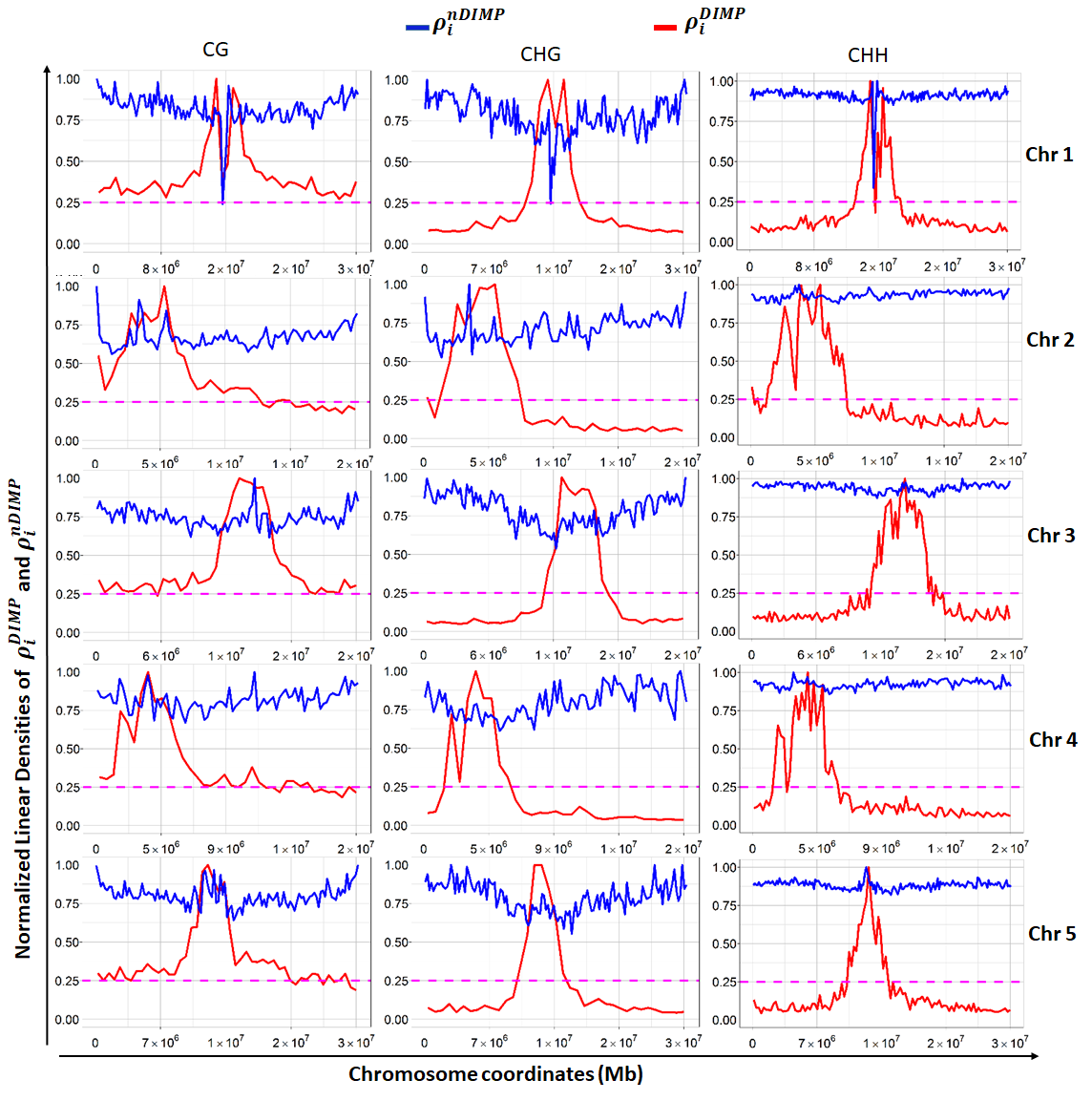
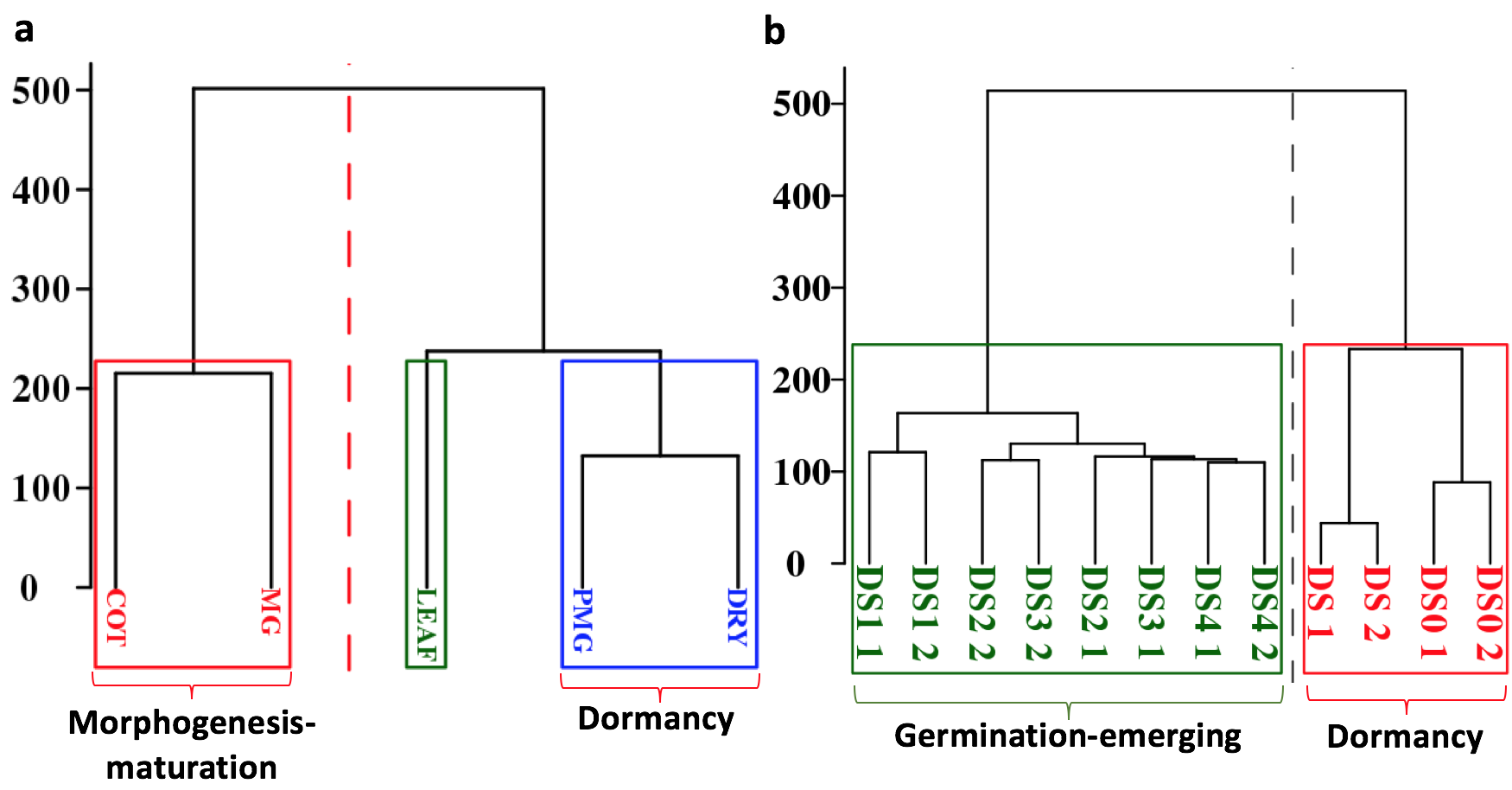
# Additional file 1. Supplementary figures.

