Journal of Zhejiang University-SCIENCE B in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Review

https://doi.org/10.1631/jzus.B2600175

Population epigenomics of plant DNA methylation: perspectives for molecular breeding

Xi LONG[1*], Juyun ZHENG[2*], Lei FANG[1], Ting ZHAO✉

[1]*Zhejiang Provincial Key Laboratory of Crop Genetic Resources, the Advance Seed Institute, Key Laboratory of Plant Factory Generation-adding Breeding, Ministry of Agriculture and Rural Affairs, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, Zhejiang, China.*

[2]*Cotton Research Institute, Xinjiang Academy of Agricultural Sciences, Key Laboratory of Cotton Genetic Improvement and Smart Production of Xinjiang, Urumqi, Xinjiang, China*

Abstract: DNA methylation, a key epigenetic modification, contributes to regulatory diversity in plants and plays important roles in phenotypic plasticity and environmental adaptation. Recent population epigenomic studies have revealed extensive methylation variation across plant genomes, characterized genomic distribution and genetic architecture, and demonstrated both its dependence on and partial independence from underlying DNA sequence variation. These studies have further linked methylation variation to gene expression regulation, domestication, local adaptation, and complex trait variation through integrative analyses and epigenome-wide association research. Advances in high-throughput sequencing have enabled the generation of population-scale plant methylomes across natural accessions, wild relatives, landraces, and breeding germplasm, providing new avenues to systematically investigate epigenomic diversity and its functional consequences. In parallel, public methylome resources, epigenomic databases and multi-omics analytical platforms are accelerating comparative and cross-species analyses of plant epigenomic variation. Future efforts that integrate population methylomes with genomics, transcriptomics, chromatin accessibility, three-dimensional genome organization, deep learning-based analytical frameworks, and targeted epigenome editing are expected to facilitate the discovery of epigenetic breeding markers, promoting the application of epigenomic variation in crop improvement.

Key words: DNA methylation; Population-scale epigenomics; Single methylation polymorphism; Gene expression; Epigenome-wide association study

✉ Ting ZHAO, 0618041@zju.edu.cn
* The two authors contributed equally to this work
Ting ZHAO, https://orcid.org/0000-0001-5102-0157

Received Mar. 16, 2026; Revision accepted June 26, 2026;
Crosschecked xxx. xx, 20xx

1 Introduction

Genetics examines the mechanisms that govern the transmission of traits across generations. Traditional evolutionary theory has largely emphasized that DNA sequence variation is generated by mutation and recombination and shaped over time by genetic drift, migration and natural selection (Mueller et al., 2025). However, this sequence-based view of inheritance has been increasingly expanded by evidence demonstrating that transgenerational inheritance may also be mediated by epigenetic mechanisms that do not alter the primary DNA sequence (Grossniklaus et al., 2013).

DNA methylation is one of the most extensively studied among epigenetic modifications, which is catalyzed by DNA methyltransferases that add a methyl group to the C5 position of cytosine to generate 5- methylcytosine (5mC). 5mC was first reported in bacterial nucleic acids in 1925 (Treat B and Robert D, 1925) and was later detected in plant and animal genomes in 1950 (Wyatt, 1950). DNA methylation is widely observed across diverse organisms, although its sequence contexts, molecular forms and biological functions differ substantially among bacteria, plants and animals (Zemach and Zilberman, 2010).

In plants, cytosine methylation occurs in CG, CHG and CHH contexts, where H represents A, C or T (Guseinov and Vanyushin, 1975). These methylation states are established, maintained and removed by coordinated methyltransferase and demethylase pathways, including *MET1*, *CMTs*, *DRM2*, *DME*, and *ROS1* (Law and Jacobsen, 2010; Erdmann and Picard, 2020; Xie et al., 2025). CG methylation is generally the most stable across mitotic and meiotic divisions, whereas CHH methylation is often more dynamic and tissue- or development-specific (Bouyer et al., 2017; Kawakatsu et al., 2017). Through these context-dependent patterns, DNA methylation has many regulatory roles such as in transposon silencing, gene expression, genomic imprinting, developmental transitions, stress responses, and agriculturally important traits. Representative examples of plant genes and traits regulated by DNA methylation are summarized in Table 1, illustrating the broad involvement of methylation in development, stress adaptation, metabolism, and crop domestication (**Table 1**).

