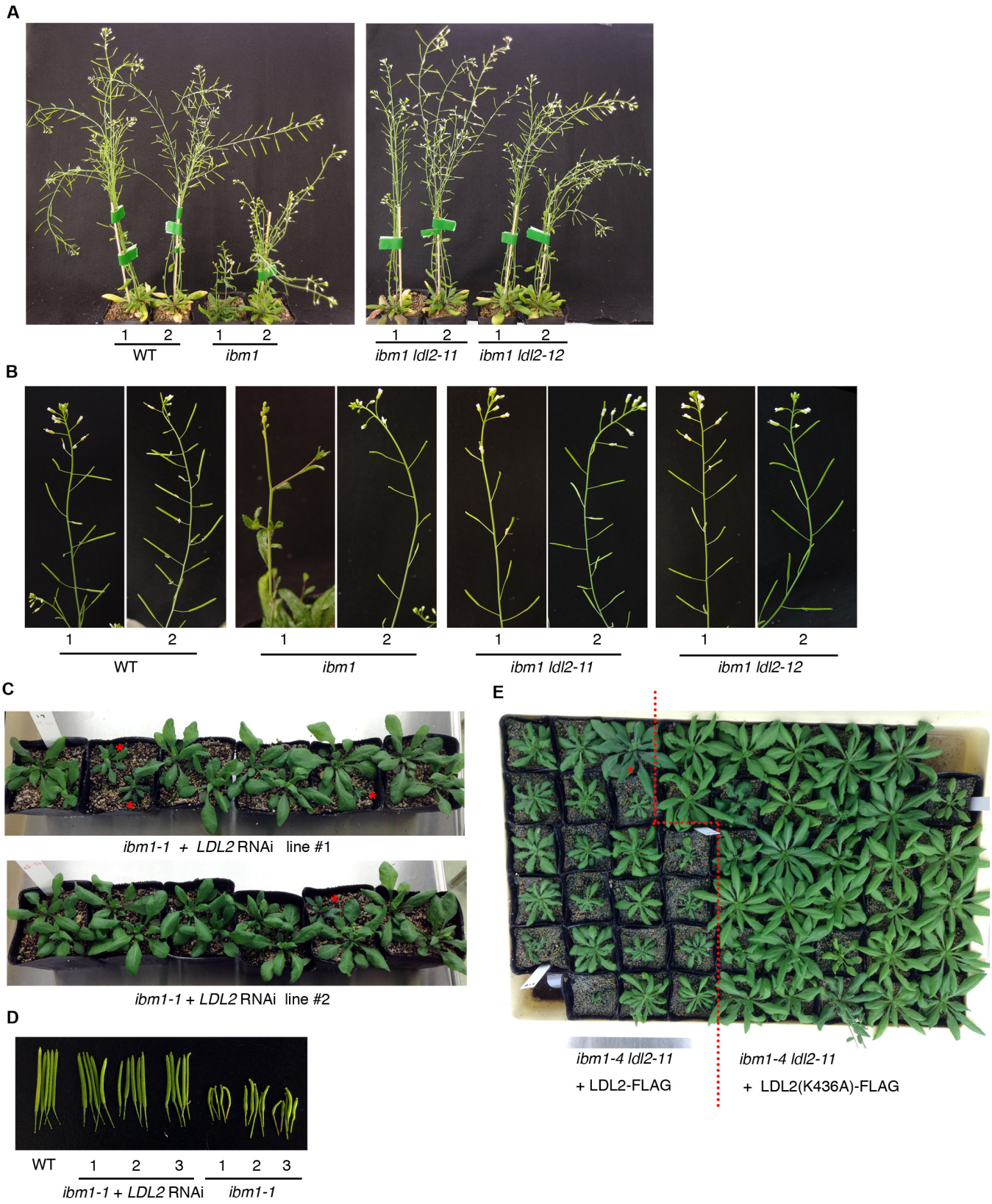
The gene-body chromatin modification dynamics mediates epigenome differentiation in Arabidopsis

Soichi Inagaki, Mayumi Takahashi, Aoi Hosaka, Tasuku Ito, Atsushi Toyoda, Asao Fujiyama, Yoshiaki Tarutani, Tetsuji Kakutani

Appendix Table of Content



Appendix Figure S1. Effects of LDL2 function on the *ibm1*-induced developmental defects.