# Supplementary Figure 1



**Figure S1.** Linear density of cytosine in Arabidopsis *msh1*-memory line. Linear density is defined as the measure of physical quantity of any characteristic value per unit of length. In this case, linear density measures DIMPs per nucleotide base. Since the number of DIMPs along the DNA sequence varies, local density of DIMPs at a fixed interval is defined by the quotient , where is the amount of DIMPs at the fixed interval . Likewise, local density of non-DIMPs is defined as . For a specified cytosine context like CG, , where is the number of CG positions at the given interval. The linear densities were normalized as: and , where and are the maximum linear densities of DIMPs and non-DIMPs observed in the given chromosome. The pink dashed line is uniformly (and arbitrarily) positioned at 0.25 to demonstrate the consistently higher level of CG DIMPs relative to non-CG.

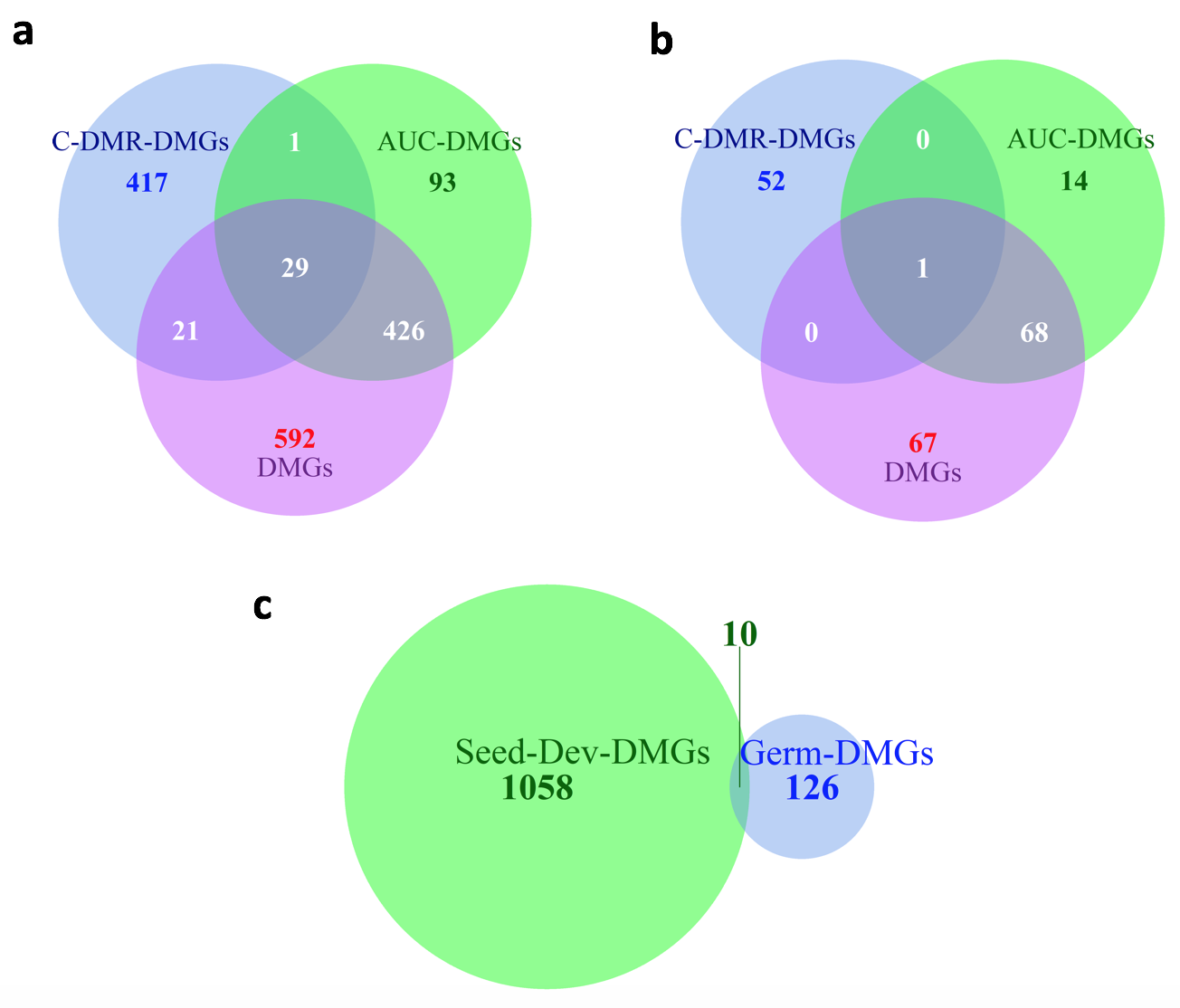
# Supplementary Figure 2



**Figure. S2.**

Classification of seed-developmental and germination methylomes from Arabidopsis seeds. (**a**). Hierarchical cluster built on the set of 7006 selected DIMPs associated genes based on AUC criteria classified the seed developmental stages into two groups: morphogenesis-maturation phase and dormancy phase. (**b**). Hierarchical cluster built on the set of 3864 selected genes based on AUC criteria classified the stages of seed previous to germination into two groups: germination-emerging and dormancy phases.