Table 1 Representative plant genes and traits regulated by DNA methylation

Gene	Species	Phenotype
ACT1 (Song et al., 2025)	*Oryza sativa* (Rice)	Promoter hypomethylation enhances cold tolerance and adaptability to high-latitude regions in rice
COL2 (Song et al., 2017)	*Gossypium hirsutum* (Cotton)	Methylation dynamics regulate photoperiod sensitivity
FWA (Soppe et al., 2000)	*Arabidopsis thaliana* (Thale cress)	Epigenetic allele causes a late-flowering phenotype
CMT2 (Jiang et al., 2024)	*Arabidopsis thaliana*	Reduced genome-wide CHH methylation, improving thermotolerance and UV-B resistance
OsSPL14 (Miura et al., 2010)	*Oryza sativa*	Promoter methylation changes are associated with altered chromatin structure, enhanced *OsSPL14* expression, and improved plant architecture and yield
tb1 (Xu et al., 2020)	*Zea mays* (Maize)	DMR methylation regulates tiller number, affecting plant architecture
Lcyc (Cubas et al., 1999)	*Linaria vulgaris* (Yellow toadflax)	Promoter hypermethylation leads to radial flower symmetry

Karma (Ong-Abdullah et al., 2015)	*Elaeis guineensis* (African oil palm)	Loss of methylation causes mantled somaclonal variation, affecting clonal propagation
SlDML2 (Lang et al., 2017)	*Solanum lycopersicum* (Tomato)	Genome-wide demethylation acts as a switch for fruit ripening initiation
Cnr (Manning et al., 2006)	*Solanum lycopersicum*	Promoter hypermethylation inhibits ripening
CAMTA2 (Song et al., 2024)	*Pyrus pyrifolia* (Asian pear)	DNA hypermethylation at a regulatory region is associated with increased *CAMTA2* expression and delayed fruit ripening
VTE3 (Quadrana et al., 2014)	*Solanum lycopersicum*	Methylation dynamics of a promoter SINE transposon regulate vitamin E synthesis
Ipa1-241 (Song, et al., 2017)	*Zea mays*	DNA methylation-mediated paramutation leads to phytic acid synthesis deficiency

Despite extensive knowledge of DNA methylation pathways and individual epialleles, it is incompletely understood how naturally occurring methylation variation is distributed, inherited, regulated, and selected at the population scale. Population epigenomics addresses this issue by treating DNA methylation as a quantitative molecular phenotype across diverse germplasm. In this review, we summarize recent progress in plant population methylomics, discuss the genetic and functional architecture of methylation variation, and highlight emerging opportunities for epigenetic marker discovery, functional validation, and molecular breeding applications.

2 Population-scale methylome resources in plants

Population-scale methylome resources provide the foundation for characterizing natural epigenomic diversity across wild accessions, landraces, domestication panels, and breeding germplasm (Fig 1a). By profiling DNA methylation across diverse individuals, these datasets enable the discovery of methylation polymorphisms and their integration with genetic, transcriptomic, environmental, and phenotypic variation (Fig 1a). Whole-genome bisulfite sequencing remains the dominant approach because it quantifies cytosine methylation at the single-base resolution in CG, CHG and CHH contexts.

