

Arabidopsis SDG proteins mediate Polycomb removal and transcription-coupled H3K36 methylation for gene activation

Yicong Wang, Masato Abe, Yuka Kadoya, Takeru Saiki, Kanae Imai, Xuejing Wang, Taiko Kim To, Soichi Inagaki, Takamasa Suzuki, Tetsuji Kakutani, Toshiro Ito, Nobutoshi Yamaguchi 

Nara Institute of Science and Technology, Japan • Tokyo Institute of Technology, Japan • The University of Tokyo, Japan • Chubu University, Japan • The University of Tokyo Graduate School of Science, Japan

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Polycomb Repressive Complex 2 (PRC2) recognizes Polycomb response elements (PREs) and catalyzes trimethylation of histone H3 on lysine 27 (H3K27me3) for gene silencing. This silencing is counteracted by H3K36 methylation for epigenetic activation of gene expression. Here, we show that the *Arabidopsis thaliana* H3K36 methyltransferases SET DOMAIN-CONTAINING PROTEIN 7 (SDG7) and SDG8 antagonize PRC2-mediated silencing and establish H3K36 methylation patterns with the general transcription machinery. The *sdg7 sdg8* double mutant shows developmental defects and lower H3K36me2 and H3K36me3 levels. SDG7 preferentially binds near PREs, but SDG8 is recruited to H3K36 methylation peaks. The *sdg7 sdg8* phenotypes are partially rescued by loss of Polycomb function. SDG7 overlaps with PRC2 and its recruiters on chromatin and evicts them from shared target genes when conditionally induced. SDG8 and RNA Polymerase II associate at SDG- and RNA POLYMERASE II ASSOCIATED FACTOR 1 complex-regulated targets for H3K36 methylation and transcription. These results suggest that SDG proteins evict PRC2 from PREs to prevent H3K27me3 deposition and activate target genes via transcription-coupled H3K36 methylation.

eLife assessment

This **important** study advances our understanding of the mechanisms underlying chromatin-mediated gene regulation by SET DOMAIN-CONTAINING PROTEIN 7 (SDG7). The evidence supporting the author's claims – centered on a combination of imaging approaches with molecular and genetic experiments – is **convincing**, although certain aspects can be improved. The work will be of broad interest to molecular biologists studying epigenetic regulation of gene expression.

<https://doi.org/10.7554/eLife.100905.1.sa2>

Post-translational modifications of histone proteins play critical roles in regulating the physical properties of chromatin and defining the transcriptional state¹. Among various histone modifications, histone methylation primarily occurs on lysine or arginine residues and can affect

both chromatin condensation and transcriptional states. Trimethylation of lysine 27 on histone H3 (H3K27me3) is associated with downregulation of nearby genes, leading to the formation of heterochromatic regions that are responsive to certain conditions². Methylation of lysine 4 and 36 on histone H3 (H3K4 and H3K36) is present on nucleosomes of transcriptionally active chromatin². Histone modifications are complex and reflect the topological constraints caused by nearby modifications. Indeed, the presence of particular histone modifications can promote the addition of the same type of modification or the removal of different types: for example, higher H3K36 levels hinder H3K27me3 deposition³. Such interconnected epigenetic systems, characterized by bistable states, determine the transcriptional activity of each gene.

Methylation marks can be added to histones by specific enzymes called methyltransferases. In all eukaryotic organisms, the activity of H3K27 methyltransferase is controlled by Polycomb Repressive Complex 2 (PRC2)^{4–6}. In *Drosophila melanogaster*, PRC2 contains four core subunits, including Enhancer of zeste [E(z)] with a catalytic SET domain and Extra sex comb (Esc)⁶. *Arabidopsis* (*Arabidopsis thaliana*) has three E(z) methyltransferases homologous to E(z)—CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA)—that catalyze H3K27me3⁷. Because these methyltransferases act redundantly, higher-order mutants exhibit more severe developmental defects than any single mutant⁸. *Arabidopsis* FERTILIZATION INDEPENDENT ENDOSPERM (FIE) is a WD40 homolog of *D. melanogaster* Esc. PRC2 interacts with multiple transcription factors through its subunits and is recruited to *cis*-regulatory regions of several hundred base pairs in length called Polycomb Response Elements (PREs). Proper PRC2 placement by *trans*-acting factors, such as BASIC PENTACYSTEINE (BPC), APETALA2 (AP2)-like or C2H2 zinc-finger (ZnF) family members, triggers the initial deposition of H3K27me3^{9,10}. The opposing H3K36 trimethylation is regulated by SET DOMAIN GROUP8 (SDG8)¹¹. SDG8 interacts with the transcription machinery, namely RNA polymerase II (RNA Pol II) and the polymerase-associated factor 1 (Paf1) complex (PAF1C)¹². PAF1C components include VERNALIZATION INDEPENDENCE2 (VIP2, also named EARLY FLOWERING 7 [ELF7]), VIP3, VIP4, VIP5 and VIP6 (also named ELF8)¹³. Despite the critical importance of H3K36me3 and the general transcription machinery, *sdg8* single mutants show relatively weak developmental defects¹⁴, suggesting that the full function of SDG8 may be masked by partial redundancy with closely related protein(s). However, the identity of such factor(s) and their roles in the epigenetic systems have not been established.

In this study, we focus on one SDG8 homolog, SDG7, to consolidate the existing knowledge of the epigenetic interaction network along with the general transcription machinery. We show that the *sdg7 sdg8* double mutant exhibits developmental abnormalities and a notable decrease in both H3K36me2 and H3K36me3 levels, supporting the partial redundancy of SDG7 and SDG8. We also examine the genomic regions to which SDG7 and SDG8 are recruited and explore the connection between the *sdg7 sdg8* mutant phenotypes, the PRC2 complex, and RNA Pol II. We propose that SDGs displace PRC2 from PREs, preventing the deposition of H3K27me3 and activating the expression of target genes through transcription-coupled H3K36 methylation.

Results

H3K36me2 and K36me3 levels are decreased by the double loss of SDG7/SDG8 genes

To identify potential factors that function together with SDG8, we examined the five class II SDG genes for overlapping expression patterns¹⁶. Visualization of expression data from the *Arabidopsis* eFP website as a heatmap revealed that *SDG8* is expressed throughout development (Fig. 1a). Among the five members, *SDG7* emerged as a promising candidate for synergistic interaction with *SDG8*, as illustrated by promoter reporter lines in which *b-GLUCURONIDASE* was driven by the *SDG7* or *SDG8* promoter (Supplementary Fig. 1). AlphaFold predicted similar

structures for the SET, post-SET and Associated With SET (AWS) domains of SDG7 and SDG8 (**Fig. 1b,c** [↗](#) and **Supplementary Fig. 2** [↗](#)). Comparison between predicted SDG7 and SDG8 and the crystal structure of the SET domain from ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) in complex with a K36me3 peptide suggested that these proteins have highly similar catalytic core pockets¹⁷ [↗](#) (**Fig. 1d** [↗](#)). These results suggest a conserved biochemical function for SDG7 and SDG8 in H3K36me3 recognition (**Fig. 1d** [↗](#)).

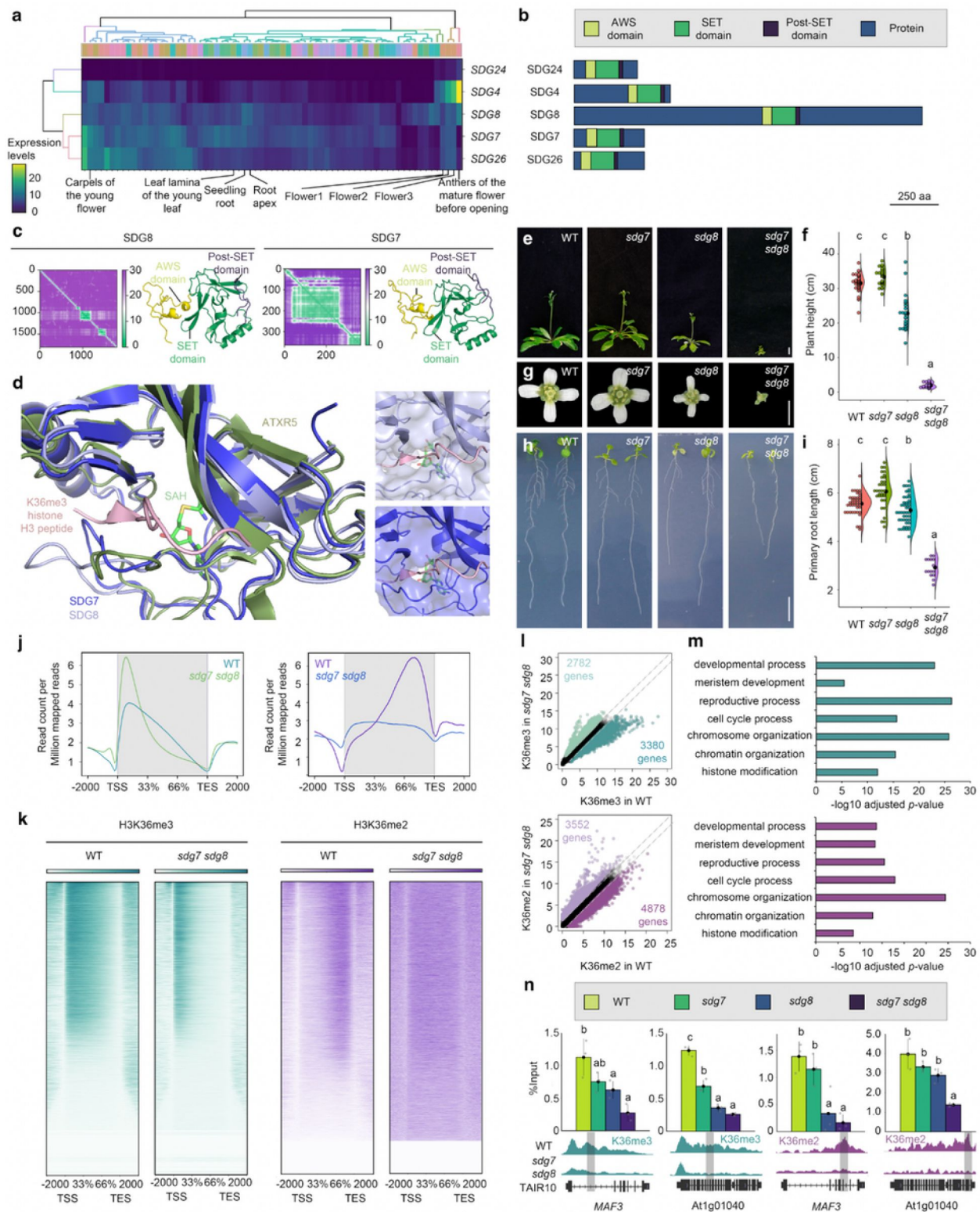


Figure 1

SDG7 and SDG8 mediate H3K36me2 and K36me3 deposition

a, Heatmap representation of gene expression levels for class II SDG family genes in specific tissues. **b**, Diagrams of Arabidopsis class II SDG family proteins. **c**, 3D structure prediction of the conserved domain of SDG7 and SDG8 by AlphaFold 2. Left, Inter-chain predicted aligned error (PAE) plot for the best-ranked model structures of SDG7 and SDG8 by AlphaFold 2. Right, Structure models of SDG7 and SDG8 predicted by AlphaFold 2. Yellow, green and purple lines indicate the AWS, SET and post-SET domains, respectively. **d**, Overlay of the crystal structure for ATXR5 SET domain in complex with K36me3 histone H3 peptide (Dark green: 5VAC), SDG8 (pale purple) and SDG7 (purple). The right panels show the interfaces of SDG7 and SDG8 with histone tails. **e**, Representative photographs of the wild type (WT), *sdg7*, *sdg8* and *sdg7 sdg8* at 25 days after germination. Scale bar, 1 cm. **f**, Plant height of the WT, *sdg7*, *sdg8* and *sdg7 sdg8* (WT: $n = 35$; *sdg7*: $n = 25$; *sdg8*: $n = 25$; *sdg7 sdg8*: $n = 18$). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on post-hoc Tukey's HSD test ($p < 0.05$; See the Data Source file for all combinations of the exact p -values). **g**, Top view of stage 14 flowers from soil-grown WT, *sdg7*, *sdg8* and *sdg7 sdg8*. Scale bar, 0.5 mm. **h**, Representative photographs showing the roots of WT, *sdg7*, *sdg8* and *sdg7 sdg8* 7-day-old seedlings. Scale bar, 1 cm. **i**, Primary root length of WT, *sdg7*, *sdg8* and *sdg7 sdg8* (WT: $n = 41$; *sdg7*: $n = 33$; *sdg8*: $n = 46$; *sdg7 sdg8*: $n = 15$). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on post-hoc Tukey's HSD test ($p < 0.05$; See the Data Source file for all combinations of the exact p -values). **j**, Averaged profiles of H3K36me3 (left) and H3K36me2 (right) signal intensity around genes in WT and *sdg7 sdg8*. **k**, Heatmaps of H3K36me3 (left) and H3K36me2 (right) accumulation across genes in WT and *sdg7 sdg8*. **l**, Scatterplot of H3K36me3 (above) and H3K36me2 (below) levels in WT and *sdg7 sdg8*. Top, shoot samples. Bottom, root samples. Hypermethylated and hypomethylated genes in *sdg7 sdg8* relative to WT are shown in light and dark colors, respectively. **m**, Selected enriched GO terms determined by their $-\log_{10}$ -adjusted p -values. H3K36me3 (top) and H3K36me2 (bottom) hypomethylated genes in *sdg7 sdg8* relative to WT. Top, shoot samples. Bottom, root samples. **n**, H3K36me3 and H3K36me2 ChIP-qPCR at *MAF3* and *At1g01040* in the WT and *sdg7 sdg8*. ChIP-seq results and the diagram of the genomic loci and PCR amplicons are shown below. Values are means $\pm$ SE ($n = 4$ independent experiments). Significant differences were calculated based on one-way ANOVA tests (K36me3 at *MAF3*: $p = 3.6 \times 10^{-3}$; K36me3 at *DCL1*: $p = 4.8 \times 10^{-7}$; K36me2 at *MAF3*: $p = 2.9 \times 10^{-3}$; K36me2 at *DCL1*: $p = 7.0 \times 10^{-3}$). Different letters indicate significant differences based on post-hoc Tukey's HSD test ($p < 0.05$; See the Data Source file for all combinations of the exact p -values).

To assess the biological function of SDG7 and SDG8 during plant development, we generated three independent *sdg7 sdg8* double mutant using different *sdg7* alleles (**Supplementary Fig. 3**). Although the *sdg7* single mutants did not show any clear developmental defects, the *sdg7 sdg8* double mutant flowered significantly earlier and exhibited a dwarf phenotype compared to the *sdg8* single mutant, which itself had an earlier flowering time and smaller than the wild type (WT) (**Fig. 1e-i** and **Supplementary Figs. 3** and **4**) ($p < 0.05$, one-way ANOVA tests). We closely examined cells from leaves and petals in the *sdg7 sdg8* double mutant, which revealed that the observed dwarfism is due to defects in both cell size and cell number (**Supplementary Fig. 4**). Consistent with these phenotypic changes, we detected lower expression levels for marker genes of auxin homeostasis, cell division and key developmental regulators in the *sdg7 sdg8* mutant using promoter reporter lines driving *GUS*, green fluorescent protein (*GFP*) or yellow fluorescent protein (*YFP*) expression (**Supplementary Fig. 5**).