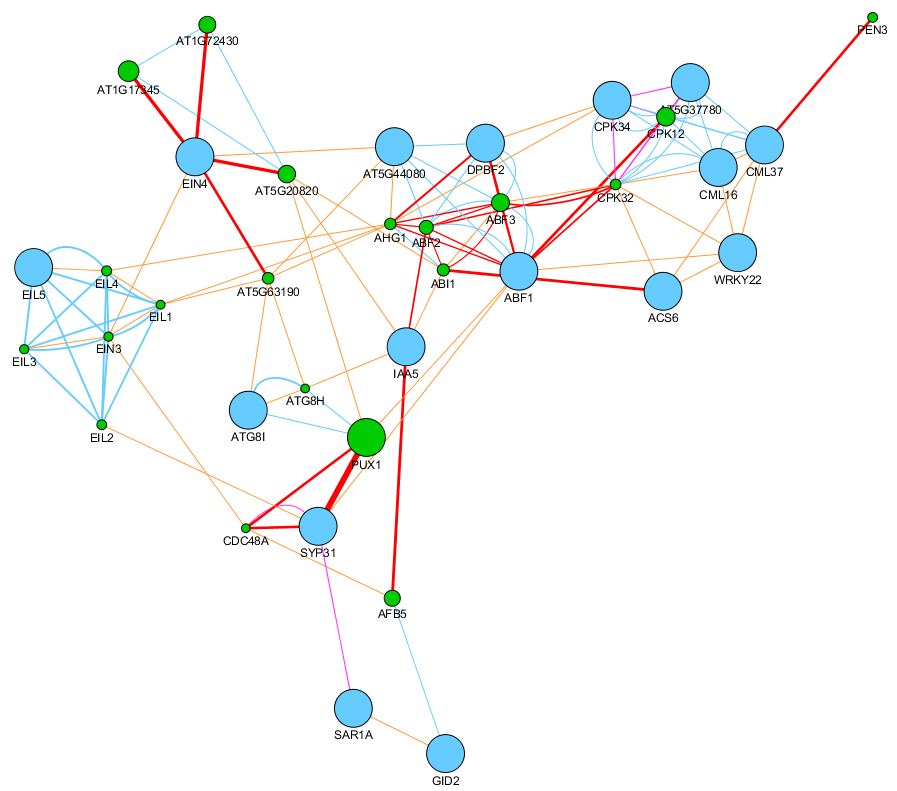
# Supplementary Figure 3



**Figure. S3.**

DMGs comparison (**a**) Overlaps of DMGs obtained from seed-developmental data set for the group comparison morphogenesis versus maturation phases based on the application of GLM to the sets: 1) union of CG, CHG and CHH DMRs (C-DMR-DMGs) derived from reference [18], 2) AUC selected genes (AUC-DMGs) by Methyl-IT, 3) All DMGs by Methyl-IT. (**b**) Overlaps of DMGs obtained from seed germination dataset for the group comparison dormancy versus germination-emerging phases. In this case, C-DMR-DMGs stand for the union of CHG and CHH DMRs (CG DMRs for germination were not found in the study from reference [18]). (**c**) comparison of the DMGs from obtained from seed-developmental data set (a) and seed germination dataset (b) by Methyl-IT.

# Supplementary Figure 4



**Figure S4.** Interaction network built for the seed development DMGs in networks identified with NEAT using GeneMANIA. Orange edges designate co-expression, such that two genes are linked if their expression levels are similar across conditions in a gene expression study. Blue edges designate physical interaction, where two gene products are linked if they were found to interact in a protein-protein interaction study. Red edges designate genetic interaction, so that two genes are functionally associated if the effects of perturbing one gene were found to be modified by perturbations to a second gene. Light blue edges designate shared protein domains, where two gene products are linked if they have the same protein domain. Networks were combined based on GO biological processes to form the final composite network. Blue nodes are DMGs. The GeneMANIA prediction server allows biological network integration for gene prioritization and predicting gene function. The network graphic was built with GeneMANIA's Cytoscape plugin <http://www.cytoscape.org/>

