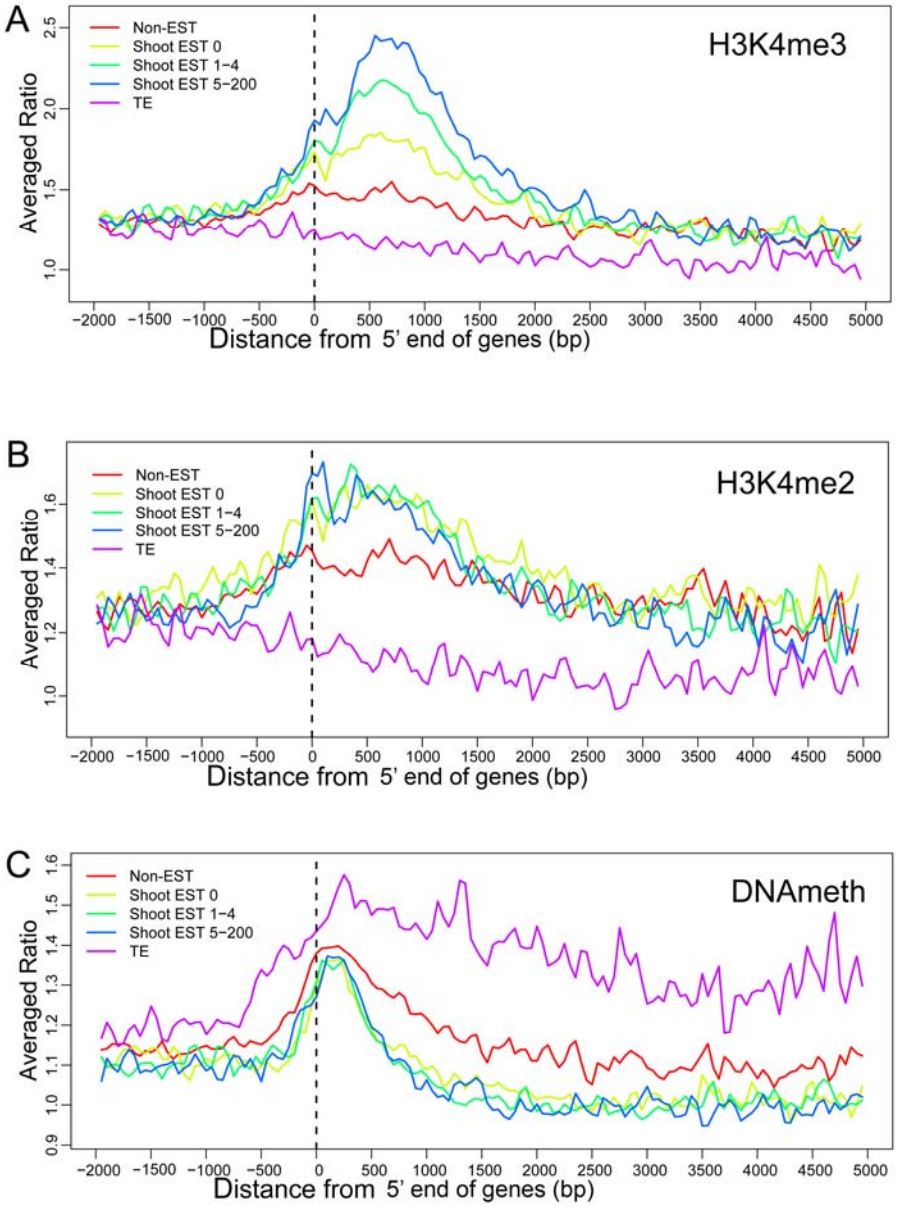
(A) Numbers of genes or TEs that have only promoters modified, only bodies modified or both promoters and bodies modified by H3K4me3, H3K4me2 or DNA methylation.

(B) Proportions of genes modified only in their bodies. All the protein-coding genes were classified into three pairs of categories: fl-cDNA or EST supported genes (SG) vs. unsupported predicted genes (UG); genes with known or predictable functions (Known) vs. genes of unknown function (Unknown) and genes with high homology to Arabidopsis (HH) vs. genes with low homology (LH).

(C) Proportions of genes modified only in their promoters

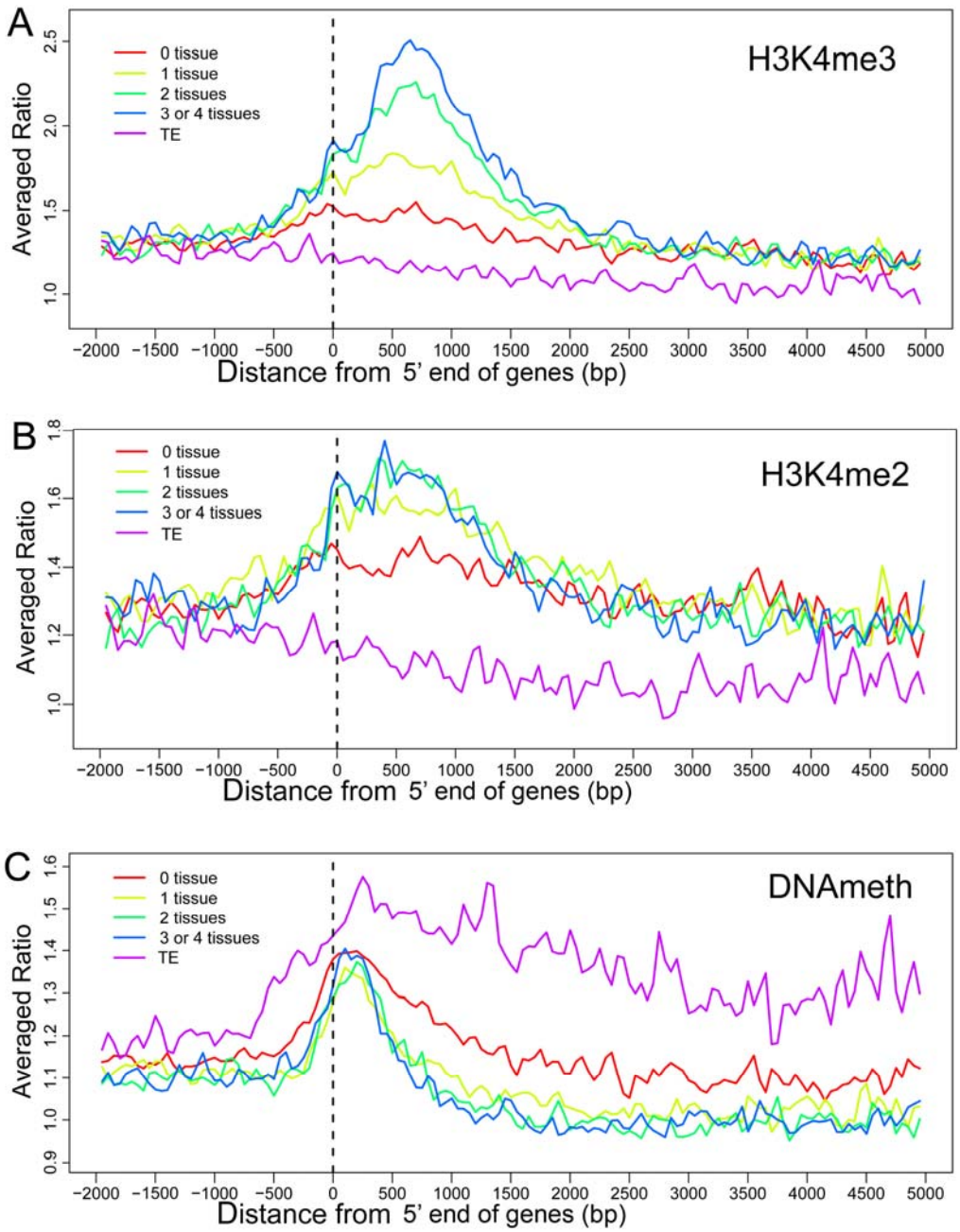
(D) Proportions of genes modified in both bodies and promoters

(E) Expression levels of genes modified in both body and promoter, body only, promoter only or unmodified by H3K4me3, H3K4me2 or DNA methylation



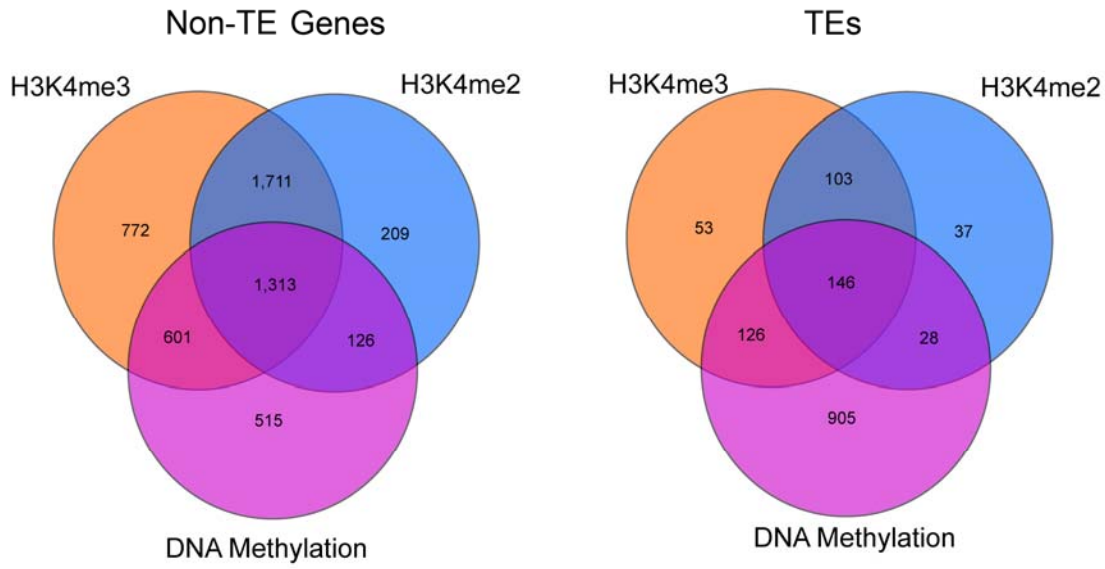
Supplemental Figure 9. Correlation between H3K4me3, H3K4me2 and DNA methylation and gene expression estimated from EST frequencies in rice shoots.

Genes on rice chromosomes 4 and 10 were divided into three categories: TE (transposable elements); Non-EST supported genes (no EST evidence in any rice tissues); and EST-supported genes. EST-supported genes were further divided into three groups: EST 0 (EST was detected in other tissues but not in leaf); EST 1-4; and EST 5-200 according to their EST frequency in the TIGR Rice Gene Expression Anatomy Viewer, http://www.tigr.org/tdb/e2k1/osa1/dnav.shtml. (A), (B) and (C). Distribution of H3K4me3, H3K4me2, and DNA methylation within aligned highly expressed, moderately expressed, weakly expressed and non-expressed rice genes. TEs were used as reference. Y-axis = averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA.

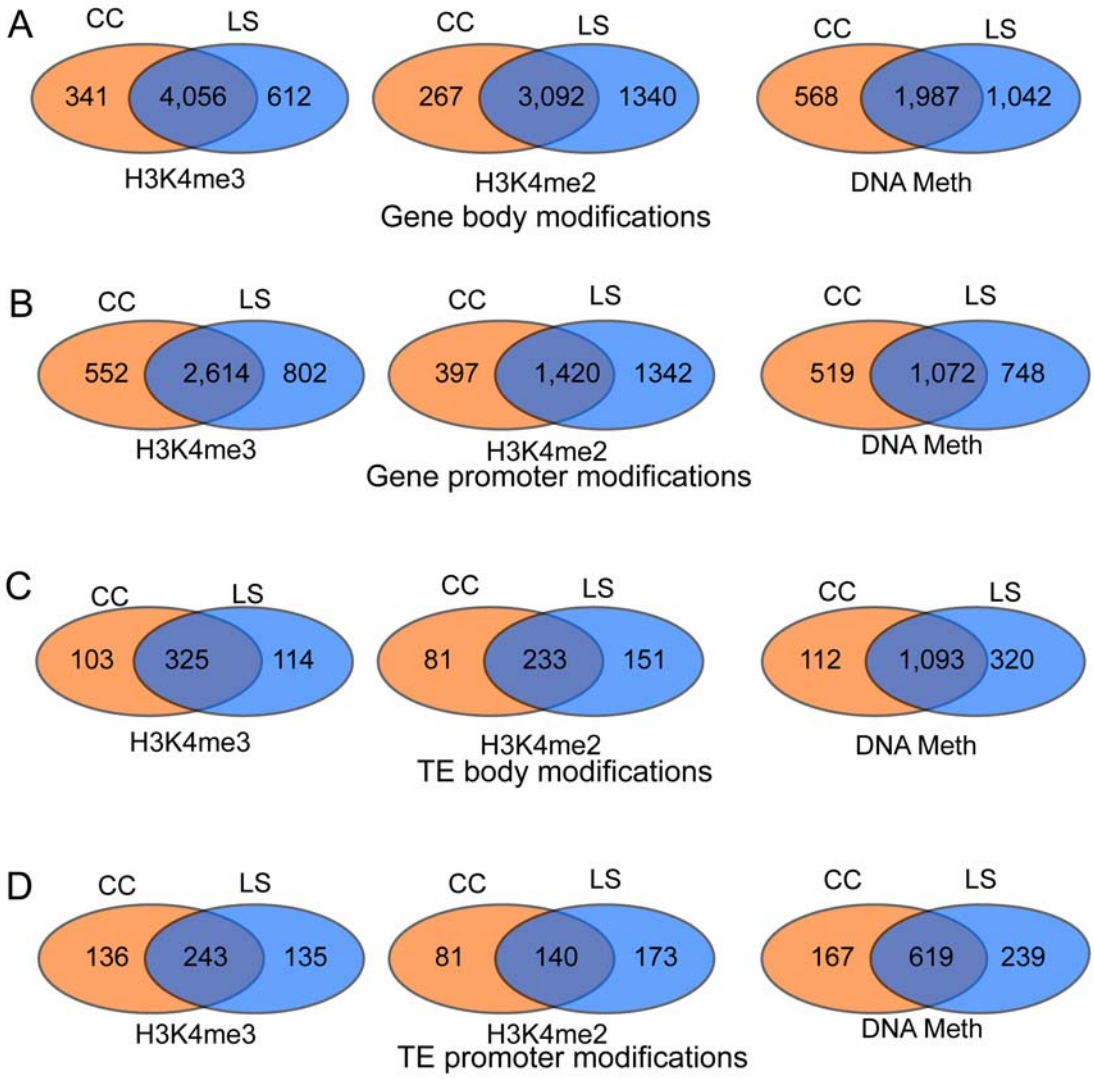


Supplemental Figure 10. Correlation of H3K4me3, H3K4me2 and DNA methylation with tissue-specificity of gene expression.

