TITLE: Dynamic DNA methylation turnover in gene bodies is associated with enhanced gene expression plasticity

Authors: Williams, Clara J.[1], Dai, Dawei[1], Tran, Kevin A.[1], Monroe, J. Grey[2], Williams, Ben P[1].

[1]Department of Plant & Microbial Biology, University of California, Berkeley.

[2]Department of Plant Sciences, University of California, Davis.

Correspondence: benwiliams@berkeley.edu

ABSTRACT

Background: In diverse eukaryotes, DNA methylation is deposited within the coding regions of many genes, termed gene body methylation (GbM). Whereas the role of DNA methylation on the silencing of transposons and repetitive DNA is well understood, gene body methylation is not associated with transcriptional repression, and its biological importance has remained unclear.

Results: We report a newly discovered type of GbM in plants, which is under constitutive addition and removal by dynamic methylation modifiers in all cells, including the germline. Methylation at Dynamic GbM genes is removed by the DRDD demethylation pathway, and added by an unknown source of de novo methylation, most likely the maintenance methyltransferase MET1. We show that the Dynamic GbM state is present at homologous genes across divergent lineages (>100 million years), indicating evolutionary conservation. Furthermore, we demonstrate that Dynamic GbM is tightly associated with the presence of a promoter/regulatory chromatin state within the gene body, in contrast to other gene body methylated genes. Lastly, we find Dynamic GbM is associated with enhanced gene expression plasticity across development and diverse physiological conditions, whereas stably methylated GbM genes exhibit reduced plasticity.

Conclusions: This study proposes a new model for GbM in regulating gene expression plasticity, including a newly discovered type of GbM in which increased gene expression plasticity is associated with the activity of DNA methylation writers and erasers and the enrichment of a regulatory chromatin state. We propose that Dynamic GbM is a hallmark of genes that can navigate a greater landscape of transcriptional states by maintaining regulatory flexibility.

KEYWORDS

DNA methylation, DNA demethylation, gene body methylation, gene expression plasticity.

INTRODUCTION

In diverse eukaryote species, DNA methylation is important for transcriptional silencing of transposable elements [1,2], repressing recombination at repetitive regions [3] and establishing genomic imprinting [4,5]. DNA methylation is a reversible and dynamic mark, which can be added or removed by a number of writer or eraser enzymes [6]. Many of these writer pathways have evolved to promote highly stable inheritance of DNA methylation over cell divisions and reproductive generations [7,8]. For example, the enzyme MET1 (also known as DNMT1 in animals) acts on hemimethylated DNA at the replication fork, faithfully copying existing methylation patterns to the daughter strand at symmetrical CG sites [6]. Consequently, the vast majority of methylated CG sites are highly stable and consistently methylated across all cells [9]. In plants, many additional writers also maintain and methylate DNA de novo at CG, CHG and CHH sites, including the CHROMOMETHYLASE (CMT) family and the DOMAINS-REARRANGED METHYLTRANSFERASE (DRM) family [6]. Additionally, a family of plant-specific DNA glycosylase enzymes (collectively named DRDD enzymes) act as methylation erasers, removing DNA methylation from thousands of loci within the genome [10]. This family is comprised of the enzymes DEMETER (DME) [11], REPRESSOR OF SILENCING 1 (ROS1) [12], and DEMETER-LIKE2&3 (DML2/DML3) [13]. The DRDD pathway is critical for reproductive development [11,14,15], and also maintains epigenetic homeostasis by protecting transcribed genes from silencing by pathways targeting proximal repetitive DNA [12,16–20]. Additionally, the DRDD pathway can also act during somatic development to generate differences in the levels of DNA methylation between tissues [10].

These methylation writers and erasers together shape DNA methylation patterns across the genome, into two broad categories: 1) “heterochromatic” DNA methylation, in which cytosines are methylated in CG, CHG and CHH sequence contexts, abundant at repetitive DNA, heterochromatin and some intergenic sequences within euchromatin [6], and 2) “gene body” DNA methylation (GbM), in which the coding region of genes are methylated solely in the CG sequence context [9]. Whereas the function of heterochromatic DNA methylation is broadly understood, primarily functioning in the transcriptional silencing of repetitive DNA [1,21,22], or the potent regulation of certain specialized target genes

[17,18,23,24], the functional role of GbM has remained more enigmatic [9,25,26]. GbM is associated with moderate-to-highly expressed genes [26,27], is inherited with high accuracy [28] and is conserved at homologous target genes across lineages [25]. However, a mechanistic link between GbM and gene regulation has been difficult to establish. Some lineages have dispensed with GbM altogether [9], raising questions about its necessity for gene regulation. Indeed, DNA methylation writer mutants or naturally-occurring genomes with reduced GbM do not display widespread changes to the transcriptional output of GbM genes, or strongly perturbed phenotypes [27,28]. Why GbM occurs in so many eukaryotic genomes, and at conserved genes in distinct lineages is therefore yet to be fully resolved.

The majority of GbM genes in the genome exhibit consistent methylation of CGs within all cells. In this study, we show that a subset of GbM genes exhibit heterogeneous methylation patterns between cells, indicating dynamic modification to methylation during development. This dynamic methylation heterogeneity is due to the activity of both writer and eraser pathways in all cell and tissue types, is evolutionarily conserved, and represents a newly discovered epigenetic state associated with regulatory histone modifications not typically found within gene bodies. Lastly, we show that Dynamic GbM genes exhibit dramatically higher gene expression plasticity than their stably methylated counterparts, suggesting that active modification to GbM may permit a greater exploration of the gene expression landscape.

RESULTS

DNA demethylation targets a subset of gene bodies.

We recently generated a quadruple somatic mutant (*drdd*) of the four active DNA demethylases (DRDD enzymes) in the model plant *Arabidopsis thaliana* [10]. Unlike previously studied DNA demethylation mutants, *drdd* mutants exhibited increased CG methylation at a subset of gene bodies. To understand this genic demethylation further, we compared the 581 genes most evidently targeted by DRDD to previously described gene-body methylated (GbM) genes [9] (Fig. 1A). These genes were identified

based on two cutoffs: the presence of a WT vs *drdd* differentially methylated region (DMR) within the gene body (see methods), and an additional filter requiring at least 5 CG dinucleotides (10 CG cytosines total) that gained >20% methylation in *drdd* mutants. We found that the genes targeted by DRDD were a mostly distinct group compared to previously identified GbM genes (Fig. 1B), exhibiting a number of different properties. Previously identified GbM genes exhibit high levels of CG methylation throughout the gene body (Fig. 1A, C), and each individual CG is consistently highly methylated across the majority of cells (Fig. 1A, E), much like the vast majority of methylated CGs in the genome (Fig. 1D). As the methylation state of these methylated genes are stable across development, we hereafter refer to these genes as Stable GbM genes. In contrast, genes targeted by DRDD display reduced methylation overall, in large part due to the fact that the majority of methylated CGs are methylated in only a *subset* of WT cells (in *drdd* mutants, these CGs are methylated in all cells, much like Stable GbM genes (Fig. 1A,E)). We refer to these genes as Dynamic GbM genes, due to the active removal of methylation that is occurring during plant development. The intermediate methylation state of CGs within Dynamic GbM genes (methylated in 10-85% of cells) is unusual and distinct from the vast majority of the genome (Fig. 1D-E), as the methylation maintenance by MET1 typically copies existing methylation states during DNA replication to ensure consistency across cell populations.

DNA demethylation generates cellular heterogeneity in all tissues and cell types.

As Dynamic GbM genes exhibit DNA demethylation in a subset of cells, we sought to determine the underlying developmental context of these methylation dynamics. It is possible that DRDD removes methylation in a specific subset of differentiated cell types within somatic tissues, or that the reduced methylation is developmentally specific to the tissue type (rosette leaves) of our sample. To test this, we assembled whole-genome methylation sequencing data from a number of studies that have defined the methylation states of a number of cell and tissue types within Arabidopsis [10,29–32], and examined the cellular heterogeneity of DNA methylation patterns of both Stable GbM and Dynamic GbM genes. To our surprise, we discovered that the intermediate methylation of Dynamic GbM genes was not specific to active demethylation of a specific tissue or cell type (Fig. 2A-B) within somatic

tissues, but rather a feature of all the tissue and cell types we examined. This suggested that the removal of methylation by DRDD is ubiquitous across somatic development. By contrast, and as expected, Stable GbM genes were consistently methylated in the majority of somatic cells (Fig. 2B). As Dynamic GbM genes show cellular heterogeneity in DNA methylation patterns (a signature of active demethylation) in all tissues, we next wished to examine if the stem cell population in shoot meristems and germline cells also display cellular heterogeneity, or if the Dynamic GbM state is restricted to somatic development. By analyzing the cellular heterogeneity of CG methylation in shoot apical meristem stem cells (labelled with the *CLV3* promoter [32]), meiocyte cells from the male gametophyte [31] and sperm cells [30], we clearly observe intermediate methylation of CG sites across cells in all cases (Fig. 2C-D). To confirm that the cellular heterogeneity of CG methylation was due to the active removal by the DRDD pathway, we next examined DNA methylation in sperm nuclei mutant for one or both of the major DRDD genes *DME* and *ROS1* [30]. In demethylase mutant sperm, Dynamic GbM CG methylation patterns were homogenous across 100% of cells, conclusively demonstrating that heterogeneity between cells in WT sperm is due to active demethylation by DME and ROS1 (Fig. 2C). Together, these data demonstrate that Dynamic GbM loci are actively demethylated in a subset of all cells in somatic, meristem and germline tissues, and consequently are not demethylated in a specific lineage or developmental context.

De novo methylation is required to maintain dynamic GbM

As Dynamic GbM genes are demethylated in the germline, we reasoned that they must also be subject to *de novo* methylation to counteract DRDD activity. If Dynamic GbM genes are not also targeted by one or more methyltransferases, then DRDD activity would ultimately remove all methylation, and the methylation patterns of dynamic GbM loci would not be stable over inheritance. We observe Dynamic GbM in WT samples from multiple independent laboratories [10,23,33], as well as distinct ecotypes [34] (Figure S1), so the methylation patterns of Dynamic GbM loci must therefore be robust against many generations of inbreeding. To identify the methyltransferase pathway that counteracts DRDD to maintain Dynamic GbM, we examined Dynamic GbM methylation patterns in mutants of the major de

novo methyltransferase pathways, RNA-directed DNA methylation (RdDM) [33] and Chromomethylases 2 and 3 (CMT2/CMT3) [33,35]. Unexpectedly, we observed no evidence that these de novo pathways act at Dynamic GbM loci, with quadruple *ddcc* mutants exhibiting near-identical methylation levels and cellular heterogeneity to wild-type (Fig. 3A-B). *De novo* methylation of Dynamic GbM loci must therefore occur by an unknown mechanism. We therefore hypothesized that the maintenance methyltransferase MET1 may exhibit *de novo* activity at these loci. *De novo* methylation by MET1/DNMT1 is not unprecedented – it has been observed in mammalian cells [36], and *de novo* CG methylation has been observed at an introduced transgene within Arabidopsis [37]. Recently, *de novo* CG methylation was also reported in inbred *ddcc* lines [38]. Heterozygous *met1* mutants lose the vast majority of gene body methylation [33], including at most Dynamic GbM genes (Fig. 3C). We therefore chose to exploit this system to perform a genetic experiment to test whether reintroduction of homozygous WT MET1 alleles is sufficient to incur *de novo* methylation at Dynamic GbM loci. Heterozygous *met1* mutants were self-fertilized, and two subsequent generations of WT-segregant (+/+) progeny were selected for whole genome methylation sequencing by enzymatic methyl-seq (EM-seq). As has been previously reported, the memory of DNA methylation patterns maintained by MET1 is initially lost in WT progeny of heterozygous mutants [33]. Similarly, we did not see complete reintroduction of Dynamic GbM in WT segregant plants (Fig. 3C). While we did observe a modest gain in methylation at a subset of CGs within Dynamic GbM genes in two out of three wild-type segregant lines, this enrichment was not statistically significant (Fig. 3C). Furthermore, inbreeding WT segregant lines for an additional generation did not result in further restoration of methylation at Dynamic GbM loci (Fig. 3C). We then further assessed the number of putative de novo methylation events per CG across all cells (e.g. the presence of one or more methylated reads per CGs) at Dynamic GbM and Stable GbM gene bodies. As expected, CG methylation events were very rare at both types of GbM genes in homozygous *met1* mutants (Figure S2). In *met1* heterozygotes and WT-segregants, we observed that de novo methylation was approximately 25% more common at Dynamic GbM genes that Stable GbM genes (Figure S2). We also observed rare but heritable CG methylation events at individual GbM genes, which varied from individual CGs to near-WT restoration of methylation patterns (Figure S3).

