

## DNA Methylation Profiling Identifies CG Methylation Clusters in *Arabidopsis* Genes

Robert K. Tran, Jorja G. Henikoff,  
Daniel Zilberman, Renata F. Ditt,  
Steven E. Jacobsen, and Steven Henikoff

### Supplemental References

- S1. Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* 13, 2212–2217.
- S2. Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220.
- S3. Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719.
- S4. Tompa, R., McCallum, C.M., Delrow, J., Henikoff, J.G., van Steensel, B., and Henikoff, S. (2002). Genome-wide profiling of

- DNA methylation reveals transposon targets of CHROMO-METHYLASE3. *Curr. Biol.* 12, 65–68.
- S5. van Steensel, B., Delrow, J., and Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* 27, 304–308.

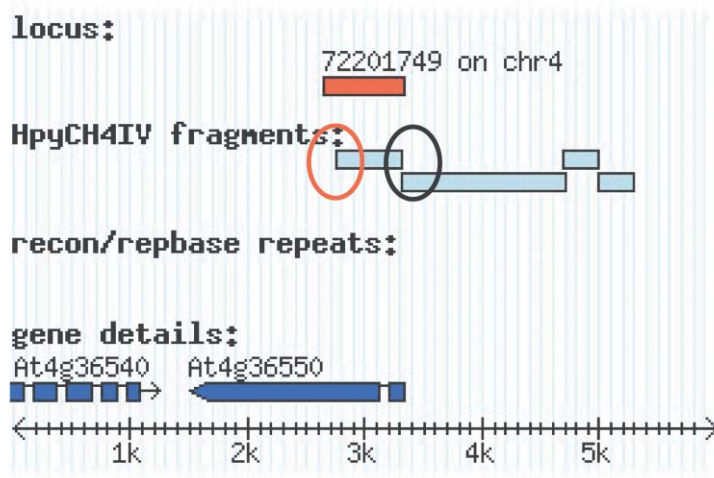
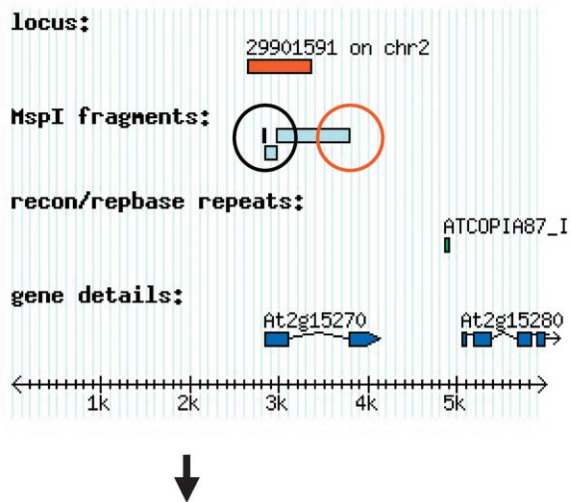


Figure S1. Map and CG Methylation Occupancy of a Representative Gene-Rich Fragment Showing CG Hypomethylation for the HpyCH4 IV Dataset

(A) Schematic display of locus 4:17246549–17247226.

(B) Table shows bisulfite sequencing results for individual CG sites in the region circled in red in (A). No methylation was detected at all for the region circled in black in (A). The arrow marks the HpyCH4 IV restriction site. Asterisks mark sites showing changes in methylation occupancy. Primers used for amplification were as follows: 5'-CCRTAATATATA ATCCARTCATCATCTCCTCARACAACACCT TTA-3' and 5'-GAGAATTTTYGAAATTAGYG ATGGAAATAGAGAAGTYGATAT-3'; and 5'-CCTCARCTTCTRCARCATTAAACATACARAA AATCATCRATAA-3' and 5'-GTTGATTTTAG AATTTTYGGAAGAAGAAGTYGGGAAAAGT-3'.