We chose the four tissues with the largest pools of ESTs in the TIGR Rice Gene Expression Anatomy Viewer, http://www.tigr.org/tdb/e2k1/osa1/dnav.shtml: panicle (150845), leaf (204353), root (79340) and callus (184189), and divided the genes on chromosomes 4 and 10 into four groups based on whether their ESTs were found in 0, 1, 2, or 3-4 of these tissue types. (A), (B) and (C). Distribution of H3K4me3, H3K4me2, and DNA methylation within aligned rice genes expressed in 0, 1,2 or 3-4 tissue types. TEs were used as reference. Y-axis = averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA.

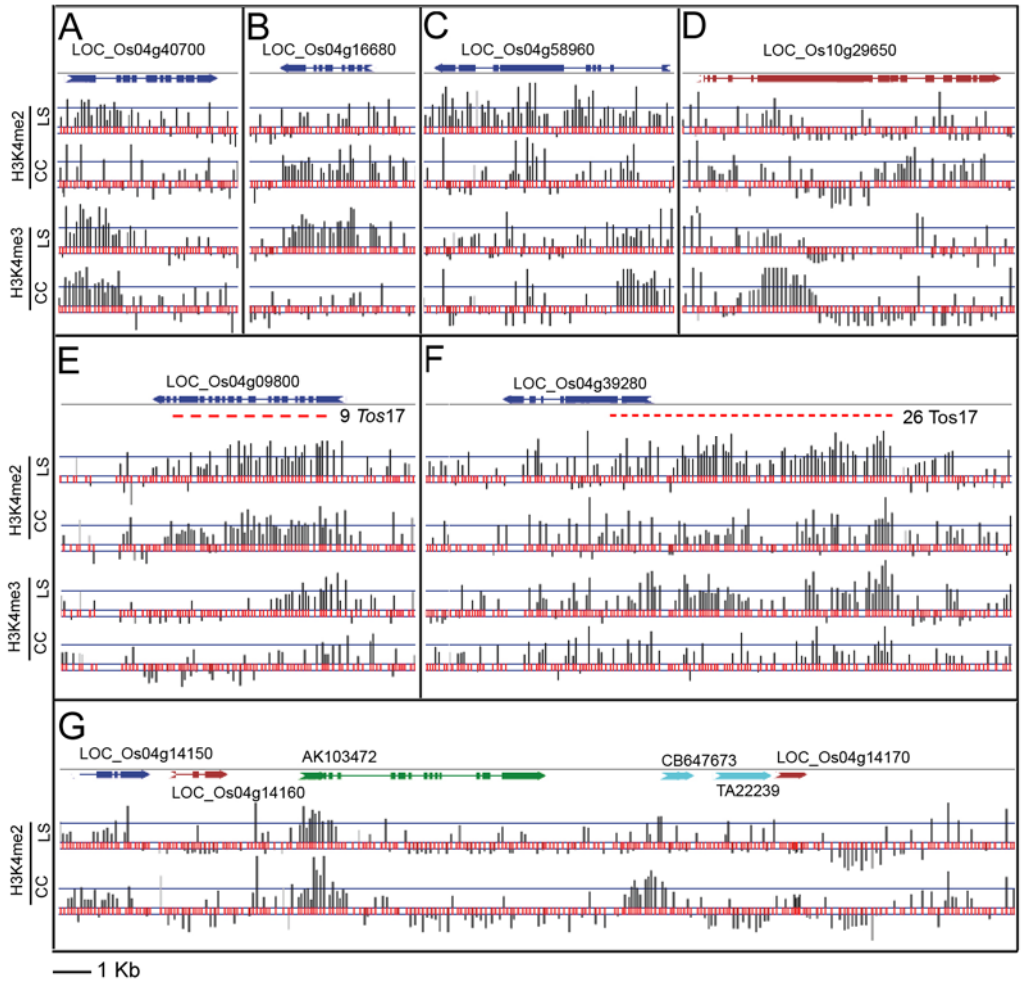


Supplemental Figure 11. Numbers of Transposable Elements and non-TE genes containing H3K4me3, H3K4me2 and DNA methylation regions and various combinations thereof in cultured rice cells



Supplemental Figure 12. Numbers of TEs and non-TE genes that are differentially modified by H3K4me3, H3K4me2 or DNA methylation in light-grown rice shoots (LS) and cultured cells (CC). Cyan = modified only in LS, orange = modified only in CC, intersection = modified in both.

A) Non-TE genes with modified bodies.
B) Non-TE genes with modified promoters.
C) TEs with modified bodies.
D) TEs with modified promoters.



Supplemental Figure 13. Examples of differentially-modified genes, Tos17 insertion sites and intergenic modifications in light-grown rice shoots (LS) and cultured cells (CC).

Y-axis (black vertical lines) = modification level, i.e. the ratios of the signal from the samples enriched for each modification to the signal from total genomic DNA for each probe. X-axis = position on chromosome. Red lines represent positions covered by probes.

(A) Gene LOC_Os04g40700: putative nuclear steroid receptor

(B) Gene LOC_Os04g16680: sedoheptulose-1,7-bisphosphatase

(C) Gene LOC_Os04g58960: Ran GTPase binding protein

(D) Gene LOC_Os10g29650: an endogenous copia-like retrotransposon Tos17 which has been reported to be activated during tissue culture but inactivated again in regenerated plants (Miyao et al., 2003).

(E), (F) Examples of *Tos*-17 retrotransposon insertions in regions of elevated H3K4me2.

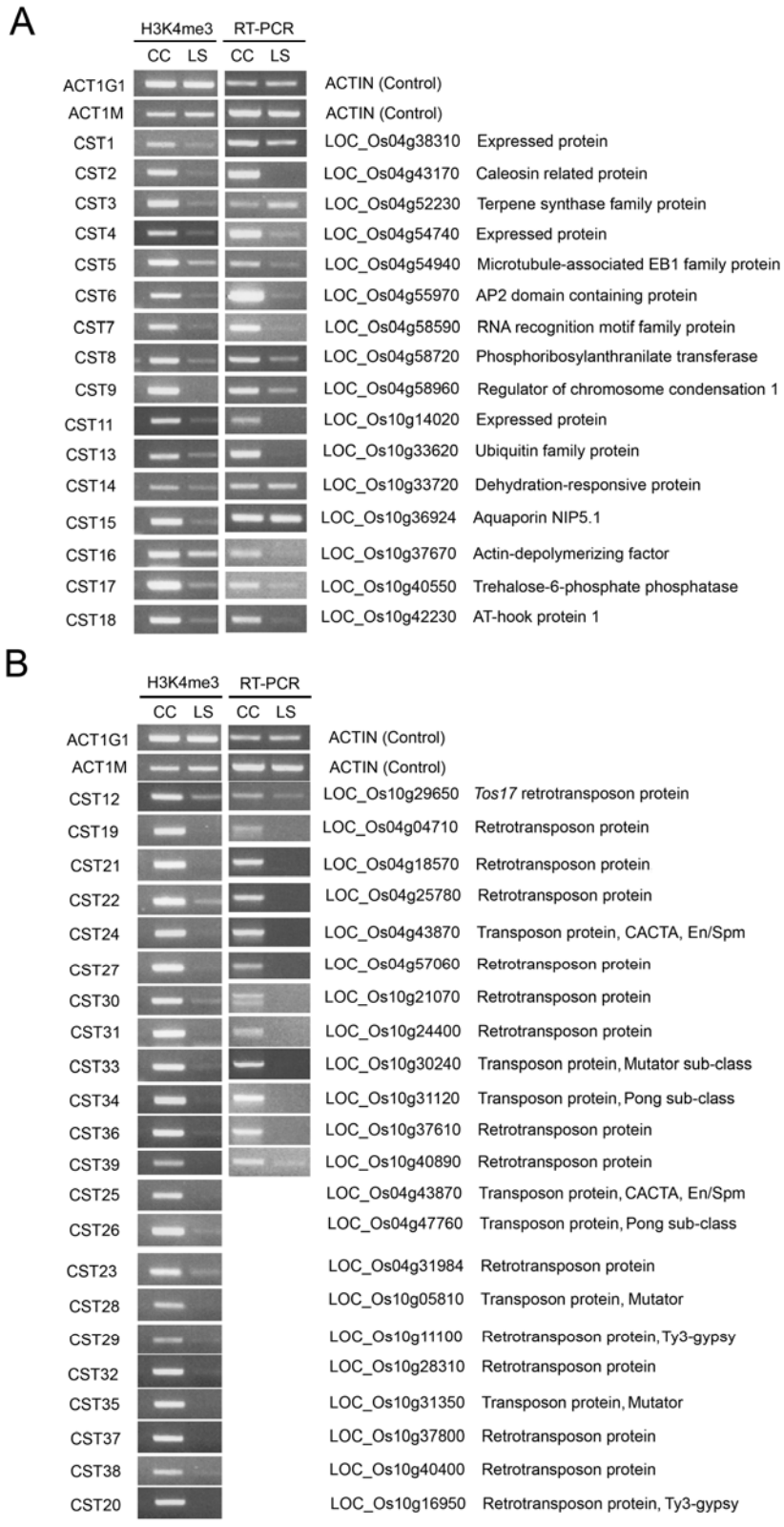
(G) An example of intergenic H3K4me2 regions corresponding to an un-annotated full length cDNA (AK103472) and an intergenic EST (CB647673).

Supplemental Figure 14. Confirmation of the positive correlation between H3K4me3 and transcript abundance. More H3K4me3 in light-grown shoots (LS) than in cultured cells (CC) correlate with more transcripts in LS than in CC.

Left column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me3

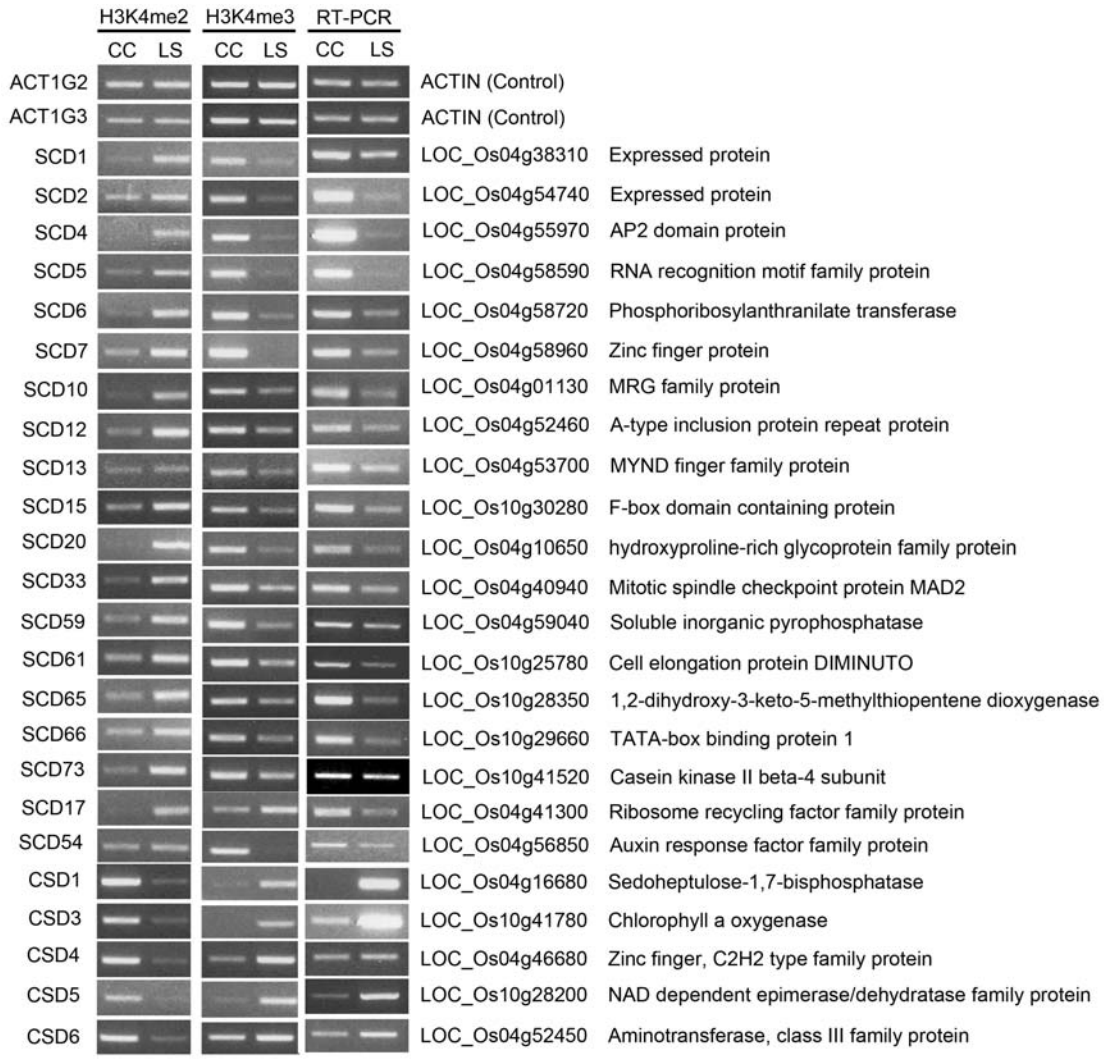
Right column = RT-PCR analysis of each gene on samples prepared from LS and CC