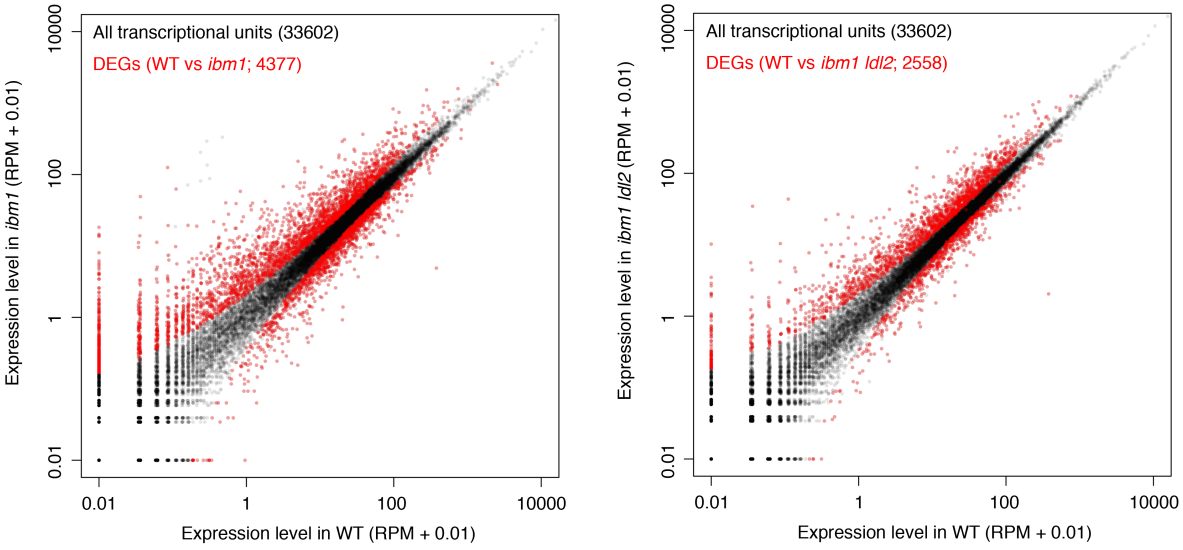
(**A**) Developmental phenotypes of *ibm1* were suppressed by *ibm1 ldl2* mutations. Eight-week-old plants are shown. (**B**) Close-up view of inflorescences of plants in **A**. (**C**) T2 lines segregating the *LDL2* RNAi transgene in the *ibm1-1* background. Plants labeled with red asterisks do not have transgene and other plants have the *LDL2* RNAi transgene. Top and bottom pictures represent two independent transgenic lines. All plants are four-week-old. (**D**) Siliques from eight-week-old plants with or without the *LDL2* RNAi transgene. Three independent transgenic lines and their non-transgenic siblings are shown. Siliques from a control WT plant are also shown in left. (**E**) Complementation of *ibm1 ldl2* by introducing pLDL2::LDL2-FLAG transgene. Five-week-old T1 plants with pLDL2::LDL2-FLAG and pLDL2::LDL2(K436A)-FLAG in *ibm1 ldl2* background are shown. Plants with pLDL2::LDL2-FLAG showed *ibm1*-like phenotypes except for one plant (red asterisk). Transgene with K436A amino acid substitution (pLDL2::LDL2(K436A)-FLAG) did not induce *ibm1*-like developmental phenotypes, and all plants remained *ibm1 ldl2*-like phenotypes. This amino acid (corresponds to K661 in human LSD1) substitution (Fig. 1D) is known to abolish enzymatic activity of the human LSD1 protein (Stavropoulos et al. 2006).

A

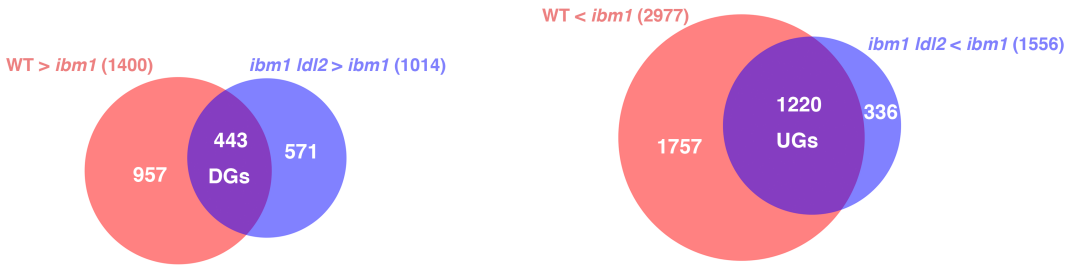
Numbers of DEGs (FDR < 0.05; b > a)