Together, these results suggest that de novo methylation by MET1 at an unmethylated locus is rare. If MET1 does maintain the Dynamic GbM state through *de novo* methylation activity, then it likely occurs when MET1 is recruited to the replication fork by the presence of hemimethylated CGs, or adjacently methylated CGs. Our data do not conclusively prove that MET1 maintains the Dynamic GbM state and thus leaves open the possibility that unknown mechanisms could operate at these loci.

Dynamic GbM loci are conserved across distant lineages

As Dynamic GbM loci are clearly maintained over many generations of inheritance and between divergent ecotypes of *Arabidopsis thaliana* (Fig. S1), we sought to determine whether the Dynamic methylation state was conserved over evolutionary time. Evolutionary conservation of the Dynamic GbM state in distant lineages would suggest that the targeting of methylation/demethylation pathways to these loci in all cells serves an adaptive benefit. We isolated the methylation profiles of the most likely homologous genes from five additional genomes representing lineages that diverged from *Arabidopsis thaliana* between 6 and 117 million years ago [39,40]. Homologs were identified by blast homology and further filtered using two criteria: 1) only homologs that were classified as one-to-one orthologs by Inparanoid orthology analysis [41] were included, to avoid misidentification of homologs in gene families with duplications. 2) homologs possessing >1% non-CG methylation across the gene body were excluded, in order to remove genes targeted by de novo methylation pathways, which could impact CG methylation heterogeneity. Even in species that diverged >100 million years ago, intermediate methylation levels at individual CGs (representing cellular heterogeneity in methylation) were clearly observable at orthologous genes to the Dynamic GbM loci observed in Arabidopsis (Fig. 4A). To verify that the frequency of intermediately methylated CGs at Dynamic GbM homologs (likely orthologs) were enriched relative to the number expected by chance, we compared homologs to the set of 581 Dynamic GbM genes against homologs to ten randomly selected sets of 581 genes. Both the number and proportion of heterogeneously methylated CGs within Dynamic GbM homolog gene bodies were substantially higher than the number and proportion observed across the gene bodies of randomly selected genes (Fig. 4B). For all species analyzed, the number of heterogeneously methylated CGs

was >2-fold higher than in randomly selected genes. These data strongly suggest that the loci targeted by DRRD and MET1 are similar across distant evolutionary lineages, and that the Dynamic GbM epigenetic state may offer an adaptive benefit to certain loci.

Dynamic GbM is associated with a regulatory chromatin state

As the Dynamic GbM state is conserved over evolutionary time, we sought to understand the biological purpose of this pattern of epigenetic modification. Similarly to previous studies on Stable GbM [9,27,28], we did not identify any strong associations between Dynamic GbM and broad patterns of gene expression. Although a small number of Dynamic GbM genes exhibited altered transcript abundance in *drdd* and *met1*+/- mutants (28 genes in *drdd* & 37 non-overlapping genes in *met1+/-*), overall transcriptional output (in standard growth conditions) was largely unaltered in the absence of Dynamic GbM (Figure S4). We also found that Dynamic GbM is associated with genes that exhibit a range of expression levels, whereas Stable GbM is predominantly associated with moderate-highly expressed loci (Figure S4). Additionally, while GO-term analysis identified a few statistically significant terms across GbM genes (Figure S5), there was no overall functional similarity that could account for more than 20% of total Dynamic GbM genes.

To better understand the possible function of Dynamic GbM beyond impacting gene expression levels, we examined the chromatin state of individual exons exhibiting Dynamic GbM, compared with exons exhibiting a Stable GbM state and genomic total exons as a control. Coding regions are typically associated with the histone marks H3K4me1 and H3K36me3, which consequently correlate with high expression levels, elevated DNA repair, and low mutation rates [42,43]. Both of these histone marks were enriched in total exons relative to the genomic average, and enriched further still in Stable GbM exons (Fig. 5A), consistent with their proposed function as stably expressed, functionally important (e.g. housekeeping) constitutive genes [25,27]. In contrast, Dynamic GbM exons exhibited reduced presence of these typical coding region histone marks, instead showing a significant enrichment in H3K4me2, as well as increased levels of H3K4me3 and histone acetylation. Exons are typically depleted for the histone mark H3K4me2 which is associated with a regulatory promoter chromatin state [44].

Consequently, we observed a much greater degree of overlap between Dynamic GbM genes and intergenic chromatin states (Fig. 5B-C) – including those normally associated with distal promoter sequences – compared to Stable GbM genes. The majority of genes in the genome, including Stable GbM genes, display a single peak of H3K4me2 at the promoter and transcriptional start site, which is replaced by H3K4me1 within the coding region. In Dynamic GbM genes, we observed elongated H3K4me2 domains, which encompassed the entire gene body, or occasionally additional distinct H3K4me2 peaks that coincided with the cytosines displaying dynamic methylation turnover (Fig. 5F). While Dynamic GbM genes are typically shorter than Stable GbM genes (Fig. 5D), the enrichment of H3K4me2 is independent of gene length (Fig. 5E). Together, these data show that Dynamic GbM is associated with a regulatory, or “promoter” chromatin state, suggesting a novel link between gene body methylation and transcriptional regulation.

Due to the presence of regulatory histone marks within the gene bodies of Dynamic GbM genes, we sought to understand the relationship between the Dynamic GbM state and transcription. Dynamic GbM genes do not show a substantial change in transcriptional output between WT and mutants with defective MET1 maintenance of GbM or DRDD demethylation of GbM (Figure S3). We therefore hypothesized that the Dynamic GbM state does not directly modulate transcriptional output (like RNA-directed DNA methylation or other silencing mechanisms) but rather facilities a more flexible or open-ended regulatory structure, due to the presence of regulatory histone marks. To test this hypothesis, we leveraged large transcriptome datasets that compare transcript abundance across a wide range of tissues and cell types [45], biotic and abiotic stresses [46] and inter-individual variability under consistent conditions [47]. Using these datasets, we defined the coefficient of variation for both Stable and Dynamic GbM genes across conditions, to examine whether GbM patterns relate to the range, or plasticity of possible expression states a gene can exhibit. Stable GbM genes exhibited low variability across development and variable conditions, consistent with the observed correlative association between stable gene body methylation and consistently expressed housekeeping genes [25,27](Fig. 6A). In contrast, Dynamic GbM genes exhibit high variance in expression levels across developmental contexts, and physiological conditions, suggesting their altered methylation and/or

chromatin state confers increased expression plasticity. This increased variance was not observed across multiple individual WT replicates grown in consistent conditions (Fig. 6A), suggesting that high variance is not caused by high stochasticity in expression levels inherent to these loci. We therefore propose that the abundance of regulatory histone modifications within Dynamic GbM genes enables a greater range of possible expression states, in stark contrast to stable methylation, which promotes canalization towards a single, robust and consistent expression state (Fig. 6B). One prediction of this model is that disruption of methylation dynamics at dynamic GbM loci should impact gene expression plasticity. *drdd* mutants possess consistently high methylation levels at Dynamic GbM genes in the majority of all cells, similar to Stable GbM genes in WT (Fig. 1E). We therefore sought to test if *drdd* mutants exhibit reduced variance in Dynamic GbM expression levels compared to WT. Consistent with our model, *drdd* mutants exhibited a 30% lower standard deviation in Dynamic GbM gene expression compared to WT (Figure S6), whereas the standard deviation in Stable GbM gene expression was unchanged. This finding further supports our model – that the turnover of DNA methylation patterns at Dynamic GbM loci is associated with enhanced gene expression plasticity.

.

DISCUSSION

DNA methylation within the bodies of genes has been a well-known feature of eukaryotic genomes for several decades [26,27,48]. Despite this, the precise function of gene body methylation has remained unclear. A positive correlation between gene body methylation and robustly expressed genes has been well established in both animals and plants [27,48], with genes that occupy the extremes of the distribution of expression states typically being unmethylated in their gene body [27]. In this study, we have presented further evidence that gene body methylation functions by modulating the possible positions within “expression space” that genes can occupy. Stable gene body methylated genes (in which CGs are methylated in every cell) exhibit a dramatically lower gene expression plasticity across development and diverse physiological conditions than the genomic average (Fig. 6). A recent high-throughput genetic knockout study also supports this model. Genetic knockout of the CG methylation maintenance methyltransferase MET1 in 18 different wild-type Arabidopsis accessions

resulted in a dramatic increase in overall expression variability between genotypes, suggesting that CG methylation functions in canalizing transcriptional activity, or reducing variability [49]. Interestingly, studies of the responses of both animals and plants have also identified GbM as a correlate of gene expression plasticity. A study of two seagrass species showed that the genes that display low expression plasticity in response to environmental stress share properties of Stable GbM genes and are predicted to be methylated [50]. A similar association has been observed in corals transplanted between high and low fitness environments, where GbM is proposed to manage the balance between low/un-methylated environmentally-responsive genes and transcriptionally stable GbM genes [51].

Our study identifies a new sub-type of GbM locus within plant genomes, an exception that helps demonstrate the rule associating GbM with reduced gene expression plasticity. This sub-type of GbM, which we term Dynamic GbM, is distinguished by the targeting of methylated CGs by the DRDD DNA demethylation pathway in all cell and tissue types. Dynamic GbM genes display greatly elevated gene expression plasticity in comparison to their Stable GbM counterparts, strongly suggesting that enzymatic modification of the Stable GbM state disrupts its primary function in restricting expression plasticity. One consequence of this DNA demethylation activity is the generation of cellular heterogeneity in DNA methylation patterns and the production of unmethylated CGs within stem and gamete cells. In order to be stable over inheritance and evolution, Dynamic GbM loci must therefore also be re-methylated by one or more methyltransferase enzymes, most likely the maintenance methyltransferase MET1. The dynamic activity of demethylase and methyltransferase enzymes at these loci therefore functions as a constant source of methylation turnover, in stark contrast to the cellular consistency of Stable GbM genes. We therefore propose that dynamic methylation turnover in all cell types is a mechanism that could enable greater expression plasticity over the development of the entire organism. As the targeting and expression of methylation modifier pathways is not cell or tissue-type specific in plants, we reason that they target Dynamic GbM genes universally, so that increased gene expression plasticity is available when adaptively beneficial. Consequently, a Dynamic GbM gene may be targeted by methylation writers and erasers even in cell types or conditions in which the gene is not expressed.