ChIP-PCR and RT-PCR were performed as described in “methods”



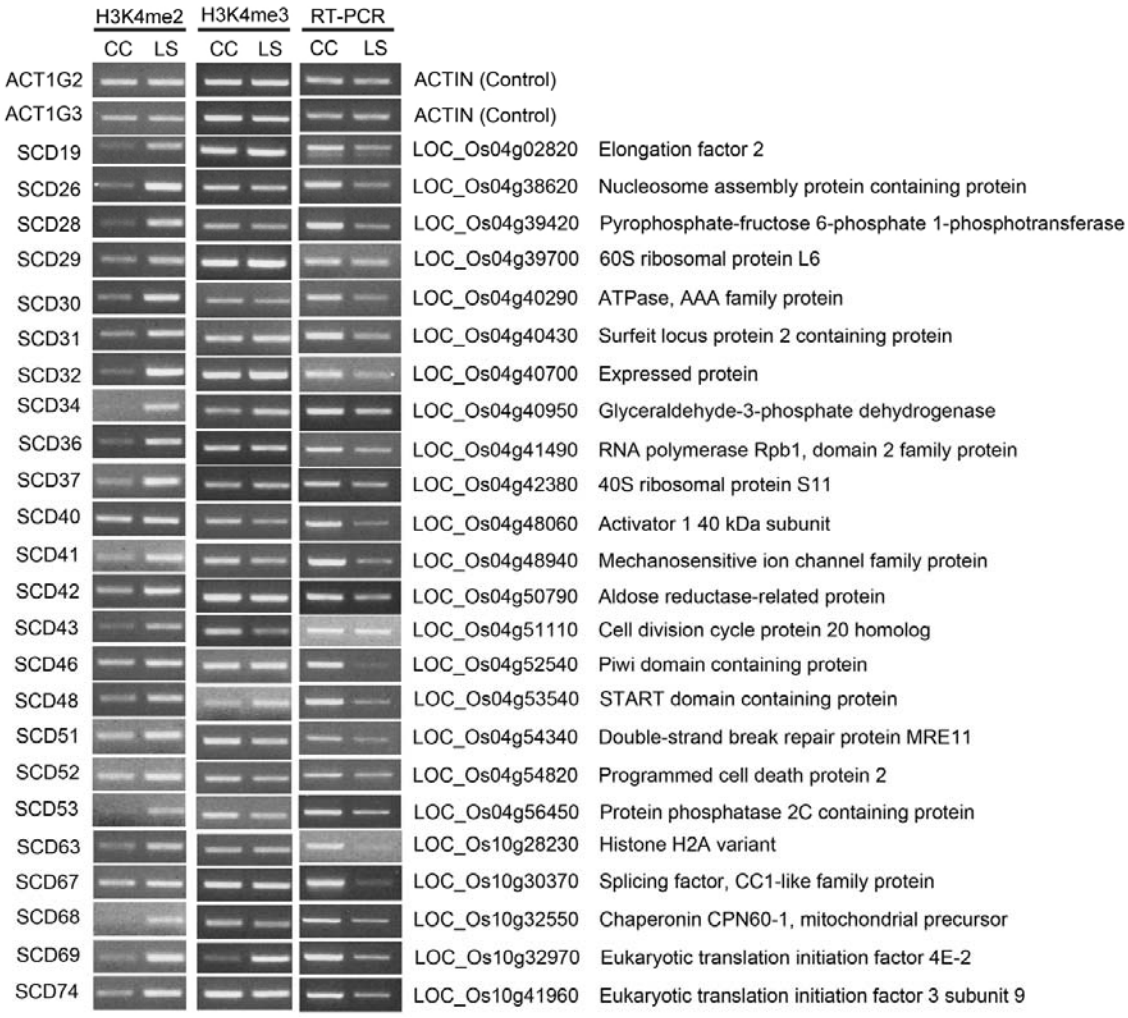
Supplemental Figure 15. Confirmation of the positive correlation between H3K4me3 and transcript abundance. More H3K4me3 in cultured cells (CC) than in light-grown shoots (LS) correlate with more transcripts in CC than in LS. (A) Protein coding genes; (B) Transposons.

Left column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me3
Right column = RT-PCR analysis of each gene on samples prepared from LS and CC
ChIP-PCR and RT-PCR were performed as described in “methods”



Supplemental Figure 16. The levels of H3K4me3 rather than H3K4me2 positively correlate with transcript abundance.

Left column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me2
Center column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me3
Right column = RT-PCR analysis of each gene on samples prepared from LS and CC
ChIP-PCR and RT-PCR were performed as described in “methods”



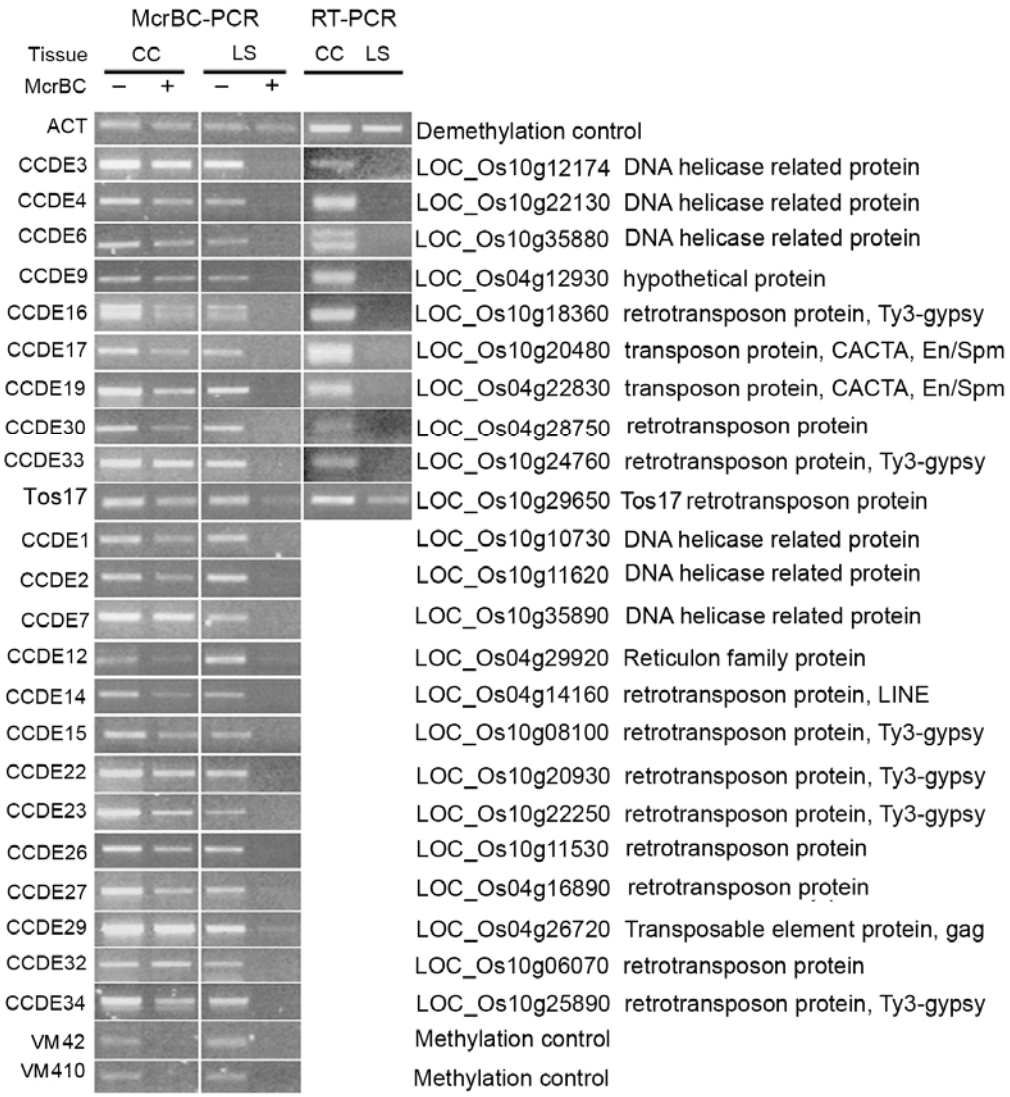
Supplemental Figure 17. The levels of H3K4me2 correlate with reduced gene expression in light-grown shoots (LS) and cultured cells (CC) if both tissues have equal levels of H3K4me3.

Left column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me2

Center column = ChIP-PCR analysis of each gene on samples prepared from LS and CC with anti-H3K4me3

Right column = RT-PCR analysis of each gene on samples prepared from LS and CC

ChIP-PCR and RT-PCR were performed as described in “methods”



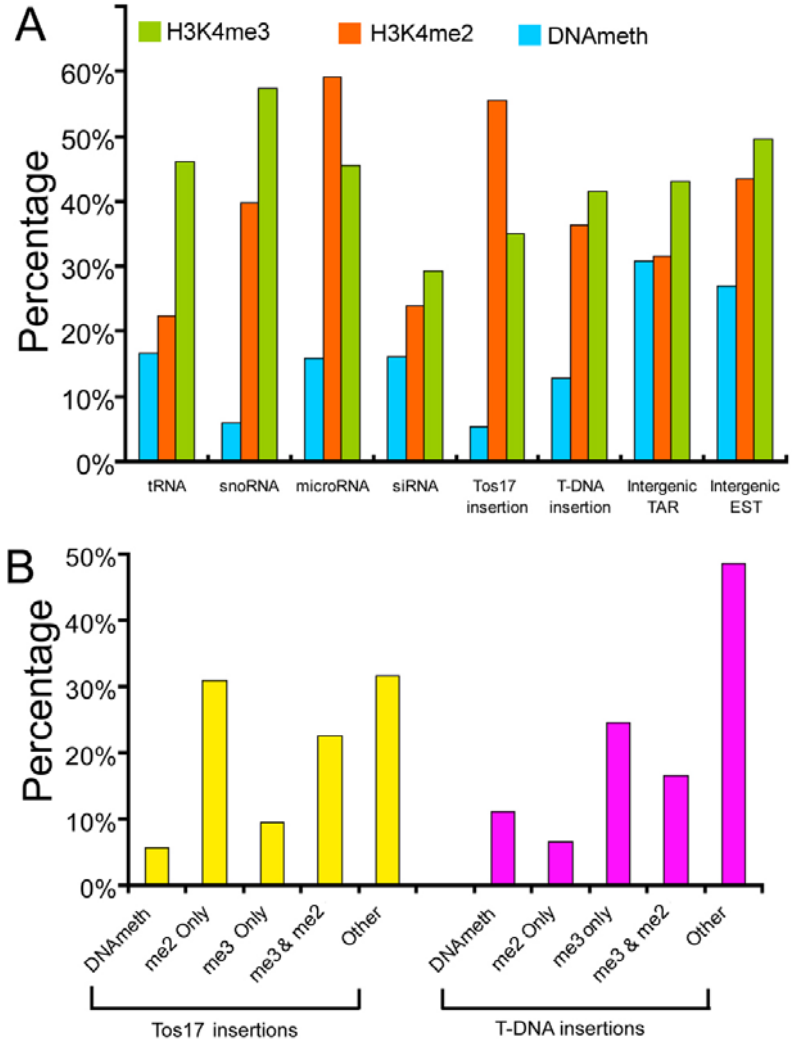
Supplemental Figure 18. DNA hypomethylation correlates with increased transcription in cultured rice cells (CC) as compared with light-grown shoots (LS).