# Supplementary Figure 5



**WT**

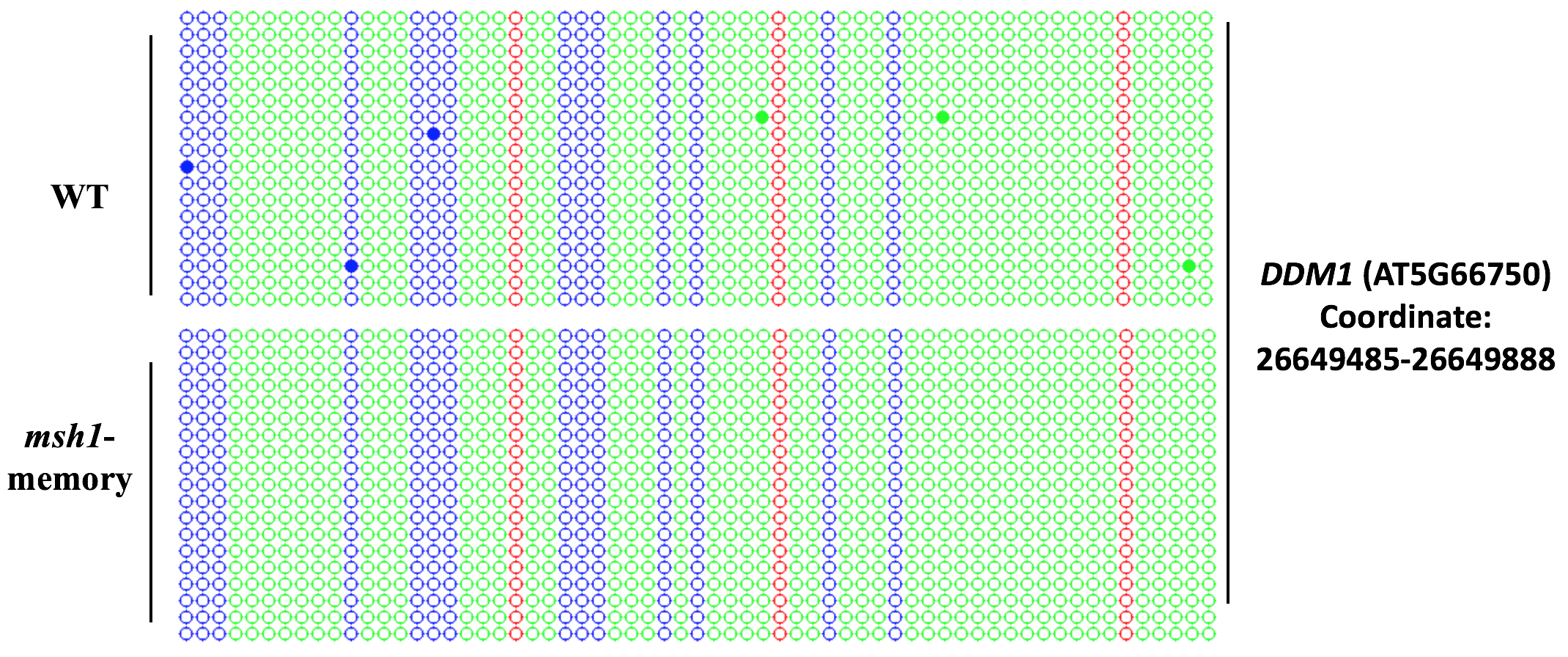
***msh1*-TDNA**

***MSH1*-RNAi (+)**

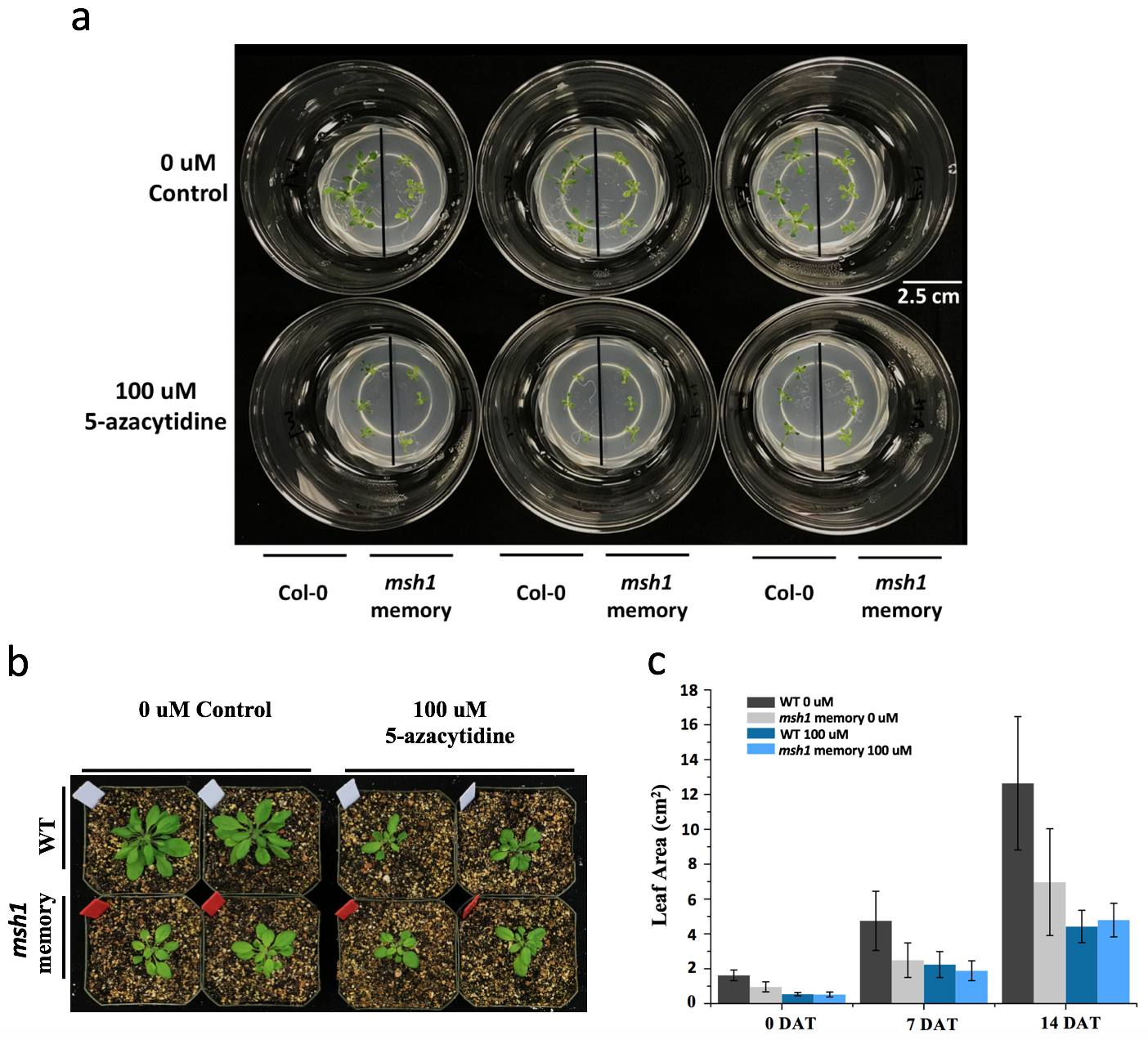
***msh1*-memory**

**Figure S5.** Different *MSH1*-derived developmentally reprogrammed populations.Selected *msh1* memory line phenotype is uniform and stable.