While our study offers strong evidence for an association between GbM and gene expression plasticity, this proposed function is largely based on a correlative association, similar to prior hypotheses of GbM function. To move towards a mechanistic understanding of the roles of Dynamic and Stable GbM, we performed a detailed examination of the chromatin states of each. Our results show a striking association between Dynamic GbM and chromatin marks typically associated with regulatory regions, such as H3K4me2. This is consistent with the increased expression plasticity of Dynamic GbM genes, as genic H3K4me2 is likely to increase the extent to which regulators of transcription can target and modify expression of these loci [52,53]. H3K4me2 has also been shown to be associated with genes with high expression variability in Arabidopsis, as well as increased tissue-specificity across development [54], consistent with our proposed model. Intriguingly, gene-body H3K4me2 has also been reported in human cells, and – consistent with our model – is enriched within genes that show high cell-type specificity across development [55], aka elevated plasticity in possible expression states. This raises the possibility that properties of the Dynamic GbM epigenetic state could be conserved across eukaryote kingdoms. In contrast to Dynamic GbM genes, Stable GbM genes show the strongest genic enrichment of histone marks associated with polymerase II elongation and the "housekeeping gene" chromatin state, namely H3K4me1 and H3K36me3 [44]. Defining the precise molecular relationship between gene body methylation, dynamic methylation turnover, expression plasticity and a regulatory histone modification state is a key future extension of this research. One finding within our study points towards a causative role for DNA methylation dynamics in impacting gene expression variability. *drdd* mutants, which cannot generate methylation dynamics at GbM genes, exhibit a reduction in the variability of Dynamic GbM gene expression states (Figure S6). As gene expression plasticity is an important determinant of how organisms adapt to stress, variable environments, and climate change [51,56], discerning the precise ways in which epigenetic states impact plasticity is an important future question.

CONCLUSIONS

In this study, we define a new feature of the epigenetic landscape in plants. A subset of gene bodies within the genome are targeted by DNA methylation writers and erasers in all cells, which create a dynamic DNA methylation turnover. This dynamic DNA methylation state is associated with a "promoter-like" chromatin state, as well as enhanced gene expression plasticity. Our study also demonstrates a clear link between typical gene body methylation (which is stable across all cells) and a canalized expression state with low plasticity. Overall, our study offers a substantial advance towards understanding the functional role of enigmatic gene body methylation in eukaryotes.

MATERIALS AND METHODS

Plant Growth Conditions

All plants were grown on a 1:1 mix of potting compost and vermiculite in a Percival AR100L3 with 16-hour days at 21-22°C and 60% humidity.

met1 mutation segregation

Seeds from a single self-fertilized heterozygous *met1-3* [57] plant (obtained from the Arabidopsis Biological Resource Center – stock number CS16394) were grown for 4 weeks in standard growth conditions as outlined above. A single rosette leaf was collected from individual plants and DNA was extracted as described [58]. The genotype of the *MET1* locus was then assayed using the following primers: METNF2, TAGCCAACAAGTTATCGCTTACTC; METNR2, TTCGCAAACCATTCTTCACAGAGC; TL-4, TAATTGCGTCGAATCTCAGCATCG. Plants that were identified as heterozygous for *met1-3* and homozygous for the WT *MET1* allele were self-fertilized and collected to repeat as described for an additional generation.

DNA extraction and Enzymatic Methyl Sequencing (EM-Seq)

Leaf tissue from 28-day old plants was collected and flash frozen in liquid nitrogen. Two leaves per sample were ground using a Qiagen TissueLyser II and immediately after 200 µl of CTAB was added and heated to 65°C for approximately 20 minutes. This was followed by the addition of 200 µl chloroform, 10 minute centrifugation and ethanol washes to isolate total genomic DNA. The DNA

concentration was quantified using a Qubit Flex fluorometer, and a diluted spike-in was added according to NEB recommendations (1 µl diluted pUC19 (0.001 ng) control DNA and 1 µl diluted unmethylated lambda DNA (0.02 ng) in 0.1X TE, pH 8.0 per 10–200 ng plant sample DNA). Genomic DNA samples with added spike-ins were sonicated to an average fragment size of 300bp using a Covaris S220 Focused Ultrasonicator (80s treatment, 140 peak power, 10 duty factor 200 cycles/burst). Subsequently, samples were incubated with an RNAse cocktail and RNAse was removed using a Qiagen spin column. EM-seq libraries were constructed following the NEBNext® Enzymatic Methyl-seq Kit protocol. Library quality was verified with a KAPA Library Quant Kit and fragment analyzer. The samples were pooled to a concentration of 4 ng/µl and 19.2 nM and sequenced using an Illumina NextSeq 2000 (paired-end 150 x 150 bp) 11 million average reads/sample. Further information on each EM-seq library sequenced in this study is available in Supplemental Table 1.

Whole genome methylation analysis

Mapping of whole genome bisulfite sequencing and EM-seq libraries was performed as described [10]. In brief, reads were processed with TrimGalore (Babraham Bioinformatics), trimming 8bp from the 5' end of reads and enforcing a 3' end quality score of >25%. Reads were mapped to the Araport11 genome using Bismark 0.20.1 [59], allowing for 1 mismatch per read in the seed region and removing PCR duplicates. Methylation values for each cytosine were calculated using the Bismark methylation extractor function. The efficiency of EM-seq conversion was verified by quantifying the percentage of methylation for reads mapped to the chloroplast. EM-seq conversion rates were >99.7% across all samples.

To classify Stable and Dynamic GbM genes, data from WT and *drdd* mutant bisulfite sequencing and EM-seq libraries [10] was used. Analysis was initially performed on a single high-depth EM-seq replicate to avoid inter-individual variation in methylation profiles. Subsequently, the classification of Dynamic GbM genes was also verified by pooling four bisulfite sequencing replicate libraries, showing that Dynamic GbM identification is robust against variation between replicates of the same genotype.

CGs exhibiting a methylation percentage <= 10% were classified as unmethylated, CGs >=85% were classified as fully methylated and CGs methylated between 10-85% were classified as heterogeneously methylated, as each deduplicated read derives from an independent cell in the original sample. Dynamic GbM genes were identified based on two cutoffs: 1) the presence of a WT vs *drdd* differentially methylated region (DMR) within the gene body (DMRs were called based on aggregation of differentially methylated CGs within overlapping 200bp windows, as described in our previous work [10,24]). 2) Dynamic GbM genes required at least 5 CG dinucleotides (10 CG cytosines total) that gained >20% methylation in *drdd* mutants. This second step removed genes possessing only a small number of heterogeneously methylated sites in an otherwise unmethylated or fully methylated gene body. Stable GbM genes were identified as described [9]. These two cutoffs resulted in a list of 581 total Dynamic GbM genes.

The evolutionary conservation of Dynamic GbM was performed by identifying most likely homologs to the 581 Arabidopsis genes using BLAST homology from 5 additional species: *Arabidopsis lyrata*, *Capsella rubella*, *Theobroma cacao*, *Citrus clementina* and *Vitis vinifera*. Homologs were further filtered using two criteria: 1) only homologs that were classified as one-to-one orthologs by Inparanoid orthology analysis were included, 2) homologs possessing >1% non-CG methylation across the gene body were excluded, in order to remove genes targeted by de novo methylation pathways. Additionally, most likely homologs were also identified for ten sets of 581 randomly selected genes, which were then averaged as a control group. The whole genome methylation profiles of each species were downloaded and remapped from other studies [39,40]. For each Dynamic GbM / random homolog, the number of heterogeneously methylated CGs was isolated using bedtools intersect. Both the number and density of heterogeneously methylated CGs was compared relative to average across 10 sets of randomly selected genes. Statistical significance was calculated using an unpaired two-tailed t-test.

RNA-seq analysis

Prior to mapping, adapters were trimmed using Trim Galore (Babraham Bioinformatics), trimming 9 bp from the 5'-end of reads, and enforcing a 3'-end quality of >25%. Reads were mapped to the Araport11 genome using STAR, permitting 0.05 mismatches as a fraction of total read length and discarding reads that did not map uniquely [60]. Differentially expressed genes were identified by running htseq-count and DESeq2 [61], ensuring a minimum of two-fold change in expression and a Benjamini–Hochberg corrected P-value <0.05. GO term analysis was performed using agriGO v2.0 with default parameters [62].

Chromatin state analysis

Chromatin state classification for the entire Arabidopsis genome was obtained from Sequeira-Mendes et al [44]. The percentage and base-pair overlap between Stable and Dynamic GbM genes and each chromatin state was calculated using Bedtools intersect. In Figure 5, "Gene body" refers to chromatin states 3 and 7 combined, "Promoter (TSS)" refers to state 1, "Promoter (distal)" refers to state 2, and "Other intergenic" refers to states 4, 5, 8 and 9 combined.

Expression variance analysis

Gene expression counts for Stable GbM genes, Dynamic GbM genes and Total genes were obtained from the Bio-Analytic Resource for Plant Biology (BAR) Expression Browser platform [63]. The coefficient of variation for each gene was then calculated using all available data across tissues/cell types (305 data points) or physiological stress conditions (165 data points). Genes that were not represented in at least 75% of all tissues/conditions were excluded. 133 genes that were identified in both the Stable and Dynamic GbM gene groups were also excluded to avoid overlap between datasets. To calculate the variability in gene expression across identical conditions, RNA-seq from 14 wild-type replicates was obtained from a study by Cortijo et al [47], and the coefficient of variation of TPM values for each gene was calculated.

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All high throughput sequencing data generated in this study will be made available (upon acceptance) at the NCBI Gene Expression Omnibus.

Competing interests

The authors declare that they have no competing interests

Funding

CW is supported by an NSF Postdoctoral Fellowship in Biology (award #2209401).

Authors' contributions

CJW and KAT performed experiments. CJW, DD, GJM and BPW performed data analysis. BPW conceived of the study and wrote the manuscript.

Acknowledgements

We are grateful to Annika Quist for her assistance with mapping RNA-seq data. We are thankful to Bob Schmitz, Steve Jacobsen and PRV Satyaki for helpful feedback and discussion.

FIGURE LEGENDS

Figure 1: Two distinct types of Gene body methylation (GbM) in the Arabidopsis genome. A) Genome browser snapshot showing representative example of a gene with Stable GbM and a gene

with Dynamic GbM in both WT and *drdd* mutants. Red bars represent methylated cytosines, and the height of the bar represents the percentage of cells in which a cytosine is methylated. B) A Venn diagram showing the number of Stable and Dynamic GbM genes and the overlap between the two groups. C) Box plot showing the overall percentage CGs methylation across entire genes (all CGs across all cells) with Stable or Dynamic GbM. Line = median, box = interquartile range, whiskers = 5th and 95th percentile. D) Violin plot showing the distribution of methylation heterogeneity levels of all CGs in the Arabidopsis genome, shown as the percentage of cells in which a CG is methylated. E) The proportion of fully methylated, heterogeneously methylated and unmethylated CGs within both Stable and Dynamic GbM gene bodies, in both WT and *drdd* mutants.