To evaluate the biochemical function of SDG7 and SDG8 in the context of H3K36me3 and H3K36me2 deposition, we conducted chromatin immunoprecipitation followed by deep sequencing (ChIP-seq). Metaplot profiles in the wild type indicated that H3K36me3 and H3K36me2 deposition reaches their highest levels near the 5' and 3' ends of transcribed genes, respectively (**Fig. 1j,k**). The *sdg7 sdg8* mutant showed higher levels of the H3K36me3 mark at the 5' ends of genes, downstream of the transcription start site (TSS), compared to the wild type, and lower

levels for the same mark located in the middle of the genes, as reported previously¹¹ (Fig. 1j,k). By contrast, the H3K36me2 marks were absent across the entire gene body of transcribed regions in the *sdg7 sdg8* mutant (Fig. 1j,k). In the *sdg7 sdg8* mutant, we identified 3,380 genes as being H3K36me3-hypomethylated and another 4,878 as H3K36me2-hypomethylated relative to the wild type (Fig. 1l and Supplementary Tables 1 and 2). Based on gene ontology (GO) term enrichment analysis, these hypomethylated genes were associated with similar developmental processes, such as meristem and reproductive development as well as cell cycle process (Fig. 1m and Supplementary Tables 3 and 4). At selected targets, H3K36me3 and H3K36me2 levels in *sdg7 sdg8* were either similar to or significantly lower than those in the *sdg8* single mutant (Fig. 1n). These results indicate that SDG7 and SDG8 are required to maintain proper H3K36me3 and H3K36me2 levels.

SDG7 occupies TSS and TES while SDG8 occupies gene body

We addressed the genome-wide binding pattern of SDG8 and SDG7 by ChIP-seq (Supplementary Tables 5 and 6). We detected SDG8 binding along the gene body, particularly in introns, exons and their junctions, whereas SDG7 binding was primarily located near the TSS and transcription end site (TES) (Fig. 2a,b). We determined that more than 50% of SDG7-bound regions are within promoter regions (Fig. 2b). We conducted a GO term enrichment analysis using the predicted direct SDG8 and SDG7 targets based on their binding peaks. In good agreement with a previous publication¹¹, we observed an enrichment for light-related GO terms associated with SDG8 target genes such as ‘photosynthesis’, ‘light reaction’ and ‘circadian rhythm’ (Fig. 2c and Supplementary Table 7). The top 20 significantly enriched GO terms for SDG7 targets included ‘biological regulation’, ‘response to stimulus’, ‘response to chemical’ and ‘developmental process’ (Fig. 2c and Supplementary Table 8). An exploration of semantic similarity among all significant GO terms with REVIGO revealed that SDG8 and SDG7 are involved in similar biological processes, such as response to stimulus and circadian rhythm (Fig. 2d and Supplementary Tables 9 and 10).

We also examined the degree of genome-wide overlap in chromatin occupancy for SDG8, SDG7, H3K36me3 and H3K36me2 at the same developmental stages (Fig. 2e). We observed a co-localization between significant SDG8- and SDG7-bound peaks ($p < 1.1 \times 10^{-17}$, hypergeometric test) and with peaks for H3K36me3 and H3K36me2, both on a global scale and at specific loci (Fig. 2e and Supplementary Table 11). Among 331 direct SDG8 targets, 103 genes (31%) met the criteria of decreased H3K36me3 deposition, with another 147 genes (44%) showing lower H3K36me2 levels in the *sdg7 sdg8* mutant (Fig. 2f). Consistent with the general distribution patterns, SDG8 bound to genes enriched with H3K36me3, showing a greater preference for the 5' region of genes, encompassing the gene body, promoter and 5' untranslated region (5' UTR) (Fig. 2f). For both H3K36me3- and H3K36me2-decorated genes, we observed most SDG8 binding peaks within the gene body (Fig. 2f). Looking at H3K36me2 deposition, we determined that SDG8 primarily binds to the 3' ends of genes, including the 3' UTR and intergenic regions (Fig. 2f), suggesting that SDG8 is recruited to H3K36 methylation peaks. We also established that 34% of all SDG7-bound genes overlap with H3K36me3-hypomethylated genes, while another 42% overlapped with H3K36me2-hypomethylated genes. By contrast, SDG7 binding peaks exhibited the same patterns along genomic features for both H3K36me2- and H3K36me3-hypomethylated genes, unlike SDG8. Taking these results together, we propose that SDG7 and SDG8 regulate key shared target genes via H3K36 methylation by distinct mechanisms.

SDG7 displaces PRC2 from PRE to overcome H3K27me3-mediated gene silencing

Permissive histone modifications influence the access of transcription factors and/or the transcriptional machinery by controlling accessibility to their binding sites along chromatin. Since SDG7 is preferentially localized at gene regulatory regions on the genome, we conducted *de novo* motif profiling and identified motifs recognized by known transcription factors within the SDG7-

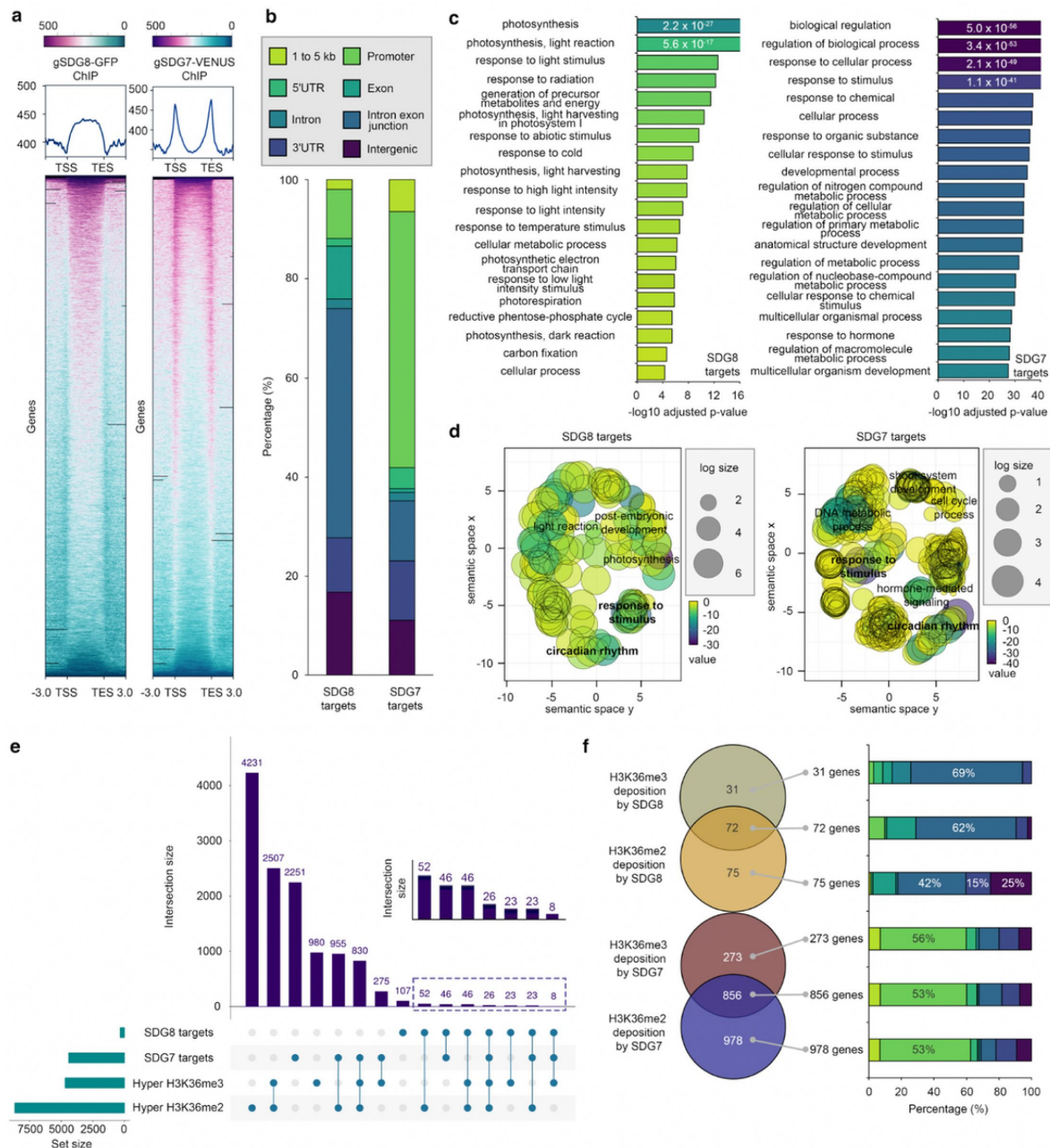


Figure 2

SDG7 is present at TSS and TES, whereas SDG8 occupies gene body

a, Averaged profiles of SDG8-GFP and SDG7-VENUS signal intensity around genes in wild type. **b**, Distribution of peak position for SDG8-GFP and SDG7-VENUS. The percentages of binding peaks at each region type are shown. **c**, Top 20 GO terms as determined by their $-\log_{10}$ -adjusted p -values using SDG8-(left) or SDG7-(right) bound genes. **d**, Scatterplot showing representative clusters in a two-dimensional space based on semantic similarities between GO terms. Color and size indicate p -value and the frequency of the GO term in the underlying gene ontology annotation database, respectively. Left: SDG8-bound genes. Right: SDG7-bound genes. **e**, Upset bar plots showing the number of overlapping genes among SDG8- and SDG7-bound genes and H3K36me3- and H3K36me2-hypermethylated genes in the wild type. **f**, Left: Venn diagram showing the overlap between H3K36me3- and H3K36me2-hypomethylated genes in *sdg7 sdg8* and direct SDG8 (top) and SDG7 (bottom) targets. Right, distribution of peak position for SDG8-(left) and SDG7-(right) bound peaks according to their functional genomic distribution, using the same color legend as in **b**. AnnotatePeak was used for assignment.

bound peaks (**Fig. 3a** and **Supplementary Table 12**). We noticed the overrepresentation of GA repeat (recognized by the BPC family), CTCC, CCG (recognized by AP2/ETHYLENE-RESPONSE FACTORS [ERFs]), G-box (recognized by ZnF proteins), and CARG box (recognized by MADS-box transcription factors) (**Fig. 3a** and **Supplementary Table 12**). Except for the CARG box, these motifs have previously been reported to be associated with Polycomb occupancy⁹. Overrepresentation of the CARG box may be due to BPC-mediated recruitment of MADS domain complex to DNA for tissue-specific regulation of target genes¹⁸. Indeed, we observed a significant overlap between *cis*-elements located at SDG7 peaks and previously identified *cis*-elements located at PREs⁹ ($p < 1.2 \times 10^{-58}$, hypergeometric test) (**Fig. 3b** and **Supplementary Table 13**). The top 10 significantly enriched shared *cis*-elements included binding motifs for Polycomb-recruiting or interacting transcription factors, such as AP2/ERF, Barley B Recombinant (BBR)-BPC and ZnF (**Fig. 3c**). After recognizing these motifs as common elements for SDG7 and PRC2, we investigated the mechanism underlying the antagonism between SDG7 and PRC2. As reported previously, enrichment of PRC2 components (FIE and CLF), and their recruiters (AZF1 and BPC1) was seen at key target loci⁹, such as *PETAL LOSS* (*PTL*), *SHORT VEGETATIVE PHASE* (*SVP*) and *AGAMOUS* (*AG*) (**Fig. 3d**). Indeed, SDG7 binding peaks were present near the PRC2 and recruiter peaks at these loci (**Fig. 3d**). To examine the possible competition between SDG7 and PRC2 for their binding sites, we performed chromatin immunoprecipitation–quantitative PCR (ChIP-qPCR) in *35S:SDG7-HA-GR proCLF:CLF-GFP* plants 8 hours after mock or dexamethasone (DEX) treatment. DEX induction resulted in rapid recruitment of SDG7-HA-GR (a fusion between SDG7, the hemagglutinin tag and the glucocorticoid receptor) to known PREs, the regions occupied by PRC2, as evidenced by ChIP-qPCR with an anti-HA antibody (**Fig. 3e**). An anti-GFP and anti-H3K27me3 ChIP-qPCR performed on the same sample uncovered a concomitant decrease in CLF and H3K27me3 occupancy at *PTL*, *SVP*, and *AG* (**Fig. 3e**), suggesting that SDG7 evicts PRC2 complex from PREs to overcome H3K27me3-mediated gene silencing.



Figure 3

SDG7 displaces PRC2 complex from PRE to overcome Polycomb repression

a, *De novo* motif analysis under each SDG7 peak summit. *p*-values obtained from HOMER are shown. **b**, Venn diagram showing the overlap between SDG7-bound *cis*-elements and all PREs bound by transcription factors. **c**, Classification of transcription factor families seen binding to SDG7 peak summit and PREs. **d**, ChIP-seq signals for H3K27me3, FIE, CLF, AZF1, BPC1, and SDG7 at the selected target loci. The gene models are shown as black bars and lines at the bottom of each panel. **e**, SDG7-HA-GR, CLF-GFP, and H3K27me3 ChIP-qPCR at PREs of *PTL*, *SVP*, and *AG* genes in mock- and dexamethasone (DEX)-treated *pro35S:SDG7-HA-GR* seedlings. ChIP-seq assay and diagram of the genomic loci and PCR amplicons are shown below. Values are means $\pm$ SE ($n = 3$ independent experiments). Significant differences were calculated based on one-tailed Student's *t*-test (SDG7-HA at *PTL*: $p = 0.039$; SDG7-HA at *SVP*: $p = 0.046$; SDG7-HA at *AG*: $p = 4.7 \times 10^{-3}$; CLF-GFP at *PTL*: $p = 6.5 \times 10^{-3}$; CLF-GFP at *SVP*: $p = 0.011$; CLF-GFP at *AG*: $p = 0.01$; H3K27me3 at *PTL*: $p = 0.040$; H3K27me3 at *SVP*: $p = 0.029$; H3K27me3 at *AG*: $p = 0.018$). **f**, Representative photographs of *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* at 25 days after germination. Scale bar, 1 cm. **g**, Top view of stage 14 flowers from soil-grown *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf*. Scale bar, 0.5 mm. **h**, Representative photographs of roots from *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf*. Scale bar, 1 cm. **i**, Plant height in *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* (*sdg7 sdg8* $n = 18$; *clf*: $n = 10$; *sdg7 sdg8 clf*: $n = 22$). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on a post-hoc Tukey's HSD test ($p < 0.05$: See the Data Source file for all combinations of the exact *p*-values). **j**, Root length in *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* (*sdg7 sdg8*: $n = 15$; *clf*: $n = 29$; *sdg7 sdg8 clf*: $n = 46$). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on a posthoc Tukey's HSD test ($p < 0.05$: See the Data Source file for all combinations of the exact *p*-values). **k**, Volcano plot showing differentially expressed genes (DEGs). Top, WT vs *clf*. Middle, WT vs *sdg7 sdg8*. Bottom, *sdg7 sdg8* vs *sdg7 sdg8 clf*. **l**, Venn diagram showing the overlap between differentially expressed genes identified by RNA-seq. Top, upregulated genes in *clf* vs WT, downregulated genes in *sdg7 sdg8* vs WT, and upregulated genes in *sdg7 sdg8 clf* vs *sdg7 sdg8*. Bottom, downregulated genes in *clf* vs WT, upregulated genes in *sdg7 sdg8* vs WT, and downregulated genes in *sdg7 sdg8 clf* vs *sdg7 sdg8*. **m**, Heatmap representation of gene expression levels in wild type (WT), *clf*, *sdg7 sdg8*, and *sdg7 sdg8 clf* at 7 days after germination. The Z-scores are based on FPKM values. **n**, Interactive graph view of 827 genes (genes shown by while color in **Fig. 3m**) generated by REVIGO.