Left column = McrBC-PCR analysis of each gene on samples prepared from CC. "-" = treated with heat-inactivated McrBC, "+" = treated with intact McrBC.

Center column = McrBC-PCR analysis of each gene on samples prepared from LS. - "-" = treated with heat-inactivated McrBC, "+" = treated with intact McrBC.

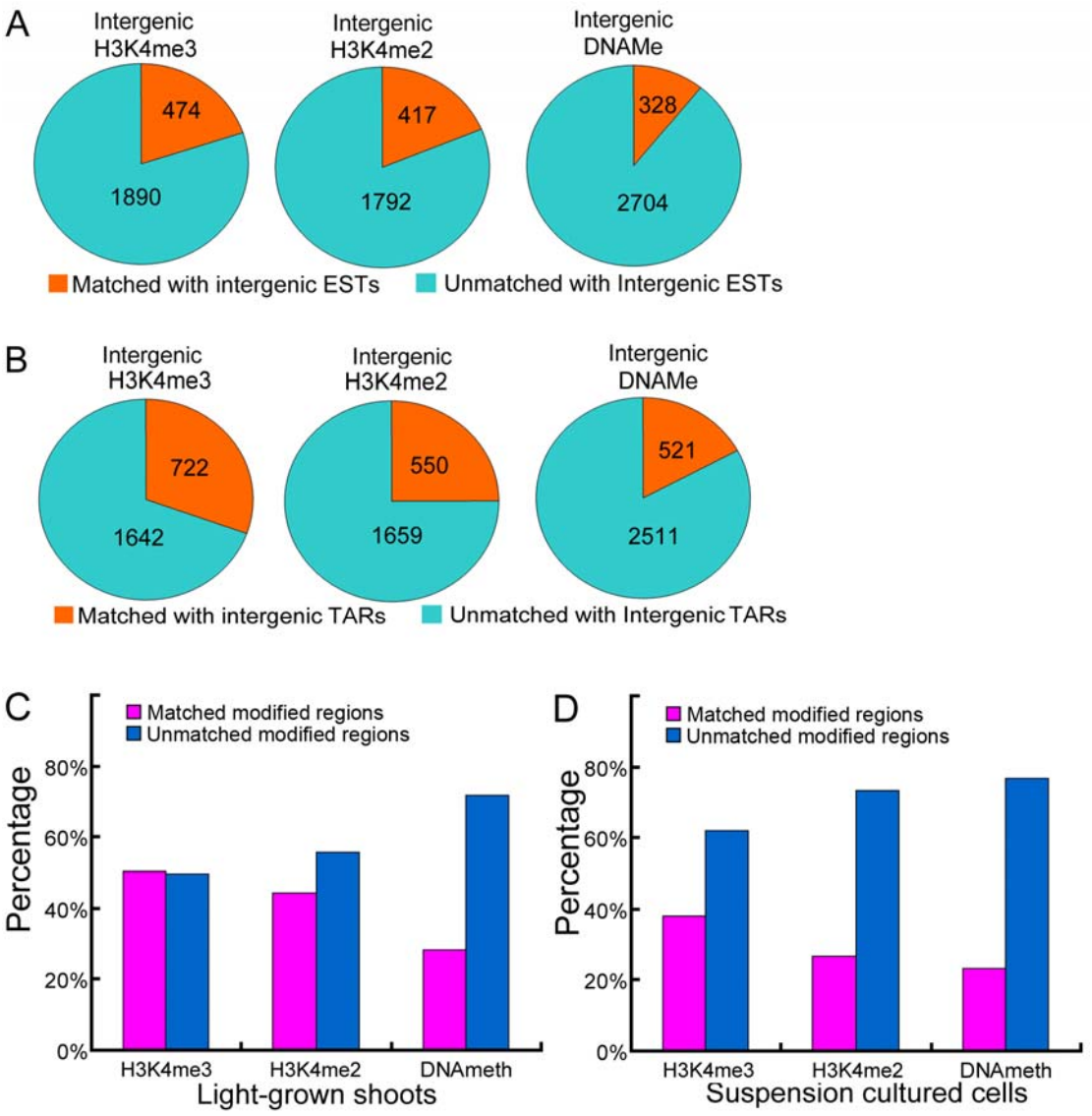
Right column = RT-PCR analysis of each gene on samples prepared from LS and CC

McrBC-PCR and RT-PCR were performed as described in "methods."



Supplemental Figure 19. Mapping H3K4me3, H3K4me2 and DNA methylation regions onto other genomic elements in cultured rice cells.

(A) Percentages of tRNA, snoRNA, microRNA, siRNA, Tos17 and T-DNA insertions, intergenic TARs and ESTs that overlap with H3K4me3, H3K4me2 and DNA methylation regions.
(B) Frequencies of Tos17 and T-DNA insertions occurring in H3K4me3, H3K4me2 and DNA methylation regions

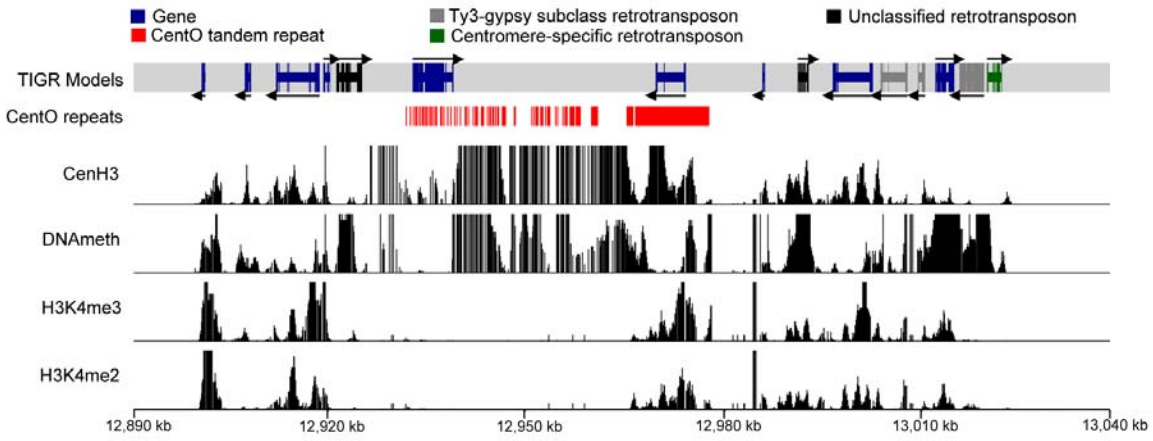


Supplemental Figure 20. Mapping intergenic epigenetically-modified regions against intergenic TAR and EST.

(A) Proportions of intergenic H3K4me3, H3K4me2 and DNA methylation regions that match intergenic ESTs (including full-length cDNAs).

(B) Proportions of intergenic H3K4me3, H3K4me2 and DNA methylation regions that match intergenic transcriptionally-active regions (TAR).

(C) and (D) Percentages of intergenic H3K4me3, H3K4me2 and DNA methylation regions that match intergenic transcribed sequences (EST, TAR and ncRNA) and those that do not in rice shoots and cultured cells.



Supplemental Figure 21. Epigenetic modifications of rice centromere 8. Rows 3-6 = $\log_{10}$ -transformed p values for each probe.

The locus ID numbers for the TE and non-TE gene models are summarized in Supplemental Table 3.

Supplemental Table 1. Coverage of three epigenetic modifications on chromosome 4 and 10.

H: Heterocrhomatin; E: Euchromatin; W: whole chromosome; LS: light shoot; CC: Cultured cells

		DNA Methylation (CC)			H3K4me2 (CC)			H3K4me3 (CC)		
	Genomic length	Region number	Region length	Percent	Region number	Region length	Percent	Region number	Region length	Percent
Chr.4H	17,500,000	2,394	2,682,539	15.33%	1,541	1,531,634	8.75%	1,880	2,253,821	12.88%
Chr.4E	17,744,269	2,640	2,262,198	12.75%	3,387	3,541,006	19.96%	3,905	6,108,126	34.42%
Chr.4W	35,244,269	5,034	4,944,737	14.03%	4,928	5,072,640	14.39%	5,785	8,361,947	23.73%
Chr.10H	12,500,000	1,782	1,848,022	14.78%	1,467	1,534,010	12.27%	1,736	2,129,529	17.04%
Chr.10E	10,376,596	1,410	1,240,139	11.95%	1,782	1,839,527	17.73%	2,026	3,100,646	29.88%
Chr.10W	22,876,596	3,192	3,088,161	13.50%	3,249	3,373,537	14.75%	3,762	5,230,175	22.86%

		DNA Methylation (LS)			H3K4me2 (LS)			H3K4me3 (LS)		
	Genomic length	Region number	Region length	Percent	Region number	Region length	Percent	Region number	Region length	Percent
Chr.4H	17,500,000	2,703	3,172,193	18.13%	1,694	1,960,473	11.20%	1,780	2,315,408	13.23%
Chr.4E	17,744,269	2,791	2,516,995	14.18%	3,443	5,477,057	30.87%	3,270	6,261,685	35.29%
Chr.4W	35,244,269	5,494	5,689,188	16.14%	5,137	7,437,530	21.10%	5,050	8,577,093	24.34%
Chr.10H	12,500,000	1,965	2,150,596	17.20%	1,487	1,875,041	15.00%	1,579	2,132,991	17.06%
Chr.10E	10,376,596	1,517	1,452,326	13.99%	1,865	2,859,182	27.55%	1,873	3,366,749	32.45%
Chr.10W	22,876,596	3,482	3,602,922	15.75%	3,352	4,734,223	20.69%	3,452	5,499,740	24.04%

Supplemental Table 2. Percentages of different classes of genes that are modified only in body, only in promoter or in both body and promoter.

Supplemental Table 2.1 Only body modified genes (Cultured cells)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	235	10.45%	213	9.48%	621	27.62%
Gene	6329	1642	25.94%	2087	32.98%	1664	26.29%
SG	2934	887	30.23%	1359	46.32%	867	29.55%
UG	3395	755	22.24%	728	21.44%	797	23.48%
Known	4136	1220	29.50%	1692	40.91%	1183	28.60%
Unknown	2193	422	19.24%	395	18.01%	481	21.93%
HH	1684	485	28.80%	757	44.95%	490	29.10%
LH	4645	1157	24.91%	1330	28.63%	1174	25.27%

Supplemental Table 2.2 Only body modified genes (Light-grown shoots)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	256	11.39%	236	10.50%	753	33.50%
Gene	6329	1592	25.15%	2038	32.20%	1915	30.26%
SG	2934	836	28.49%	1233	42.02%	956	32.58%
UG	3395	756	22.27%	805	23.71%	959	28.25%
Known	4136	1134	27.42%	1573	38.03%	1337	32.33%
Unknown	2193	458	20.88%	465	21.20%	578	26.36%
HH	1684	446	26.48%	668	39.67%	557	33.08%
LH	4645	1146	24.67%	1370	29.49%	1358	29.24%

Supplemental Table 2.3 Only promoter modified genes (Cultured cells)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	186	8.27%	120	5.34%	202	8.99%
Gene	6329	411	6.49%	545	8.61%	700	11.06%
SG	2934	76	2.59%	206	7.02%	303	10.33%
UG	3395	335	9.87%	339	9.99%	397	11.69%
Known	4136	184	4.45%	339	8.20%	435	10.52%
Unknown	2193	227	10.35%	206	9.39%	265	12.08%
HH	1684	50	2.97%	117	6.95%	164	9.74%
LH	4645	361	7.77%	428	9.21%	536	11.54%

Supplemental Table 2.4 Only promoter modified genes (Light-grown shoot)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	195	8.67%	165	7.34%	198	8.81%
Gene	6329	340	5.37%	368	5.81%	706	11.16%
SG	2934	45	1.53%	67	2.28%	318	10.84%
UG	3395	295	8.69%	301	8.87%	388	11.43%
Known	4136	134	3.24%	157	3.80%	442	10.69%
Unknown	2193	206	9.39%	211	9.62%	264	12.04%
HH	1684	44	2.61%	57	3.38%	163	9.68%
LH	4645	296	6.37%	311	6.70%	543	11.69%

Supplemental Table 2.5 Both promoter and body modified genes (Cultured cells)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	193	8.59%	101	4.49%	584	25.98%
Gene	6329	2755	43.53%	1272	20.10%	891	14.08%
SG	2934	1762	60.05%	766	26.11%	319	10.87%
UG	3395	993	29.25%	506	14.90%	572	16.85%
Known	4136	2187	52.88%	1014	24.52%	468	11.32%
Unknown	2193	568	25.90%	258	11.76%	423	19.29%
HH	1684	994	59.03%	418	24.82%	200	11.88%
LH	4645	1761	37.91%	854	18.39%	691	14.88%

Supplemental Table 2.6 Both promoter and body modified genes (Light-grown shoots)

Type		H3K4me3		H3K4me2		DNA Methylation	
	Total	Modified	Percentage	Modified	Percentage	Modified	Percentage
TE	2248	183	8.14%	148	6.58%	660	29.36%
Gene	6329	3076	48.60%	2394	37.83%	1114	17.60%
SG	2934	1904	64.89%	1431	48.77%	422	14.38%
UG	3395	1172	34.52%	963	28.37%	692	20.38%
Known	4136	2432	58.80%	1890	45.70%	592	14.31%
Unknown	2193	644	29.37%	504	22.98%	522	23.80%
HH	1684	1071	63.60%	806	47.86%	253	15.02%
LH	4645	2005	43.16%	1588	34.19%	861	18.54%

Supplemental Table 3: Annotated gene models in the core region of centromeres 4 and 8.