Figure 2: **Dynamic GbM genes exhibit methylation heterogeneity in all cell and tissue types**. A) Representative genome browser snapshot showing heterogeneously methylated CGs at a Dynamic GbM locus in a variety of cell and tissue types. (B and C) Violin plots showing the distribution of methylation heterogeneity levels of CGs within Stable and Dynamic GbM genes (methylated domains only) from various somatic cell and tissue types (B), and meristem and germ cells (C). Solid lines represent the median and dashed lines represent the interquartile range. All Dynamic GbM violins are significantly different to Stable GbM violins (p=<0.0001), with the exception of *ros1; dme* (+/-) sperm cells. D) Proportion of fully methylated, heterogeneously methylated and unmethylated CGs across all tissue and cell types for both Stable and Dynamic GbM gene bodies.

Figure 3: The Dynamic GbM state is maintained by MET1. A) Representative genome browser snapshot of a Dynamic GbM locus in *met1*, *ddcc* and *drdd* mutants. Scale bar = 200 bp. B) Boxplots showing the cellular heterogeneity of CGs within Dynamic GbM genes in WT, *drdd* mutants and multiple de novo methylation mutants. C) Boxplots showing the cellular heterogeneity of CGs within Dynamic GbM genes in homozygous (-/-) and heterozygous (+/-) *met1* mutants, as well as multiple WT-segregant progeny from a selfed *met1* heterozygote. In B and C, horizontal lines = median, box =

interquartile range and whiskers = 95th percentile. Horizontal dashed lines represent the cutoffs used to identify heterogeneously methylated CGs (10-85% methylation across all cells).

Figure 4: Evolutionary conservation of the Dynamic GbM state. A) Representative Dynamic GbM gene homologs from six different species. Each red bar represents a single CG, and the bar height is the percent methylation heterogeneity across cells. Dashed lines represent the cutoffs used to call heterogeneously methylated cytosines (10-85% of cells). Rightmost panel: the estimated divergence time between species and *Arabidopsis thaliana*, obtained from TimeTree [64]. Points represent the median divergence time estimate, and vertical dashed lines represent the 95% confidence interval. B) Violin plots showing the number (left panel) and proportion (right panel) of heterogeneously methylated CGs across all homologs of Dynamic GbM genes, and ten identically-sized groups of randomly selected genes. Solid lines represent the median and dashed lines represent the interquartile range. *p = <0.05, ***p = <0.0005.

Figure 5: **The association between Dynamic GbM and regulatory chromatin marks**. A) Boxplots showing the normalized read depth of ChIP-seq data of fourteen histone modifications (data from [42]) within exons overlapping Dynamic GbM domains (blue), Stable GbM domains (green) or total exons (grey). Horizontal lines = median, box = interquartile range, whiskers = 5th and 95th percentiles. The red box highlights model gene body (H3K4me1) and promoter (H3K4me2) histone marks. B) The percentage overlap between Stable and Dynamic GbM gene bodies and four chromatin states identified by [44]. C) The total base pair overlap between Stable and Dynamic GbM gene bodies and a chromatin state typically enriched in distal promoters. D) Boxplot showing the distribution of gene body lengths of both Stable and Dynamic GbM genes. E) Boxplot showing the normalized read depth of H3K4me2 ChIP-seq data at Stable and Dynamic GbM genes separated into bins based on gene body length. F) Genome browser snapshot of raw H3K4me1 and H3K4me2 ChIP-seq data at Stable and Dynamic GbM genes. Tracks represent WT data from four independent studies. Data obtained from [65].

Figure 6: Dynamic GbM is associated with increased gene expression plasticity. A) Boxplots showing the coefficient of variation in transcript abundance for all genes, Stable GbM, and Dynamic GbM genes across development, diverse physiological stresses and WT replicates in standard growth conditions. Horizontal lines = median, box = interquartile range, whiskers = 5th and 95th percentile. ***p=<0.0005, *p=<0.05. B) Schematic showing the proposed impact of Stable and Dynamic GbM on gene expression plasticity, represented as Waddington's landscape. Stable GbM genes exhibit low variance, suggesting canalization into a single, robust gene expression state. Dynamic GbM genes exhibit high variance, suggesting the capacity to access multiple possible gene expression states.

Figure S1: Dynamic GbM is consistent across WT samples and different ecotypes. Genome browser snapshots of four dynamic GbM genes in Col-0 WT Arabidopsis from three independent studies [10,33,66] and four additional ecotypes [34].

Figure S2: *De novo* methylation events in *met1* mutants and WT segregants. The rate of de novo CG methylation events per CG across all cells. Bars represent the mean, error bars represent standard deviation. Individual data points represent independent genetic lines that were sequenced. Chloroplast CG methylation is included a control for the false positive rate from incomplete conversion or sequencing errors.

Figure S3: Example de novo methylation events in *met1* mutant segregants at GbM loci. Genome browser snapshots of four Dynamic GbM genes in WT, drdd, met1 heterozygotes (+/-) and three independent lines of WT-segregant progeny (+/+).

Figure S4: RNA-seq analysis of Stable and Dynamic GbM genes. A) Boxplots showing distribution of Log2 fold-change in transcript abundance for Stable and Dynamic GbM genes in WT v *met1* heterozygous mutants (which lose GbM) and WT v *drdd* mutants (which exhibit elevated GbM at

Dynamic GbM genes). Horizontal line = median, box = interquartile range, whiskers = 1st and 99th percentile. B) Distribution of Stable GbM genes across WT transcript abundance deciles. C) Distribution of Dynamic GbM genes across WT transcript abundance deciles. In B and C, 1 = highest expression decile, 10 = lowest expression decile.

Figure S5: Enriched GO-terms in Dynamic GbM genes. Symbol color is a heat map scale for statistical significance and symbol size is proportional to the number of genes for the GO term. The x axis indicates the enrichment ratio for DEGs relative to all genes with the GO term.

Figure S6: Expression variance of Dynamic and Stable GbM genes in WT and drdd mutants. Barplots show the number of genes in the extreme ends of the expression distribution. Extreme high = top expression decile, extreme low = bottom expression decile + silent genes. Asterisks represent the total standard deviation in TPM expression values for each group.

REFERENCES

1. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature. 2001;411:212–4.

2. Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, et al. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. Cell. 2018;175:1259-1271.e13.

3. Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev. 2015;29:2183–202.

4. Gehring M, Bubb KL, Henikoff S. Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting. Science. 2009;324:1447–51.

5. Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet. 2011;12:565–75.

6. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11:204–20.

7. Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, et al. Century-scale Methylome Stability in a Recently Diverged Arabidopsis thaliana Lineage. PLOS Genet. 2015;11:e1004920.

8. Catania S, Dumesic PA, Pimentel H, Nasif A, Stoddard CI, Burke JE, et al. Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. Cell. 2020;180:263-277.e20.

9. Bewick AJ, Ji L, Niederhuth CE, Willing E-M, Hofmeister BT, Shi X, et al. On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci. National Academy of Sciences; 2016;113:9111–6.

10. Williams BP, Bechen LL, Pohlmann DA, Gehring M. Somatic DNA demethylation generates tissue-specific methylation states and impacts flowering time. Plant Cell. 2022;34:1189–206.

11. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, et al. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell. 2002;110:33–42.

12. Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu J-K. ROS1, a Repressor of Transcriptional Gene Silencing in Arabidopsis, Encodes a DNA Glycosylase/Lyase. Cell. 2002;111:803–14.

13. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci. 2007;104:6752–7.

14. Ono A, Yamaguchi K, Fukada-Tanaka S, Terada R, Mitsui T, Iida S. A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. Plant J. 2012;71:564–74.

15. Gent JI, Higgins KM, Swentowsky KW, Fu F-F, Zeng Y, Kim DW, et al. The maize gene maternal derepression of r1 encodes a DNA glycosylase that demethylates DNA and reduces siRNA expression in the endosperm. Plant Cell. 2022;34:3685–701.

16. Yamamuro C, Miki D, Zheng Z, Ma J, Wang J, Yang Z, et al. Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation. Nat Commun [Internet]. 2014 [cited 2014 Oct 16];5. Available from: http://www.nature.com/ncomms/2014/140605/ncomms5062/full/ncomms5062.html?message-global=remove

17. Williams BP, Pignatta D, Henikoff S, Gehring M. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. PLoS Genet. 2015;11:e1005142.

18. Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad Sci U S A. 2015;112:3553–7.

19. Halter T, Wang J, Amesefe D, Lastrucci E, Charvin M, Singla Rastogi M, et al. The Arabidopsis active demethylase ROS1 cis-regulates defence genes by erasing DNA methylation at promoter-regulatory regions. Zilberman D, Weigel D, Zilberman D, Zhu J-K, Ton J, editors. eLife. eLife Sciences Publications, Ltd; 2021;10:e62994.

20. Williams BP, Gehring M. Principles of Epigenetic Homeostasis Shared Between Flowering Plants and Mammals. Trends Genet. Elsevier; 2020;36:751–63.

21. Henderson IR, Jacobsen SE. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev. 2008;22:1597–606.

22. Gallego-Bartolomé J, Liu W, Kuo PH, Feng S, Ghoshal B, Gardiner J, et al. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. Cell. 2019;176:1068-1082.e19.

23. Rigal M, Kevei Z, Pélissier T, Mathieu O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. EMBO J. 2012;31:2981–93.

24. Williams BP, Gehring M. Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit. Nat Commun. 2017;8:2124.

25. Zilberman D. An evolutionary case for functional gene body methylation in plants and animals. Genome Biol. 2017;18:87.

26. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. Nature Publishing Group; 2012;13:484–92.

27. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet. 2007;39:61–9.

28. Picard CL, Gehring M. Proximal methylation features associated with nonrandom changes in gene body methylation. Genome Biol. 2017;18:73.

29. Kawakatsu T, Stuart T, Valdes M, Breakfield N, Schmitz RJ, Nery JR, et al. Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants. 2016;2:16058.

30. Khouider S, Borges F, LeBlanc C, Ungru A, Schnittger A, Martienssen R, et al. Male fertility in Arabidopsis requires active DNA demethylation of genes that control pollen tube function. Nat Commun. Nature Publishing Group; 2021;12:410.

31. Long J, Walker J, She W, Aldridge B, Gao H, Deans S, et al. Nurse cell-–derived small RNAs define paternal epigenetic inheritance in Arabidopsis. Science. American Association for the Advancement of Science; 2021;373:eabh0556.

32. Gutzat R, Rembart K, Nussbaumer T, Hofmann F, Pisupati R, Bradamante G, et al. Arabidopsis shoot stem cells display dynamic transcription and DNA methylation patterns. EMBO J. John Wiley & Sons, Ltd; 2020;39:e103667.

33. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell. 2013;152:352–64.

34. Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell. 2016;166:492–505.

35. Zhong Z, Feng S, Duttke SH, Potok ME, Zhang Y, Gallego-Bartolomé J, et al. DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2021;118:e2023347118.

36. Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, et al. Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol. Nature Publishing Group; 2021;28:594–603.

37. Zubko E, Gentry M, Kunova A, Meyer P. De novo DNA methylation activity of methyltransferase 1 (MET1) partially restores body methylation in Arabidopsis thaliana. Plant J Cell Mol Biol. 2012;71:1029–37.

38. Briffa A, Hollwey E, Shahzad Z, Moore JD, Lyons DB, Howard M, et al. Unified establishment and epigenetic inheritance of DNA methylation through cooperative MET1 activity [Internet]. bioRxiv; 2022 [cited 2022 Oct 20]. p. 2022.09.12.507517. Available from: https://www.biorxiv.org/content/10.1101/2022.09.12.507517v1

39. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLOS Genet. 2014;10:e1004785.

40. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 2016;17:194.

41. Sonnhammer ELL, Östlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2015;43:D234–9.

42. Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, et al. Mutation bias reflects natural selection in Arabidopsis thaliana. Nature. Nature Publishing Group; 2022;602:101–5.

43. Quiroz D, Lopez-Mateos D, Zhao K, Carbonell-Bejerano P, Yarov-Yarovoy V, Monroe JG. Biased mutagenesis and H3K4me1-targeted DNA repair in plants [Internet]. bioRxiv; 2022 [cited 2022 Nov 8]. p. 2022.05.28.493846. Available from: https://www.biorxiv.org/content/10.1101/2022.05.28.493846v2

44. Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, et al. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell. 2014;26:2351–66.

45. Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J. 2016;88:1058–70.

46. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell. 2017;29:1806–21.

47. Cortijo S, Aydin Z, Ahnert S, Locke JC. Widespread inter-individual gene expression variability in Arabidopsis thaliana. Mol Syst Biol. John Wiley & Sons, Ltd; 2019;15:e8591.

48. Hellman A, Chess A. Gene Body-Specific Methylation on the Active X Chromosome. Science. American Association for the Advancement of Science; 2007;315:1141–3.

49. Srikant T, Yuan W, Berendzen KW, Garrido AC, Drost H-G, Schwab R, et al. Canalization of genome-wide transcriptional activity in Arabidopsis thaliana accessions by MET1-dependent CG methylation [Internet]. bioRxiv; 2022 [cited 2022 Nov 3]. p. 2022.07.14.500095. Available from: https://www.biorxiv.org/content/10.1101/2022.07.14.500095v1

50. Entrambasaguas L, Ruocco M, Verhoeven KJF, Procaccini G, Marín-Guirao L. Gene body DNA methylation in seagrasses: inter- and intraspecific differences and interaction with transcriptome plasticity under heat stress. Sci Rep. Nature Publishing Group; 2021;11:14343.

51. Dixon G, Liao Y, Bay LK, Matz MV. Role of gene body methylation in acclimatization and adaptation in a basal metazoan. Proc Natl Acad Sci. National Academy of Sciences; 2018;115:13342–6.

52. Wang Y, Li X, Hu H. H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics. 2014;103:222–8.

53. Liu L, Jin G, Zhou X. Modeling the relationship of epigenetic modifications to transcription factor binding. Nucleic Acids Res. 2015;43:3873–85.

54. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. 2009;10:R62.

55. Pekowska A, Benoukraf T, Ferrier P, Spicuglia S. A unique H3K4me2 profile marks tissue-specific gene regulation. Genome Res. 2010;20:1493–502.

56. Rivera HE, Aichelman HE, Fifer JE, Kriefall NG, Wuitchik DM, Wuitchik SJS, et al. A framework for understanding gene expression plasticity and its influence on stress tolerance. Mol Ecol. 2021;30:1381–97.

57. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, et al. The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. Curr Biol. 2007;17:379–84.

58. Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 1991;19:1349.

59. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.

60. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

61. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

62. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 2017;45:W122–9.

63. Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. Plant J. 2005;43:153–63.

64. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al. TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol. 2022;39:msac174.

65. Liu Y, Tian T, Zhang K, You Q, Yan H, Zhao N, et al. PCSD: a plant chromatin state database. Nucleic Acids Res. 2018;46:D1157–67.

66. Rigal M, Becker C, Pélissier T, Pogorelcnik R, Devos J, Ikeda Y, et al. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis thaliana F1 epihybrids. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2016;113:E2083–92.