Considering the deposition of opposing histone modifications by SDG7/SDG8 and PRC2, these two types of proteins may function antagonistically in plant development and gene expression. Indeed, introduction of the *clf* mutant causing loss of Polycomb function in the *sdg7 sdg8* background partially rescued the defects seen in the *sdg7 sdg8* mutant (**Fig. 3f-h**). Most phenotypic defects (plant height, flower size, primary root length and lateral root density) in *sdg7 sdg8* were partially suppressed in *clf sdg7 sdg8* compared to *sdg7 sdg8* (**Fig. 3i,j**) ($p < 0.05$, one-way ANOVA tests). Some phenotypes, however, such as flowering time and leaf size, were enhanced in the triple mutant (**Supplementary Fig. 6**), possibly due to secondary effects caused by expression changes in many regulatory genes as reported previously¹⁹. To assess the effect of introducing the *clf* mutation in *sdg7 sdg8* for gene expression patterns, we performed transcriptome deep sequencing (RNA-seq) of wild type, *clf*, *sdg7 sdg8* and *sdg7 sdg8 clf* 7-day-old seedlings (**Supplementary Table 14**). Based on a principal component analysis (PCA) and heatmap analysis with clustering of the RNA-seq data, each genotype formed a different group, reflecting the high quality of our RNA-seq data (**Supplementary Fig. 6**).

We next defined DEGs that were upregulated in the *clf* mutant relative to WT (195 genes), downregulated in the *sdg7 sdg8* mutant compared to WT (467 genes) and downregulated in *clf sdg7 sdg8* relative to *sdg7 sdg8* (838 genes) (**Fig. 3k**). The genes that fulfilled these criteria included the floral repressor gene *FLOWERING LOCUS C*²⁵ (*FLC*) and the antisense transcript originating from the 3' end of *FLC*, *COOLAIR*²⁵ (**Fig. 3l**). DEGs that were downregulated in the *clf* mutant compared to WT (77 genes), upregulated in the *sdg7 sdg8* mutant relative to WT (1,287

genes), and upregulated in *clf sdg7 sdg8* compared to *sdg7 sdg8* (802 genes) may also be involved in phenotypic rescue (**Fig. 3k,l**). We next defined genes that may contribute to the partial phenotypic rescue of *sdg7 sdg8* by the *clf* mutation. *K*-means clustering of all differentially expressed genes (DEGs) identified four main patterns and 16 subclusters (**Fig. 3m**). *SMALL AUXIN UPREGULATED RNAs*²⁰ (*SAURs*) and *YUCCA4*²¹ (*YUC4*), which affect plant size, clustered together (cluster 1 in **Fig. 3m**) and were upregulated in *clf sdg7 sdg8* compared with *sdg7 sdg8*. *GIGANTEA*²² (*GI*), whose protein product regulates the circadian clock, phytochrome B signaling and photoperiod-mediated flowering, clustered with these genes (**Fig. 3m**). By contrast, the genes encoding the calmodulin-like proteins *TOUCH3*²³ (*TCH3*) and *TCH4*, and the serine/threonine kinase *PINOID*²⁴ (*PID*), were downregulated in *clf sdg7 sdg8* compared to *sdg7 sdg8*. A GO term enrichment analysis linked the differentially expressed genes (DEGs) antagonistically regulated by *CLF* and *SDG7/SDG8* (828 genes) to GO terms related to development, phytohormone response, phytohormone-mediated development, indole metabolism, and cell wall organization (**Fig. 3n**). Hence, the partial rescue of the pleiotropic defects and altered gene expression seen in *sdg7 sdg8* by loss of Polycomb function underscores the antagonistic relationship between *SDG7/SDG8* and *PRC2*.

SDG8 and PAF1C propagate H3K36 methylation together with RNA Pol II for shared target activation

SDG8 was previously shown to interact with the transcription machinery components RNA Pol II and its associated factor *PAF1C*. We integrated multiple transcriptome datasets in *sdg7 sdg8* and five *PAF1C*-deficient mutants¹³ (*elf7*, *vip3*, *vip4*, *vip5* and *vip6*) and identified shared downstream genes regulated by both *SDG8* and *PAF1C*. To this end, we selected DEGs with decreased transcript levels in *sdg7 sdg8* and at least one *PAF1C*-deficient mutant (**Fig. 4a,b** and **Supplementary Table 15**). We determined that 230 out of 467 genes (49%) are commonly downregulated in these mutants (**Fig. 4a**), suggesting that *SDG8* primarily mediates target gene expression in conjunction with *PAF1C*. We next focused on *PAF1C*-regulated DEGs that exhibit *CLF*-dependent repression and contributed to the partial phenotypic rescue of the *clf sdg7 sdg8* triple mutant relative to *sdg7 sdg8* (**Fig. 4c**). Several genes, including *FLC*, were significantly repressed by *CLF* (**Fig. 4c**). Indeed, *CLF* and *SDG7* binding peaks were present near the TSS of these loci (**Fig. 4d**). In addition, we observed H3K27me₃ deposition along the gene body at these loci in wild type, while the spreading of the silencing modification disappeared in the *clf* mutant. By contrast, we noticed the higher enrichment of *SDG8*, RNA Pol II and its phosphorylated versions for elongation²⁶ near the TSS and 3' end pause site of RNA Pol II (**Fig. 4d**). We also detected weaker *SDG8* and RNA Pol II enrichment along the gene body for these loci (**Fig. 4d**). The presence of peaks near the TSS and 3' end pause site of RNA Pol II correlated well with the positions of H3K36me₃ and H3K36me₂ peaks, respectively (**Fig. 4d**). Consistent with the levels of H3K36me₃, H3K36me₂ and H3K27me₃ at these loci in the *sdg7 sdg8* mutant, the expression levels of *FLC*, *At4g13572*, *At4g13575*, *UMAMIT14* and *COOLAIR* were lower in *sdg7 sdg8* compared to wild type and *clf* (**Fig. 4d,e**). In addition, the expression levels of most of these genes returned to wild-type levels in the *clf sdg7 sdg8* triple mutant (**Fig. 4e**). Our combined data suggest that *SDG* proteins evict *PRC2* from PREs to prevent H3K27me₃ deposition and activate target gene expression via transcription-coupled H3K36 methylation (**Fig. 4f**).

Discussion

Here we identify *SDG7* and *SDG8* as key regulators that overcome *PRC2*-mediated silencing, leading to a switch from H3K27 to H3K36 methylation during growth and development. We provide a molecular framework for the antagonistic roles of *SDG7* and *PRC2*, which relies on their competition at PREs in the regulatory regions of their shared target loci. Key targets, under dual

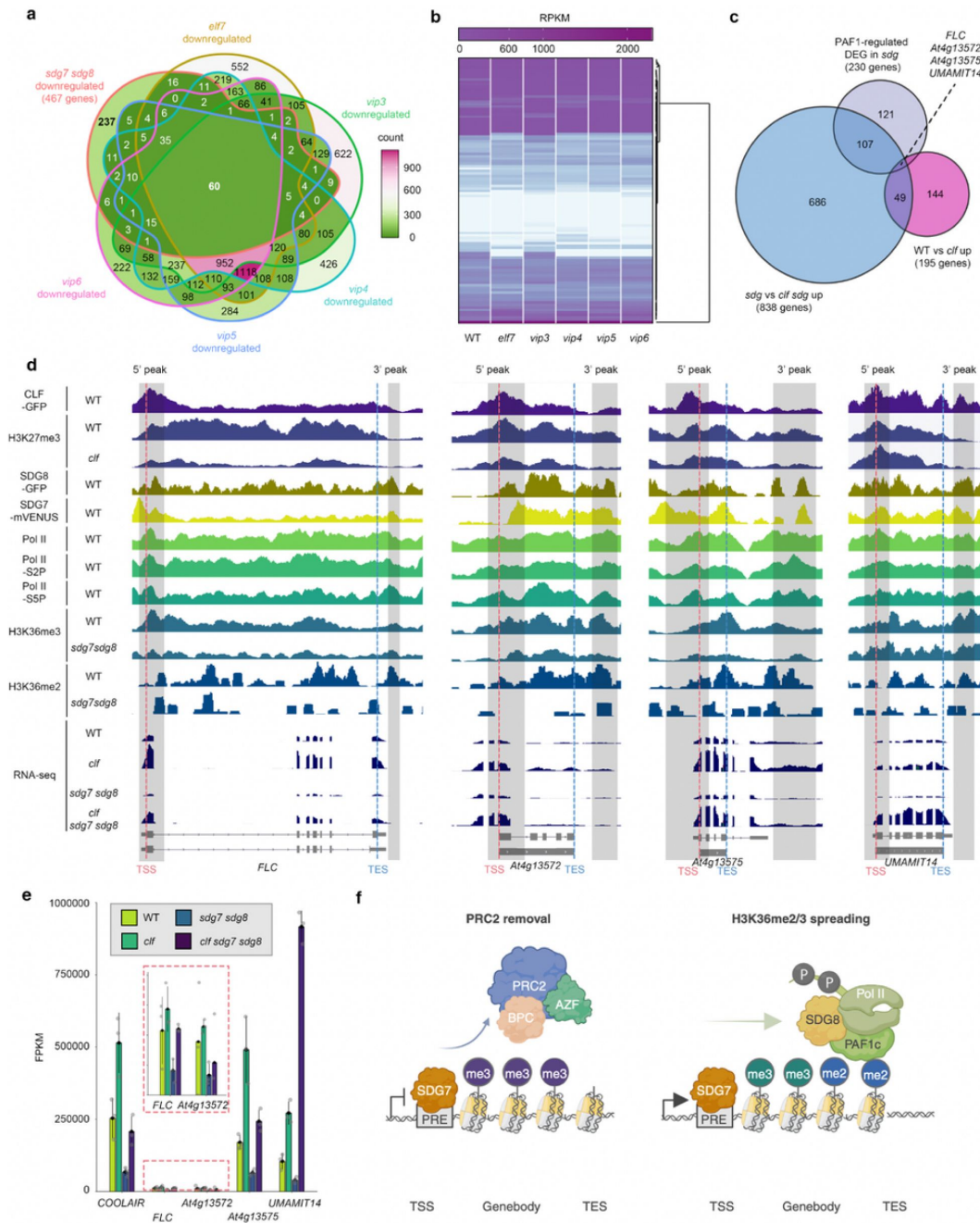


Figure 4

SDG8 and PAF1C propagate H3K36 methylation together with RNA Pol II

a, Venn diagram showing overlapping downregulated genes in *sdg7 sdg8* vs wild type (WT), *elf7* vs WT, *vip3* vs WT, *vip4* vs WT, *vip5* vs WT and *vip6* vs WT. **b**, Heatmap representation of gene expression levels for 230 PAF1C-regulated genes in WT, *elf7*, *vip3*, *vip4*, *vip5* and *vip6*. RPKM values are shown. **c**, Venn diagram showing the overlap between 230 PAF1C complex-regulated genes, upregulated genes in *clf* vs WT, and upregulated genes in *sdg7 sdg8* *clf* vs *sdg7 sdg8*. **d**, ChIP-seq signals for CLF, H3K27me3, SDG8, SDG7, RNA Pol II, RNA Pol II-S2P, RNA Pol II-S5P H3K36me3 and H3K36me2 at the selected target loci and their expression levels in WT, *clf*, *sdg7 sdg8* and *sdg7 sdg8 clf*. The gene models are shown as black bars and lines at the bottom of each panel. **e**, Expression levels of COOLAIR, FLC, At4g13572, At4g13575 and UMAMIT14 in WT, *clf*, *sdg7 sdg8* and *sdg7 sdg8 clf*. Values are means $\pm$ SE ($n = 3$ independent experiments). **f**, Molecular mechanism of K27/K36 methylation propagation by PRC2 and SDG8/SDG7. H3K27me3; purple circles. H3K36me3; green circles. H3K36me2; blue circles.

opposite transcriptional regulation by SDG7 and PRC2, are controlled by SDG8 and the RNA Pol II elongation complex. The binding of these factors balances global distribution patterns between methylation of H3K36 and H3K27 and determines expression levels.

In *Arabidopsis*, PRC2 binds to PREs consisting of GAGA, telobox and G-box motifs. *Trans-acting* factors such as BPC, *Arabidopsis* ZINC-FINGER PROTEINS (AZFs), SCREAM and MUTE recruit PRC2 to PREs for H3K27me3 deposition and gene silencing^{9,27}. We uncover a PRC2 eviction strategy that enables the alteration of epigenetic states. SDG7 is recruited to PREs and displaces PRC2, facilitating H3K36 methylation and activation of the target gene. Hence, the opposing dynamics of H3K27 and H3K36 methylation is regulated by *cis*-localized DNA motifs. It has been reported that the MADS-box protein AG can evict PRC2 from PREs at specific loci²⁸. Interestingly, CArG boxes, which are important for MADS box binding, were significantly enriched under the SDG7 binding peaks. The relationship between SDG7 and MADS-box proteins in the context of PRC2 displacement needs to be examined in the future. Although we identified many *cis*-elements associated with PRC2 occupancy, we did not identify the telobox among them. Hence, additional factors that can displace PRC2 from PREs besides SDG7 likely exist and may act at different stages in different tissues.