a \ b	WT	*ibm1*	*ibm1 ldl2*
WT		2977	1816
ibm1	1400		1556
ibm1 ldl2	742	1014	

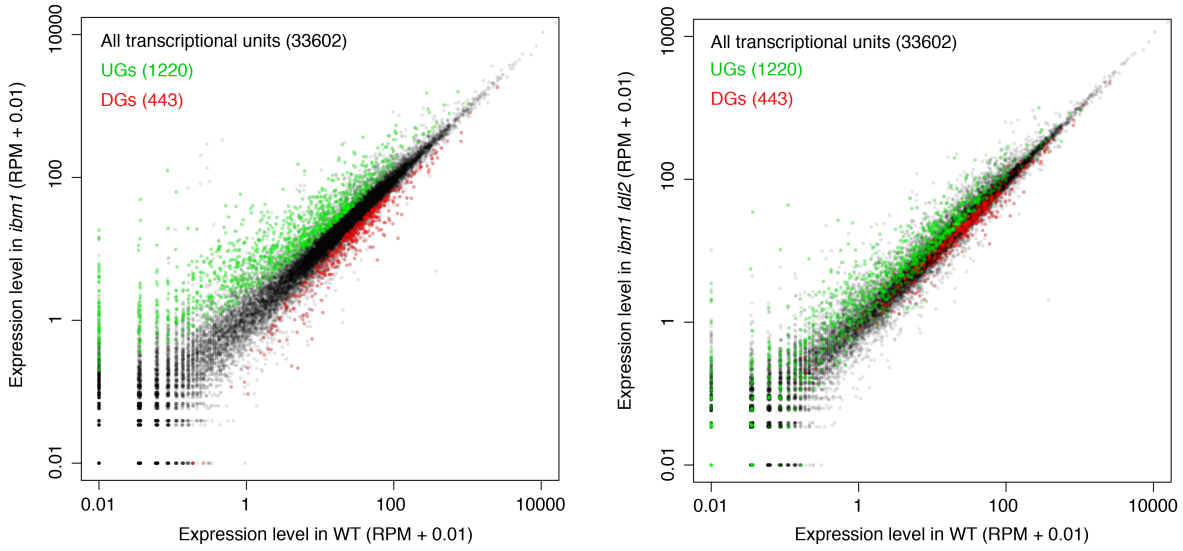
B



C

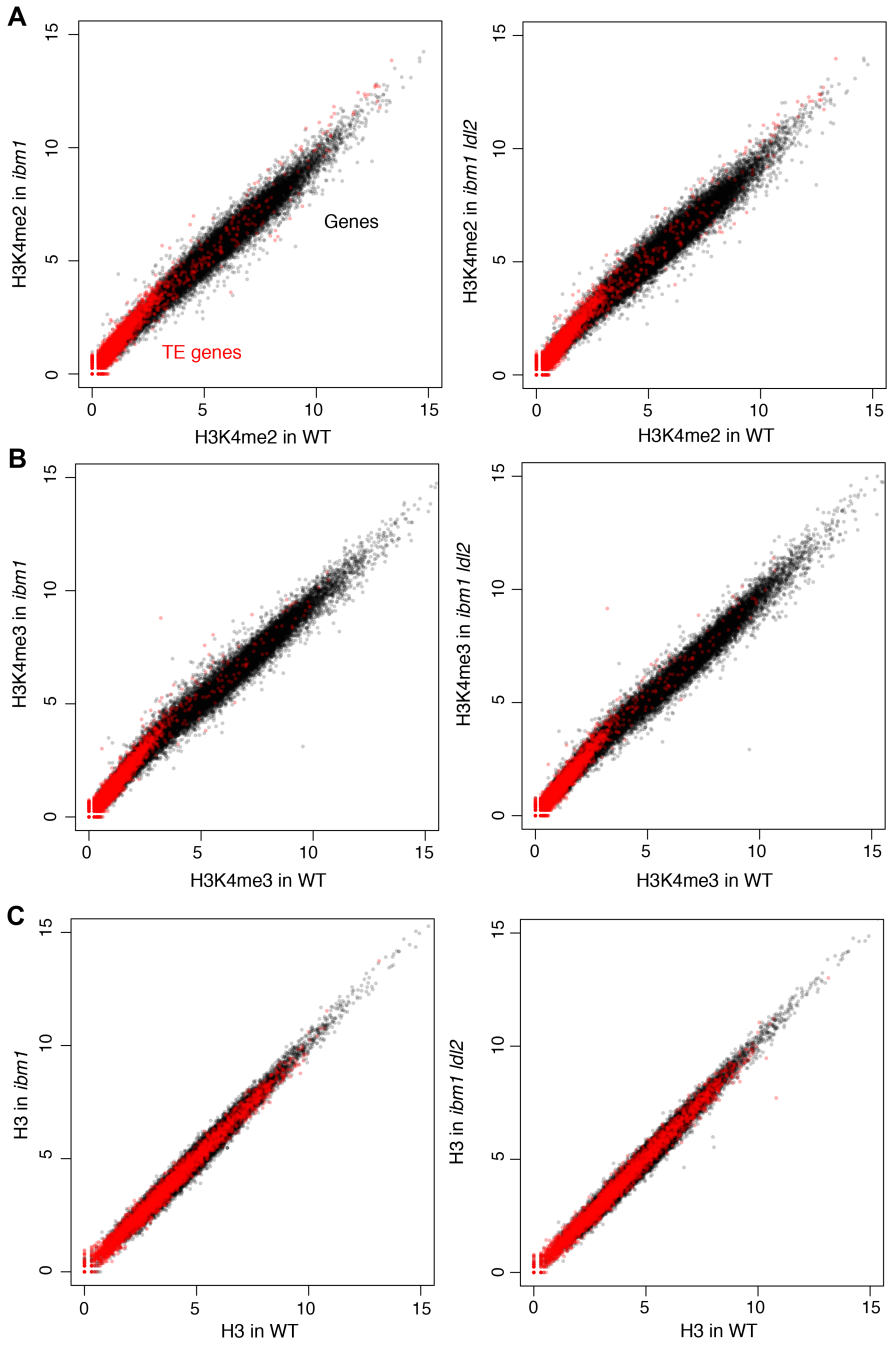


D



Appendix Figure S2. mRNA-seq analysis for *ibm1* and *ibm1 ldl2*.