Lines	Sites	1*	11*	34	41	43*
<i>Ler</i>		60	85	70	95	0
<i>clk-st</i>		5	0	20	80	40
<i>cmt3</i>		0	0	95	80	100
<i>kyp</i>		0	5	10	75	75
<i>ago4</i>		0	0	20	90	80
<i>drm1/2</i>		95	90	90	75	80
<i>drm1/2/cmt3</i>		0	0	0	85	60



↓

Lines	Sites	1*	79*	81	96	129	196	204	213	286
<i>Ler</i>		0	5	80	65	0	90	85	85	0
<i>clk-st</i>		75	85	65	95	0	75	80	75	0
<i>cmt3</i>		0	95	80	90	0	95	95	90	5
<i>kyp</i>		80	90	80	90	0	80	90	40	0
<i>ago4</i>		0	80	80	80	0	90	100	95	0
<i>drm1/2</i>		0	95	80	95	95	95	85	95	0
<i>drm1/2/cmt3</i>		85	85	65	95	50	100	85	90	0

Figure S2. Map and CG Methylation Occupancy of a Representative Gene-Rich Fragment Showing CG Hypomethylation for the Hpa II Dataset

(A) Schematic display of locus 2:6644864–6645576.

(B) Table shows bisulfite sequencing results for individual CCGG sites in the region circled in red in (A). No methylation was detected at all for the region circled in black in (A). The arrow marks the HpaII restriction site. Asterisks mark sites showing changes in methylation occupancy. Primers used for amplification were as follows: 5'-CCRTAATATATAATCCARTCATCATCTCCTCARACAACACCTTTA-3' and 5'-GAGAATTTTYGAAATTAGYGATGGAAATAGAGAAGTYGATAT-3'; and 5'-CCTCARCTTCTRCARCATTACATACARAAAATCATCRATAA-3' and 5'-GTTGATTTTGAATTTTYGGAAGAAGAAGTYGGGAAAGT-3'.

Table S1. Loci Scored as CG Methylation Targets in Mutants for the HpyCh4 IV Dataset

Genomic Location	Gene ID	TIGR Designation
3:14733875-14734551	At3g42630	pentatricopeptide (PPR) repeat-containing
4:17246549-17247226	At4g36550	U-box domain-containing protein low similarity
4:8205834-8206535	At4g14240	expressed protein
3:3231584-3232308	At3g10390	amine oxidase family protein/SWIRM
1:3223159-3223839	At1g09910	expressed protein
1:12534603-12535283	At1g34350	expressed protein
2:6233102-6233813		intergenic
1:28714082-28714798	At1g76510	ARID/BRIGHT DNA binding domain
2:6045801-6046476	At2g14255	zinc-finger (DHHC-type) family
2:7312243-7312952	At2g16860	GCIP-interacting family protein/intergenic
2:16966613-16967288	At2g40630	expressed protein
1:27768548-27769233	At1g73840	hydroxyproline-rich glycoprotein family
4:13688128-13688849	At4g27340	Met-10+ -like family protein; contains Pfam
4:16600925-16601611	At4g34820	expressed protein
4:12942908-12943588	At4g25290	deoxyribodipyrimidine photolyase family
3:2097754-2098459	At3g66658	betaine-aldehyde dehydrogenase, putative
5:9500433-9501138	At527000	kinesin motor protein-related
2:15461676-15462373	At2g36850	glycosyl transferase family 48 protein
5:10782036-10782722	At5g28740	transcription-coupled DNA repair
2:18550734-18551430	At2g44950	zinc finger (C3HC4-type RING finger) family
3:5606751-5607433		intergenic
5:3178219-3178942	At5g10140	MADS box protein flowering locus F (FLF)
5:6134705-6135399	At5g18490	expressed protein
3:1356227-1356941	At3g04910	protein kinase family protein
1:21600119-21600842	At1g58250	similar to SABRE protein
2:8251706-8252384	At2g19010	GDSL-motif lipase/hydrolase family protein
3:3241769-3242455	At3g10420	sporulation protein-related
4:18226382-18227080	At4g39120	inositol monophosphatase family protein

*Arabidopsis thaliana* mutants and wild-type strains used for sample preparation were previously described [S1–S3]. Whole 5-week-old plants were used for preparing genomic DNA for microarray analysis as described [S4]. Samples were dissolved in 10 mM Tris 1 mM EDTA (pH 8), and 50–60  $\mu$ g aliquots were subjected to digestion by the addition of 200 units of restriction endonuclease for 3–4 hr. Endonucleases Hpa II and HpyCH4 IV were obtained from New England Biolabs (Beverly, MA). Digested DNAs were size-fractionated on sucrose gradients and labeled by random priming, and oppositely labeled mutant and wild-type samples were hybridized to microarrays as described [S5]. Data processing and analysis were performed on 3–4 biological replicates with dye swaps as described [S4]. “Self versus self” control experiments were used to define a lower-bound threshold of 3 standard deviations from the mean of the distributions of  $\log_2$ -transformed ratios. Gene IDs and TIGR designations were obtained from TAIR (<http://www.arabidopsis.org>), Repbase libraries from <http://www.girinst.org>, and RECON libraries from Zhirong Bao.