Our data suggest that SDG8 and RNA Pol II (as well as its phosphorylated versions Ser2P and Ser5P) are located near the TSS and 3' ends of target genes. We observed H3K36me3 and H3K36me2 peaks at the TSS and 3' end pause sites of RNA Pol II, and these signals were significantly weaker in *sdg7 sdg8*, suggesting the critical role of SDG8 and the RNA Pol II elongation complex for H3K36me3 and H3K36me2 deposition. Differentially H3K36me3-methylated genes maintained their H3K36me3 peaks at their 5' ends in the wild type and the *sdg7 sdg8* double mutant. Additional support for this pattern comes from previous ChIP assays in the *sdg8* single mutant¹¹, suggesting that SDG8 may primarily function in the propagation of H3K36me3 rather than its initial deposition. Thus, additional histone methyltransferases may contribute to H3K36me3 deposition near the TSS, even in the *sdg7 sdg8* double mutant background. Notably, H3K36me2 was almost absent from differentially H3K36me2-methylated genes in the *sdg7 sdg8* double mutant. H3K36me2 functions not only in the elongation and/or termination steps of mRNA transcription from the 5' end, but also in the initiation of transcription from the 3' end, as observed with *COOLAIR*.

Changes in the relative balance of mutually exclusive histone modifications occur during plant development. Due to the different distribution of histone modification patterns along the genome between plants and animals, gaining mechanistic insight into the antagonism in plants will benefit not only plant development but also the evolution of gene expression systems.

Methods

Plant materials and growth conditions

All *Arabidopsis* (*Arabidopsis thaliana*) lines analyzed in this study are in the Columbia (Col-0) accession. *sdg8-2*²⁹ (SALK_026442), *sdg7-2*²⁹ (SALK_131218), *sdg7-3* (SALK_072682), *sdg7-4* (WiscDsLox430F09), *clf-28* (SALK_139371), *proLBD16:GUS30 proPLT2:PLT2-YFP3*, *proSHR:SHR-GFP3*², *DR5:GUS33*, *proWOX5:NLS-GFP3*⁴, *proCYCB1;2:CYCB1;2-NLS-YFP5*, *proCLF:CLF-GFP3* were used. Primers used for genotyping are listed in [Supplementary Table 16](#).

Arabidopsis seedlings were grown on half-strength Murashige and Skoog (MS) medium plates. MS salts (Duchefa Biochemie), 2-(*N*-morpholino)ethanesulfonic acid (Nacalai Tesque, Kyoto) and 0.8% (w/v) agar (Nacalai Tesque, Kyoto) were added to distilled water. The pH was adjusted from 4.3 to 5.6 with KOH. After autoclaving at 121°C for 20 min, 50 ml medium was poured into individual sterile No. 2 square culture dishes (Eiken Chemical) on a clean bench.

Seeds were surface sterilized with 70% (v/v) ethanol for 10 min, washed three times with sterile distilled water and sown on half-strength MS plates. After stratification at 4°C in the dark for up to 7 days, the plates were placed in dish drainer trays located in a growth chamber at 22°C under constant light conditions (200 $\mu\text{mol m}^{-2}\text{s}^{-1}$).

Cloning and plant transformation

To generate the construct harboring the *SDG8* promoter, the genomic region preceding the gene was amplified from Col-0 genomic DNA with primers proSDG8-FW and proSDG8-RV using PrimeSTAR GXL DNA polymerase (TaKaRa) through gradient PCR. The resulting PCR product was subjected to electrophoresis on an agarose gel, purified from the gel using a gel/PCR extraction kit (NIPPON Genetics) and then inserted into pENTR/D-TOPO (Thermo Fisher Scientific) using Gateway cloning. The identity of the inserted sequence was confirmed via Sanger sequencing on an Applied Biosystems 3130 genetic analyzer (Applied Biosystems). Supplementary Table 16 provides the list of primers employed for the promoter cloning. The promoter fragments were recombined into the pBGWSF7.0 GUS-GFP³⁷ vector via LR reaction facilitated by LR Clonase II mix (Thermo Fisher Scientific). Following verification of the junction between the promoter insertion and *GUS-GFP*, the resulting plasmids were introduced into *Agrobacterium* (*Agrobacterium tumefaciens* strain GV3101. Positive *Agrobacterium* colonies were used for transformation of *Arabidopsis* plants through the floral dip method³⁸. For *gSDG7-VENUS* and *gSDG7-GUS*, a 4461-bp *SDG7* genomic fragment (from -1051 bp to +3410 bp; the A of the start codon was set as +1) amplified with the primers gSDG7-FW and gSDG7-RV was cloned into the pCR8/GW/TOPO vector (Invitrogen). An *SfoI* restriction site was introduced upstream of the *SDG7* stop codon through mutagenesis PCR using the primers mSDG7-SfoI-FW and mSDG7-SfoI-RV. *VENUS* and *GUS* fragments were then inserted into the *SfoI* site of the pCR8/GW/TOPO-*gSDG8* plasmid. Finally, the *gSDG7-VENUS* or *gSDG7-GUS* cassette was recombined into pEarleyGate303 using LR Clonase II (Invitrogen). To generate the pro35S:*SDG7-GR* plasmid, a *SDG7* cDNA fragment of 1174 bp with 85 bp of the 5' untranslated region (5' UTR) was amplified by the primers SDG7-KpnI-FW and SDG7-ApaI-RV. After digestion, the *SDG7* cDNA without the stop codon was inserted into pGreen0281 in-frame with the sequence encoding GR-HA.

The resulting plasmids were introduced into *Agrobacterium* (strain GV3101. Positive *Agrobacterium* colonies were used for transformation of *Arabidopsis* plants through the floral dip method³⁸).

Data comparison

Publicly available expression datasets were obtained from The Bio-Analytic Resource for Plant Biology database (<http://bar.utoronto.ca>) or from published work¹³. After reorganizing gene expression levels in Microsoft Excel (Version 16.79.2), a heatmap was generated with the packages heatmaply (Version 1.4.2) or superheat (Version 0.1.0) in R (Version 4.3.1). Venn diagrams were generated using Venny² (https://bioinfogp.cnb.csic.es/tools/venny/) or BioVenn (https://www.biovinn.nl) on the respective websites, or ggVennDiagram (Version 1.2.3) in R (Version 4.3.1). *p*-values were calculated based on the hypergeometric test (http://nemates.org/MA/progs/overlap_stats_prog.html). Upset plots were generated using the UpSet R Shiny App (https://gehlenborglab.shinyapps.io/upsetr/).

Protein structure prediction and comparison

Protein structure was predicted using Alpha fold 2.0³⁹ (Version 1.5.3) on Google Colab (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>). The amino acid sequences for five class II SDG family members were obtained from the TAIR website (<https://www.arabidopsis.org>). The crystal structure data of the ATXR5 SET domain in complex with

K36me3 (PDB entry 5VAC¹⁷) was obtained from the Protein Data Bank (<https://doi.org/10.2210/pdb5VAC/pdb>). Comparison between predicted SDG proteins and ATXR5 was conducted using the PyMOL Molecular Graphics System (Version 2.5.2).

Phenotyping

For plant height analysis, plants were grown on soil. The final height of the main stem bearing an inflorescence was measured. Photographs of the entire plant were taken, and plant height was measured with ImageJ software (NIH).

For root phenotypic analysis, seedlings were grown on vertically oriented MS plates. The root length of seedlings 7 days after germination (7 DAG) was measured. Photographs of the entire seedling were taken. The length from the base of the hypocotyl to the root tip was measured with ImageJ software (NIH). To quantify lateral root number, the roots of 7-DAG seedlings were used. The lateral root number was counted using a microscope (Olympus SZ).

For leaf phenotypic analysis, the 5th true leaves from plants grown on soil were used. Leaves were excised with forceps and placed into glass vials containing FAA (formalin: ice-cold acetic acid:70% [v/v] ethanol at a volume ratio of 1:1:18) for fixation. After 30 min in Faa under vacuum, the leaves were transferred to a chloral hydrate/glycerol/H₂O mixture solution (8 g: 1 mL: 2 mL) and incubated for at least 16 h until clear. Leaves were then placed onto glass slides (Matsunami Glass Industry Co., Ltd.), and covered with coverslips (Matsunami Glass Industry Co., Ltd.). Images were taken with an AXIO Zoom V16 stereo microscope (Zeiss). Leaf area and leaf length from the leaf tip to the petiole were measured with ImageJ software (NIH). To quantify palisade cell size and number, the same leaf samples were used. The cells were imaged using an AxioScope A light microscope (Zeiss). The epidermal cell area was measured with ImageJ software (NIH). Cell numbers were calculated from leaf and cell areas.

For petal phenotypic analysis, plants were grown on soil. Petals from stage 15 flowers were used. The petal area was measured with ImageJ software (NIH). To quantify cell size and number, petal epidermal cells were imaged by scanning electron microscopy (SEM). For SEM, petals were placed in FAA (45% [v/v] ethanol, 2.5% [v/v] formaldehyde, and 2.5% [v/v] acetic acid), vacuum-infiltrated until the tissues sank and incubated at room temperature for at least 16 h. The fixed tissues were then passed through a gradient ethanol series (50% [ethanol:water, v/v], 60%, 70%, 80%, 90%, 95%, 100% × 2) for 20 min each, followed by a gradient acetone series (25% [acetone:ethanol, v/v], 50%, 75%, 95%, 100% × 2) for 30 min each. Then, the tissues were dried with an EM CPD300 critical point dryer with liquid CO₂ (Leica Microsystems). Samples were gold-coated with a coating time of 45 s using a gold coater instrument (Hitachi E 1010). The tissues were imaged with a S-4700 scanning electron microscope (Hitachi) with an accelerating voltage of 15 kV. At least 10 petals for each genotype were observed; representative images are shown. Cell area was measured with ImageJ software (NIH). Cell numbers were calculated from petal and cell areas.

All plots were generated using ggplot2 (Version 3.4.3) in R (Version 4.3.1). Significant differences were determined through one-way analysis of variance (ANOVA) tests, followed by a post-hoc Tukey's honest significant difference (HSD) test (https://astatsa.com/OneWay_Anova_with_TukeyHSD/).

ChIP-seq

ChIP-seq was conducted following the procedures reported previously with minor modifications⁴⁰. Two grams of 5-day-old seedlings were used. Frozen tissues were ground to a fine powder using a mortar and pestle and transferred to nuclei isolation buffer (10 mM HEPES, 1 M sucrose, 5 mM KCl, 5 mM MgCl₂, 5 mM EDTA, 16% [v/v] formaldehyde, 20% [v/v] Triton X-100, β-mercaptoethanol, 100 mM Pefabloc) for 10 min to crosslink proteins and DNA. The reaction was quenched by adding 0.125 M glycine. After removal of debris by filtration through two layers of

Miracloth, the resulting solution was centrifuged at 3,000g for 10 min at 4°C. Pellets were resuspended in nuclei isolation buffer. Chromatin was separated using nuclei separation buffer (1 M HEPES (pH = 7.6), 3 M KCl, 1 M MgCl₂, 0.5 M EDTA) by centrifugation. The pellets were dissolved in SDS lysis buffer (1 M Tris-HCl (pH = 7.8), 10% [w/v] SDS, 0.5 M EDTA) and ChIP dilution buffer (50 mM Tris-HCl (pH = 7.8), 0.167 M NaCl, 1.1% [v/v] Triton-X, 0.11% [v/v] sodium deoxycholate). Then, sonication (Time: 20 min, Duty cycle: 5%, Intensity: 4, Cycles per burst: 200, Temperature: 5°C, Power mode frequency: Sweeping) was conducted with a Covaris M2 (M&S instruments) to obtain approximately 300-bp sheared DNA fragments. Immunoprecipitation was conducted using anti-H3K36me3, anti-H3K36me2 and anti-H3 antibodies (H3K36me3, ab9050; H3K36me2, ab9049; H3, abcam), anti-GFP antibody (SAB4701015; Sigma-Aldrich), and Dynabead Protein A (Thermo Fisher) for at least 16 h at 4°C. Beads were washed with low salt RIPA buffer (50 mM Tris-HCl (pH = 7.8), 0.15 M NaCl, 1 mM EDTA, 0.1% [w/v] SDS, 1% [v/v] Triton-X, 0.1% [v/v] sodium deoxycholate), high-salt RIPA buffer (50 mM Tris-HCl (pH = 7.8), 0.3 M NaCl, 1 mM EDTA, 0.1% [w/v] SDS, 1% [v/v] Triton-X, 0.1% [v/v] sodium deoxycholate), LND buffer (0.25 M LiCl, 1% [v/v] IGEPAL, 1% [v/v] sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl (pH = 7.8)) and 1× TE (10 mM Tris-HCl, 0.1 mM EDTA pH 8.1). Then, ChIP dilution buffer was added to beads and incubated at 65°C overnight to reverse the crosslinking. RNase and Proteinase K were then added to digest residual RNA and proteins. ChIP DNA was purified with a Monarch PCR & DNA Cleanup Kit (Monarc). To generate libraries for sequencing, a ThruPLEX DNA-seq kit (RUBICON GENOMICS) was used following the manufacturer's instructions. Libraries were purified with AMPure beads (Beckman Coulter) and sequenced on a NextSeq 500 or NovaSeq 6000 instrument (Illumina).

Mapping of ChIP-seq data for histone modification was performed as reported previously³⁹. Using Bowtie (version 1.2.2), bam files were created by selecting the same sequences matching the *Arabidopsis thaliana* TAIR10 reference genome (<https://www.arabidopsis.org>). Based on the bam files, bed files were created with bedtools (Version 2.27.0). The coverage was determined in R (Version 3.5.2). Mapping of binding peaks for histone modification enzymes was also conducted as described previously²⁸. The raw sequencing reads underwent quality checks and trimming using FastQC (Version 0.11.7) and Trimmomatic (Version 0.38), respectively. Subsequently, mapping was carried out with Bowtie2 (Version 2.3.4.2) using the resulting fastq files. Peaks were then identified using SICER (Version 1.1). Read counting was performed through featureCounts (Version 1.6.3). Heatmaps were generated using deeptools (Version 3.2.1). Motif analysis was conducted using HOMER (Version 4.1). Binding peaks were visualized in the Integrative Genomics Viewer (IGV) Web App (<https://igv.org/app/>). The ChIP-seq data have been submitted to the DDBJ database (SDG7/SDG8: DRA17907, H3K36me2: DRA17938, and H3K36me3: DRA17939).

Gene Ontology (GO) term enrichment analysis

GO term analysis was conducted with the online tool g:Profiler⁴¹ (https://biit.cs.ut.ee/gprofiler_beta/gost) with g:GOS function. To reduce duplicated GO terms, the agriGO⁴² database was used (<http://bioinfo.cau.edu.cn/agriGO/>). The *p*-values produced by the GO analysis were filtered by REVIGO⁴³ (<http://revigo.irb.hr/>). Original files of the network and tree maps were generated by the REVIGO website (<http://revigo.irb.hr/>). The graphs of the tree maps were created using ggplot2 (Version 3.4.3) in R (Version 4.3.1). The final network figure was generated in Cytoscape (Version 3.10.1).

Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR)

ChIP-qPCR was conducted following the procedures reported previously with minor modifications⁴⁴. To conduct ChIP up to 0.8 g (fresh weight) of 5-day-old seedlings grown on half-strength MS plates was harvested in vials filled with phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 2 mM KH₂PO₄, pH 7.4). Fixation was performed in 1% (v/v) formaldehyde for 15 min. During fixation, tissues were vacuum-infiltrated three to five times until they sank. Afterwards, samples were moved to PBS with 0.125 M glycine for 5 min to quench the fixation. Seedlings were rinsed with icecold PBS twice, frozen in liquid nitrogen, and stored in a

deep freezer until use. Tissues were ground to a fine powder using a mortar and pestle and dissolved in Nuclei Extraction Buffer (100 mM MOPS pH 8.0, 10 mM MgCl₂, 0.25 M sucrose, 0.5% [w/v] dextran T-40, 2.5% [w/v] Ficoll 400, 9.36 μ L/mL β -mercaptoethanol and 10 μ L/mL protease inhibitors) to extract protein–DNA complexes. After filtration over two layers of Miracloth and centrifugation, chromatin pellets were resuspended in nuclei lysis buffer (10 mM EDTA pH 8.0, 50 mM Tris-HCl, pH 8.0, and 1% [w/v] SDS). ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 1.1% [v/v] Triton X-100 and 0.01% [w/v] SDS) was then added to each sample. Sonication was conducted with an ultrasonic disruptor (TOMY) to obtain 200–700-bp sheared DNA. ChIP dilution buffer with 22% (v/v) Triton X-100 was mixed in with the sonicated samples and centrifuged to remove debris. After preclearing by Dynabeads™ Protein A (Thermo Fisher) and collection of input DNA, immunoprecipitation was conducted at 4°C overnight using Dynabeads and commercial antibodies (GFP SAB4301138; HA, 12C5; H3K27me3, ab6002; H3K36me3, ab9050; H3K36me2, ab9049 Abcam). Beads were successively washed in low salt buffer (2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl, 0.1% [w/v] SDS and 1% [v/v] Triton X-100), LiCl buffer (0.25 M LiCl, 1% [w/v] Nonidet P-40, 1% [w/v] deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1), and Tris EDTA buffer (TE, 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.1) twice each for 10 min at 4°C. Elution of protein–DNA from beads was conducted in nuclei lysis buffer by incubation for 25 min at 65°C twice. After combining both eluates, crosslinking was reversed with NaCl overnight at 65°C. DNA was purified with a QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions with slight modifications. qPCR was performed as described above. Primer sequences are listed in [Supplementary Table 16](#). Significant differences were calculated based on one-way ANOVA tests followed by a post-hoc Tukey's HSD test.

RNA-seq

Total RNA was extracted from the entire root of 5-day-old MS-grown seedlings. An RNeasy plant mini kit (Qiagen) was employed to extract RNA following the manufacturer's instructions. Using total RNA treated with DNase I (Qiagen), sequencing libraries were prepared using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and NEBNext Ultra TMII Directional RNA Library Prep Kit (NEB). After confirmation of library quality on a bioanalyzer (Agilent), sequencing was conducted using a NovaSeq 6000 instrument (Illumina).

Prior to mapping, the raw sequencing reads underwent quality checks and trimming using FastQC (Version 0.11.7) and Trimmomatic (Version 0.38), respectively. Subsequently, mapping against TAIR10 genome was carried out with HISAT2 (Version 2.1.0) using the resulting fastq files. Read counting and calculation of FPKM and TPM were performed through featureCounts (Version 1.6.3). Hierarchical clustering and principal component analysis were conducted with the R package stat (Version 3.6.1). Differentially expressed genes were identified using DESeq2 (Version 1.24.0). Volcano plots were generated by plotly (Version 4.9.2.1). Reads were visualized in the Integrative Genomics Viewer (IGV) Web App (<https://igv.org/app/>). The RNA-seq data have been submitted to the DDBJ database (DRA17906).

β -glucuronidase (GUS) staining

For GUS staining, 5-day-old seedlings grown on MS plates were fixed in ice-cold 90% (v/v) acetone for 10–15 min. The samples were washed with sterile water, followed by GUS buffer twice (50 mM phosphoric acid buffer, 0.5 mM K₃Fe[CN]₆ [Nacalai Tesque, Kyoto], 0.5 mM K₄Fe[CN]₆ [Nacalai Tesque, Kyoto]). Subsequently, tissues were incubated at 37°C for 30 min to 1 h in GUS buffer with 1 mM X-Gluc (5-bromo-4-chloro-3-indolyl β -D-glucuronide) (Nacalai Tesque, Kyoto). To stop the staining reaction, samples were transferred into 70% (v/v) ethanol.

To observe GUS-stained tissues, samples were washed in sterile water three times, incubated in 4% (v/v) HCl with 20% (v/v) methanol for 15 min at 55°C. Then, the solution was changed to 70% (v/v) ethanol with 7% (w/v) NaOH and incubated for another 15 min at room temperature. Subsequently, tissues were successively incubated in 40% (v/v), 20%, 10% ethanol for 10 min.

Finally, samples were incubated in 10% (v/v) ethanol with 50% (v/v) glycerol for 30 min and observed with an upright microscope (Zeiss Scope.A1) by differential interference contrast microscopy.

Confocal microscopy

Fluorescence observation of roots and seedlings was conducted using an FV1000 (Olympus), LSM710, or SP8 microscope (Leica). To generate seedling sections, tissues were placed into DISPOSABLE BASE MOLDS M475-2 (Simport) containing 5% (w/v) agar (Nippon Gene). The molds were kept for at least 5 min to form agar blocks. Sections of 50 μ m thickness were then prepared using a Linear Slicer PRO7 (Dosaka-em). Tissues were sandwiched between a NEO MICRO cover glass (MATSUNAMI) and a slide glass (MATSUNAMI) for GFP fluorescence observation. During the observation, a water-immersion lens (x60) was used as the objective lens, and the chosen channel was the GFP channel (500–570 nm).

Fluorescence observation of the inflorescences was conducted using a LSM900 microscope (Carl Zeiss) as described previously⁴⁵. After cutting the tip of the primary inflorescence to a length of approximately 0.5 cm with a razor blade, the primary inflorescence was immersed in a 1 mg/ml solution of propidium iodide (PI) for 2 min to stain the cell walls. The stained inflorescences were then inserted upright into halfstrength MS medium, ensuring the plant surface was submerged in deionized water. After preparing the samples, fluorescence was promptly observed, and both 2D and 3D images were acquired. For observation, a water-immersion lens (x20) was used as the objective lens. In terms of the optical path setup, the Airy scan mode was selected, and the channels chosen were the GFP channel (500–570 nm) and the PI channel (590–700 nm). Additionally, the acquisition of 3D images involved moving the z-axis continuously to scan and create images.

Declarations

Data Availability

Protein structure data was obtained from Protein Data Bank (5VAC: <https://www.rcsb.org/structure/5vac>).

ChIP-seq and RNA-seq data were collected from the National Center for Biotechnology Information GEO database (CLF/H3K27me3 ChIP: GSE108960 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108960>]; AZF1/BPC1/H3K27me3 ChIP: GSE84483 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84483>], Pol II/Pol II-S2P/Pol II-S5P ChIP: GSE112443 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112443>], and *vip3-6/elf7* RNA-seq :GSE171778 [<https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE171778>]). The RNA-seq and ChIP-seq data have been deposited in the DDBJ database (DRA017906 [<https://ddbj.nig.ac.jp/resource/sra-submission/DRA017906>], DRA017907 [<https://ddbj.nig.ac.jp/resource/sra-submission/DRA017907>], DRA017938 [<https://ddbj.nig.ac.jp/resource/sra-submission/DRA017938>], and DRA017939 [<https://ddbj.nig.ac.jp/resource/sra-submission/DRA017939>]). Source data are provided with this paper.

Code Availability

Main source codes that are used in this work are available on the GitHub repository [<https://zefeng2018.github.io/Plant-Gene-Expression-Prediction/>]. Please refer to “Methods” above for details of key parameters and package versions.

Author contributions

Conceptualization: W.Y., M.A., Y.K., Y.X., T.I. and N.Y. (lead) and all other authors (supporting); data curation: N.Y.; formal analysis: W.Y., M.A., Y.K., T.S., K.I., T.T., S.I., T.S. and N.Y.; funding acquisition: T.I. and N.Y.; investigation: W.Y., M.A., Y.K., T.S., K.I. and N.Y. (lead) and T.T., S.I., T.S., T.K., Y.X. and T.I. (supporting); project administration: T.I. and N.Y.; software: N.Y.; supervision: T.I. and N.Y.; validation: N.Y.; visualization: N.Y.; writing: N.Y. (original draft) and all authors (review and editing).

Competing interests

The authors declare no competing interests.

Additional Declarations:

There is **NO** Competing Interest.

Acknowledgements

We thank Akie Takahashi, Hiroko Egashira, Mizuki Shinoda Taeko Kawakami and Rie Shimano for technical assistance, TAIR for providing *sdg7-2*, *sdg8-2*, *sdg8-3*, *sdg8-4*, *clf-28* and *DR5:GUS* seeds, Dr. Keiji Nakajima (Professor in Nara Institute of Science and Technology) for giving *proPLT2:PLT2-YFP*, *proSHR:SHR-GFP* and *proWOX5:NLS-GFP* and Dr. Masaki Ito (Professor in Kanazawa University), Dr. Caroline Dean (Professor in John Innes Centre) and Dr. Hidehiro Fukaki (Professor in Kobe University) for sharing *proCYCB1;2:CYCB1 ;2-YFP*, *proSDG8:SDG8-GFP* and *proLBD16:GUS* seeds, respectively. Computations were partially performed on the National Institute of Genetics (NIG) supercomputer at the Research Organization of Information and Systems, NIG. This work was supported by a grant from a JSPS KAKENHI Grant-in-Aid for Scientific Research B (No. 23H02503), a Grant-in-Aid for challenging Exploratory Research (No. 19K22431), a Grant-in-Aid for Transformative Research Area (A) (No. 21H05663, 23H04968), a grant from the Daiichi Sankyo Foundation of Life Science, a grant from the LOTTE Foundation, and a grant from Ohsumi Frontier Science Foundation to N.Y., a JSPS KAKENHI Grant-in-Aid for Scientific Research C (No. 22K06180) to TKT, and a JSPS KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas (No. 21K19266), a JSPS KAKENHI Grant-in-Aid for Scientific Research A (No. 20H00470), and a Grant-in-Aid for Transformative Research Area (A) (No. 22H05176) to T.I.

Note

This reviewed preprint has been updated to use the correct the author order.

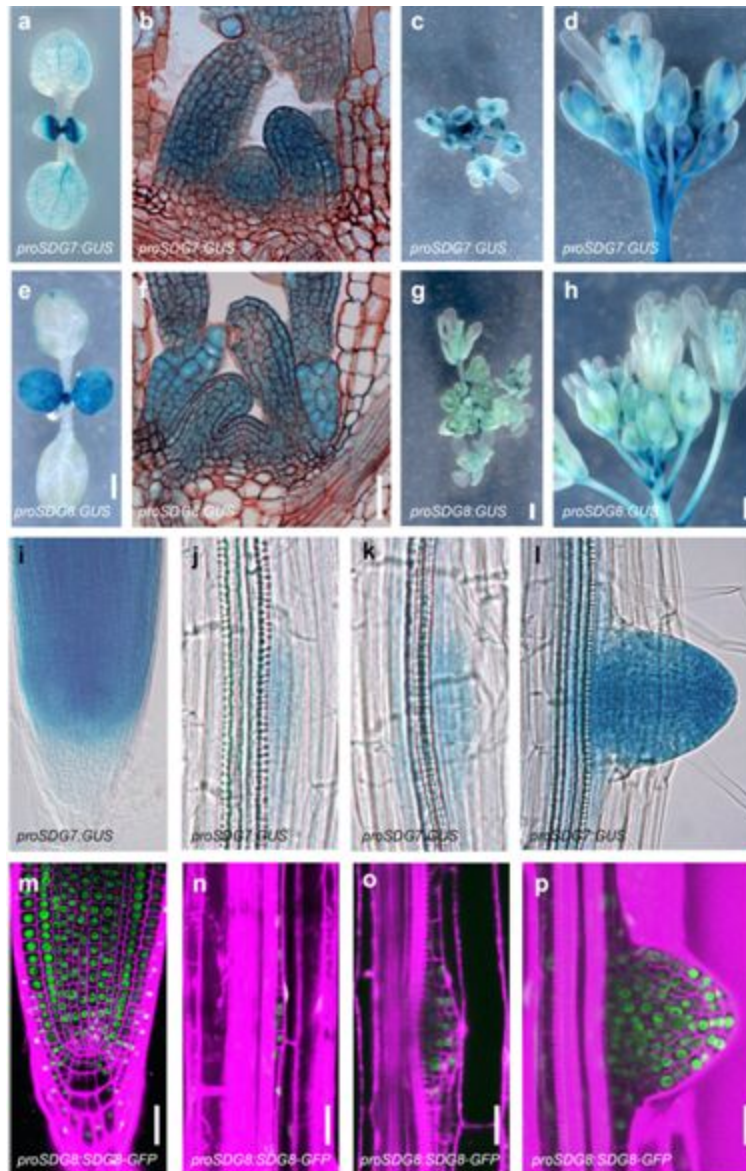


Fig. S1

Overlapping expression patterns of *SDG7* and *SDG8* in *Arabidopsis thaliana*.

a-h, Expression pattern of *SDG7* and *SDG8* using *proSDG7:SDG7-GUS* (a-d) and *proSDG8:GUS* (e-h) transgenic lines in cotyledons (a,e), shoot apical meristems (b,f), top and side views of primary inflorescences (c,g and d,h, respectively). Scale bars, 500 μ m (e, g, h) or 50 μ m (f). **i-p**, Expression pattern of *SDG7* and *SDG8* using *proSDG7:SDG7-GFP* (i-l) and *proSDG8:SDG8-GFP* (m-p) transgenic lines in primary root tips (i,m), stage II lateral roots (j,n), stage IV lateral roots (k,o) and emerged lateral roots (l,p). Scale bars, 50 μ m (m-p).