Population-scale methylome resources have been established for the model species *Arabidopsis thaliana* (Schmitz et al., 2013; Kawakatsu et al., 2016), including global natural accessions, and for several major crops, such as rice (Cao et al., 2023), maize (Xu, et al., 2020), soybean (Jiang et al., 2026), tomato (Guo et al., 2023), lettuce (Cao et al., 2024), pear (Song, et al., 2024), and cotton (Zhao et al., 2024)(Table 2). These datasets represent a wide range of biological materials, including wild and cultivated accessions, domestication panels, diversity collections, and representative breeding materials. In many cases, methylomes are generated from tissues directly relevant to agronomic traits, including seedling leaves, shoot apices, cotton fibers, fruit flesh, and pericarp. Such tissue-specific sampling strategies enhance the utility of population methylomes for linking epigenomic variation to developmental processes and agriculturally important traits. Collectively, these datasets reveal a clear transition from single-reference methylome profiling toward population-level, tissue-informed and multi-omics epigenomic resources.

Table 2 Summary of population-scale DNA methylation datasets in plants

Species	Sample number	Germplasm	Tissue	Multi-omics integration

Arabidopsis thaliana (Schmitz, et al., 2013; Kawakatsu, et al., 2016)	1,107	Natural accessions	Whole seedling	WGBS, RNA-seq
Oryza sativa (Cao, et al., 2023)	95	Wild rice, cultivated *indica*, cultivated *japonica*, Aro, Aus, weedy *indica*, weedy *japonica*	Third leaf (30-day)	WGBS
Zea mays (Xu, et al., 2020)	37	Modern inbreds, landraces, teosintes / wild relatives	Seedling	WGS, WGBS
Glycine max (Shen et al., 2018)	45	Wild, landraces, cultivars	Leaf	WGBS
Solanum lycopersicum (Guo, et al., 2023)	96	Wild, landraces, cultivars	Pericarp	WGBS, RNA-seq, Metabolome
Lactuca sativa (Cao, et al., 2024)	52	Cultivars, wild relatives	Whole seedling	WGBS, RNA-seq, Chromatin Accessibility, Interaction Data
Pyrus pyrifolia (Song, et al., 2024)	41	Wild, landraces, improved cultivars	Fruit flesh	WGBS, RNA-seq, WGS
Gossypium hirsutum (Zhao, et al., 2024)	207	Cultivars	20 days post-anthesis fibers	WGBS, RNA-seq, WGS
Glycine max (Jiang, et al., 2026)	1,102	Wild soybeans, landraces, and improved cultivars	Second trifoliate leaves	WGBS

3 Origin and architecture of methylation variation

Population-level methylation variation can arise from three major sources (Becker et al., 2011). First, spontaneous epimutations result from the imperfect maintenance, establishment or removal of DNA methylation and occur at rates substantially higher than DNA sequence mutations (Van Der Graaf et al., 2015). Second, genetically associated methylation variation can be caused by local or distal variants, such as SNPs, transposable element (TE) insertions, structural variants, copy-number changes, as well as mutations in methylation-related genes such as *MET1* (Reinders et al., 2009) and *ROS1* (Tang et al., 2024). Third, environmentally or developmentally induced methylation changes may occur in response to stress, biotic interactions, or developmental transitions. Although many such changes are transient, a subset may persist across generations, especially when reinforced by genetic background or recurrent environmental selection.

Population-scale BS-seq allows methylation variation to be quantified at both regional and single-cytosine resolution. At the regional level, differentially methylated positions and regions—DMPs and DMRs—are used to compare individuals, populations, or domestication groups (Richards, 2006; Xu, et al., 2020; Song, et al., 2024; Zhao, et al., 2024). At the single-cytosine level, single methylation

polymorphisms (SMPs) describe methylation-state variation at individual cytosines and can be encoded as methylated, unmethylated, or partially methylated states (Zhao, et al., 2024; Jiang, et al., 2026). This encoding enables population-genetic analyses of methylation diversity, including minor epiallele frequency, methylation disequilibrium, population structure, and methylation-based epigenotypes, or "menotypes" (Zhao et al., 2018).