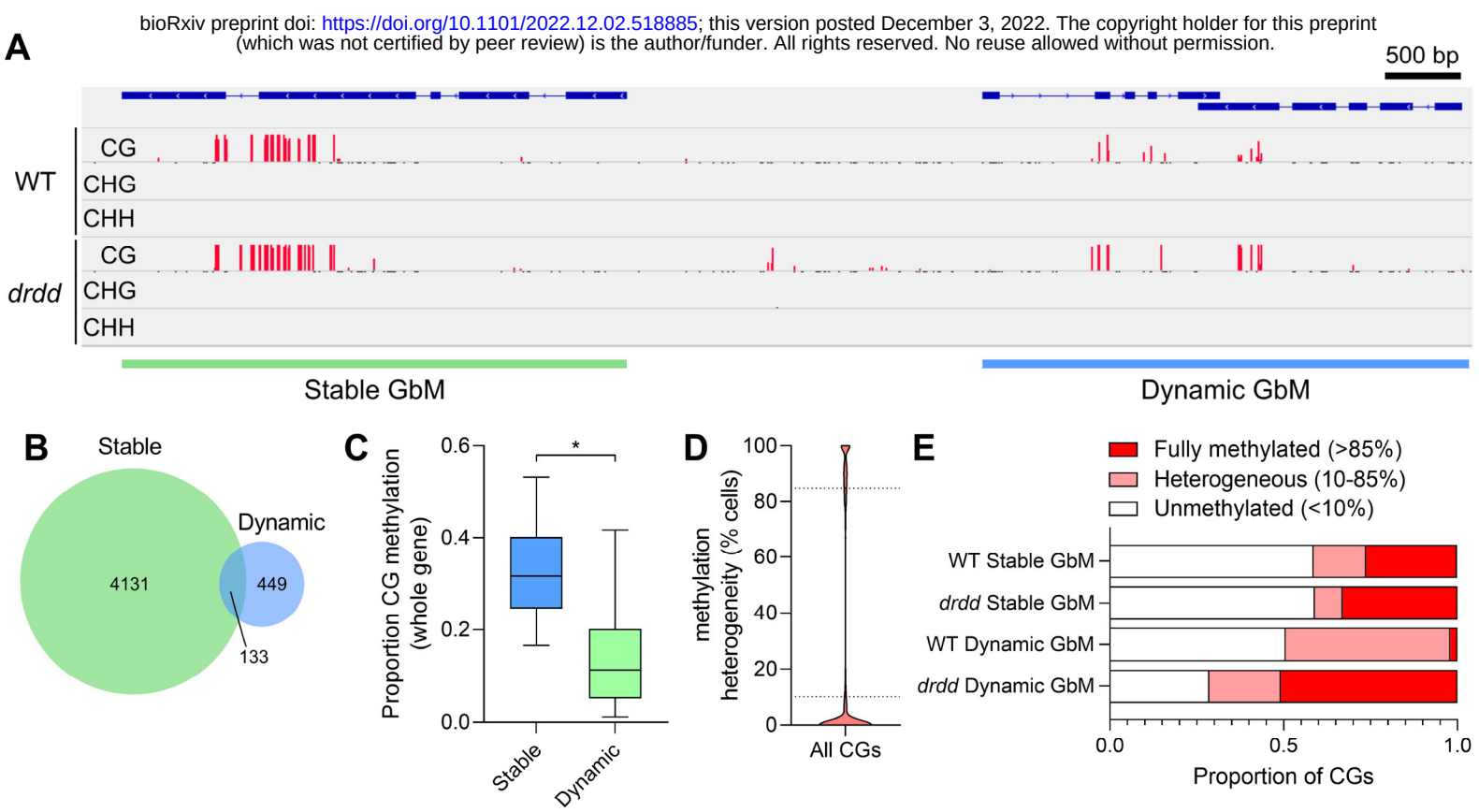


Figure 1

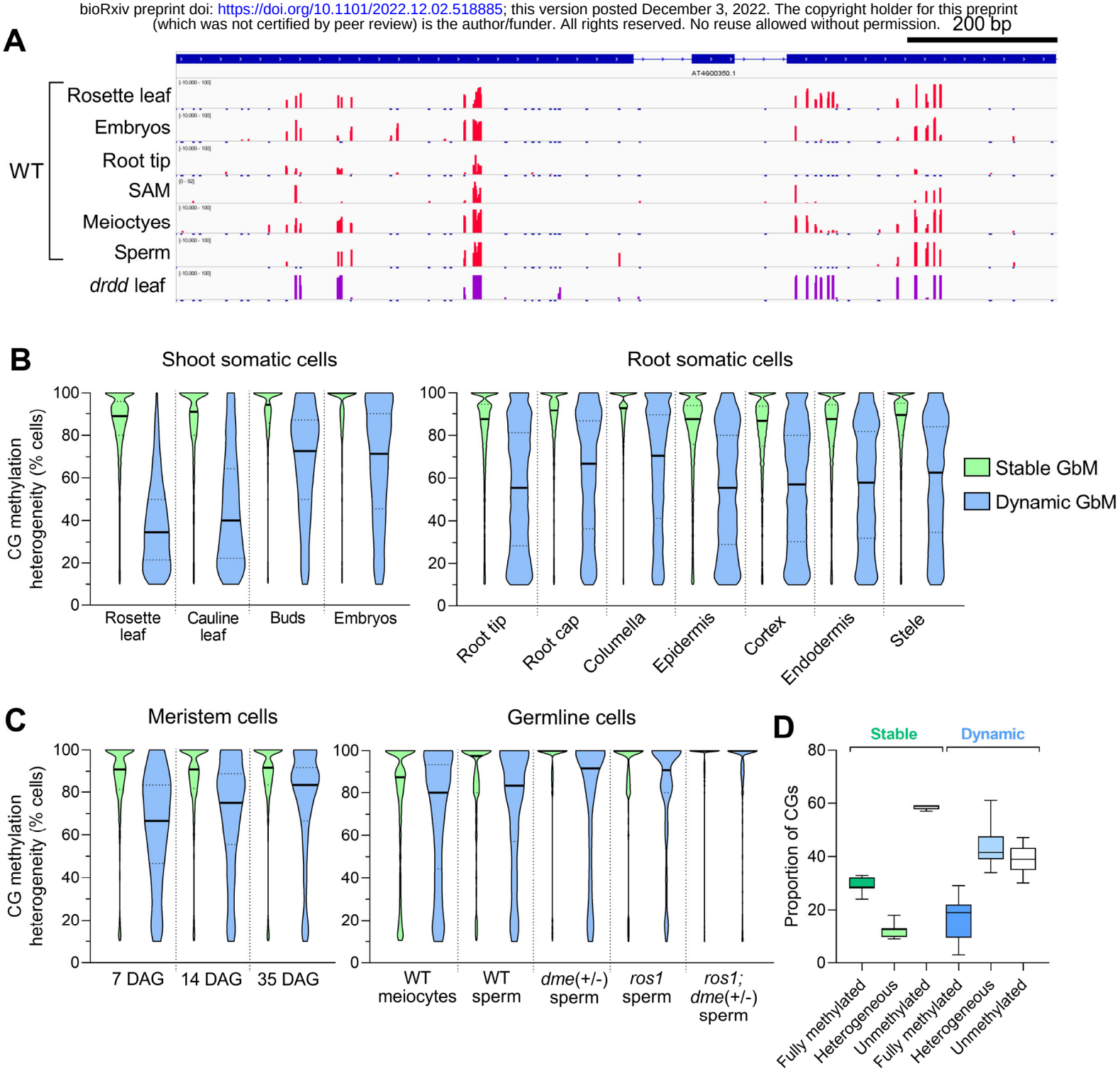


Figure 2

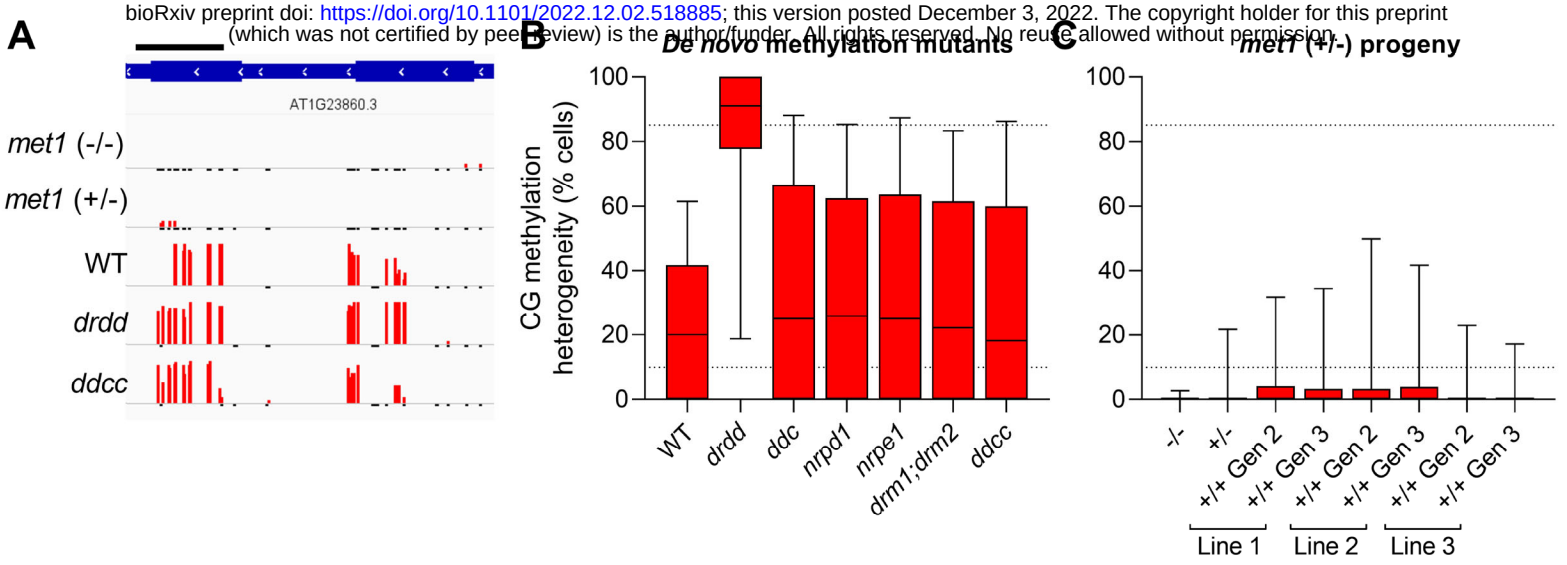
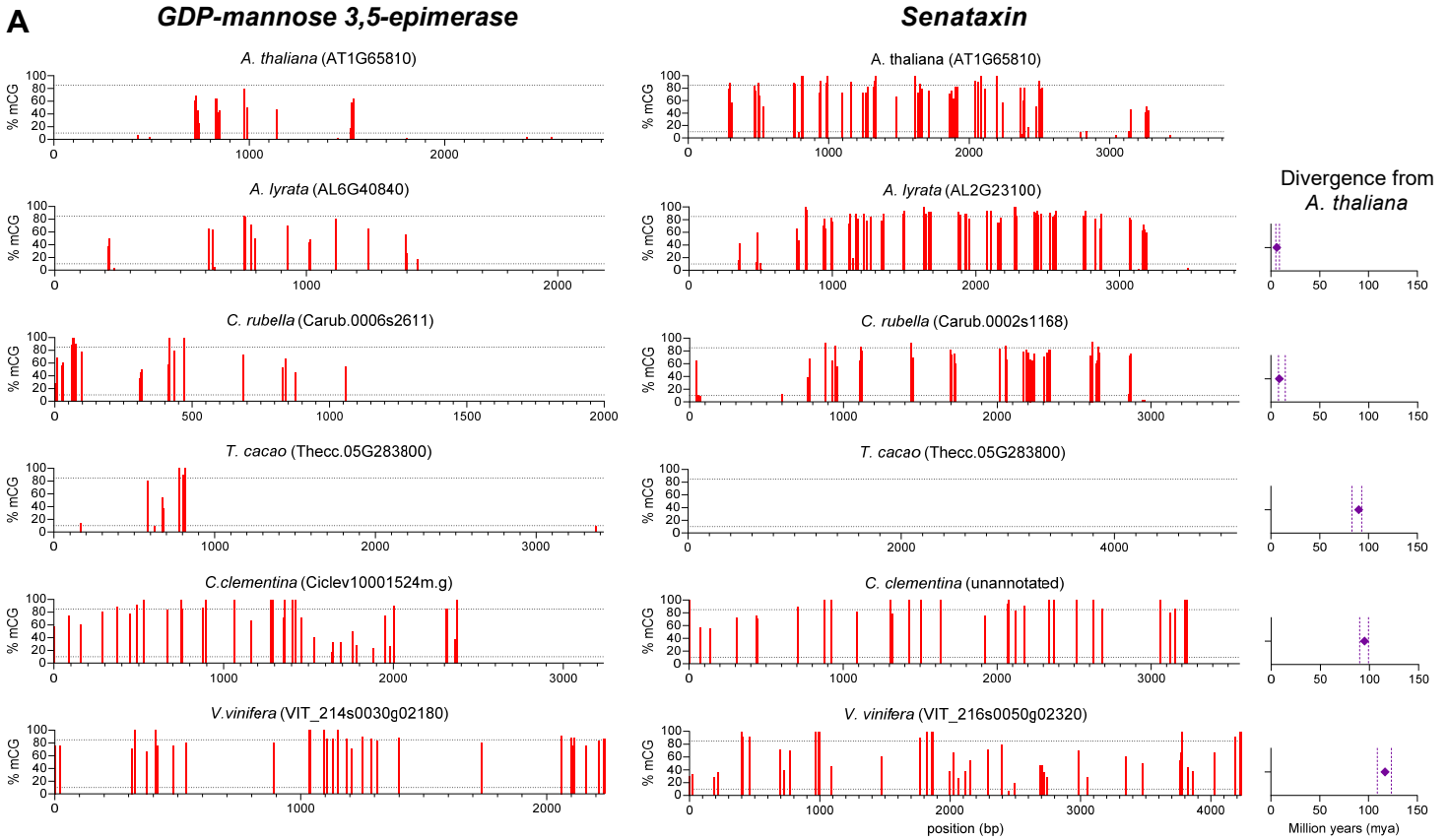


Figure 3



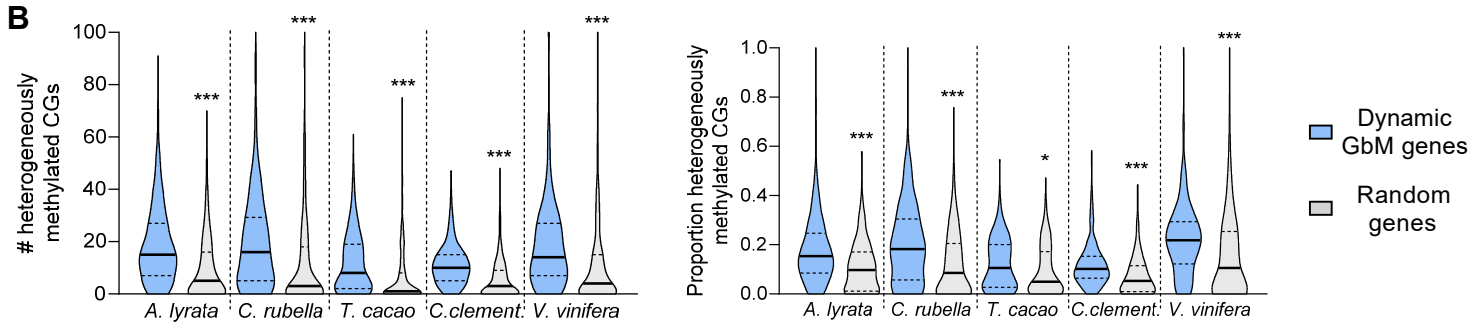
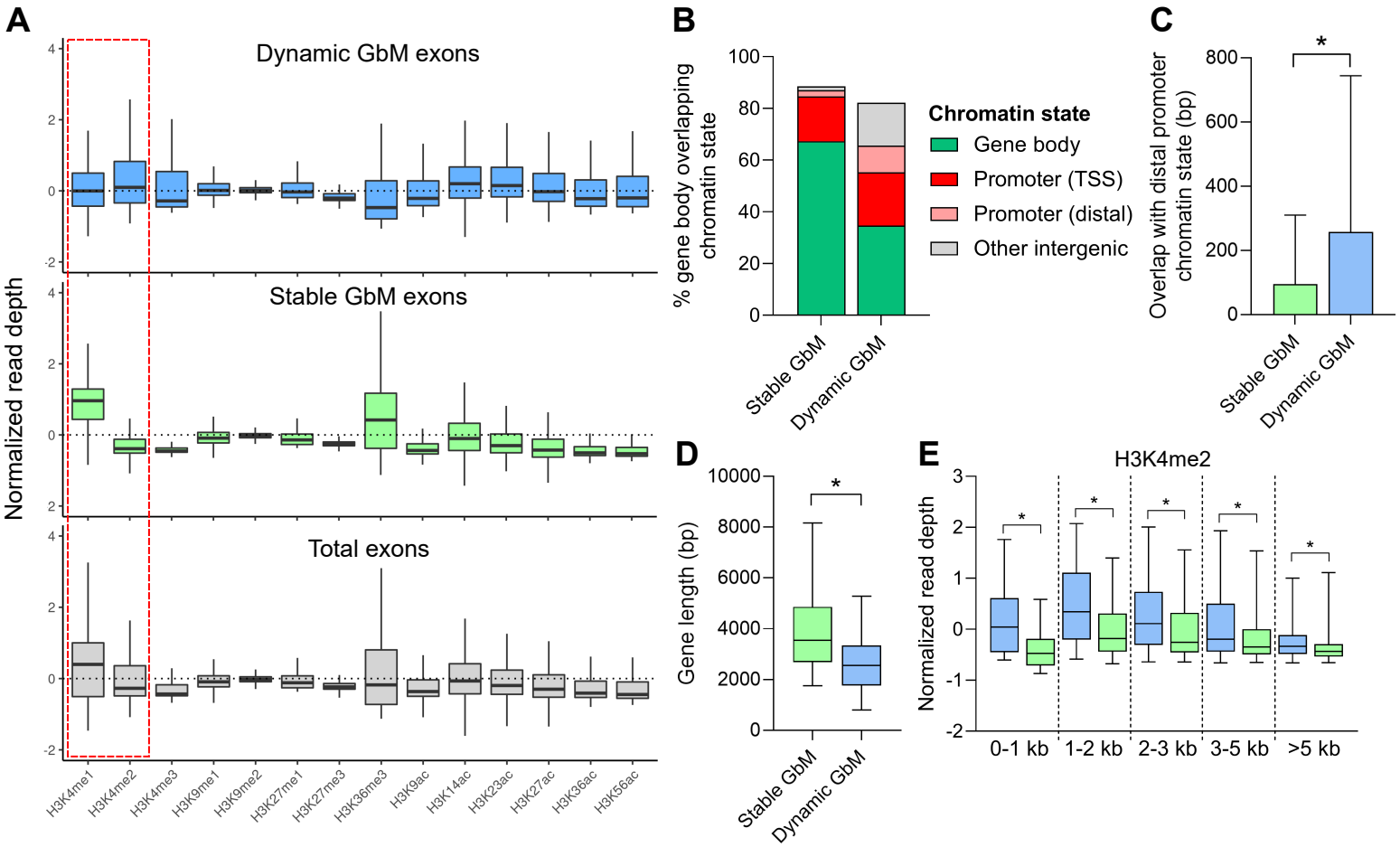