Chromosome	Locus ID	5'End	3'End	Strand	Annotation
Chr04	LOC_Os04g17780	9718249	9722555	-	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17789	9724700	9727793	-	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17798	9729263	9729974	-	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17810	9742138	9742852	+	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17820	9744325	9745743	+	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17830	9750966	9752316	+	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17840	9756597	9762105	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17850	9770720	9776997	+	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17860	9781226	9785710	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17870	9797368	9797898	+	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17880	9798286	9800256	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17890	9805511	9813368	+	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17900	9820140	9823640	+	hypothetical protein
Chr04	LOC_Os04g17910	9824244	9825961	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17920	9827937	9830041	+	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17930	9831403	9832481	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17940	9837666	9842628	-	retrotransposon protein, putative, Ty3-gypsy
Chr04	LOC_Os04g17950	9846312	9847451	-	retrotransposon, putative, centromere-specific
Chr04	LOC_Os04g17960	9855307	9857450	+	retrotransposon protein, putative, unclassified
Chr04	LOC_Os04g17970	9858207	9859185	+	retrotransposon protein, putative, unclassified
Chr04	LOC_Os04g17980	9859516	9862122	+	hypothetical protein
Chr08	LOC_Os08g21650	12890422	12890983	-	hypothetical protein
Chr08	LOC_Os08g21654	12897059	12898002	-	hypothetical protein
Chr08	LOC_Os08g21660	12901844	12908539	-	Eukaryotic translation initiation factor 3 subunit2
Chr08	LOC_Os08g21670	12909107	12910163	+	hypothetical protein
Chr08	LOC_Os08g21680	12911115	12915059	+	Transposable element protein, Retrotrans_gag
Chr08	LOC_Os08g21690	12922825	12929089	+	retrotransposon protein, Ty3-gypsy
Chr08	LOC_Os08g21700	12960165	12964786	-	DCL protein, chloroplast precursor, expressed
Chr08	LOC_Os08g21710	12969782	12971899	+	hypothetical protein
Chr08	LOC_Os08g21720	12976544	12976962	-	MYST1, putative
Chr08	LOC_Os08g21740	12981981	12983647	+	retrotransposon protein, putative, unclassified
Chr08	LOC_Os08g21760	12987359	12993517	-	RER1C protein, putative, expressed
Chr08	LOC_Os08g21770	12994697	12998846	-	retrotransposon protein, putative, Ty3-gypsy
Chr08	LOC_Os08g21780	12999660	13001544	-	retrotransposon protein, putative, Ty3-gypsy
Chr08	LOC_Os08g21790	13003127	13006000	+	hypothetical protein
Chr08	LOC_Os08g21800	13006829	13010601	-	retrotransposon protein, putative, Ty3-gypsy
Chr08	LOC_Os08g21804	13011043	13013330	+	retrotransposon, putative, centromere-specific

Supplemental Table 4. Primers used in McrBC -PCR, ChIP-PCR, RT-PCR and genomic bisulfite sequencing.

Supplemental Table 4.1 McrBC-PCR primers for validation of DNA methylated regions

Name	Primer L	Primer R
UBQ1	TAGTTGCTGACCATGCTTCG	TCCATTGGGAGACAATAGGC
ACT1	GCCGTCCTCTCTCTGTATGC	GGCAAGCAACATTGTAAGCA
VM11	TTTTGTTTGCTTGAGCATCG	GGGCTCCACAGTCATTTGTT
VM24	TTAATTCAACGGCTCCTTGG	AGACTCGCATCCAAATCCAG
VM29	CGGCGAACTCTTGTCAGTTA	CCACCTGCTACACTGCTGAC
VM37	GCGATGTTTTCGACAGGAAA	CGTTTCGACAACACACTCCA
VM39	TACTGCTGCCGAAGTGATTG	AAGTGCTTCACCGGCATAAC
VM42	TCCAGATTGAATCCGACCTC	GCTATGCTAAGGCTCGGTTG
VM43	GGGAACTCACTGAGGAGCAG	GACGCAAATTTGGATTCGAT
VM44	ATCAGATTGCAGCACATGGA	CGACGGAGTGGAGGTAGGTA
VM46	CACGGCCCGACTACAACTAT	GCTCGTGCACGTCAACTAAA
VM211	TCACATGTCCAAAACCGAGA	CGACAGTGTGTTGTGTGCTG
VM212	GAAGGCTTGCTTTTCAACCA	TTAGTCGGCGCTTTGCTAAT
VM213	ACTTGTGCCGAACTCCAGAC	TTGGTTTTGCATGGTAGACG
VM216	AGCTGTCGACTCCTCCAAAA	CGATAGCACAGGCCATATCA
VM221	GTCCCATCGATCGAAAAGC	CGGCATGGAAACGGTAGTAA
VM224	TCCTCTCCAAATCCCACATC	CCCCAACCAAAAGTCTAACG
VM226	CTCGGTTGCCTTCTTCTCAC	TTTTGCAACAGAACGTCAGG
VM310	CACGAGATTGGAAGGGAAGA	ACGGTTGTAGTCGGATCGTC
VM311	GATGCCAACGCTCATCTGTA	CTGACCTCCTTGTGGACGAT
VM312	AATCAGCTCCGCAACCATAG	CCCAAACCCTAACCCTAACC
VM313	GGCTAGTCAACAGGGTCGAA	TTGATCTCTCATGCGTTTGC
VM314	CGGTTGAGGAGGAGAGAGAA	GAGTACGTCCCCCAATGAGA
VM316	GTGGAGATGAAATGGCCTTG	TGCAGGTATTGACGGTCTGA
VM317	CCTCTGGGATTGGGAAAAAT	TTTGACGAGCACATCTCAGG
VM318	TCTTGCAAGCATTGGTCATC	AAATCGAGAGAGAGGGATCG
VM319	CGTCTAACCCCACTTGCTTC	TTCTGGGCGTGACAGTGTAG
VM320	AAGATTCATCACCACGCACA	GGAGTCCACCTCTCCTCCTT
VM321	AAGCCCACTAATTTCGGTGT	TAGGGCACTGCTCTGTTGTG
VM322	CTGAAACCGGTGTCCTCAAC	TAGGAAGAGGCTCCGTCAAA
VM323	CAAGCCTCCTCTCATCCTCA	GTATTTATGCGCCATTGCTG
VM324	AAGCCGAGCCAATGTTTTAG	ATATCAAGCCAGCGAGCAAG
VM328	CCACAAGAAGTGGGTGAGGT	CGACACCTCACACTCGGTAA
VM329	GGAGGGTTAGGAATGGGAGA	GGTTAACCCCACTTGCTTCC
VM330	CAGGAGAAGGTTGCTTTTGC	ACTTGAGTCCGGCATGTCTT
VM331	GCTGAGGTCATTGCTCCTTC	CCATTTTTCACCTTGGGTTG
VM332	CGGATGGTGTCGGAATAAAA	TTTCAACCATGGCAACGTAA
VM333	CGGGCAAACCCTAATCACTA	AGGTTTCGGCATGGTTGATA
VM334	AATGCGCCATTAGTTCAACC	GGTCGAAACACTGTCTTGCTC
VM335	ACCCGGATGCAAGATTACAG	CTAGCCATATTGTGCCGTTG
VM410	CCTTCTTGGGGGTTTCTCTC	CAAGTTCGGCAAACTGTCAA

Supplemental Table 4.2 ChIP-PCR primers for validation of H3K4me2 modified regions

Name	Primer L	Primer R
VDM1	CCAGGAAGTCGAGCTTGTGT	CATCATCACCCACCACCAC
VDM2	GGTGCCGTAGATCGAGGATA	CCAACTATGCACCCATCACA
VDM3	CTCTTTAACGGCGCTACCTG	AAAAACAAACGGTGGATTGC
VDM4	GGCACTTGGTCATGTTGTTG	CCTCCCCACAAGATCGTAGA

VDM5	ACATATGGGTCCGTCTCAGG	TCTCCGCTCTCCCTTGTCTA
VDM6	GGCTCAGCCAAACCAAAATA	GGCTCTTGCACGTGACATC
VDM7	TGTGTGCTGCGATAGGTAGC	TTGTACTGCTCGGGGAAGAT
VDM8	CTGCAGGAAATTCAGAAGCA	CTTGGTTGTGCTGAGGATGA
VDM9	CATTTTTCCCATCGCTTTTG	AAACGGAGAAGCGAAGAACA
VDM10	ATGATGCAGAGCAAGTGCAG	AAGACGGATAGGGGAGAGGA
VDM11	GTCGCCATGGGGAGTAGTAA	CGTCGTTTTCTTGGTTTGGT
VDM12	GGAGCGGAAGTAATTGTCCA	GCGTGTGGAATGAAGCTTTT
VDM13	GTTTGTCCCCCACAAACATC	CTCGCTGAAATGTTGCAGAA
VDM14	TGCCAGCCAATAGTGAACAA	GCTCACCCATGGAGAGGTAA
VDM15	TTCTCATGTGGGTTAGGAGCTT	ACCATGGCGAATTCTTTGTC
VDM16	AAGCGCATACACACAAGCAC	TGCCAGGAGGTGAGAGTACA
VDM17	TTGTCCTTGACCTCCTCCAC	GCTTTCATGGCGTACGTTTT
VDM18	AGAGAAGCCCGGTGAATTTT	GGCCATAAGAAACGAGGTCA
VDM19	CAAACAAGGCAACAGCTTCA	GCGCAAAGTTCTCGGTATTC
VDM20	CTCAAGCCCATCTCCTCAAC	TTTTGACCCTGGTTTTGTCC
VDM21	GGTTTCCCTGGTTACAAACG	TCATCAACCTTTTGCAGGTG
VDM22	AGCTTAGGAGCACCACCAGA	ATAGCCAGCATTTTGCTTCG
VDM23	ATTGTGTGCCACCATCAAGA	GGCTTCAGAAGGACCATGAG
VDM24	GAAGAGCAGCTTGATTTGGTG	GGCTGTGTGGGAGTAGGAGA
VDM25	CCTCGAGCATGCAATCAAAT	GAATGCAGGACACTGCGATA

Supplemental Table 4.3 ChIP-PCR primers for validation of H3K4me3 modified regions

Name	Primer L	Primer R
VTM1	GTCGTGCAGGAGGCAGTC	TTTAGAGCGCACCCTCAGTT
VTM2	TGATGGCAACCAAGAAGTCA	CACATCGAGAGGTGCTCGTA
VTM3	GCAATCCACCGTTTGTTTTT	GGCCAAAATCTGGCAACTTA
VTM4	GATGGTGTTGAGGAGGAGGA	ACCCCTACCCTGACACTTCC
VTM5	CGAGAGCCTATGGGCTACAA	AAAACGCAAGCACTTGATGA
VTM6	GCAAACCGAACAGGCATTAT	GCTTTGATTGTGGGGAGCTA
VTM7	TCTGGGCCTGCAGTTAATTC	TGGCGACTATCTTCCCCATA
VTM8	CAACTGCCCACTTGATTCAC	CTTGGTTGTGCTGAGGATGA
VTM9	ATCCCAACCTGCAATTTTTG	TCTCCTGCTTAGTGCCACCT
VTM11	TCCACCCCACAAGAGAGAAG	ACTCGTCGAGGTGAATCTGG
VTM12	CATCGAGAAAAATGCCCACT	CTTTGGTGCTGTGCAAATGT
VTM13	AACCCATTCACCTGTTCAGC	CTGACCATACGGGCTCTTCT
VTM14	GACTCCATGGGGAGTAGCAA	TCCCGGTCGTCAAAGAATTA
VTM15	TGGCGATTTGTGTTCTTCTG	CATGCTCCCGCATATCTTTT
VTM16	TTCCACAATCCGTCTCATCA	AGTGAATTGCCGGTGCTATC
VTM17	GCTAGATTGGGAGGACACCA	CAACACGGACTCCAAAACATT
VTM18	TGGTACCTCCTCCTGTCACC	CCAACCGCAATGGAGACTAT
VTM19	AAATGCAAACCTGCCATACC	TGAATTCAAGCGCTCCTTTT
VTM20	GCAGCCATGCCAAGTATTTT	TGCTATACGCAATGCTGGAG
VTM21	TTTTCCAGATCCTGTTTGACC	GCATGATAAAAATTGCATACGG
VTM22	GCTTTTGATCGGCATCTCTC	AAGGTGGCTGGAAACATTTG
VTM23	TGTACAGCCATCCATGCACT	GTGGAGGAGGTCAAGGACAA
VTM24	CTGCTGCTAGCTTTCGGTCT	ATCGCATCGATCAAGCTACC
VTM25	AACGAGGTCTCTCAGCGGTA	CGGAGATAGGGAAGGGGATA

Supplemental Table 4.4 PCR primers for bisulfite genomic sequencing validation of DNA methylated regions

Name	Primer L	Primer R	LOC ID
A	TTAAGTGGATTGGTTGAATT	CTAATAAAAACCTAAAATTACACC	Os10g24100