The persistence of methylation variation in natural populations depends strongly on sequence context. CG methylation is generally the most meiotically stable epigenetic modification, which can be transmitted across generations, allowing methylation variants to accumulate over time (Van Der Graaf, et al., 2015). In contrast, CHH methylation is often erased and re-established during reproductive or embryonic development, making it more dynamic and less stably inherited across generations (Hollwey et al., 2023). SMP-based analyses have revealed that methylation polymorphisms differ markedly among sequence contexts. In cotton, CHH-SMPs have shown the highest minor allele frequency, followed by SNPs, CHG-SMPs and CG-SMPs, suggesting that CHH methylation is more dynamic at the population level (Zhao, et al., 2024). Methylation disequilibrium analyses further revealed that CG methylation variation can be correlated across regions of up to approximately 50 bp, whereas CHG and CHH methylation tend to show shorter-range associations (Zhao, et al., 2024). These patterns are consistent with the relatively stable inheritance of CG methylation and the more dynamic regulation of non-CG methylation.

Genomic localization provides another important dimension for characterizing population-level methylation variation. SMPs and DMRs are often enriched within or near genes, where they may affect transcriptional regulation and phenotypic variation (Schmitz, et al., 2013; Cao, et al., 2024; Zhao, et al., 2024). However, comparative analyses across genomic compartments indicate that many SMPs originate from TE regions. TE-associated SMPs often show lower minor allele frequencies and higher methylation conservation than SMPs in genes or single-copy regions, consistent with the role of DNA methylation in transposon silencing and genome defense (Zhao, et al., 2024; Jiang, et al., 2026). Therefore, the genomic distribution of methylation variants reflects a balance between epigenetic plasticity in regulatory regions and evolutionary constraint in transposon-rich regions.

4 Genetic control versus epigenetic independence

A central question in plant population epigenomics is how to distinguish methylation differences driven by underlying DNA sequence variation from methylation variants that arise and persist relatively independently of genotype. For an epigenetic variant to contribute independently to adaptive evolution, it should satisfy several stringent criteria: it must arise independently of genotype, be stably inherited across generations, and induce heritable phenotypic changes that can be subject to natural selection (Mueller, et al., 2025).

DNA methylation levels can be treated as quantitative molecular phenotypes that mediate or reflect the effects of genetic variation (Smith et al., 2014)(Fig 1b). Many DNA methylation polymorphisms are closely associated with local sequence variants, including *cis*-SNPs, TE insertions, structural variants, copy-number variants, and variants in regulatory elements. Mutations in genes involved in methylation establishment, maintenance or removal may also generate broader methylome changes (Lindroth et al., 2001; Lang, et al., 2017; Zhao et al., 2021; He et al., 2022). These observations indicate that a substantial fraction of methylation variation has no independent epigenetic cause but is rather the consequence of

genetic differences.

The proportion of methylation variation explained by genetic factors varies markedly among species, methylation contexts, genomic regions, and analytical methods. In cotton, *cis*-acting meQTLs account for 5.82% of CG sites, 2.64% of CHG sites and 1.52% of CHH sites (Zhao, et al., 2024). In maize, fewer than 40% of DMRs are associated with nearby SNPs, suggesting that much of the observed methylation variation cannot be explained by local sequence variants alone (Xu et al., 2019). Similarly, local sequence variants explain approximately 22.5% of DMRs in soybean (Shen, et al., 2018) and roughly 20% of methylation variation in *Arabidopsis thaliana* (Schmitz, et al., 2013). These estimates suggest that genetic effects are widespread but incomplete. The broad range of reported values likely reflects true biological differences among species, as well as differences in sample size, sequencing depth, methylation context, statistical model, marker density, and definitions of cis-regulatory windows. Moreover, estimates based only on nearby SNPs may underestimate the contribution of structural variants, TE polymorphisms, rare variants, and distal trans-acting loci.