(**A**) The numbers of differentially expressed genes (DEGs) in each comparison between samples. The number indicates the number of up-regulated genes in sample b compared to sample a. (**B**) Scatter plots showing the comparisons of expression level between WT, and the *ibm1* (left panel) or *ibm1 ldl2* (right panel) mutant. Red dots representing DEGs in each comparison are superimposed over black dots representing all transcriptional units. Value represents the number of reads per million mapped reads (RPM) + 0.01 to show all genes including genes with 0 read in either or both sample(s). (**C**) Venn diagrams to identify genes regulated by IBM1-LDL2 pathway (See text for details). (**D**) Scatter plots showing the expression levels of the UGs and the DGs identified in **C**.

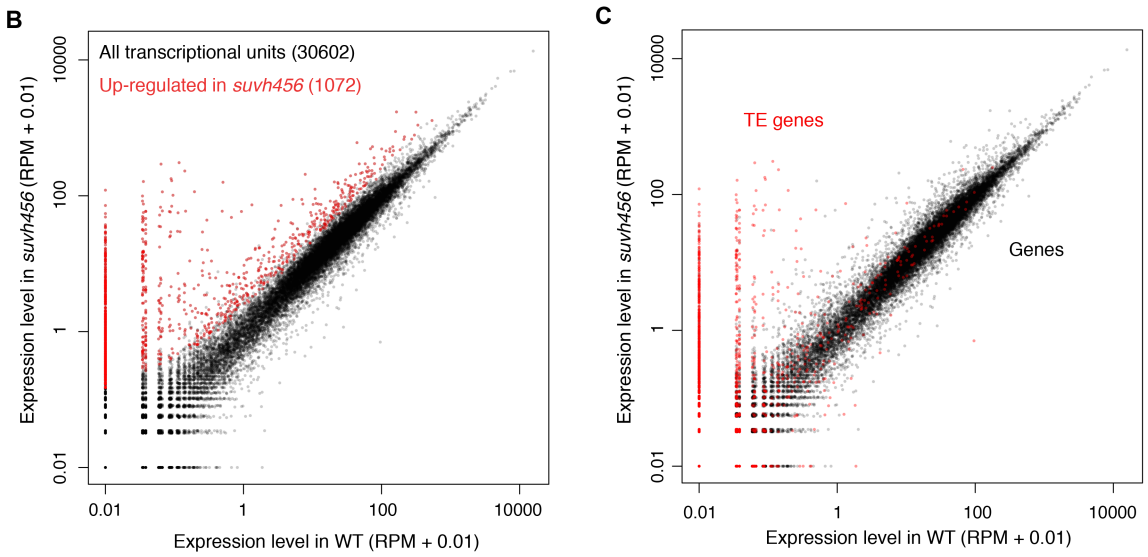


Appendix Figure S3. ChIP-seq for H3K4me2, H3K4me3 and H3 in *ibm1* and *ibm1 ldl2*.

(**A**-**C**) Scatter plots showing the comparison of the level of H3K4me2 (**A**), H3K4me3 (**B**) and H3 (**C**) in each gene between WT, and *ibm1* (left panel) or *ibm1 ldl2* (right panel). Black dots indicate protein-coding genes and red dots indicate TE genes. Value represents the square root of RPM and the results from one of two biological replicates are shown.

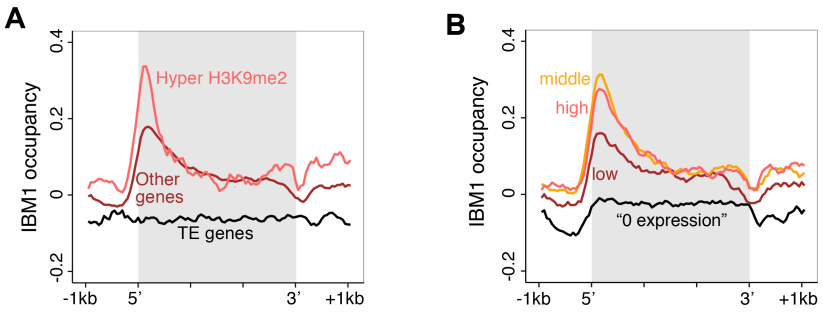
A

	Protein-coding genes	Transposable element genes
Up-regulated in *suvh456*	523	* 549
Not up-regulated in *suvh456*	26871	3321