B	TGATTAGTTTGGTTGTGTTAAGT	AAATTACTTCCTATATCCCT	Os04g29580
C	TAGTAGAGAAGGTTTTTAAG	AATACTAAAAAACACCCTCCT	Os04g16980
D	GATGTTGATTTAAGAAAGTA	CAAACAAAAAAAACRAAAAAACC	Os10g39930
E/F	TATGGTGTTATTTTGTTTYGTGAAG	CTATCACATCTTTTTCTTATT	Os10g22130
G	GGATTTGTATTTGAGTATTAAG	CCTTCTTTCAACATTAAAAA	Os04g53630
H/I	GAGTAATTTATGTGATTATAAGG	RCCAATTATACTTAACCCAA	Os04g07670

Supplemental Table 4.5 ChIP-PCR primers for validation of differential H3K4me3 modification between two tissues

Name	Primer L	Primer R
ACT1G1	CGTCTGCGATAATGGAACTG	CACACGGAGCTCGTTGTAGA
ACT1M	GCCGTCCTCTCTCTGTATGC	GCAATGCCAGGGAACATAGT
SCT1	AGCAATCGAGCTTCTTCAGC	CAGAAGCCATGGTTTTGTGA
SCT2	CCCATCAAGGTAGGCATCTC	CATTGCGTTACACTGGAGGA
SCT3	TAGGAGGGACCTGAACGTGA	CAACCACAGCCTGCAAATTA
SCT4	GCCAAACAGAGTGCACATGA	TTGATCGCTAGCTGTTGCAT
SCT5	TTTCTTCCCCACCTCATCTG	CGGTCTCGCTCCTAATTCAT
SCT6	TGACTACTGCCATTGGCTTG	GTTGTCGACGCTCAATCAGA
SCT7	AACTGTCCGTCTTCCCATGT	TGGATCATAGAACGCACCAG
SCT8	CCTGTGCTTGATCCATCGTA	GCAATCCTTTTTGCCACATT
SCT9	AAAAAGGAAGCCTCGTAGCC	GATGGTTCTTAGCCCTGCAA
SCT10	TGATGAGCAGGGATTGTTGT	CACAGCACCAACGTTACCAA
SCT11	TCTTGTCTTTGGGGAAATCG	TGCTGGAGTTGAAGCTGATG
SCT12	CTAGGGGGAAAAGGCAAAAG	GGGAATTCTGTGTGCTGGTT
SCT13	ATCTACGGACCGTGTTCTGG	AACAGCATTGCAGCACACAT
SCT14	TCTGCGGCTGGTCTAGAGAT	GTTCATCAAGCTTGGCTTCC
SCT15	CAAAGTTATGGGACGTGGAA	TCCGTGACAAGGATGACAAA
SCT16	AAGCCATAACCCAACAACCA	AATCTTGTGTTTCGGGATGC
SCT17	GACAATGTCAGGGGCTTGTT	TCAAAATGCCAGGCTAATCA
SCT18	TTTGCAACAAAGCATGTGGT	CAACCCATCAAGAACTGCAA
SCT19	TTTTTCCGTCGAGTTCCATC	CAGTAGCAGCCTCCCTTGAA
SCT20	CTGGACCAGACACGTTCAGA	TGGCAAAAGCTCACAATCAG
CST1	TCAGCGAATATCCCACTTCC	TTTATCTTCCTCTGCGGTTGA
CST2	GACCAGAGCAATGAGGAAGG	AGATGGATGCGTCTCTCGTT
CST3	AGATTGCTGCTCCTGTTGGT	AAAAGAATGCAGGGGGAACT
CST4	GGCAACAAGGCCATTTTAAG	TGAGAGGCAGGTAAGATGGTG
CST5	ACTCGGACTGTTGTGTGCTG	CACTCATTTGTGAGGCCAGA
CST6	GGACGCCAAGGTCACTACTC	GCCGCTTTCTCCTCCTTATC
CST7	AGCACGGTGATACCATGACA	TTTGTTTGTGCTGTGCGATT
CST8	GCAATGGCGGCATATAAGTT	CAAGGAAGGACCTTGAACCA
CST9	GGAAGGATCTCCAAATGCAA	CATCTCTTCGCGCCTTTTAC
CST11	TGTTCAACGTCTTGGACTGC	GCAAACATGCATGAGCAGAT
CST12	TCAACTGGGACACGGATTTT	ACGGTGGCACCATTTTTAAC
CST13	GATGAGGCGAGGAGTGAAAG	TCCTGAACAACCGAATCACA
CST14	CTTTGGCACCCAATGATGTA	CTCTGCAATTTTCCAGCACA
CST15	TGATTTGATGGCTTTTGCAG	CCAAGTCCCAAAAGCATTGT
CST16	GCTCTTCCTGCAGTTGTCGT	GAGGAGGAGGAAGACGTGTG
CST17	CACCGATCTGATTGCCCTAC	AGGTGCTCCTGTTGAGTGCT
CST18	CCTTCAGCAAATTCATCTGGA	CCACACCTGTGGACTGTGAC
CST19	CCCCTGTGGGAGGATTACTT	AATTATTCCGAGCCGACCTT
CST20	GGCCACTTGGTTTGAATTGT	AGTAGGGGTGGGTCTGCTTT
CST21	ATCGCGGAGAATGAGAAGAA	CTTGCAAGAATTGGGTCGAT
CST22	TAACACGTATCCGGGGTTGT	AGCAACAGCCAAAGCAGACT

CST23	CAACATCCTCCCCATTGAAC	TCTTTTGGCCCAGTTTTCAC
CST24	CCATGATGAATCCACCGAGT	CAACAGCCAGCCAAGGTAAT
CST25	AAGAGAGTGGAAGGGGAGGA	GAGATGGGGTTTCAGGTTCA
CST26	GTGTGCATTTGGTGTTTTGC	TGCCTCCAAATTTGTTCCAT
CST27	TCCAGGAGTTCTGGGTGTTC	TGACACTGCAGAGAGGGTTG
CST28	CCTCATCATCAACACCCACA	GAGGTTAGATGGGGTGAGCA
CST29	GGAATTCAAAAGAGGCACCA	CTCCCCACCTCAGGACATAA
CST30	CCTCCTCATGAAGGTTGGAC	TTGGGGGAATGCCATAGATA
CST31	ATTTGGGGGAGAGCTCTGAT	GCTCCTCTAAGTGCTCCTGGT
CST32	AAATCAAGGGGTTGCTGTTG	GAATTCCCCATCCTTCCTGT
CST33	GCAGTTCCCAGCTCATGAAT	ATCTATAGGGCCACGAAGCA
CST34	GCAGGAAACAAGCAAATTCC	GGTTTGGCTTTTGACCTTGA
CST35	GTTGGATCCGATGGAATTTG	GGGAACAACCAATGCAACAT
CST36	CGTTGAATCCACCTGTTCCT	CTGTTGCGACTGCCATTCTA
CST37	GTCACCCTCACCCAAACAAT	CCAGGACTTGGTTGCATTTT
CST38	CAATCAACTGGTTTCGCACA	GTGGGAGACATCAAGGAGGA
CST39	AAGAAACGGAAGGGGAAAAA	GCACCCTAGAATTCCCATGA

Supplemental Table 4.6 ChIP-PCR primers for validation of differential H3K4me2 modification between two tissues

Name	Primer L	Primer R
ACT1G2	GCCGTCCTCTCTCTGTATGC	ACAGTGTGGCTGACACCATC
ACT1G3	GACGGAGCGTGGTTACTCAT	GCAATGCCAGGGAACATAGT
SCD1	ATGGTCACAACTGCAGGTCA	AATTTCAGGGTGGTTTGCTG
SCD2	TCAGGAATAATCGTACTTCTGTCG	TTCAATCCGCAAAACGGTAT
SCD4	TACAGTGCGTCGTCGTGTCT	TTCACCTGCACAAACACACA
SCD5	ATTGCTTCCCAGCAAGAAGA	CCTCAGTTCTCAACGCACAA
SCD6	CATTTTTGGCCCCTTCTACA	AGGCGTTTCTCTTCCCTCTC
SCD7	GGAAGGATCTCCAAATGCAA	CATCTCTTCGCGCCTTTTAC
SCD10	CTTTGGGTGTTCCCTTACGA	AGATTGAGGCCAGTGCAGAT
SCD12	GGGTGGTTCTTGAAAGGTCA	AGCAACAGCTGGAGTCCAAT
SCD13	GCGAAAGCAATAACTGCACA	TGAACTTCAGCAGCAAGCAT
SCD15	TTGCTACAGGCGATTTTCCT	ATCCTGGAAAAGGCTCCAAT
SCD17	TGAACTCGTTGGCATCAAAG	GGTATGCCGAAGGATGATGT
SCD19	GGCAGGTTTGCCTTGTATGT	ACCAAACCAACCAAATCGAG
SCD20	CACTTTTCTCTTGGCCTTCG	CAGAGCATGAACTGCCACAT
SCD26	CCACATGCATCTCAAAGTGC	GGAACGAACGGTTGATGAAT
SCD28	GAAAGCTGGCCAATTTTGTT	CCCGGTGGAAAAAGAAACTC
SCD29	TCAACCAAACATTTCAAACCA	CGCACTGAATCTGTGATGAA
SCD30	TTCCGGAAGAGAGTTTCTGG	GGCCTGACTTGCCAAAATAA
SCD31	TGGGAAGAGGTTCCTCAACA	CAAGGCACATCATCCACTTG
SCD32	CCTGTGTGTCACCACTCACC	AGAGGCCAAGAACCAACAAA
SCD33	CCTTGTCGCTCTTCTCCTTG	AGAAGTACGGCCTCACGATG
SCD34	TCAGATCGAAATCGCAAATG	CCATGTAGTCGGTGGTGATG
SCD36	TGCCCTTCGAGTTTCTCATT	TTCGTGCAGTGATACCTTGC
SCD37	AGAATGCCCTCTCAGTCTGC	GTTAGCGGCGTTGTGGTAGT
SCD40	AGTGGCTGCTCCATCTCTGT	GCATGTGAGGCAAACTCAGA
SCD41	ACGAATTCTAACCCGTCGTG	TGCACAAAAATTCGAATCCA
SCD42	AACATTTCCAGCACCCAGAG	GTGCTTTGGTATCGGAGCTT
SCD43	ATCTCTGCGTCGAATCACCT	CAAGCAACGATGAATCAGGA
SCD46	CCCAGTTGCTGTGTCTCGTA	GTGAACGGACTTTGACACGA
SCD48	CTTGCGCTGCTTATCATCTG	ATTTCTCTTGCCCGTCCATT
SCD51	ATTGCTTTGGGCTTTGCTTA	TATCCCCCAACAGAATGAGC

SCD52	CTTGCTCCAATGCGGTTTAT	GAACTTTGCCACAGGAGCAT
SCD53	CAGTACCAAATGCCCAAACC	CTCAAAAGCAAGGCTGGAGT
SCD54	AACTGAGATGGCAGGTTTGG	TCTTCAGGTGCCAAGAAGGT
SCD59	ACAGGATCCTCTCGTTGAGC	AGTTCCGGTAGGGGAAAAGA
SCD61	TAAGGGGTAACAGGGCAATG	GTGAGCGAGAACGCATGATA
SCD63	TGGCTTTGCGCTTCTATTTT	AGAAGGTTCCGCTCTCCATT
SCD65	TTAACCCCCAATCAGCATTC	GTTTAGGGTTGGGGTTGGTT
SCD66	CCAACAACCCCCAAGTTAAA	TCGTTCTTGGTGCTTTGATG
SCD67	CCCCTAGGCCCTTGTGTAAT	AAGACTGCCCTAGCGAATTG
SCD68	TTGCTGGCACTCAAGAAAGA	AAGCCTGCTTCCAACCTACA
SCD69	GCTGCATCGTGGATTACTGA	TCCCGGAACATCCATAAAAG
SCD73	TCTGGATGTTCACCGCATTA	CAACCTTGTACCGCTTAGCC
SCD74	CATCAGCCGCCATAAGCTAC	ATCAACGCGGCAGCTAGTAT
CSD1	TCCTCCAGTGTAACGCAATG	CGCTCTCAAAGATTGTCCTG
CSD3	TGGGGTCTGAAGAATCACAG	TTGCTCATCAAGCCTTCCTT
CSD4	GCAAGCATCTGACGAAATGA	GCTCGTTCAACGCATATCAA
CSD5	GACCATTATTGGCCGTTGAC	TCATGGACCACATGAGAAAAA
CSD6	TGGCTGCTTCTTGTTTTTCA	CCCCTTGAACCTGGAAAGAT

Supplemental Table 4.7 McrBC-PCR primers for validation of differential DNA methylation between two tissues