In contrast, independent epialleles generally refer to methylation states that show no obvious association with local DNA sequence variation but can be stably maintained across cell divisions or generations. Such epialleles may arise from spontaneous epimutations, environmentally induced methylation changes, TE silencing, RNA-directed DNA methylation, developmental epigenetic reprogramming, or stochastic failures in methylation maintenance and demethylation pathways. If these methylation states are meiotically stable and alter gene expression or agronomic traits, they may represent a novel class of heritable variation that is potentially useful for crop improvement.

When a methylation polymorphism is controlled by nearby or distal genetic variants, it may function as an intermediate molecular phenotype through which genetic variation influences gene expression and downstream traits. Alternatively, it may simply be a linked passenger signal that is correlated with the causal genetic variant but has no direct functional effect. Therefore, associations between methylation and traits should not be automatically interpreted as evidence for an independent epigenetic effect. Without accounting for genotype, population structure and linkage disequilibrium, methylation–trait associations may overestimate the causal contribution of DNA methylation to phenotypic variation.

5 Functional interpretation through eQTM and multi-omics

Expression quantitative trait methylation (eQTM) analysis provides a population-scale framework for linking DNA methylation variation with gene expression variation (Fig 1c). By associating methylation states or methylation levels with transcript abundance, it can identify methylation loci whose variation is correlated with expression changes in nearby or distal genes. This suggests that methylation variation, particularly CG methylation in upstream regulatory regions, can contribute substantially to transcriptional diversity. The subsequent sections elaborate on the mechanistic basis of these associations, starting with eQTM.

Joint multi-omics analysis is essential for filtering out noise and anchoring core methylation-sensitive regulatory genes (Fig 1c). Large-scale paired methylomic and transcriptomic datasets allow researchers to quantify how methylation state fluctuations at specific regulatory elements, including promoters, enhancers, silencers, and TE-derived regulatory regions, affect target gene expression. When combined with chromatin accessibility data, transcription factor binding profiles, and three-dimensional chromatin interaction maps, these datasets can reveal whether methylation variation alters regulatory element activity, transcription factor occupancy or enhancer–promoter communication (Fig 1c).

The construction of methylation-mediated gene regulatory networks further facilitates the integration of population-scale methylomes, transcriptomes, chromatin accessibility profiles, genotypes, and phenotypes. Such networks can identify key regulatory elements, hub transcription factors, and downstream genes through

which methylation variation contributes to complex quantitative traits, including stress resistance, metabolite synthesis, fiber development, and fruit quality. This integrative framework bridges the mechanistic gap between epigenetic markers and complex phenotypes and provides a more precise basis for evaluating the functional contribution of methylation variation in plant populations.

Promoter methylation is often associated with transcriptional repression in both plants and animals (Zhang et al., 2018), although its impact depends on the type, proportion, genomic position, and density of methylation (Yu et al., 2013). Alterations in methylation can also reshape chromatin states, modifying the three-dimensional genome architecture (Zhong et al., 2021). In an *in vitro Arabidopsis* binding assay, more than 75% (248/327) of tested TFs showed methylation-sensitive DNA binding preferences (O'malley et al., 2016; Heberle and Bardet, 2019). In maize, trait-associated DMRs such as *vgt1*-DMR and *tb1*-DMR form chromatin loops with their downstream target genes, as demonstrated by Hi-C and ChIP analyses in the B73 reference genome (Xu, et al., 2020).