# Supplementary Figure 6

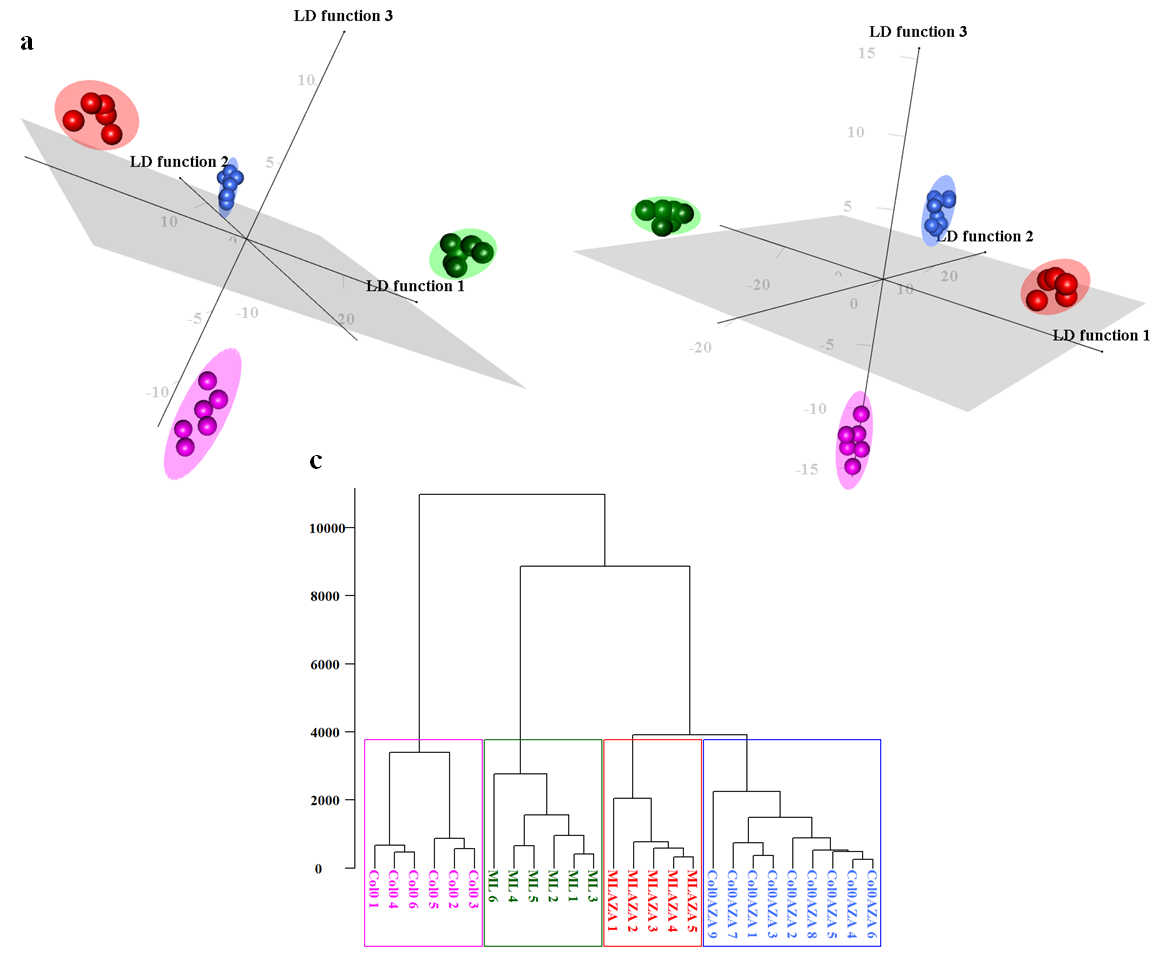
****

**Figure S6.** Control for bisulfite PCR sequencing. The conversion rate of bisulfite treatment was confirmed with *DDM1* (AT5G66750) as control. The overall calculated bisulfite conversion rate was 99.47% for WT and 100% for *msh1* memory line. Dot-plot analysis was applied to the bisulfite sequencing result. Red, blue, and green circles represent CG, CHG and CHH respectively (methylation solid, no methylation blank). Each line represents one clone sequenced, and at least 15 clones were sequenced for each PCR reaction.