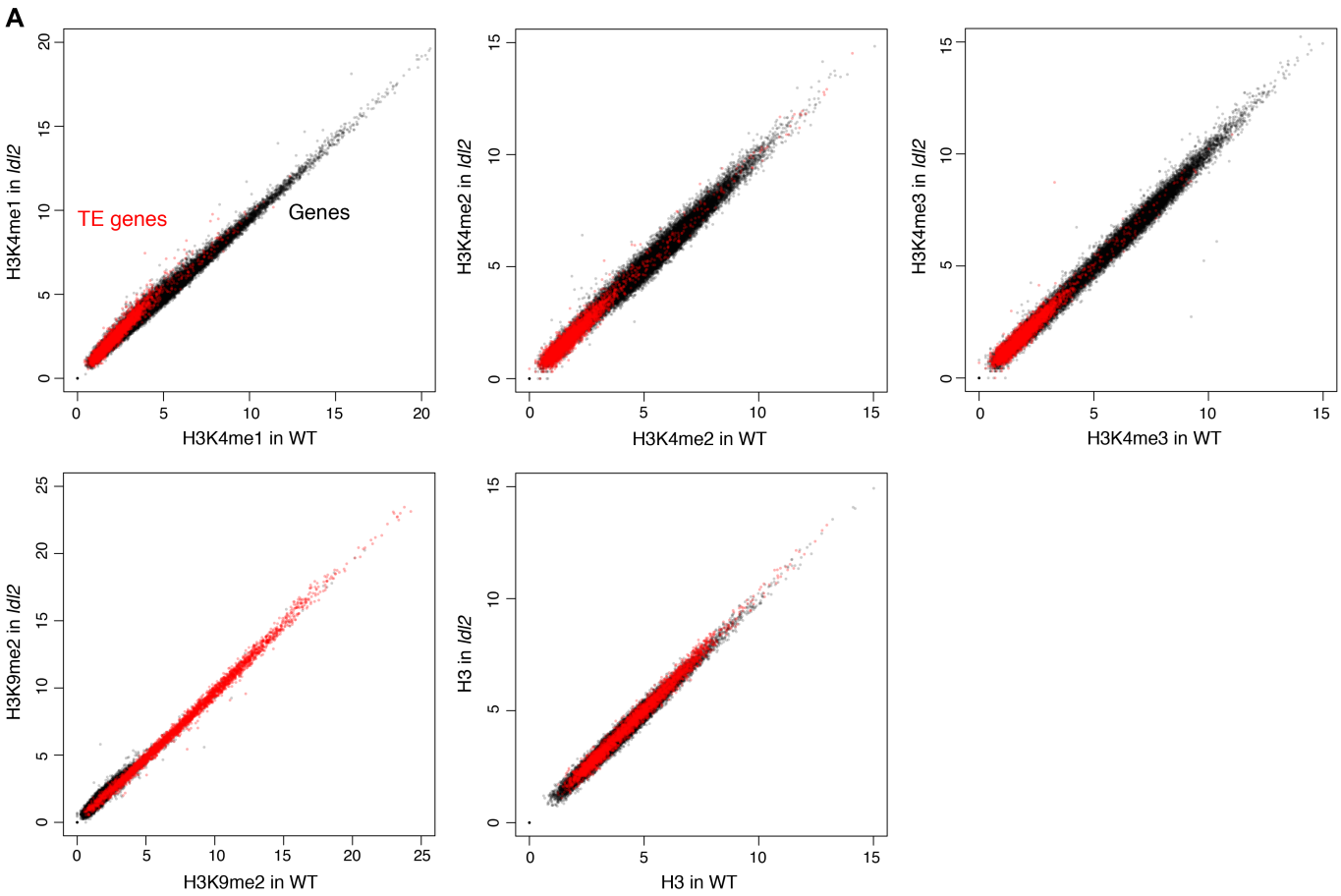
Appendix Figure S4. mRNA-seq analysis for *suvh456*.

(**A**) The numbers of genes up-regulated in *suvh456* and the other (not up-regulated) genes, with protein-coding genes and TE genes shown separately. Asterisk denotes a significant enrichment of TE genes in the group of up-regulated genes in *suvh456* (P = 1.0 x 10^{-219}, hypergeometric test). (**B**) Scatter plot showing the comparisons of expression level between WT and *suvh456*. Genes up-regulated in *suvh456* are shown by red dots superimposed over black dots, which represent all genes. Value represents RPM + 0.01 to show all genes including genes with 0 read in either or both sample(s). (**C**) The same scatter plot as **B**, indicating TE genes by red dots and protein-coding genes by black dots.



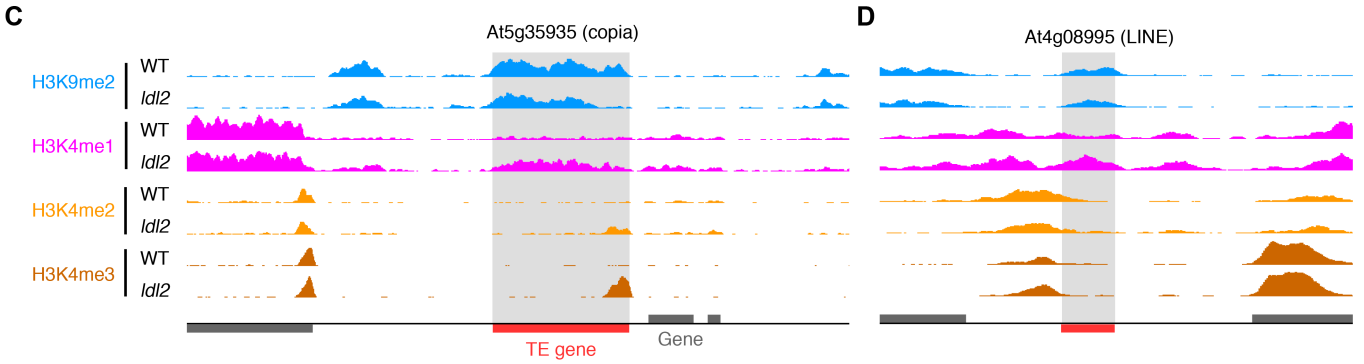
Appendix Figure S5. ChIP-seq for IBM1 localization.