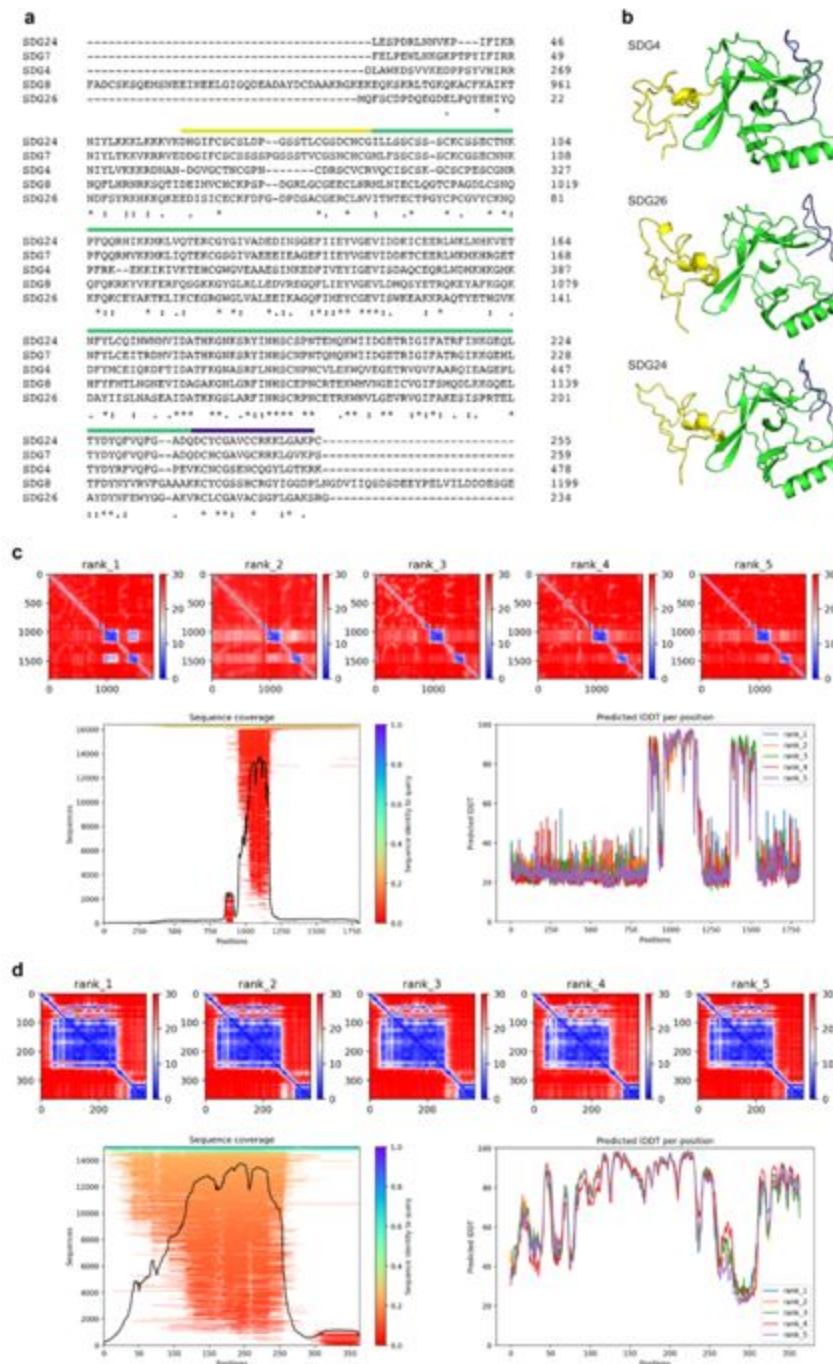


Fig. S2

Structural prediction of class II SDG family proteins.

a, Sequence alignment of the conserved domains in SDG7, SDG8 and other SDG members. Blue, green and red lines indicate the AWS, SET and post-SET domains, respectively. Asterisks show conserved amino acids. **b**, 3D structure prediction of the conserved domains of SDG4, SDG26 and SDG24 by AlphaFold 2. Blue, green and red lines indicate the AWS, SET and post-SET domains, respectively. **c,d**, Structure prediction of full-length SDG8 and SDG7 by AlphaFold 2. Top, Inter-chain predicted aligned error (PAE) plot for the best-ranked model structures of SDG7 (**c**) and SDG8 (**d**) by AlphaFold 2. Bottom left, Multiple Sequence Alignment (MSA) sequence coverage by AlphaFold 2. Bottom right, Predicted local distance difference test (LDDT) per position of the five models by AlphaFold 2.

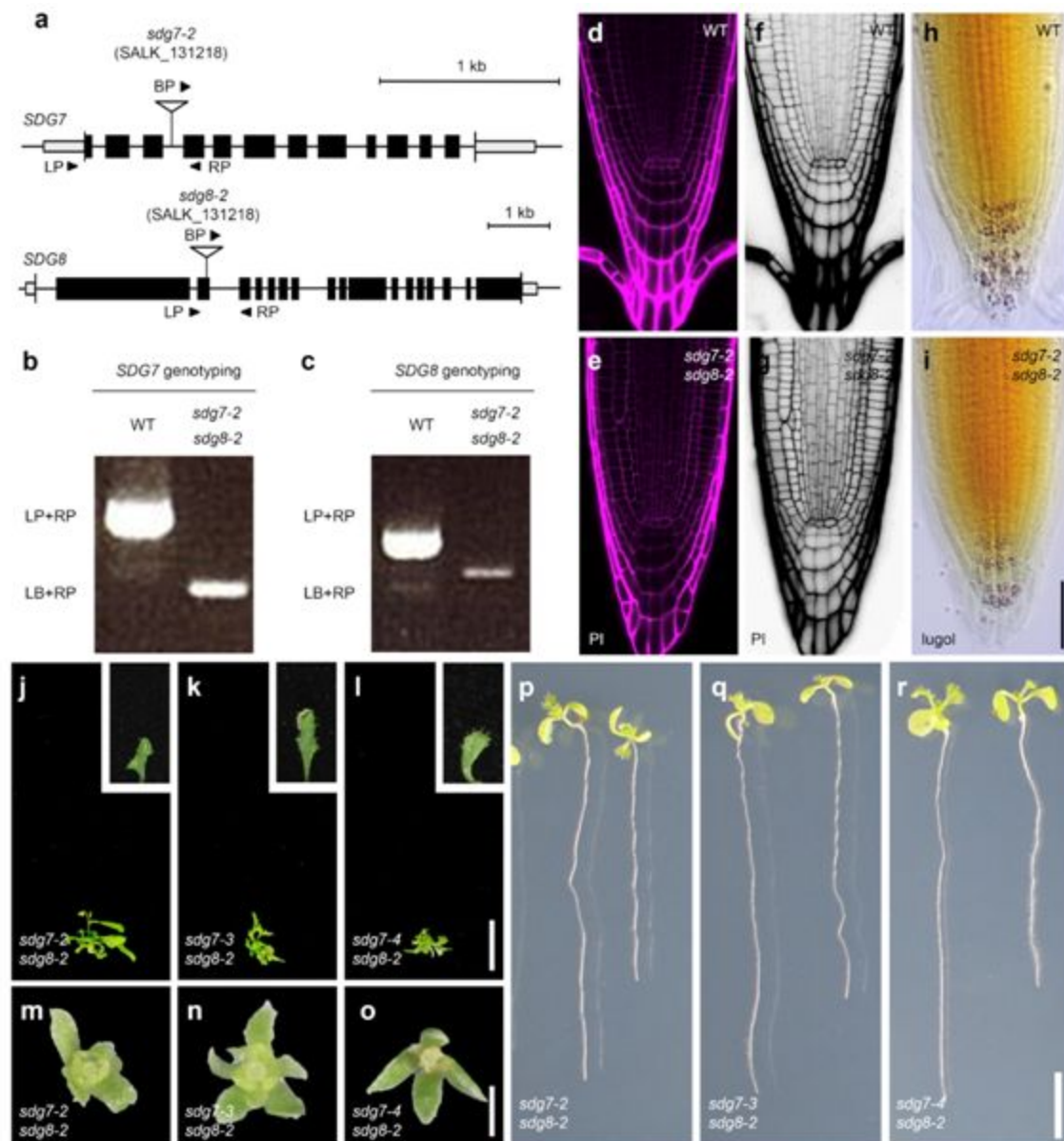


Fig. S3

Isolation of *sdg7 sdg8* double mutants.

a, Diagram of the *SDG7* and *SDG8* loci. The black boxes, gray boxes and short lines inside gene bodies indicate exons, untranslated regions, introns and qPCR-amplified regions, respectively. White triangles indicate the position of T-DNA insertions; black arrowheads denote genotyping primers (*sdg7-2* LP, *sdg7-2* RP, *sdg8-2* LP and *sdg8-2* RP); and white arrowheads denote RT-PCR primer pairs. **b**, **c**, Genotyping results of WT and *sdg7-2 sdg8-2* double mutant. **b**. *SDG7* genotyping. **c**. *SDG8* genotyping. **d-g**, Cellular organization of the root meristem, as visualized by propidium iodide (PI) staining in wild type (WT; **d**, **f**) and *sdg7 sdg8* (**e**, **g**). **h,i**, Cellular organization of the columella root cap, as visualized by Lugol staining in WT (**h**) and *sdg7 sdg8* (**i**). Scale bar, 50 μ m. **j-l**, Representative photographs of the *sdg7-2 sdg8-2* (**j**), *sdg7-3 sdg8-2* (**k**) and *sdg7-4 sdg8-2* (**l**) double mutants at 25 days after germination. Insets: top view of 1st true leaves. Scale bar, 1 cm. **m-o**, Top view of stage 14 flowers from soil-grown *sdg7-2 sdg8-2* (**m**), *sdg7-3 sdg8-2* (**n**) and *sdg7-4 sdg8-2* (**o**). Scale bar, 0.5 mm. **p-r**, Representative photograph of the roots from the *sdg7-2 sdg8-2* (**p**), *sdg7-3 sdg8-2* (**q**) and *sdg7-4 sdg8-2* (**r**) double mutants. Scale bar, 1 cm.

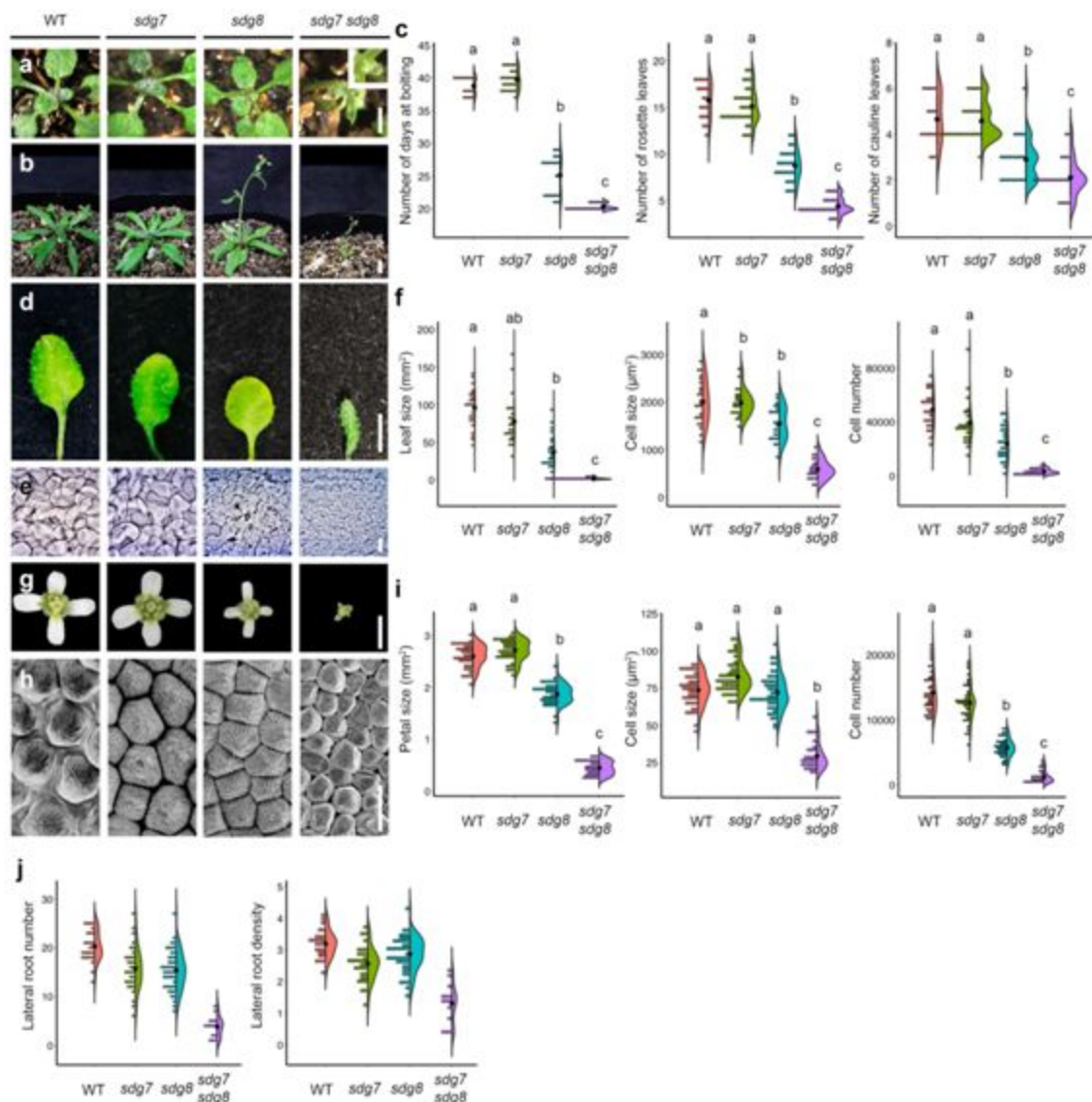


Fig. S4

Phenotypes of the *sdg7 sdg8* double mutant.

a, Top views of the wild type (WT), *sdg7*, *sdg8* and *sdg7 sdg8* at 10 days after germination. **b**, Side views of the WT, *sdg7*, *sdg8* and *sdg7 sdg8* at 25 days after germination. **c**, Number of rosette leaves (left) and cauline leaves (middle), and days to bolting (right) in the WT, *sdg7*, *sdg8* and *sdg7 sdg8* (WT: $n = 20$; *sdg7*: $n = 34$; *sdg8*: $n = 28$; *sdg7 sdg8*: $n = 26$). **d**, Top view of 5th true leaf in the WT, *sdg7*, *sdg8* and *sdg7 sdg8*. Scale bars (a, b, d), 1 cm. **e**, Appearance of palisade cells from 5th true leaves in the WT, *sdg7*, *sdg8* and *sdg7 sdg8*. Scale bar, 20 μm . **f**, Leaf size (left), cell size (middle), and cell number (right) in the WT, *sdg7*, *sdg8* and *sdg7 sdg8* (WT: $n = 24$; *sdg7*: $n = 24$; *sdg8*: $n = 24$; *sdg7 sdg8*: $n = 24$). **g**, Top view of stage 14 flowers from soil-grown WT, *sdg7*, *sdg8* and *sdg7 sdg8*. Scale bar, 0.5 mm. **h**, Petal epidermal cells in the WT, *sdg7*, *sdg8* and *sdg7 sdg8*. Scale bar, 10 μm . **i**, Petal area (left), petal cell number (middle), and petal cell size (right) in the WT, *sdg7*, *sdg8*, and *sdg7 sdg8* (WT: $n = 50$; *sdg7*: $n = 50$; *sdg8*: $n = 50$; *sdg7 sdg8*: $n = 40$). **j**, Lateral root number (left) and lateral root density (right) in the WT, *sdg7*, *sdg8* and *sdg7 sdg8* at 7 days after germination (WT: $n = 27$; *sdg7*: $n = 33$; *sdg8*: $n = 46$; *sdg7 sdg8*: $n = 15$). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on post-hoc Tukey's HSD test ($p < 0.05$; See the Data Source file for all combinations of the exact p -values).