Figure 4



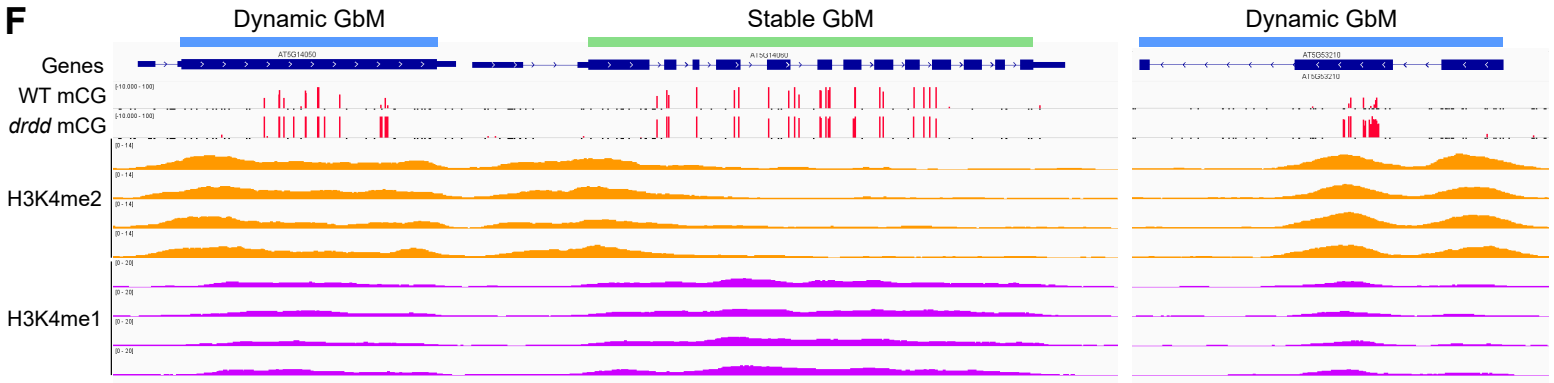


Figure 5

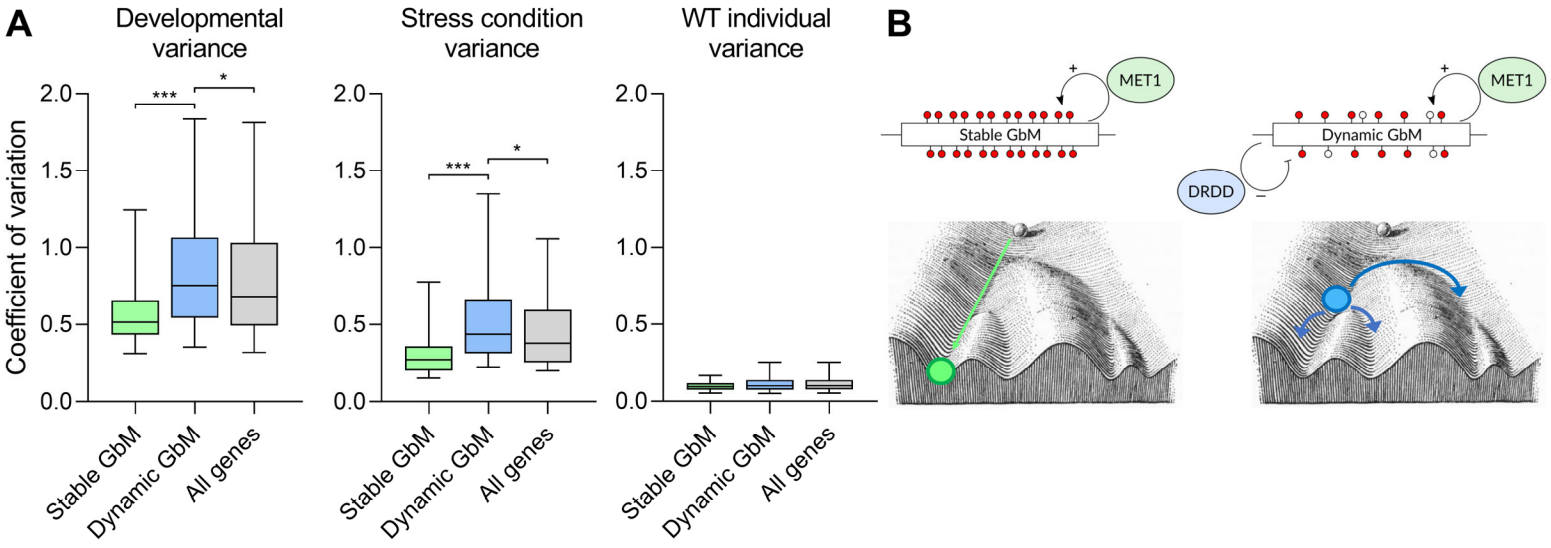


Figure 6

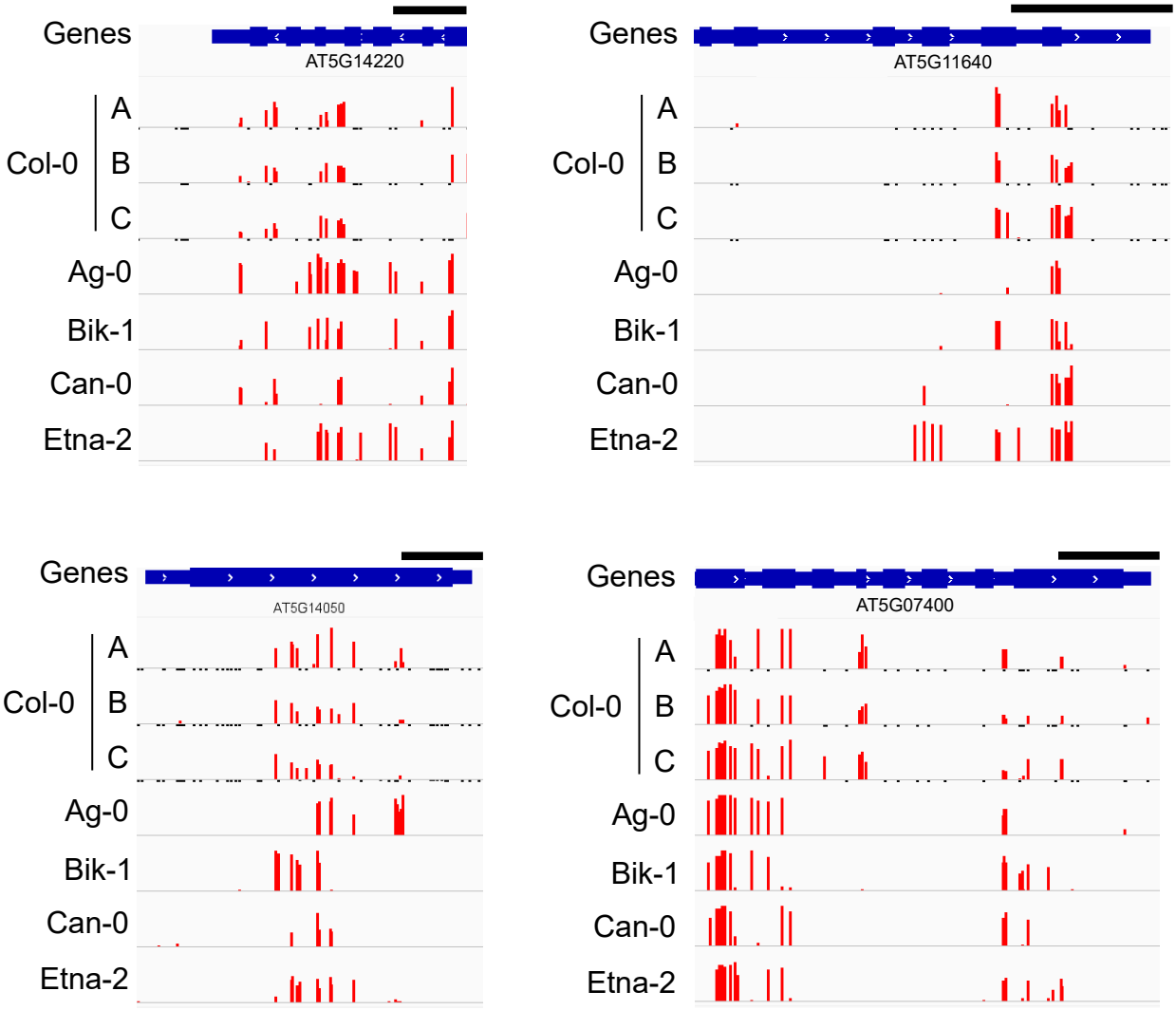


Figure S1

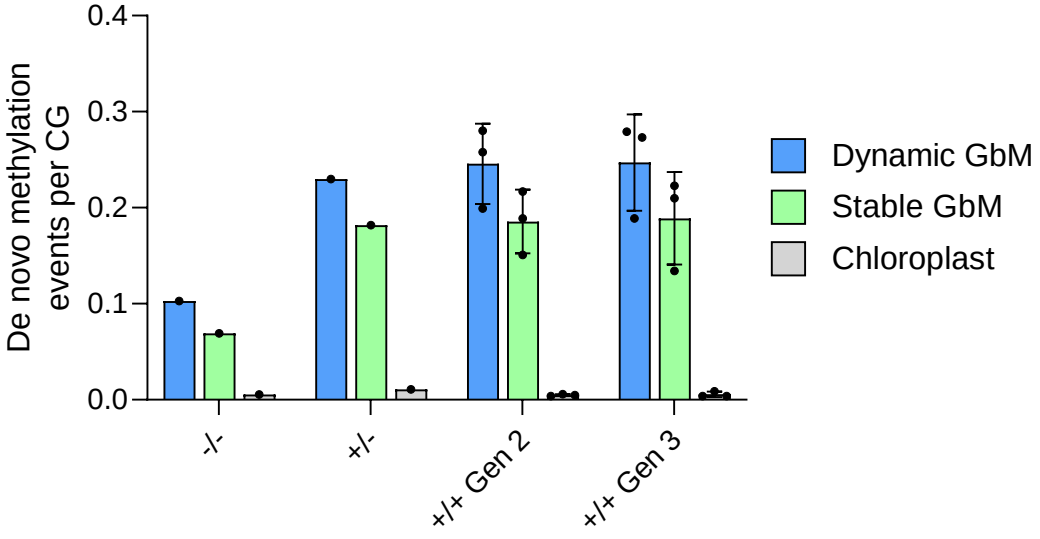

De novo methylation events per CG
0.4
0.3
0.2
0.1
0.0
-/-
+/-
+/+ Gen 2
+/+ Gen 3
Dynamic GbM
Stable GbM
Chloroplast