(**A**) IBM1 occupancy profile around genes with hyper H3K9me2 identified in Fig. 2B, other protein-coding genes and TE genes determined with ChIP-seq. The value in y-axis represents normalized read count (RPM) for HA-IBM1 subtracted by that for the negative control WT. (**B**) IBM1 occupancy profile around groups of genes divided by their expression levels. "0 expression" represents protein-coding genes with 0 RPKM in an mRNA-seq data, and other protein-coding genes are equally divided into 3 groups according to their RPKM values.



B

	Protein-coding genes	Transposable element genes
H3K4me1 up in *idl2*	721	* 197
No change	26695	3706

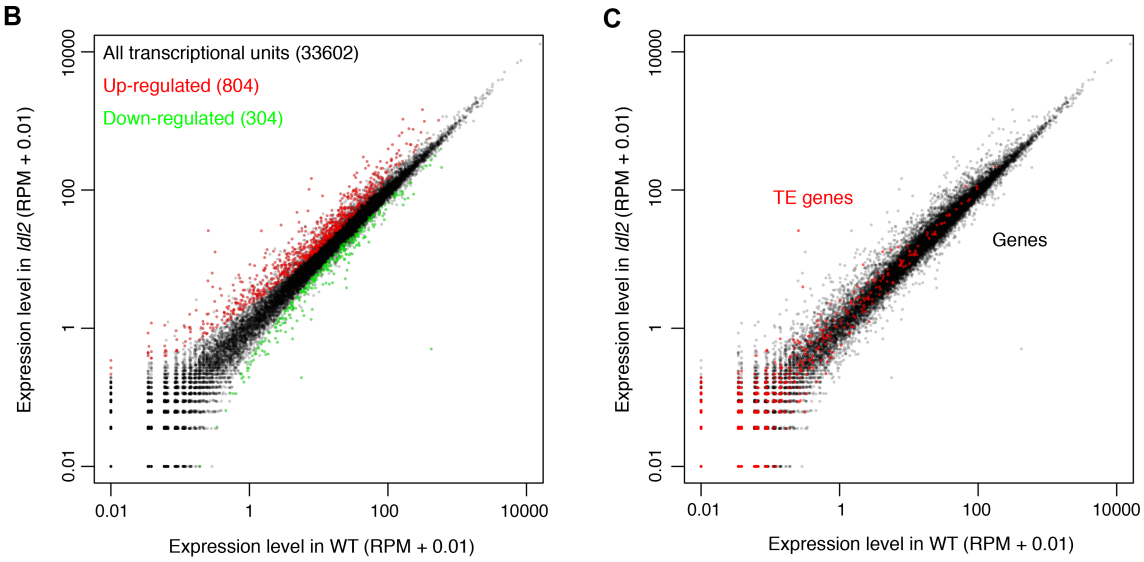


Appendix Figure S6. Patterns of H3 modifications in the *Idl2* single mutant.

(**A**) Scatter plots showing the comparison of the level of H3K4me1, H3K4me2, H3K4me3, H3K9me2 and H3 in each gene between WT and *Idl2*. Black dots indicate protein-coding genes and red dots indicate TE genes. Value represents the square root of RPM and the results from one of two biological replicates are shown. (**B**) The numbers of genes with increased level of H3K4me1 in *Idl2* compared to WT, and the other (not increased) genes, with protein-coding genes and TE genes shown separately. Asterisk denotes a significant enrichment of TE genes in the group of H3K4me1-increased genes in *Idl2* (P = 4.3 x 10^{-15}, hypergeometric test). (**C-D**) Patterns of H3K9me2 (range; 0-8), H3K4me1 (0-5), H3K4me2 (0-10) and H3K4me3 (0-15) around representative TE genes (shaded) with increased levels of H3K4me1 in *Idl2* compared to WT. While the mRNA level of At5g35935 (**C**) was significantly up-regulated in *Idl2* (Fig. S7), that of At4g08995 (**D**) was not. Normalized coverage (per million mapped reads) is shown. Genes (dark gray) and TE genes (red) in forward and reverse orientations are shown above and below the line, respectively.