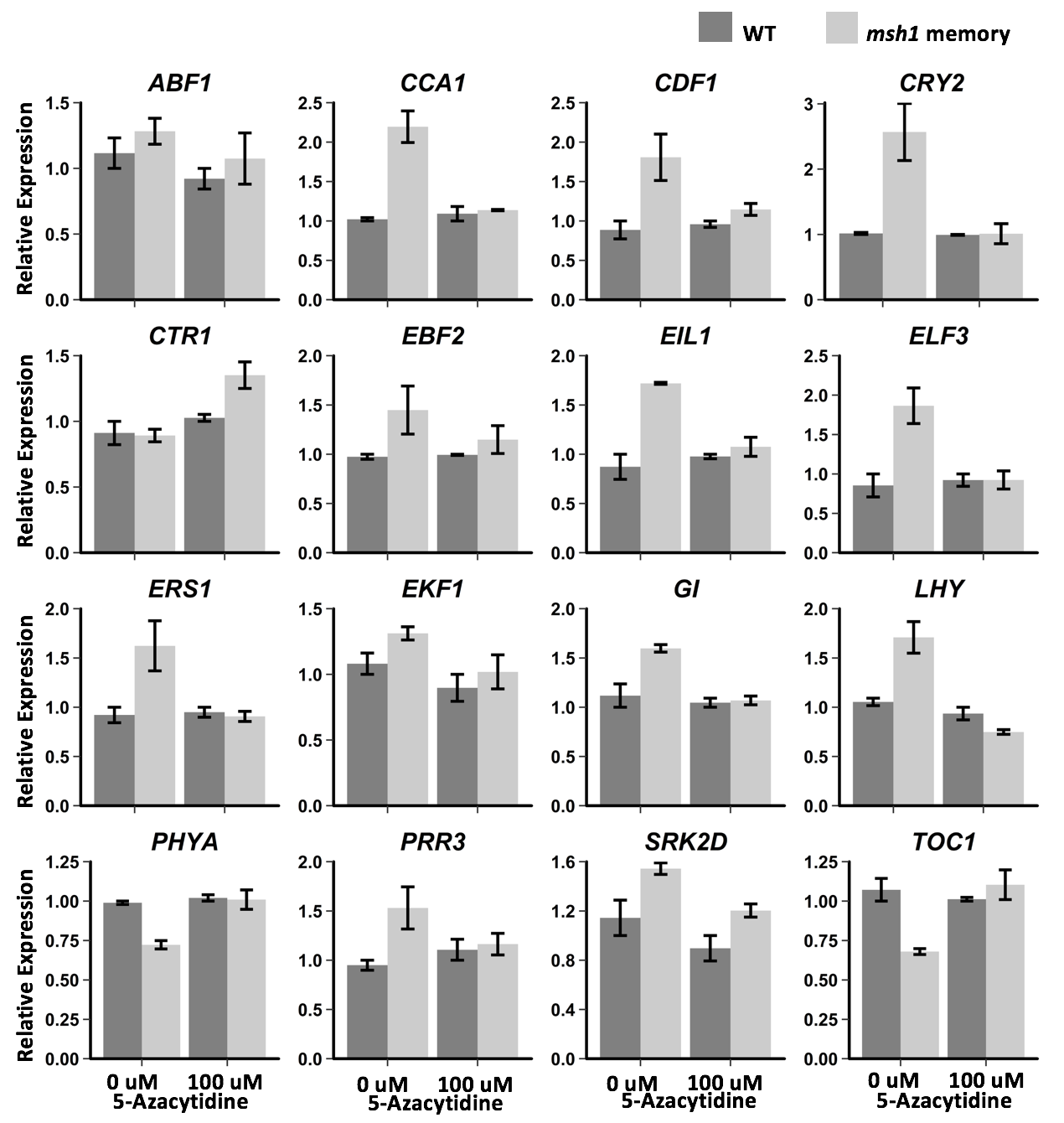
**Figure S7. (a)**100 uM 5-azacytidine treatment of wild type and the *msh1* memory line.Plants were germinated on MS medium containing 0uM or 100 uM 5-azacytidine. Photo was taken 10 days after germination. **(b)** Plant phenotypes following 100 uM 5-azacytidine treatment of the *msh1* memory line and isogenic wild type control. Plants were first grown on MS medium containing 0uM or 100 uM 5-azacytidine for 10 days, then moved to normal soil growth conditions; photos taken two weeks after transplanting. **(c)** Leaf area measurements of wild type (WT) and *msh1* memory plants after 5-azacytidine treatment. Bars represent means ± SD, n=18. Results shown in (c) are from one representative experiment out of three replicates. DAT= Days After Transplanting.