Figure S2

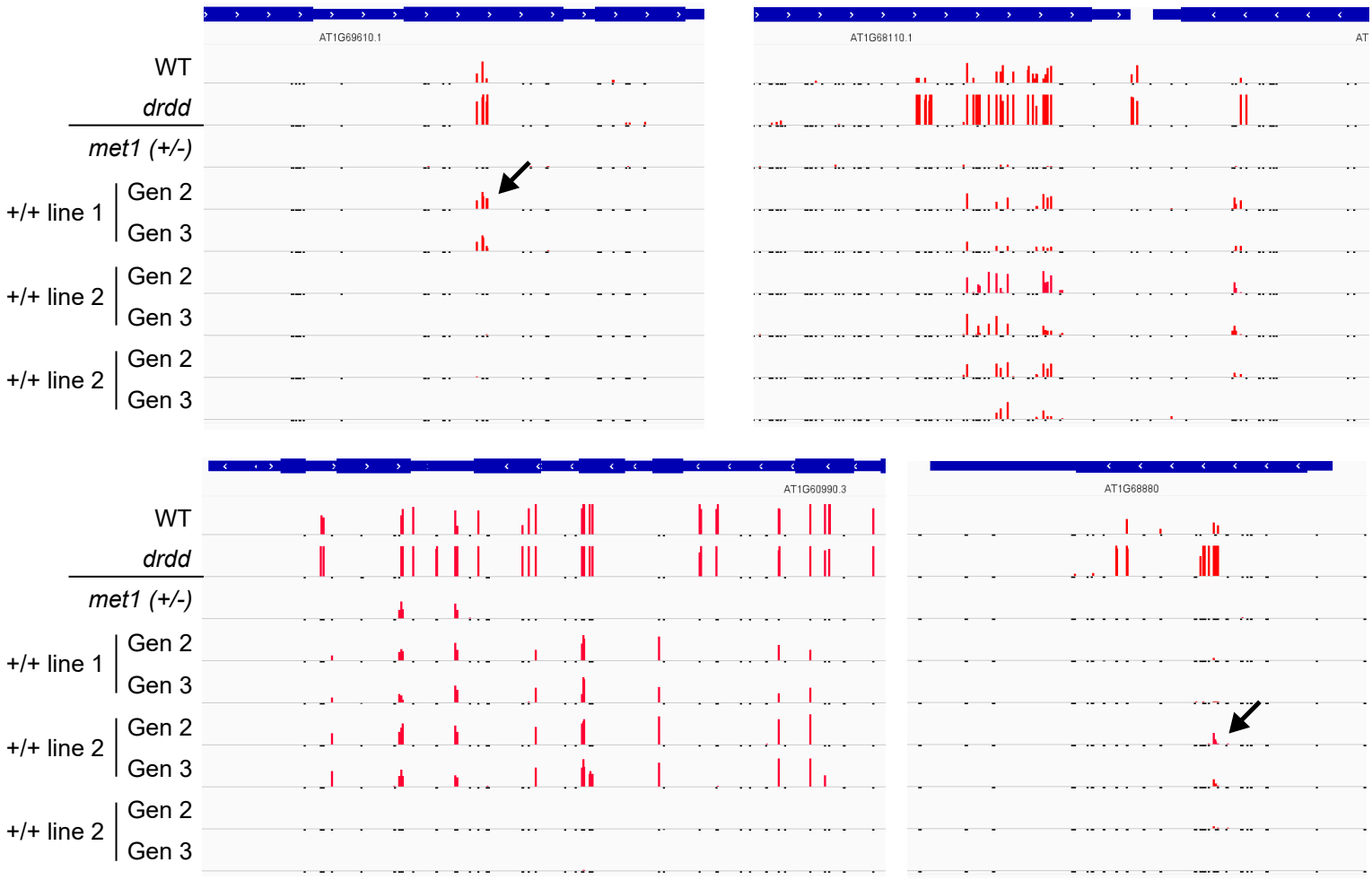


Figure S3

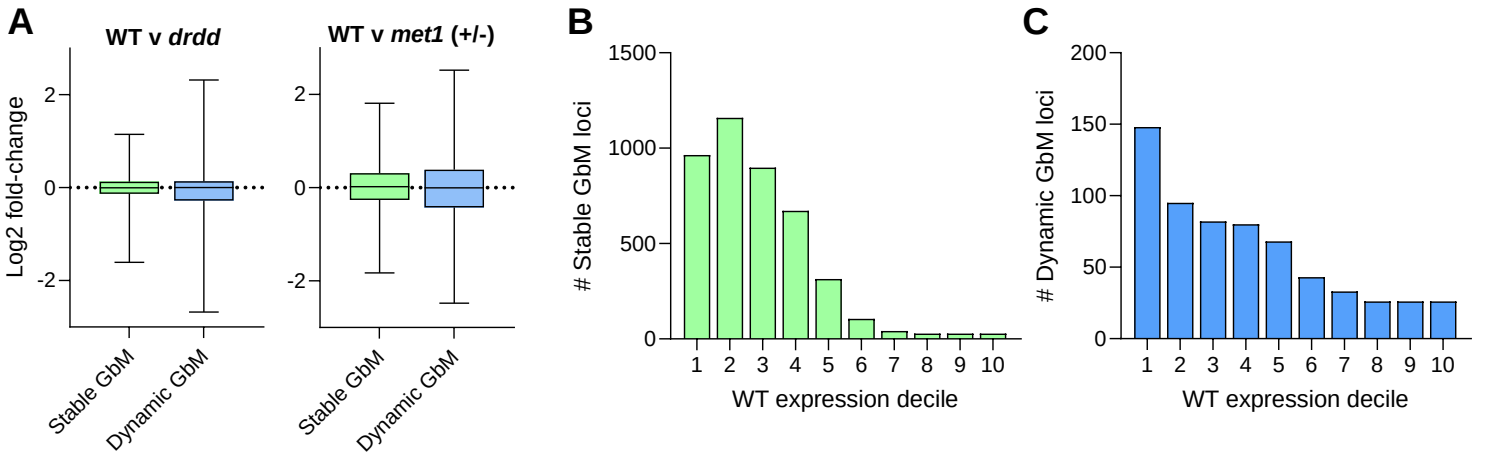

A
WT v drdd
WT v met1 (+/-)
Log2 fold-change
2
0
-2
Stable GbM
Dynamic GbM
B
Stable GbM loci
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10
WT expression decile
C
Dynamic GbM loci
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10
WT expression decile

Figure S4

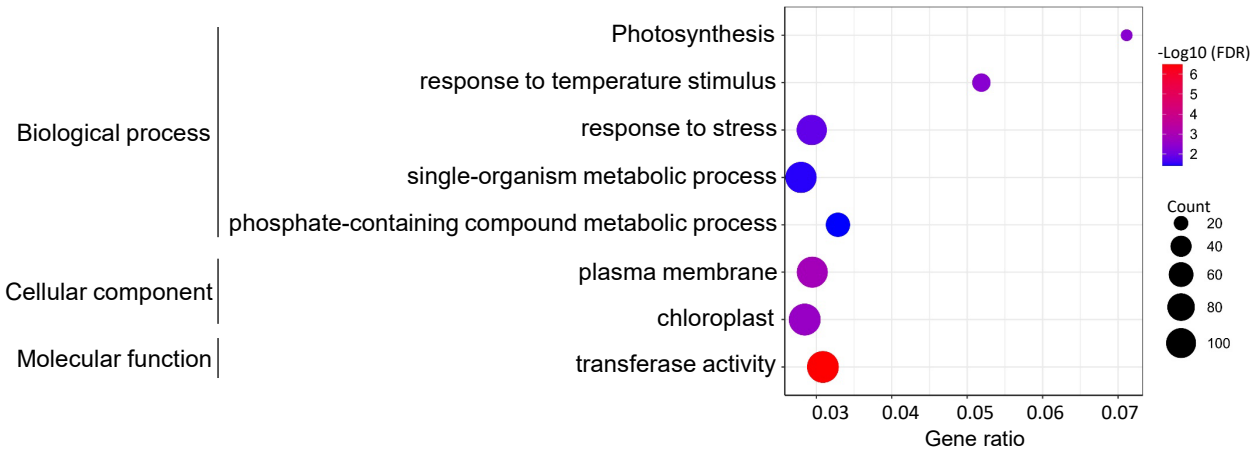

Biological process
Cellular component
Molecular function
Photosynthesis
response to temperature stimulus
response to stress
single-organism metabolic process
phosphate-containing compound metabolic process
plasma membrane
chloroplast
transferase activity
0.03
0.04
0.05
0.06
0.07
Gene ratio
-Log10 (FDR)
6
5
4
3
2
Count
20
40
60
80
100

Figure S5

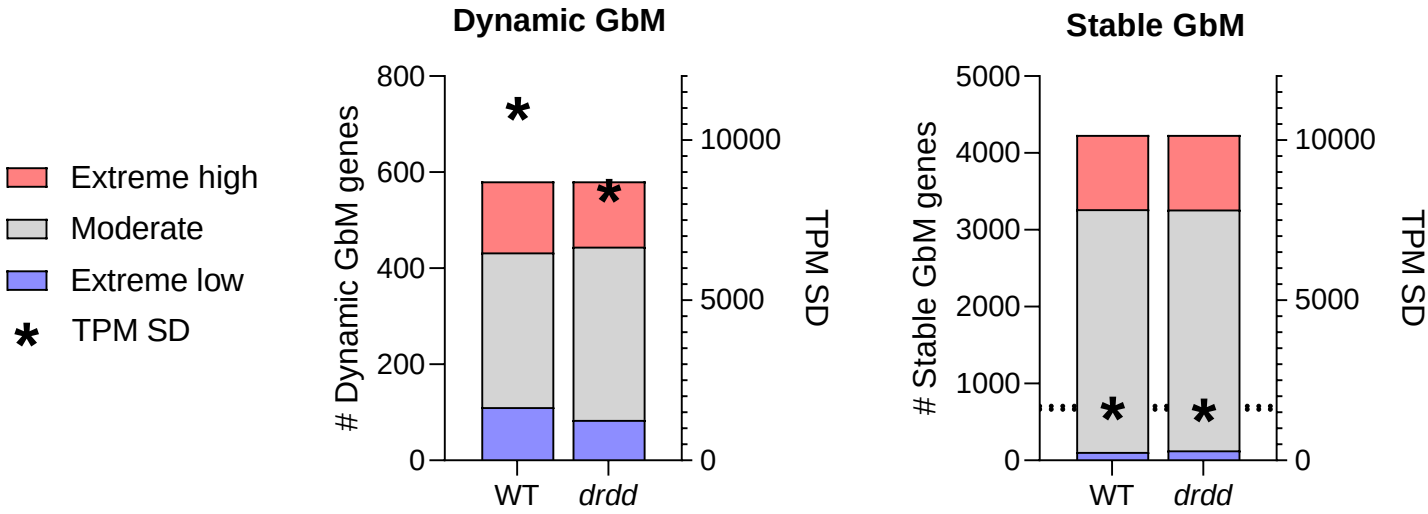

Extreme high
Moderate
Extreme low
TPM SD
Dynamic GbM
Dynamic GbM genes
800
600
400
200
0
10000
5000
0
TPM SD
WT
drdd
Stable GbM
Stable GbM genes
5000
4000
3000
2000
1000
0
10000
5000
0
TPM SD
WT
drdd

Figure S6