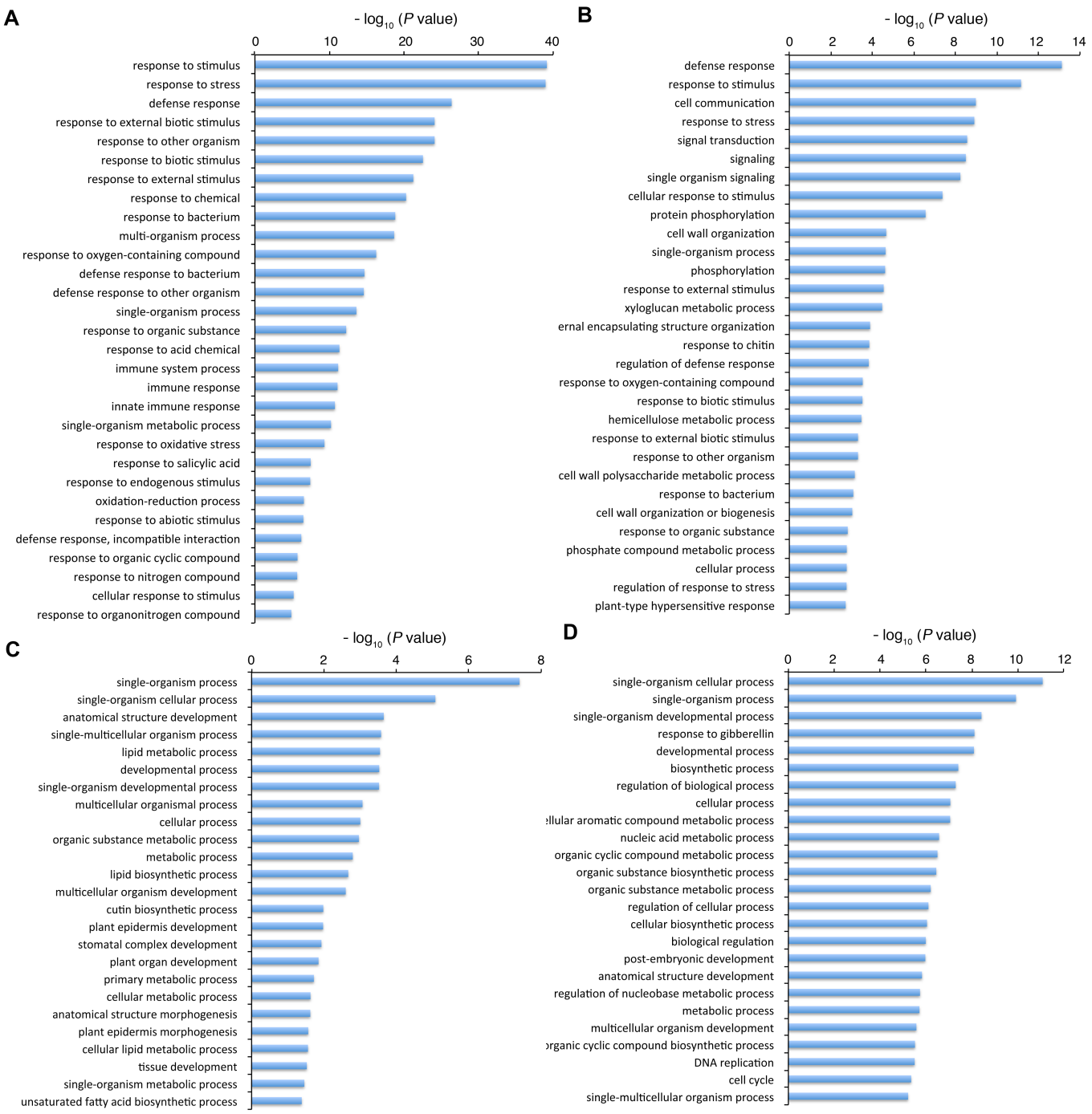
A

	Protein-coding genes	Transposable element genes
Up-regulated in *ldl2*	785	* 19
No change	26327	3884
Down-regulated in *ldl2*	304	0



Appendix Figure S7. mRNA-seq analysis for the *ldl2* single mutant.

(**A**) The numbers of genes up-regulated and down-regulated in *ldl2*, and the other (not significantly changed) genes, with protein-coding genes and TE genes shown separately. Asterisk denotes an over-representation of the up-regulated TE genes in the group of TE genes with increased levels of H3K4me1 in *ldl2* identified in Fig. S6 ($P = 0.00024$, hypergeometric test). But this is not the case for protein-coding genes ($P = 0.50$). (**B**) Scatter plot showing the comparisons of expression level between WT and *ldl2*. Genes up-regulated and down-regulated in *ldl2* are shown by red and green dots, respectively, superimposed over black dots, which represent all genes. Value represents RPM + 0.01 to show all genes including genes with 0 read in either or both sample(s). (**C**) The same scatter plot as **B**, indicating TE genes by red dots and protein-coding genes by black dots.



Appendix Figure S8. Gene ontology analysis of differentially expressed genes in *ibm1* and *ldl2*.