Table S2. Loci Scored as CG and CNG Methylation Targets in Mutants in HpalI

Genomic Location	Gene ID	TIGR Designation	<i>cmt3</i>	<i>kyp</i>	<i>ago4</i>	<i>drm1</i> <i>drm2</i>	<i>drm1</i> <i>drm2</i> <i>cmt3</i>
3:22953865-22954546	At3g61960	protein kinase family protein	-	-	-		-
1:23802117-23802796	At1g64110	AAA-type ATPase family protein contains Pfam	-	-	-		-
4:8205834-8206535	At4g14240	expressed protein	-	-			
3:768104-768819	At3g03300	DEAD/DEAH box helicase carpel factory-related	-	-			
2:11845828-11846546		intergenic	-				-
3:18785875-18786564	At3g50590	transducin family protein/WD-40 repeat	-				-
4:11925388-11926109		intergenic	-				-
4:1491599-1492315	At4g03390	leucine-rich repeat transmembrane protein	-				
4:6096768-6097492		intergenic	-				
2:154330-155049	At2g01320	ABC transporter family protein	-				
2:7998672-7999372	At2g18410	expressed protein	-				
4:17274397-17275104	At4g36630	expressed protein	-				
4:5366995-5367697	At4g08450	disease resistance protein (TIR-NBS-LRR	-				
3:22978783-22979500	At3g62010	expressed protein		-			
4:9167760-9168478		intergenic		-			
1:27685632-27686347	At1g73620	thaumatin-like protein, putative			-		
4:5847849-5848556	At4g09180	basic helix-loop-helix (bHLH) family protein					-
4:908834-909527	At4g02070	DNA mismatch repair protein MSH6-1 (MSH6-1)					-
3:3576446-3577140		intergenic, recon repeat At000118					-
4:1388174-1388849	At4g03130	BRCT domain-containing protein contains Pfam		-		+	
1:28812703-28813401	At1g76740	expressed protein				-	+
1:21205676-21206378	At1g56590	clathrin adaptor complexes medium subunit	+	+		-	
3:1356227-1356941	At3g04910	protein kinase family	+	+			+
2:6644864-6645576	At2g15270	expressed protein	+	+			
1:5592558-5593239	At1g16350	inosine-5'-monophosphate dehydrogenase	+				+
4:5849840-5850517		intergenic, recon repeat At000416	+				
2:12856520-12857228	At2g30100	ubiquitin family protein low similarity to	+				
4:374944-375664		intergenic	+				
2:5695309-5696016	At2g13650	GDP-mannose transporter (GONST1) identical to		+			
3:1788503-1789226	At3g05970	long-chain-fatty-acid-CoA ligase/long-chain		+			
4:11599656-11600378		intergenic		+			
4:11992655-11993335	At4g22840	bile acid:sodium symporter family protein					+
1:8671027-8671725	At1g24460	myosin-related contains TIGRFAM TIGR01612:					+
2:17095412-17096127	At2g40950	bZIP transcription factor family protein					+
4:6263240-6263933	At4g10000	protein kinase family protein contains Pfam					+
4:6318858-6319535	At4g10120	sucrose-phosphate synthase, putative similar					+
4:6431266-6431955	At4g10380	major intrinsic family protein / MIP family				+	
1:19842566-19843266	At1g53190	zinc finger (C3HC4-type RING finger) family				+	
2:18187361-18188066	At2g43900	endonuclease/exonuclease/phosphatase family				+	
3:14733875-14734551	At3g42630	pentatricopeptide (PPR) repeat-containing				+	
4:8225569-8226262	At4g14290	expressed protein				+	

See legend to Table 1.