# Supplementary Figure 8



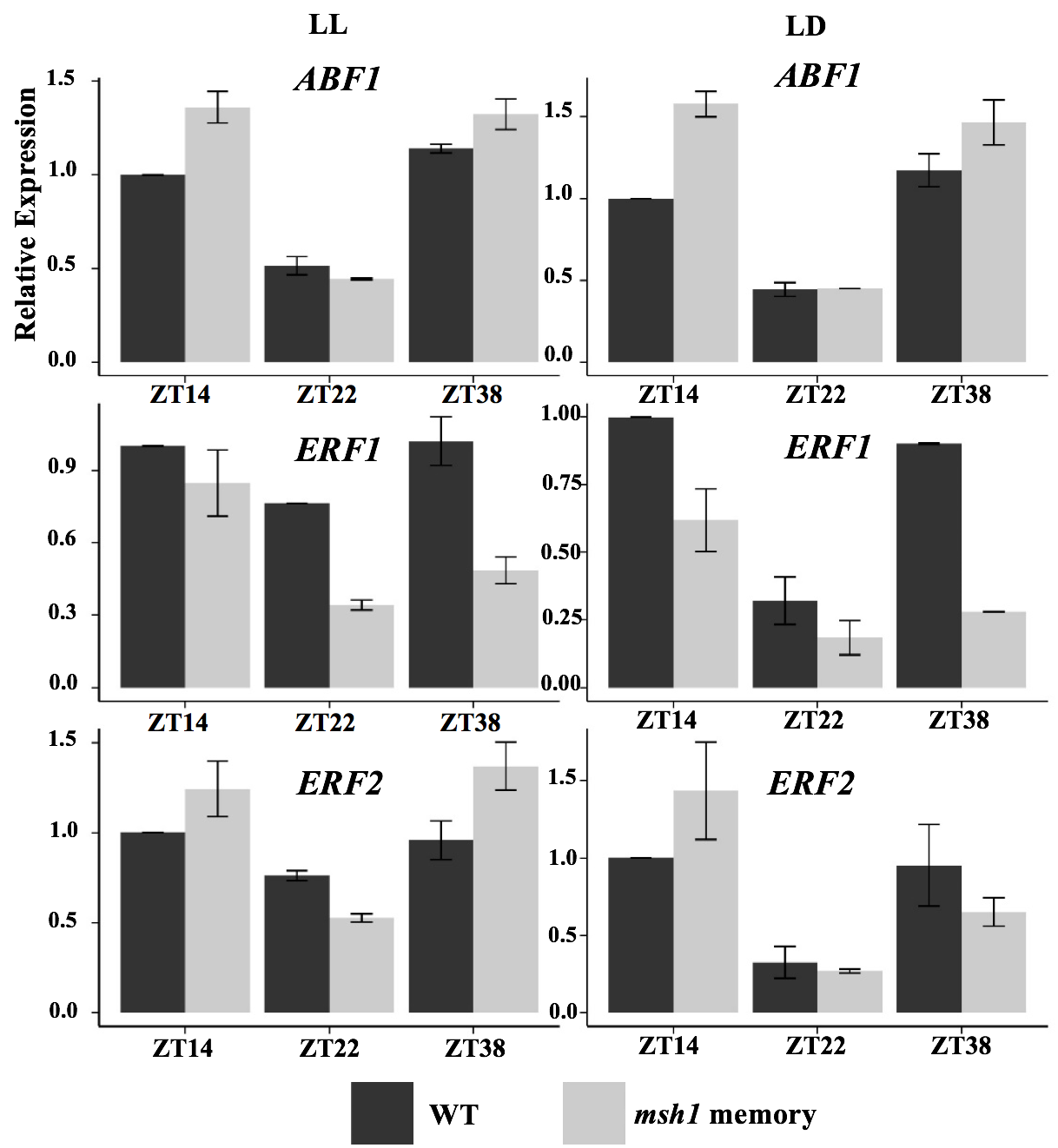
**Figure S8.** Linear discriminant analysis (LDA) of memory line and Col-0 wild type gene expression under 5-azacytidine treatment.Individual samples (ML memory line untreated, green, Col-0 untreated, pink, MLAZA memory line treated, red, and Col-0AZA wild type treated, blue) were represented as vectors of log2-of read-counts obtained from the processing of RNASeq experimental data. Only genes that were identified as DEGs in at least one of the group comparisons (Col-0 vs ML, ML-AZA, Col0AZA) were considered. This analysis yielded a sample space consisting of 26 individuals, with each represented as a vector with 11590 dimensions/variables (genes). **(a)** PCA+LDA performed for the 26 samples in the gene space of 11590 dimensions. **(b)** PCA+LDA performed for the 26 samples in the space of 2660 dimensions. LDA was performed after dimension reduction based on: 1) feature selection:

# Supplementary Figure 9

****

**Figure S9.** Expression of 16 signature genes under 5-azacytidine treatment. Relative expression of the 16 signature genes in Table 2 were tested by quantitative real-time PCR under 100 uM 5-azacytidine treatment. Relative expression was calculated by normalizing to the wild type under corresponding treatment. Error bars represent mean ± SD of three independent biological replicates.

# Supplementary Figure 10

****

**Figure S10.** Altered expression of circadian clock-regulated genes in Arabidopsis *msh1* memory line confirmed by qPCR.Expression patterns of three genes regulated by circadian clock, *ABF1* (AT1G49720), *ERF1* (AT3G23240), and *ERF2* (At5g47220) were assayed by quantitative real-time PCR under both LL (constant light) and LD (12 hours light, 12 hours day) conditions. RNA from three time points was used. Relative expression was calculated by normalizing to the highest value of corresponding wild type in each biological replicate. Error bars represent mean ± SD of three independent biological replicates.