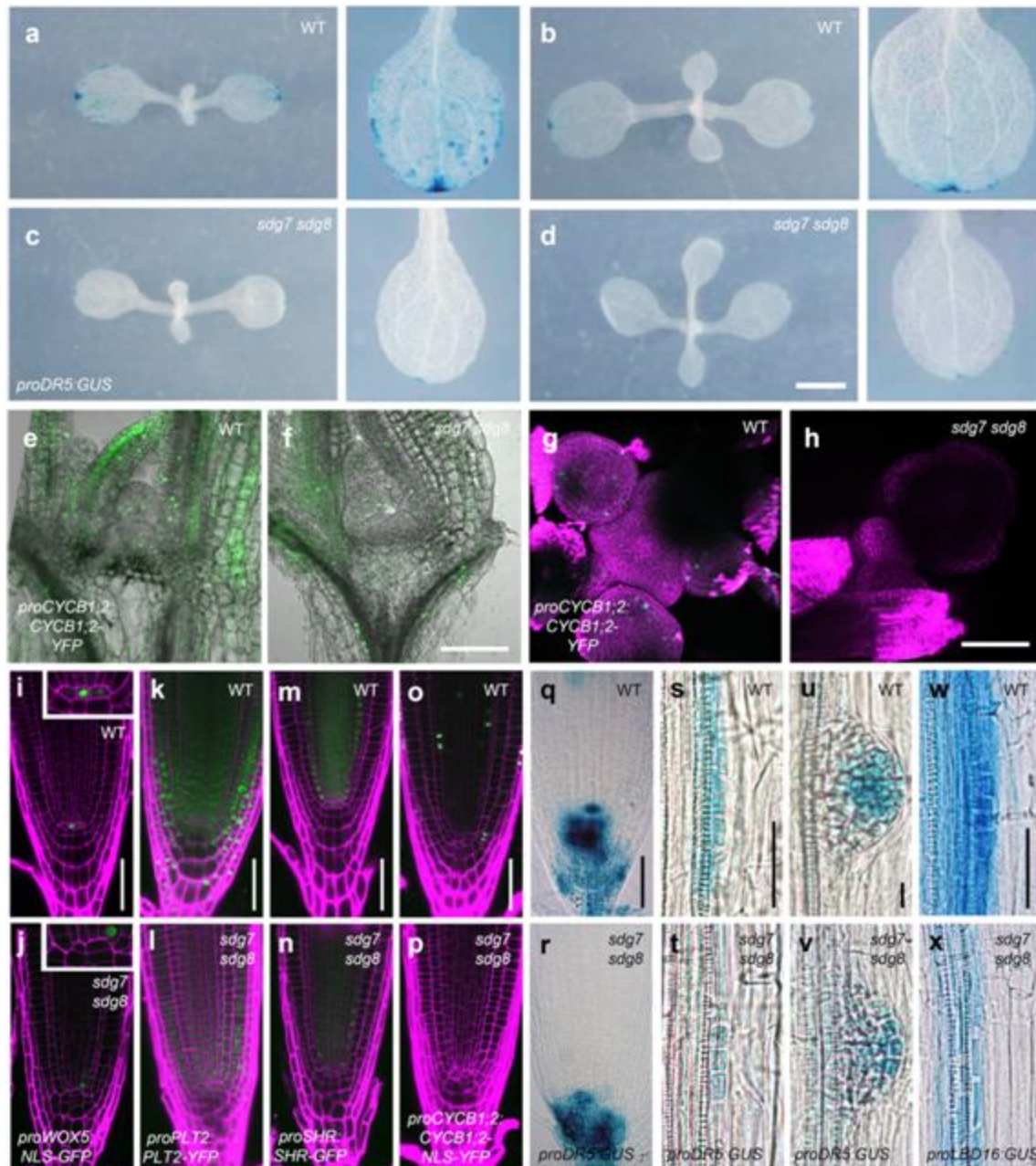


Fig. S5

Molecular defects of the *sdg7 sdg8* double mutants.

a–d, GUS staining pattern from the *DR5:GUS* reporter construct in the WT (**a**, **b**) and *sdg7 sdg8* (**c**, **d**). Seven-day-old (**a**, **c**) and 13-day-old (**b**, **d**) seedlings are shown. Left, top views of seedlings. Right, Close-up views of cotyledons. Scale bar, 3 mm. **e**, **f**, YFP fluorescence pattern from the *proCYCB1;2:CYCB1;2-YFP* transgene in the shoot apical meristems from WT (**e**) and *sdg7 sdg8* (**f**). Scale bar, 50 μ m. **g**, **h**, YFP fluorescence pattern from the *proCYCB1;2:CYCB1;2-YFP* transgene in the inflorescence meristems of WT (**g**) and *sdg7 sdg8* (**h**). Scale bar, 100 μ m. **i–r**, GFP fluorescence or GUS staining pattern from the indicated cell markers in the primary roots of WT (**i**, **k**, **m**, **o**, **q**) and *sdg7 sdg8* (**j**, **l**, **n**, **p**, **r**). *proWOX5:NLS-GFP* (**i**, **j**), *proPLT2:PLT2-YFP* (**k**, **l**), *proSHR:SHR-GFP* (**m**, **n**), *proCYCB1;2:CYCB1;2-NLS-YFP* (**o**, **p**) and *DR5:GUS* (**q**, **r**). **s–x** GUS staining pattern from the indicated cell markers in lateral roots of WT (**s**, **u**, **w**) and *sdg7 sdg8* (**t**, **v**, **x**). *DR5:GUS* (**s–v**) and *proLBD16:GUS* (**w**, **x**). Stage II (**s**, **t**, **w**, **x**) and Stage IV (**u**, **v**). Scale bars, 50 μ m.

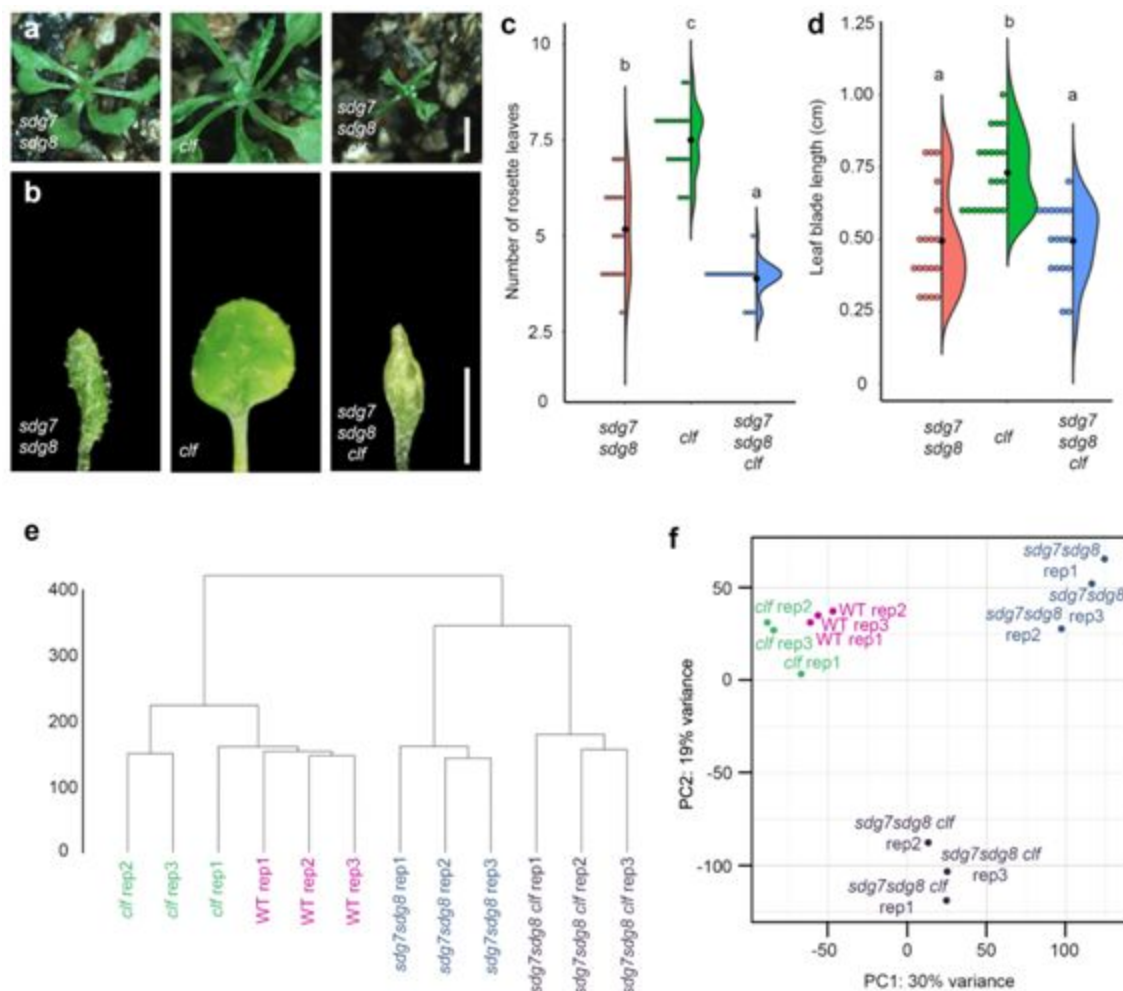


Fig. S6

Phenotypic and molecular defects in the *clf sdg7 sdg8* triple mutants.

a, Top views of the *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* at 20 days after germination. Scale bar, 1 cm. **b**, Top views of the 1st leaf in the *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* at 20 days after germination. Scale bar, 1 cm. **c**, **d**, Leaf number at flowering and size of 1st true leaf in the *sdg7 sdg8*, *clf* and *sdg7 sdg8 clf* (*sdg7 sdg8*: *n* = 18; *clf*: *n* = 20; *sdg7 sdg8 clf*: *n* = 17). Significant differences were calculated based on one-way ANOVA tests ($p = 1.1 \times 10^{-16}$). Different letters indicate significant differences based on post-hoc Tukey's HSD test ($p < 0.05$; See the Data Source file for all combinations of the exact *p*-values). **e**, Hierarchical clustering. The dendrogram of all RNA-seq samples obtained with the R package stat is shown with the bootstrap values on each node. **f**, Principal component analysis (PCA). The PCA plot illustrates all twelve RNA-seq samples, with replicates represented by the same color.

References

1. Liu C., Lu F., Cui X., Cao X. (2010) **Histone methylation in higher plants** *Annu Rev Plant Biol* **61**:395–420
2. Xiao J., Lee U. S., Wagner D. (2016) **Tug of war: adding and removing histone lysine methylation in Arabidopsis** *Curr Opin Plant Biol* **34**:41–53
3. Yang H., Howard M., Dean C. (2014) **Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC** *Curr Biol* **24**:1793–1797
4. Förderer A., Zhou Y., Turck F. (2016) **The age of multiplexity: recruitment and interactions of Polycomb complexes in plants** *Curr Opin. Plant Biol* **29**:169–178
5. Xiao J., Wagner D. (2015) **Polycomb repression in the regulation of growth and development in Arabidopsis** *Curr Opin. Plant Biol* **23**:15–24
6. Mozgova I., Hennig L. (2015) **The polycomb group protein regulatory network** *Annu. Rev. Plant Biol* **66**:269–296
7. De Lucas M., et al. (2016) **Transcriptional Regulation of Arabidopsis Polycomb Repressive Complex 2 Coordinates Cell-Type Proliferation and Differentiation** *Plant Cell* **28**:2616–2631
8. Chanvivattana Y., et al. (2004) **Interaction of Polycomb-group proteins controlling flowering in Arabidopsis** *Development* **131**:5263–5276
9. Xiao J., et al. (2017) **Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis** *Nat Genet* **49**:1546–1552
10. Zhou Y., et al. (2018) **Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis** *Nat Genet* **50**:638–644
11. Li Y., et al. (2015) **The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants** *Genome Biol* **16**
12. Yang H., Howard M., Dean C. (2016) **Physical coupling of activation and derepression activities to maintain an active transcriptional state at FLC** *Proc Natl Acad Sci U S A* **113**:9369–9374
13. Nasim Z., Susila H., Jin S., Youn G., Ahn J. H. (2022) **Polymerase II-Associated Factor 1 Complex-Regulated FLOWERING LOCUS C-Clade Genes Repress Flowering in Response to Chilling** *Front Plant Sci* **13**
14. Zhao W., Neyt P., Van Lijsebettens M., Shen W. H., Berr A. (2019) **Interactive and noninteractive roles of histone H2B monoubiquitination and H3K36 methylation in the regulation of active gene transcription and control of plant growth and development** *New Phytol* **221**:1101–1116

15. Lee J., Yun J. Y., Zhao W., Shen W. H., Amasino R. M. (2015) **A methyltransferase required for proper timing of the vernalization response in Arabidopsis** *Proc Natl Acad Sci U S A* **112**:2269–74
16. Baumbusch L. O., et al. (2001) **The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes** *Nucleic Acids Res* **29**:4319–4333
17. Bergamin E., et al. (2017) **Molecular basis for the methylation specificity of ATXR5 for histone H3** *Nucleic Acids Res* **45**:6375–6387
18. Petrella R., et al. (2020) **BPC transcription factors and a Polycomb Group protein confine the expression of the ovule identity gene SEEDSTICK in Arabidopsis** *Plant J* **102**:582–599
19. Shafiq S., Berr A., Shen W. H. (2014) **Combinatorial functions of diverse histone methylations in Arabidopsis thaliana flowering time regulation** *New Phytol* **201**:312–322
20. Stortenbeker N., Bemer M. (2019) **The SAUR gene family: the plant's toolbox for adaptation of growth and development** *J Exp Bot* **70**:17–27
21. Cao X., et al. (2019) **The Roles of Auxin Biosynthesis YUCCA Gene Family in Plants** *Int J Mol Sci* **20**
22. Mishra P., Panigrahi K. C. (2015) **GIGANTEA - an emerging story** *Front Plant Sci* **6**
23. Cazzonelli C. I., et al. (2014) **A chromatin modifying enzyme, SDG8, is involved in morphological, gene expression, and epigenetic responses to mechanical stimulation** *Front Plant Sci* **5**
24. Christensen S. K., Dagenais N., Chory J., Weigel D. (2000) **Regulation of auxin response by the protein kinase PINOID** *Cell* **100**:469–4678
25. Whittaker C., Dean C. (2017) **The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation** *Annu Rev Cell Dev Biol* **33**:555–575
26. Yu X., Martin P. G. P., Michaels S. D. (2019) **BORDER proteins protect expression of neighboring genes by promoting 3' Pol II pausing in plants** *Nat. Commun* **10**
27. Kim E. D., et al. (2022) **Dynamic chromatin accessibility deploys heterotypic cis/trans-acting factors driving stomatal cell-fate commitment** *Nat. Plants* **8**:1453–1466
28. Pelayo M. A., et al. (2023) **AGAMOUS regulates various target genes via cell cycle-coupled H3K27me3 dilution in floral meristems and stamens** *Plant Cell* **35**:2821–2847
29. Xu Y., et al. (2014) **Arabidopsis MRG domain proteins bridge two histone modifications to elevate expression of flowering genes** *Nucleic Acids Res* **42**:10960–10974
30. Okushima Y., Fukaki H., Onoda M., Theologis A., Tasaka M. (2007) **ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis** *Plant Cell* **19**:118–130
31. Santuari L., et al. (2016) **The PLETHORA Gene Regulatory Network Guides Growth and Cell Differentiation in Arabidopsis Roots** *Plant Cell* **28**:2937–2951