Expression quantitative trait methylation (eQTM) creates statistical associations between methylation variation and gene expression (Fig 1c). Compared with single-genome analyses, eQTM provides a more comprehensive and accurate understanding of the relationship between DNA methylation and gene expression. Studies within individual populations typically identify hundreds of genes subject to methylation-mediated regulation, often enriched in specific functional categories. Moreover, increasing the sample size markedly expands the number of detectable eQTMs (Meng et al., 2016). For instance, a large-scale cotton population study identified over 5,000 *cis*-acting eQTMs among tens of thousands of protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs). These eQTMs regulate approximately 5.7% of expressed PCGs and up to 29% of expressed lncRNAs, suggesting that lncRNA expression is particularly sensitive to methylation variation (Zhao, et al., 2024). The relationship between DNA methylation and gene expression is influenced by both sequence context and the spatial proximity of DMRs to transcription start sites (TSSs). In maize and cotton, CG and CHG methylation levels are typically negatively correlated with gene expression, whereas CHH methylation can show positive correlations (Xu, et al., 2019; Zhao, et al., 2024).

In cotton, over 90% of expression quantitative trait methylations (eQTMs) have been significantly associated with CG methylation, underscoring its predominant role in transcriptional regulation. Approximately 31% of PCGs and nearly 60% of long non-coding RNAs (lncRNAs) are concurrently influenced by all three methylation contexts. Notably, 90% of *cis*-eQTMs are positioned upstream of their target genes (Zhao, et al., 2024). While eQTM mapping successfully establishes widespread statistical associations between methylation and transcription, deciphering the exact underlying molecular mechanisms requires multi-omics integration. Population-scale ATAC-seq and cistrome studies have already been reported for some crops (Engelhorn et al., 2025; Zhu et al., 2025). Furthermore, integrating ATAC-seq with DAP-seq in future research could provide mechanistic insights into how DNA methylation regulates gene expression.

6 Roles in domestication, adaptation and complex traits

DNA methylation can contribute to phenotypic variation by modulating gene expression (Zhang, et al., 2018), TE activity (Xu, et al., 2020), transcription factor binding (O'malley, et al., 2016; Heberle and Bardet, 2019), and chromatin states (Zhong, et al., 2021), thereby linking genetic background and environmental cues to complex traits.

In natural plant populations, methylation variation is frequently associated with geography, climate and ecological gradients. For example, the *Arabidopsis* 1001 Epigenomes Project revealed strong associations

between methylation patterns and environmental factors, with TE methylation positively correlated with latitude and precipitation but negatively correlated with elevated temperature, and gene body methylation associated with colder winter conditions (Kawakatsu, et al., 2016). Studies along elevation gradients and artificial selection experiments further suggest that environmental pressures can reshape population-level methylation profiles (Singh et al., 2024). In rice, cold-acclimation-related methylation sites in the *ACT* promoter exhibited a "higher in the south, lower in the north" gradient, reflecting methylation differentiation under distinct climatic conditions (Song, et al., 2025).

Domestication and crop improvement are further factors that reshape methylation landscapes (Fig 1d). The overall methylation levels tend to decrease in rice (Cao, et al., 2023) and tomato (Guo, et al., 2023) but increase in pear (Song, et al., 2024) and lettuce (Cao, et al., 2024), suggesting that domestication-associated methylation changes are species-specific and may depend on genome architecture, selected traits, tissue type, and breeding history. Such changes may generate or fix stable epialleles affecting agronomic traits, including fruit ripening, metabolite accumulation, stress responses, and developmental transitions.

Evidence from epiRILs demonstrates that heritable methylation variation can directly contribute to quantitative traits (Trerotola et al., 2015). For example, in *Arabidopsis*, epiRILs derived from wild-type and *DDM1*-mutant parents showed substantial variation in flowering time and plant height, with some epialleles stably maintained for more than eight generations (Johannes et al., 2009). In natural populations, EWAS have identified methylation loci associated with metabolic traits, fiber yield and quality, and metabolite accumulation in *Arabidopsis* (Xu, et al., 2019), cotton (Zhao, et al., 2024), and tomato (Guo, et al., 2023)(Fig 1e). These findings suggest that DNA methylation can reveal trait-associated regulatory variation missed by sequence-based mapping. Nevertheless, methylation–trait associations require cautious interpretation because methylation marks may be causal, genetically driven intermediate phenotypes, or linked passenger signals.