(**A,B**) Gene ontology analysis of 1220 UGs (**A**) and 1757 genes up-regulated in *ibm1* and not restored in *ibm1 ldl2* (**B**), showing that a large numbers of genes involved in the defense responses to pathogen are up-regulated in *ibm1*. The result that many defense-related genes were up-regulated in *ibm1* and not fully restored in *ibm1 ldl2* (**B**) might reflect partial transgenerational effect. Indeed, transgenerational effect of defense response is known, and that is reported to depend on H3K9 methylase KYP/SUVH4 (Luna et al, 2014). (**C,D**) Gene ontology analysis of 443 DGs (**C**) and 957 genes down-regulated in *ibm1* and not restored in *ibm1 ldl2* (**D**).

Appendix Table S1. List of primers used in this study.

Name	Sequence	Purpose	Note
attB1-LDL2-RNAi	5’-ggggacaagtttgtacaaaaaagcaggctGATCGGGAAAAGTAACCTCACG-3’	Construction of LDL2 RNAi	Small letters denote attB1 sequence
attB2-LDL2-RNAi	5’-ggggaccactttgtacaagaaagctgggtTCTCTTACAACTCCAGCTCG-3’	Construction of LDL2 RNAi	Small letters denote attB2 sequence
attB1-LDL2proF	5’-ggggacaagtttgtacaaaaaagcaggctTGTTGTCGAGACAGCTAAAACTGC-3’	Construction of pLDL2::LDL2-FLAG	Small letters denote attB1 sequence
attB2-LDL2R	5’-ggggaccactttgtacaagaaagctgggtgATTAAAATGCAGGGGGTTTAAGG-3’	Construction of pLDL2::LDL2-FLAG	Small letters denote attB2 sequence
LDL2-K436A-F	5’-CCTTCTGAACgcAGTTGCCATGCTTTTC-3’	Mutagenesis of LDL2	Small letters denote mutated site
LDL2-K436A-R	5’-GAAAAGCATGGCAACTgcGTTCAGAAGG-3’	Mutagenesis of LDL2	Small letters denote mutated site
ERL2-1F	5'-CAGATCATGTAAGGCTTGTCTTATGC-3'	DNA methylation analysis	
ERL2-1R	5'-ACGTACAGATTGGCAAGACTGAAG-3'	DNA methylation analysis	
KYP-F3	5'-CTGAATGGTAAGGACGTGAACTTGG-3'	DNA methylation analysis	
KYP-R3	5'-AACTACCAACCAACCTGGAAGGTC-3'	DNA methylation analysis	
ACT2 for	5'-CTGCTGGAATCCACGAGACAAC-3'	DNA methylation analysis	
ACT2R2.2	5'-GATCCCATTCATAAAACCCCAGC-3'	DNA methylation analysis	
At3g42658 Pro for	5'-AGGGATGTTAATCGGGCCTCTG-3'	DNA methylation analysis	
At3g42658 Pro rev 4	5'-GAGAACGATTGGAACGATTGTC-3'	DNA methylation analysis	
CACTA1 +186F	5'-TGACATGATAGTATGTCCCTGC-3'	DNA methylation analysis	
CACTA1 +481R	5'-CAGCATCTGCTAACTTTGCTAG-3'	DNA methylation analysis	
At3g21290 +2437F	5'-ACTGTGACAGTTCCAGTCGGAG-3'	DNA methylation analysis	
At3g21290 +3473R	5'-ACTGCACTCAGTGACTGGCAGC-3'	DNA methylation analysis	
At3g62330 +706F	5'-AGGGATGTCCATCTACACTTGG-3'	DNA methylation analysis	
At3g62330 +1420R	5'-CAGCTAGAGTAAGATACGTGATCTC-3'	DNA methylation analysis	
At2g04860 RT for	5'-GGAATGTCGTTACATGGTTATGCG-3'	DNA methylation analysis	
At2g04860 RT rev	5'-GTGATCTCGTCAGGTTTCAACC-3'	DNA methylation analysis	

Appendix Table S2. The most distorted SNPs in the bulked segregant analysis.

Chromosome	Position	Mutation	Predicted effect on protein	Gene ID	Other name
3	3308057	C to T	Synonimous	AT3G10580	No
3	4907337	C to T	Non-coding region		
3	4752971	C to T	Non-coding region		
3	5866154	C to T	Non-coding region		
3	3306245	C to T	Ala to Val	AT3G10572	ABERRANT PEROXISOME MORPHOLOGY 9, APEM9
3	3693561	C to T	Non-coding region		
3	3887259	C to T	Gly to Arg	AT3G12200	ATNEK7, NEK7, NIMA-RELATED KINASE 7
3	4657122	C to T	Non-coding region		
3	3147394	C to T	Arg to Gln	AT3G10180	No
3	5670417	C to T	Non-coding region		
3	4931426	C to T	Non-coding region		
3	4479841	C to T	Trp to stop	AT3G13682	LDL2, LSD1-LIKE2

Supplementary References

Luna E, López A, Kooiman J, Ton J (2014) Role of NPR1 and KYP in long-lasting induced resistance by β-aminobutyric acid. *Front Plant Sci* **5:** 184.

Stavropoulos P, Blobel G, Hoelz A (2006) Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat Struct Mol Biol* **13:** 626-632.