Name	Primer L	Primer R
CCDE1	CAAAAGAGGGGTCTTCCACA	GCATGATGGAAGCAGCTACA
CCDE2	GCGCATGCTTCTATCTAGGG	CATGCAAAAACCGCTATCCT
CCDE3	AATGTGCCACACAACCAGAA	TGGACCAGTGTGCTTGCTAC
CCDE4	ATTAAGGATGCCATGCGAAG	CTCGACAGTTCAACCAAGCA
CCDE6	GGCAATCCCTTGAACACACT	CGGCTTTATTGTCCACGATT
CCDE7	GCCCTGTTTCAAAAGCACAT	CGCTCTCAAACCCTGCTTAC
CCDE9	CGCACATATACACCCAACCTT	CCGTGCTCAGCTTGACAATA
CCDE12	GATGATGGAGGAGCACGACT	GGTAAGGCCATTCACAGTGC
CCDE14	TGCATAGGGCCATTAAGGAG	CAAGGTTCTCCCACCACTGT
CCDE15	AATCAGCGTTCGAGGAGCTA	TAAGCCCTTCGCACAACTCT
CCDE16	GGCAAGTGTTCGACCGTTAT	GGGAATCAAGACTCGTCCAA
CCDE17	TCCACTTCGGCGTTATTACC	CCTACAGGCAGGACTCAAGC
CCDE19	TGGTTGCCTGTGTGTTCAAT	GATGGACATGCTTCCAGGTT
CCDE22	GATCAAACATTGCGGTTGTG	ACCTCGCATTGACCAGGTAG
CCDE23	GCTAGCATCCGTGCCTTTAG	TGGTGTCCTGTGCCAAATTA
CCDE26	GCAATTTTTCGGTGTCCTGT	GGAGCAAAGAACGACTCCAG
CCDE27	GAGTTCACTGCTCCCTCTGC	GGGATGAAGAAATCCCGAAT
CCDE29	GGCATCGGAAAGAGGGTATT	TCAGTTCCTGCAGGGTTAGG
CCDE30	TTCAGCAGGCTATCCGTCTT	TTGGGAAAGCGAAGTAGTGG
CCDE32	TAGATGGGGTTATGCGAAGG	TCGACTCAACATGCTTTTGG
CCDE33	CGTTGTCAAGACCAAGCAGA	GGCTTCAATCCACTTGGAAA
CCDE34	TTGATTTGCTCAACGCAGAC	CGTATGGTCCACAGCATCAC
Tos17	GCTACCCGTTCTTGGACTAT	CTGAAATCGGAGCACTGACA
ACT1	GCCGTCCTCTCTCTGTATGC	GGCAAGCAACATTGTAAGCA
VM42	TCCAGATTGAATCCGACCTC	GCTATGCTAAGGCTCGGTTG
VM410	CCTTCTTGGGGGTTTCTCTC	CAAGTTCGGCAAACTGTCAA

Supplemental Table 4.8 RT-PCR primers for validation of differentially expressed genes between two tissues

Name	Primer L	Primer R
ACT1G1	CGTCTGCGATAATGGAACTG	CACACGGAGCTCGTTGTAGA
ACT1M	GCCGTCCTCTCTCTGTATGC	GCAATGCCAGGGAACATAGT

SCT1	GCTATCCTGTCTTCGCAGGT	TCACTTGCTTGCTCACAACC
SCT2	AGCACGTTCTACGGCGAGT	GAACGAGATCGTCCTCATCG
SCT3	GCCACCACCACAAAATAAGG	CAGGAACAGCCCAAGGTG
SCT4	GTGTCGGCCAAGAAGTCGT	GACGAAGATCCCCAGCAC
SCT5	GTGACATCCCCACCTACGTC	GTTGAGCTTCCCCTTGAGGT
SCT6	GCGATTCCGTAGAAGGGATT	GTCATGGCATCATTCAGTGC
SCT7	AGCCAAGGAAGCATCAGAAA	TTGCATCCTCTATGCTGTCG
SCT8	TCTGTAGGGGGTGAAAATGG	CCTGTGCTTGATCCATCGTA
SCT9	GGTGGATTTGGTTCTGATGG	GCTCTTCCTCCTGGGAAAAC
SCT10	TGCACATACCCGTCCTGTAA	AGAATTTCTTTCCGCCCAGT
SCT11	CAACTCGATCCAGCCAGATT	GGTTGCTCATTGTTGATTTTGA
SCT12	GTCCACAAGGAAGCTGAGGA	TTGCCTGTCAAGCATCAAAC
SCT13	AGAGGAAATCGGCAATGATG	CGTCTGGTCCACTTGGTTTT
SCT14	CCATGAACGAACCATTAGCA	GCGATGAAGACGAGGTCATT
SCT15	AGTCTGGACCGCATACAACC	GGGTATCCAACCGACTTGTT
SCT16	CGAAGCAGAAGCAATCATCA	AAAGGAATAGGGCCTGCAAT
SCT17	ACATCACCAAGGGCTACGTC	CCTCCATCGATCTTTCTTGG
SCT18	GGGTGGATGATCAAGGTGAA	CAGACCACTGGCTTGAAGTG
SCT19	ACCAGCAGATTGACGACACC	GGGCAATAATGGCGGAAG
SCT20	GTATCCGGTATCGGTTGTGG	CGTGATGCTGTCGCTAGTGT
CST1	ACGTTCAGTGAGGAGCCTGT	CAGGGCCAAGATATCAGCTC
CST2	AGGGAATCGAGTGTGTTTCG	GATGGTCAGCGAGTCCTTG
CST3	TGGAGGGAAAGCTGAACAGT	TCCTTGACGACTGCATTCAC
CST4	ATGATTGTTGCTGCCATTTG	GCCACTGGCTGAGTCAAGAT
CST5	AATAATCACCCGGTCACCAA	GCAAGTTACGAGCTGTCGAA
CST6	AAGGAAACCTTCAGGGGCTA	GGCTGCAACATTTGGGTAGT
CST7	AGGAAAGGTTGCCTGGATCT	TGGTCATCAGTTGGCATCAT
CST8	GACGAAGCAAAGCCAACTTC	ACATAGAACCCTGCCAATGC
CST9	AAATGGCTGAAAGGCTTCCT	TGCTGGGTGAGCTAGAACCT
CST11	ATCGGCTTCAACGACTGTCT	CCAAGGCCCTTTACAGATGA
CST12	GCTACCCGTTCTTGGACTAT	CTGAAATCGGAGCACTGACA
CST13	ACAGTTTACATCCGCCTTCG	CCCCTTCAGTTCAAGCAAAA
CST14	AACAATGAGGGAATGCTTCG	CTCTGCAATTTTCCAGCACA
CST15	TGAGAAAAGCCTGTGAAGCA	AGCAATGGGAGGATTGTGTC
CST16	ATATACGCCGTGTCGAGGAA	TACATCATGCACCCGAGATG
CST17	TGTATGCACGTTTGCAGTGA	GCTCGTCAGCTTGAACATCA
CST18	AAGCCAGTCGAGACAAGGAA	TGTCAGAGTCGTGCCTGTTC
CST19	AGCCATCAGCTCCTGAGGTA	TTGATGCCTAGGGAACCAAG
CST21	ATTGGCTGAATTGGTCGTTC	CAACTCGAGCAATGACAGGA
CST22	ACTGGCGGCCATATTTACAG	CAAAGGGAGTGGTACCCAGA
CST24	TTCCCTCCAGCTTTCTTTGA	TCATCAACCGAGCAATTTCA
CST27	GGTGCTGTCTTGATGCAGAA	TGAACCACAGGTGCATTGTT
CST30	TGCAACAGTCTCCTGAAACG	CCAGCAGCTTGATCAACAAA
CST31	GAGAGTTGGGCGGAGTTATG	ATGCCCCCATGTAATTGAAA
CST33	CTTGAAGGCCACCTATTGGA	CCACTCCCATGTCCTCTTGT
CST34	CCGGAAAAGGGAAGTAAAGC	ACCAAGCAGTTTCTCCCTCA
CST36	ACCTGTTCCTCCTGTGTTGG	AGGAAGGTGAGCTCCTGTCA
CST39	CTGGCTTGGACTGAAGAAGG	TCTAATGCCGAAGGCAGTTT
ACT1G2	GCCGTCCTCTCTCTGTATGC	ACAGTGTGGCTGACACCATC
ACT1G3	GACGGAGCGTGGTTACTCAT	GCAATGCCAGGGAACATAGT
SCD1	ACGTTCAGTGAGGAGCCTGT	CAGGGCCAAGATATCAGCTC
SCD2	ATGATTGTTGCTGCCATTTG	GCCACTGGCTGAGTCAAGAT

SCD4	AAGGAAACCTTCAGGGGCTA	GGCTGCAACATTTGGGTAGT
SCD5	AGGAAAGGTTGCCTGGATCT	TGGTCATCAGTTGGCATCAT
SCD6	GACGAAGCAAAGCCAACTTC	ACATAGAACCCTGCCAATGC
SCD7	AAATGGCTGAAAGGCTTCCT	TGCTGGGTGAGCTAGAACCT
SCD10	TTCTATCATTGCGCTGATGC	CTGCTTGGCTCTTTTCAAGC
SCD12	GCAAGTGAAGGAAGCACACA	TTCAACTGTGCCGTTTCAAG
SCD13	CCCTAGGTGCCAAGAAGAAA	CTGCAGCATCCAACATTAGC
SCD15	TTTCAAGCCAATCACTGCTG	CAGACCAGGGAATAAACCAAA
SCD17	CTGGTATGCCGAAGGATGAT	TGAATTCATGAGTGCCTTGC
SCD19	GGGCTTTCACACTCTCAAGC	AACCCAAACGACTCGATGAC
SCD20	CAGAGCATGAACTGCCACAT	GGACTGGTGTCTCTGCCATT
SCD26	GATGACGATGAGGAGGAGGA	ACACTAATCCAACGGCAAGG
SCD28	GCCGTGGTTTGAACATTTCT	GCAACATCTTCCAGCGGTAT
SCD29	GTGAACCAGCCCTACGTCAT	ATTCAAGAGCGCACTGAGGT
SCD30	TGTGTTCCAGTTGAGCGAAG	GCACCAGGCTCTGAATTAGC
SCD31	CAGGACAAAGCGACTGTCA	CACTGAAAGGTGCAGCGATA
SCD32	CATCACTGGATGAGCGAAAA	GCATGTCCGGTACAAAGTGA
SCD33	CGCTCGTCTCCTACAAGGTC	TCAACCCAGGATTTCAAAGG
SCD34	TGCAACTCAGAAGACCGTTG	TGGAGTAGCTCGCAAAAGGT
SCD36	GTGAACCGAACACTCGGAAT	GCAGTGCATACTTGCCTCAA
SCD37	TGACAAGAAATGCCCATTCA	ATGGGCGCAGAAACTTCTTA
SCD40	ATGTTTGCGCAGAAGAAGGT	TTAGGTTACCGTGCCAGGAC
SCD41	ACCTTCCACAGAAACCGTTG	ACCTGGCTGGAGAACAGAGA
SCD42	TGGTGACATTGTGCATGTTG	AAAGCACCAGTTACCCATGC
SCD43	CCCGATTTGAGTTGTTCGTT	TGCACAACATGGTTCCTGAT
SCD46	CAGTTCGGGATTTCTTTGGA	ATCGTCGGGGAGTAGTCCTT
SCD48	GAGTGGGACATCCTCTCCAA	CCGAATTTGCGGAAAAACTA
SCD51	GGCTCTGGTGTTGGTGAAAT	ACGTGAACCGTGAATCTTCC
SCD52	AAGGCCGATCACTACACCAC	GGGATTGACAAGGCAACACT
SCD53	GTCCTTGTTGGTGACCGTCT	TCAAATCCAGGGTCAAAAGC
SCD54	GACCTGAGCGTGTGAGCATA	AGCGATGGCAGATCTCAACT
SCD59	TCCAATCCAAGGAGGATGAG	CACAGCTGGAGGGTAAAGGA
SCD61	CGACCAATTCTGGTTCAGGT	TCCTGCACCTCCTTCTCAGT
SCD63	ATATGACGTGCAAGCAGCAG	GGTTCCTATTCGATGGAGCA
SCD65	AGTAATCAGCTGGCGCCTAA	CCCTCAGAAGCCTTCAACTG
SCD66	GCTTCCCAGCAAAGTTCAAG	CCTTTACGCAGTTGCTCTCC
SCD67	CCAAATGGTATGCCTTGCTT	GCTCAGTAACAGCCGAGACC
SCD68	CTAAGGGACGCAATGTGGTT	GTAATCCATTCCACCCATGC
SCD69	TGGTACGACATCCAGTCCAA	AAATTGCCTCGGACAACAAC
SCD73	GGGAGAGGACACATCTTGGA	TCACATCAACAGACCGCAAT
SCD74	AGCGTCTAATGCGATTTGCT	GGGCCATCACCATGAATAAC
CSD1	AGCACGTTCTACGGCGAGT	GAACGAGATCGTCCTCATCG
CSD3	GTATCCGGTATCGGTTGTGG	CGTGATGCTGTCGCTAGTGT
CSD4	TGCAAGCTTGTGGCTTACAC	TTTCAGGCTTTTTCGCTTGT
CSD5	TGAACTGGAGAGGGAGCAGT	AGGGTGTTGTCGGAGTTACG
CSD6	TCCTTTTGGAACCGTACGAC	GTGCAGTTGCAAGATGAGGA
CCDE3	GCCGTGGGAGTATCTCATGT	GCACCCATAGAGGTGAAAGC
CCDE4	GTGGACAAGCTGAATGCAAA	TTGGGATGGTTTGTCCTTGT
CCDE6	CCCGAGTATTGCTGGAACAT	TCAAACACAAGGGCATCGTA
CCDE9	CTGTGATGTCCATGGTGAGG	CCAGGAGTTCTTCAGCAACC
CCDE16	CTTTATGGCGATGAGGCAAT	GAACTCCTCGCCTTTGAGTG
CCDE17	GTCGGTGACGATGCCTTATT	TCCACTTCGGCGTTATTACC