Network-based methods may further identify core regulatory genes, methylation-sensitive transcription factors, and downstream pathways involved in complex trait regulation. Such frameworks are essential for bridging the mechanistic gap between epigenetic markers and breeding-relevant phenotypes; integrating EWAS with meQTL, eQTM, transcriptomics, environmental data, and functional validation will be essential for clarifying the true contribution of methylation variation to adaptation, domestication and agronomic trait diversity.

7 Resources and analytical platforms

The accumulation and public release of plant methylome datasets have provided important resources for epigenomic research (Fig 1f). Existing databases and analytical platforms, such as AtMAD (Lan et al., 2021), cross-species angiosperm methylome resources, and MethBank 4.0, have enabled the organization, visualization and comparative analysis of DNA methylation profiles across species, tissues and developmental stages (Niederhuth et al., 2016; Zhang et al., 2023)(Fig 1f). These resources have greatly facilitated the identification of conserved methylation features, differentially methylated regions and candidate regulatory loci.

Despite the above progress, most current methylome databases still have limited population-scale coverage and insufficient integration with other omics layers (Kawakatsu, et al., 2016). In practical research, particularly for polyploid crops or species with large and structurally complex genomes, a standalone epigenomic database has limited interpretive power. DNA methylation variation must be analyzed in the context of pangenome variation (Tao et al., 2019; Song et al., 2020), structural variants (Alonge et al., 2020), TE polymorphisms, transcriptomic diversity, chromatin accessibility, and phenotypic variation (Lu et al., 2017; Maher et al., 2018). Without these layers, it is difficult to determine whether a methylation variant is genetically driven, functionally associated with gene expression, or causally linked to complex traits.

In view of the above, the current trend is shifting toward comprehensive multi-omics databases and interactive visualization platforms that integrate pangenomes, variomes, transcriptomes, epigenomes, regulatory annotations, and phenomes (Springer and Schmitz, 2017; Gui et al., 2020; Cui et al., 2023). Such platforms should support deep integrative analyses, including meQTL mapping, eQTM analysis, EWAS,

methylation-mediated regulatory network construction, and the pangenome-aware visualization of methylation variation (Rakyan et al., 2011; Dubin et al., 2015; Kremling et al., 2019). These integrative resources will help filter noise, prioritize core regulatory genes, distinguish genetically dependent methylation changes from candidate independent epialleles, and bridge epigenetic markers with complex phenotypes. Building such resources will be essential for advancing plant population epigenomics and incorporating epigenetic information into precision breeding.

8 Advancing toward epigenetic breeding

The spontaneous formation of epialleles and their potential transmission across generations make plants an attractive system for studying epigenetic inheritance and exploring the use of epigenetic variation in crop improvement. Epigenomic modifications can strengthen, weaken or generate regulatory interactions, thereby influencing gene expression, chromatin organization and phenotypic diversity (Liu et al., 2025).

8.1 Scalable methylation profiling for breeding applications

A major prerequisite for epigenetic breeding is the ability to profile DNA methylation accurately and cost-effectively across large numbers of samples. However, accurate methylation quantification typically requires bisulfite sequencing at substantially higher coverage than conventional whole-genome sequencing used for SNP detection, making population-scale methylome profiling expensive and technically demanding.

However, targeted methylation profiling platforms may provide a more practical solution. Methylation arrays or targeted methylation sequencing panels can be designed to focus on informative differentially methylated regions, single methylation polymorphisms, methylation-sensitive regulatory elements, and trait-associated epigenetic markers (Fig 1g). However, most currently available methylation arrays are optimized for human and animal studies, and crop-specific methylation arrays remain largely unavailable. In this regard, the accumulation of plant methylome datasets provides an important foundation for designing crop-adapted methylation platforms. For breeding applications, these markers must remain informative across genetic backgrounds, tissues, developmental stages, environments, and management conditions.