32. Gallagher K. L., Paquette A. J., Nakajima K., Benfey P N. (2004) **Mechanisms regulating SHORT-ROOT intercellular movement** *Curr Biol* **14**:1847–1851
33. Ulmasov T., Murfett J., Hagen G., Guilfoyle T. J. (1997) **Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements** *Plant Cell* **9**:1963–1971
34. Pi L., et al. (2015) **Organizer-Derived WOX5 Signal Maintains Root Columella Stem Cells through Chromatin-Mediated Repression of CDF4 Expression** *Dev. Cell* **33**:576–588
35. Iwata E., et al. (2011) **GIGAS CELL1, a novel negative regulator of the anaphase-promoting complex/cyclosome, is required for proper mitotic progression and cell fate determination in Arabidopsis** *Plant Cell* **23**:4382–4393
36. Shu J., et al. (2019) **Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings** *Plant Direct* **3**
37. Karimi M., Inzé D., Depicker A. (2002) **GATEWAY vectors for Agrobacterium-mediated plant transformation** *Trends Plant Sci* **7**:193–195
38. Bent A. (2006) **Arabidopsis thaliana floral dip transformation method** *Methods Mol Biol* **343**:87–103
39. Jumper J., et al. (2021) **Highly accurate protein structure prediction with AlphaFold** *Nature* **596**:583–589
40. Yamaguchi N., et al. (2021) **H3K27me3 demethylases alter HSP22 and HSP17.6C expression in response to recurring heat in Arabidopsis** *Nat Commun* **12**
41. Kolberg L., et al. (2023) **g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update)** *Nucleic Acids Res* **51**:W207–W212
42. Tian T., et al. (2017) **agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update** *Nucleic Acids Res* **45**:W122–W129
43. Supek F., Bošnjak M., Škunca N., Šmuc T. (2011) **REVIGO summarizes and visualizes long lists of gene ontology terms** *PLoS One* **6**
44. Yamaguchi N., et al (2014) **PROTOCOLS: Chromatin Immunoprecipitation from Arabidopsis Tissues** *Arabidopsis Book* **12**
45. Reddy G. V., Roy-Chowdhury A. (2009) **Live-imaging and image processing of shoot apical meristems of Arabidopsis thaliana** *Methods Mol Biol* **553**:305–16

Editors

Reviewing Editor

Pablo Manavella

Universidad Nacional del Litoral-CONICET, Santa Fe, Argentina

Senior Editor

Jürgen Kleine-Vehn

University of Freiburg, Freiburg, Germany

Reviewer #1 (Public Review):

Summary:

In this manuscript, the authors reveal a new role for SDG7 in the regulation of H3K36me2 and me3. SDG7 appear to be functionally redundant to SDG8 as the double mutant presents lower levels of H3K36me and stronger phenotypes than either single mutant, however, their mechanisms of action might differ as the proteins displayed different localization on their target genes, with SDG7 localizing preferentially to TSS and TES while SDG8 covers the gene body. SDG7 binds preferentially to PREs, which recruit PRC2 for H3K27me3 deposition. The authors therefore present an interesting model where SDG7 evicts PRC2 from silenced genes, leading to a loss of H3K27me3. This would allow the transcriptional activation of the genes and the deposition of H3K36me3.

Strengths:

Overall, the manuscript is well-written and organized, although some paragraphs need clarifications. The figures are clear and well designed and the proposed model is compelling. While the manuscript is already interesting as it is, I think addressing the following questions would elevate it even more and refine the proposed model:

Weaknesses/potential aspects to address:

(1) It is still unclear whether SDG7 directly catalyzes H3K36me or if it promotes its deposition simply by eviction of PRC2. The AlphaFold and structure analyses show a significant similarity between the catalytic domains which would support the first possibility, but some more experiments would be required to prove this more definitively.

(2) Does SDG7 directly recognize the PRE (as suggested by the model in Figure 5F) or is it recruited by some transcription factors? Is SDG7 known to interact with any of the PRC2 recruiters?

(3) Line154/Figure 2A: The metagene plot for H3K36me3 shows a lower level on the gene body but a higher peak in *sdg7sdg8* double mutants compared to the Wild-type, which is a bit surprising, especially considering that the immunostaining in reference 19 showed a near complete loss of H3K36me3 signal in the same double mutant. Can this higher peak be an artifact from the normalization strategy, or due to the existence of different subpopulations of genes?

Indeed, on the genome tracks presented by the authors, the hypomethylated genes show a loss of signal on the entire gene body, and not a higher peak near the TSS. It might be interesting to generate metagene plots for H3K36me3 hypo and hyper-methylated genes, to see if the higher peak at the TSS is solely due to the hyper-methylated genes.

(4) Figure 2C: More than 40% of differentially methylated genes are actually hypermethylated, but the authors do not discuss this at all. What are those genes, are they targeted by SDG7 or 8? Could they be responsible for the higher peak at the TSS observed in the double mutant? (see previous comment).

(5) Figure 2C and D: The method section states that the ChIP-seq was performed on 5-day-old seedlings, while the legend of this figure mentions root and shoot samples but this does not appear in the figure itself. There is also mention of shoot and root samples in Supplementary

Tables 1 and 2. The authors should clarify which tissue was used for the data presented in Figure 2 and correct the legends or the methods accordingly.

(6) Line 270/Fig 4K and L: The text mentions looking at the 838 genes "downregulated in *clf* *sdg7* *sdg8* relative to *sdg7* *sdg8*" and in the overlap, the authors identified FLC. However, in Figure 5D, FLC is upregulated in *clf* *sdg7*-*sdg8* compared to *sdg7*-*sdg8*, not downregulated as mentioned in line 270. The Venn diagram in Figure 4L mentions "*sdg* vs *clf* *sdg* up", which would fit the pattern seen in Figure 5D, but the number of genes (838) matches the number of downregulated genes in the *sdg7**sdg8* vs *clf* *sdg7**sdg8* volcano plot.

I would actually expect the phenotype rescue to be caused by genes that are up in Wt vs *clf*, down in Wt vs *sdg7*-*sdg8*, and back up in *sdg7*-*sdg8* vs *clf*-*sdg7*-*sdg8*, not "up/down/down" as mentioned in the text: genes would be downregulated in *sdg7*-*sdg8* because of a loss of H3K36me and therefore hypermethylation of H3K27, but in the absence of CLF, this hypermethylation is reversed and the genes are upregulated in the triple mutant compared to the *sdg7*-*sdg8* mutant. This is also what the authors see and describe in their cluster analysis in Figure 4M and line 280, mentioning an upregulation in *clf*-*sdg7*-*sdg8* vs *sdg7*-*sdg8*. Could the authors please clarify these discrepancies between the different subplots and within the text itself? Was there maybe some error plotting the volcano plot and/or Venn diagram?

In general, as this part is quite complicated, maybe it would benefit from a clearer explanation from the authors as to why they look at those particular overlaps, so that the reader can more easily follow their train of thought.

(7) Figure 4N/Line 286: How were these 828 genes identified? Is it stemming from a *clf*-*sdg7*-*sdg8* vs *sdg7*-*sdg8* comparison? The legend says "genes shown by white color in Fig. 4M", do the authors mean the two clusters previously described?

(8) Line 300: "suggesting that SDG8 primarily mediates target gene expression in conjunction with PAF1C". This statement is based on overlapping genes that are downregulated in *sdg7*-*sdg8* double mutant and *paf1c* mutants but concludes only on the role of SDG8. I feel that to state that SDG8 regulates expression in conjunction with PAF1C, the authors should rather examine the genes downregulated in the *sdg8* mutant, especially considering the reduced overlap between genes downregulated in *sdg8* and *sdg7*-*sdg8* (according to Figure 2C, only 30% of the genes downregulated in *sdg8* are also downregulated in the double mutant), or this statement should be corrected to also include SDG7.

Maybe it would be easier to read the figure if the authors created a master list of genes downregulated in at least one of the *paf1c* mutants they examined (as they anyway do not examine in detail the contribution of each individual *paf1c* mutant), and overlap it with the genes downregulated in *sdg7*, *sdg8* or *sdg7*-*sdg8*.

(9) Line 326: "We also discovered that SDG7 and SDG8 overcome PRC2-mediated silencing, leading to a switch from H3K27 methylation to H3K36 methylation during growth and development." While part of this statement is supported by the ChIP data presented in Figure 4E, I think a ChIP for H3K36me₂ and/or me₃ is necessary to prove the existence of a K27me to K36me switch.

(10) Line 347: The authors state that SDG8 is located at the TSS and 3' end of genes, but on line 187 they state that it occupies the gene body (which is supported by the plot in Figure 3A).

(11) Line 351: The authors suggest a role of RNAPII in the deposition of K36me, but their data are not sufficient to support this hypothesis. The transcriptome data show that both SDGs and PAF1C regulate a similar set of genes, but they do not show data demonstrating that RNAPII is necessary for the deposition of K36me. It might be interesting to examine H3K36me levels in a *paf1c* mutant to further consolidate their hypothesis.

Reviewer #2 (Public Review):

Summary:

In this manuscript, the authors combined imaging approaches with molecular and genetic experiments to

- (i) for the first time establish a chromatin regulatory role for SDG7;
- (ii) examine its role in H3K36 methylation, along with its homolog SDG8;
- (iii) examine its potential role in mediating antagonism to PRC2, thereby mediating transcriptional activation.

Strengths:

The manuscript explores interesting and relevant mechanistic hypotheses about chromatin-mediated gene regulation by combining a range of experimental tools and a genome-wide perspective. The writing is very clear. The study makes good connections to existing data and generates datasets that complement existing datasets, providing a valuable resource to the community.

Weaknesses:

Some of the claims appear to need further supporting evidence to establish their robustness.

Review of the main conclusions and supporting evidence:

(1) SDG7 contributes to H3K36 methylation at several loci along with SDG8:

This conclusion is supported by the genome-wide differences (measured by ChIP-seq) in H3K36me3 and me2 levels observed between WT, the *sdg7* and *sdg8* single mutants, and the double mutant. The reduction in H3K36 methylation levels observed at a large number of genes in the *sdg7* mutant, and the further reduction in H3K36 methylation levels in the double mutant compared to *sdg8* (observed at multiple genes) indicates that SDG7 promotes this transcription-associated modification. The significantly larger number of H3K36 hypomethylated genes in the double mutant (3380) compared to either of the single mutants (523 and 605) indicates lower overall H3K36me3 levels in the double mutant. While direct evidence of SDG7 methyltransferase activity on H3K36 is lacking, the study reports structural comparisons using AlphaFold predictions that suggest the possibility of such a role.

(2) SDG7 influences gene expression together with SDG8:

This conclusion is supported by a range of phenotypic differences observed between *sdg8* and the double mutant *sdg7 sdg8*. The reported genome-wide differential gene expression between the WT and the double mutant as well as expression differences in specific genes detected by imaging are also consistent with SDG7 and SDG8 together influencing gene expression. However, a role specifically for SDG7 could be further supported by a direct differential expression analysis between the single mutant *sdg8* and the double mutant *sdg7 sdg8*. A role for SDG7 in gene expression is also consistent with its observed effect on H3K36 methylation, which is generally associated with productive transcription.

(3) SDG8 is exclusively nuclear, but SDG7 localises to the cytosol and the nucleus in meristematic cells:

This conclusion is supported by imaging of fluorescent-tagged SDG8 and SDG7 in the root tip, which shows nuclear localisation of SDG8-GFP, consistent with previous reports, and a combination of nuclear and cytosolic localisation for SDG7-VENUS. However, the study

presents only one replicate - more replicates are needed to establish the robustness of this conclusion.

(4) Distinct binding patterns of SDG7 and SDG8 on chromatin:

This conclusion is supported by the analysis of genome-wide binding patterns of these proteins measured by ChIP-seq (using fluorescent tagged versions). This data indicates that while SDG7 tends to localise to TSS and TES regions of genes, SDG8 tends to be more uniformly spread across gene loci. These patterns suggest that SDG7 and SDG8 may influence H3K36 methylation through distinct mechanisms, but do not indicate what these mechanisms may be.

(5) SDG7/SDG8 antagonism with PRC2:

a. SDG7 overlaps with PRC2 and its recruiters on chromatin and can bind to PREs

This conclusion is supported by statistically significant overlaps genome-wide between cis-elements at SDG7 binding peaks and those previously reported to be Polycomb associated. This is further supported by the observation of SDG7 binding peaks close to PRC2 subunit binding peaks at known PRC2 targets.

b. SDG7/SDG8 antagonism with PRC2

This conclusion is supported by the observation that the *sdg7 sdg8* double mutant phenotypes are partially rescued by disrupting CLF, one of the PRC2 methyltransferases. However, this does not necessarily suggest a direct antagonism with PRC2, but could be part of a more general antagonism between productive transcription (in which H3K36 methylation plays an active role), and PRC2-mediated silencing.

c. SDG7 can evict PRC2 from PREs to overcome H3K27me₃-mediated silencing

This conclusion - that SDG7 directly antagonises PRC2 interaction with cis-elements that mediate its targeting - is partly supported by the observation that inducing overexpression of SDG7 (through dexamethasone induction) can cause changes in CLF occupancy and H3K27me₃ levels at certain designated PREs. To establish the robustness of these conclusions, it will be necessary to examine designated regions to be used as a negative control and include a non-transgenic control for the SDG-7-HA ChIP. A suitable negative control may be a non-PRE region known to have a CLF peak and high H3K27me₃ levels, where the levels would not be expected to change in this experiment.

It remains to be established whether this apparent antagonism of PRC2 by SDG7 is only part of a more general antagonism of PRC2 by productive transcriptional activity, or whether SDG7 can evict PRC2 from PREs independent of transcripts and therefore is a precursor to switching to an active transcriptional state.

Another aspect that would be interesting to examine in the light of recent findings is the single-cell-level behaviour at these genes targeted by SDG7-to examine whether SDG7 induction is causing individual copies of these genes to stochastically switch to an active transcriptional state so that the observed changes in H3K27me₃ result from a fraction of copies completely losing silencing rather than all copies losing some H3K27me₃.

(6) H3K36me₃ levels are co-regulated by SDG8 and Paf1c, SDG8 associates with Pol II to deliver transcription-coupled H3K36 methylation

This conclusion is supported by analysis of a subset of loci previously reported to be regulated by Paf1c, which also exhibit changes in the *sdg7 sdg8* double mutant as well as a *clf* mutant. This analysis is based on a qualitative examination of ChIP-seq signal at a small number of loci, and therefore provides indirect support for this conclusion. It is, therefore, difficult to draw mechanistic insights from this analysis that go beyond correlation.

<https://doi.org/10.7554/eLife.100905.1.sa0>