CCDE19	ACGCAAGGAGGAAGAACAGA	TCCAGTTCGGCGTTATTACC
CCDE30	CTAGATGGCTTCGCTTCTGC	CAACAAAACGAGCGAACTGA
CCDE33	GCCCCTTCAGATGTGTTTGT	TGCCTTTGTGAATGTCTTGC
Tos17	GCTACCCGTTCTTGGACTAT	CTGAAATCGGAGCACTGACA

SUPPLEMENTAL METHODS

Isolation of Methylated DNA Using the McrBC Digestion Method

Total genomic DNA was extracted from light-grown shoots and suspension-cultured cells using the Plant DNeasy Mini Kit (QIAGEN). Methylated DNA was isolated using an McrBC digestion method adapted from previous report (Lippman et al., 2004). McrBC is a DNA methylation-specific restriction enzyme that cleaves between two [A/G] C sites separated by 40-3000 bp, only when the C residue in each [A/G] C motif is methylated (New England Biolabs). Due to the high frequency of [A/G] C sites in the genome, methylated regions will be cleaved into small fragments. 20 µg of genomic DNA were digested overnight with 200 units McrBC (New England Biolabs), then separated on a 1.0% agarose gel. DNA fragments smaller than 500 bp, which presumably contain mainly methylated DNA, were recovered using QIAquick Gel Extraction Kits (QIAGEN) (Supplemental Figure 1). About 2.0 µg recovered methylated DNA or sonicated total genomic DNA (< 500 bp) were labeled with Cy3 or Cy5 using the BioPrime Array CGH Genomic Labeling System (Invitrogen) and used for microarray hybridization. Two replicates of each tissue were analyzed.

Chromatin Immunoprecipitation (ChIP)

Chromatin from light-grown shoots and cultured cells was immunoprecipitated with antibodies against H3K4me2 (Upstate), H3K4me3 (Abcam) or CenH3 (Nagaki et al., 2004) as described previously (Gendrel et al., 2005). ChIP (enriched) and input (whole cell extract) DNA samples were amplified using a random amplification procedure as described (Lippman et al., 2005). 2.0 µg amplified ChIP or Input DNA was labeled with Cy3 or Cy5 using the BioPrime Array CGH Genomic Labeling System (Invitrogen) and used for microarray hybridization. Three replicates of each tissue were analyzed. Conventional ChIP-PCR was performed on DNA samples from input, mock (no antibody) and ChIP using the primer pairs listed in Supplemental Table 4.

Tiling-path microarray design

Chromosome 4 and 10 sequences used to design the tiling-path microarray were based on the rice *japonica* genome assembly release 2.0 by TIGR. Probes were selected using the NASA Oligonucleotide Probe Selection Algorithm (NOPSA) that takes probe sequence complexity, nucleotide composition, secondary structure and other factors into account (Stolc et al., 2005). To reduce the potential of cross hybridization from highly repetitive sequences, we measured the frequency in the genome of each of the 20 consecutive 17 nt sequences within each 36 nt probe, calculated the average frequency for each probe, and probes (except those in the two centromeres) with average frequencies greater than five were removed (Stolc et al., 2005). This procedure led to a resolution of 36 bp per probe plus a median gap of 82 bp between probes. The NimbleGen Maskless Array Synthesizer system was used to synthesize 380,766 oligonucleotides *in situ* on a single chip that covers 77.5% of rice chromosomes 4 (~35 Mb) and 10 (~23 Mb). The estimated coverage was calculated as follows:

$$\frac{(36bp+82bp)\times 380766}{35244269bp+22876596bp}=77.5\%$$

This tiling array platform can be found in the NCBI Gene Expression Omnibus (GEO) database with the accession number GSE9925.

Microarray Hybridization

Microarray slides were incubated in prehybridization buffer (5X SSC, 0.1% SDS, 1% BSA) for 30 min – 1 hour at 42^{o} C and washed with water. Slides were then hybridized with Cy3- or Cy5-labeled DNA targets in hybridization buffer (5X SSC, 0.1% SDS, 0.5 mg/ml salmon sperm DNA, 0.5 mg/ml BSA) for 16-20 hours at 50^{o} C. After hybridization, slides were washed sequentially with 2X SSC / 0.1% SDS, 0.2X SSC /0.1% SDS and 0.2X SSC for 10 min each at room temperature. The dried slides were scanned with a GenePix 4200A scanner (Axon), and independent TIFF images for both Cy3 and Cy5 channels were acquired at 5μm resolution**.**

Microarray data processing and analysis

Normalization of microarray raw data

Microarray raw data was first processed by the LOWESS normalization module in the LIMMA R package (Smyth et al., 2005) to balance the hybridization intensities for biases between Cy3 and Cy5 fluorescence. In the second step, the Quantile normalization module in the NMPP package (Wang et al., 2006) was used to remove the experimental variances across all the replicate arrays and samples.

Identification of modified genomic regions

Identification of genomic regions enriched for methylated DNA or modified histones was based on the One-sided Wilcoxon Signed Rank Test (Hollander and Wolfe, 1999). We performed the test on paired data points (ChIP/Input) within a sliding window of ± 500 bp across each interrogated oligonucleotide probe, on average containing 8 to 12 probes, to determine whether hybridization intensities of the methylated or ChIP-enriched DNA samples were significantly greater than those of the total genomic DNA input samples. The algorithm of the Wilcoxon Signed Rank Test is described in Affymetrix's Statistical Algorithms Description Document (2002 Affymertrix, Inc.). A "methylated DNA" or "modified histone" region was defined by combining adjacent probes with a significance threshold of $p < 0.05$, allowing a maximal gap of 150 bases, and requiring a minimal run of two consecutive probes

Mapping and functional analysis of DNA methylation, H3K4me3 and H3K4me2

Re-mapping probes to rice genome release 4.0

Since the tiling array was designed using rice genome release 2.0 and the genome was updated to release 4.0 in January, 2006 (http://www.tigr.org/tdb/e2k1/osa1/), we remapped all probes to the new version using BLAT (Kent, 2002). 388,082 of the 388,546 probes for chromosomes 4 and 10 and centromeres 4 and 8, were mapped on version 4.0 with 100% identity, and 96.5% match a unique genomic position.

Mapping modified regions to annotated genes

All sequences of genes, transposable elements, and other genomic elements studied in our analyses were obtained from TIGR rice genome release 4.0. This version annotated 6,347 protein-coding genes and 2,432 TEs, respectively, on chromosomes 4 and 10. Some of

the most repetitive transposons and a few protein-coding genes were not covered on the tiling array because we were unable to design sufficiently selective probes. After removing the uncovered gene models (a covered gene model requires a minimum density of 2 probes per Kb), 6,329 (99.7%) protein-coding genes and 2,248 (92.4%) TEs were covered in our analysis.

In rice genome release 4.0 about 60% of protein-coding genes are supported by full-length cDNA or EST evidence, and the rest are computer-predicted. Therefore, the annotated gene start may refer to either the putative transcription initiation site or the start of gene annotation. For genes supported by full-length cDNAs annotated gene starts usually refer to transcription start sites. However, for predicted genes, annotated gene starts may refer to the translation start site (ATG). Therefore, for all analyses regarding the distribution of specific modifications within genes, we calibrated gene alignments using a reference set of full-length cDNAs. We defined a gene to be carrying a particular modification if we detected an overlap of at least 150 bp beween a gene and a particular modified region. We also defined the region 1.0 kb 5' to the gene start site as the promoter region.

Analysis of gene epigenetic modifications and gene expression

We used two rice gene expression datasets to characterize the relationships between epigenetic modifications and transcriptional activity. To analyze the effects of epigenetic modifications on gene expression in light-grown shoots we used the microarray data (Affymetrix rice microarray) provided by the NSF Rice Oligonucleotide Array Project public resource (http://www.ricearray.org). We also repeated the analysis using our own shoot expression data (Jiao and Deng, 2007) and obtained the same results (data not shown). To analyze differences between light-grown shoots and suspension-cultured cells in epigenetic modification and gene expression we used a recently-published rice expression microarray dataset from our lab (Jiao and Deng, 2007).

Multivariate linear regression analysis

In order to compare the extent to which genes were modified, we defined a modification density per kb for each gene, calculated by the formula:

$$Density = \frac{\text{number of signal probes}}{\text{gene length}} \times 1 \text{ kb}$$

Multivariate linear regression analysis was performed on modification density and $\log_{10}$ transformed expression intensity with the built-in "LINEST" function in Microsoft Excel, then the "TREND" function was used to return the predicted Y values linearly distributed along the best-fitting line for each modification as a separate factor. Expression data for light-grown shoots were obtained from the NSF Rice Oligonucleotide Array Project public resource (http://www.ricearray.org).

Mapping H3K4me3, H3K4me2 and DNA methylation regions in relation to other genomic elements

To characterize the DNA methylation, H3K4me3 and H3K4me2 regions located between gene territories, we mapped these regions against other genomic elements found on rice chromosomes 4 and 10 including the sequences encoding tRNA, microRNA, snoRNA, siRNA, intergenic EST, intergenic Transcriptionally Active Regions (TARs) and two kinds of insertion sequences (T-DNA & Tos17). EST sequences were downloaded from the NCBI rice Unigene collection and mapped to rice genome release 4.0. ESTs located between the territories of annotated genes or transposable elements were designated “intergenic”. Intergenic TARs were identified in a previous high-resolution rice whole-genome tiling microarray project as actively transcribed elements that did not correspond to annotated genes (Li et al., 2007). 139 rice tRNA genes were obtained from the TIGR rice genome 4.0 annotation, 68 rice snoRNA sequences were collected from the NONCODE ncRNA database (Liu et al., 2005), 44 rice microRNA sequences were collected from miRBASE (http://microrna.sanger.ac.uk/) and 75 rice endogenous siRNA were collected from the supplemental data of Sunkar *et al*, 2005a and Sunkar *et al*, 2005b. The positions of 2442 Tos17 retrotransposons were determined from TIGR rice genome build 4.0 and 3490 T-DNA insertion sites were obtained from the supplemental data of Jeong *et al*, 2006.

Since the non-coding elements are too short to be detected by our analysis but may be processed from much longer transcripts we extended our analyses of epigenetic modifications 500 bp to either side of the chromosomal coordinates of tRNA, snoRNA, microRNA and siRNA sequences. Conversely, we used the unaltered chromosomal coordinates of intergenic TARs, ESTs and Tos17 and T-DNA insertion sites to map their epigenetic modifications. As with genes, we defined an element to be carrying a particular modification if we detected an overlap of at least 150 bp between the element and a particular modified region.

SUPPLEMENTAL REFERENCES

Gendrel, A.-V., Lippman, Z., Martienssen, R., and Colot, V. (2005). Profiling histone modification patterns in plants using genomic tiling microarrays. Nat Meth **2,** 213-218.

Hollander, M., and Wolfe, D.A. (1999). Nonparametric Statistical Methods. (New York: John Wiley and Sons, Inc.).

Jiao, Y., and Deng, X.W. (2007). A genome-wide transcriptional activity survey of rice transposable element-related genes. Genome Biol. **8,** R28.

Kent, W.J. (2002). BLAT---The BLAST-Like Alignment Tool. Genome Res. **12,** 656-664.

Li, L., Wang, X., Sasidharan, R., Stolc, V., Deng, W., He, H., Korbel, J., Chen, X., Tongprasit, W., Ronald, P., Chen, R., Gerstein, M., and Deng, X.W. (2007). Global Identification and Characterization of Transcriptionally Active Regions in the Rice Genome. PLoS ONE **2,** e294.

Lippman, Z., Gendrel, A.-V., Colot, V., and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. Nat Meth **2,** 219-224.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., Richard McCombie, W., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430,** 471-476.

Liu, C., Bai, B., Skogerbo, G., Cai, L., Deng, W., Zhang, Y., Bu, D., Zhao, Y., and Chen, R. (2005). NONCODE: an integrated knowledge database of non-coding RNAs. Nucl. Acids Res. **33,** D112-115.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. (2003). Target Site Specificity of the Tos17 Retrotransposon Shows a Preference for Insertion within Genes and against Insertion in Retrotransposon-Rich Regions of the Genome. Plant Cell **15,** 1771-1780.

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. Nat. Genet. **36,** 138-145.

Smyth, G.K., Michaud, J., and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics **21,** 2067-2075.

Stolc, V., Li, L., Wang, X., Li, X., Su, N., Tongprasit, W., Han, B., Xue, Y., Li, J., Snyder, M., Gerstein, M., Wang, J., and Deng, X.W. (2005). A Pilot Study of Transcription Unit Analysis in Rice Using Oligonucleotide Tiling-path Microarray. Plant Molecular Biology **59,** 137-149.

Wang, X., He, H., Li, L., Chen, R., Deng, X.W., and Li, S. (2006). NMPP: a user-customized NimbleGen microarray data processing pipeline. Bioinformatics **22,** 2955-2957.