Third-generation sequencing technologies, such as Oxford Nanopore Technologies, provide important advantages for methylation analysis, including long reads, rapid sequencing, and direct detection of base modifications such as 5mC and 6mA without bisulfite conversion (Liu and Conesa, 2025)(Fig 1g). These platforms enable the simultaneous acquisition of genetic and epigenetic information and are particularly useful for crops with complex genomes, abundant TEs, or extensive structural variation. Nevertheless, sequencing cost, base-modification calling accuracy, analytical standardization, and throughput remain major limitations for routine breeding-scale applications.

8.2 Deep learning and pre-trained models for epigenotype-to-phenotype prediction

As methylomes, transcriptomes, chromatin accessibility profiles, pangenomes, and phenomes continue to accumulate, epigenetic breeding faces a challenge not only in data generation but also in data interpretation. Traditional statistical models remain essential for EWAS, meQTL, eQTM, and genomic prediction, but they often have limited capacity to capture nonlinear, context-dependent, and multi-layer regulatory relationships (Albert and Kruglyak, 2015). This limitation is particularly relevant to

DNA methylation because its regulatory effect depends on sequence context, methylation context, genomic position, chromatin accessibility, transcription factor binding, tissue identity, developmental stage, and environmental exposure. Therefore, computational frameworks that can jointly model sequence features, epigenomic states, regulatory annotations, and phenotypic variation are increasingly needed for epigenotype-to-phenotype prediction.

Deep learning provides one such framework by learning regulatory patterns directly from genomic and epigenomic data. Early sequence-to-function models, such as Basset (Kelley et al., 2016) and DeepSEA (Zhou and Troyanskaya, 2015), have demonstrated that DNA sequence can be used to predict regulatory features, including chromatin accessibility and transcription factor binding. These models established the concept that regulatory grammar is learnable from sequence and provided a foundation for predicting the functional effects of non-coding variation. Although these early models were not designed specifically for plant DNA methylation, they introduced a general strategy for linking sequence variation to regulatory activity.

More recently, pre-trained genomic language models have expanded this strategy by learning generalizable representations from large-scale genome sequences. Models such as DNABERT (Ji et al., 2021), Nucleotide Transformer (Dalla-Torre et al., 2025), AgroNT (Mendoza-Revilla et al., 2024), and PlantCaduceus (Zhai et al., 2025) use Transformer or related architectures to capture sequence patterns across broader genomic contexts. These models are particularly useful for regulatory annotation, variant prioritization and sequence-to-function prediction. Plant-oriented models, including AgroNT (Mendoza-Revilla, et al., 2024) and PlantCaduceus (Zhai, et al., 2025), further increase the relevance of foundation models for crop genomics by incorporating plant genome diversity and longer-range sequence dependencies.

For population epigenomics, methylation-aware prediction models are especially important. Targeted models such as PlantDeepMeth have been developed to predict CG, CHG and CHH methylation states from plant DNA sequences, enabling genome-wide methylation imputation and the identification of sequence determinants of context-specific methylation (Guo et al., 2025). Such models can help distinguish methylation variants that are likely to be sequence-driven from those that may arise independently of local genetic variation. When combined with meQTL and eQTM analyses, methylation prediction models may therefore improve the prioritization of candidate functional epialleles (Fig 1h).

8.3 Targeted epigenome editing for epiallele creation and validation

Targeted epigenome editing offers a powerful strategy for testing the causal effects of methylation variation and creating novel epialleles with potential breeding value (Fig 1i). Catalytically inactive Cas9, or dCas9, retains guide-RNA-directed DNA binding activity but lacks nuclease function. By fusing dCas9 to DNA methyltransferases, demethylases, transcriptional repressors, or chromatin-modifying domains, researchers can induce site-specific epigenetic modifications at target loci without altering the underlying DNA sequence (Sung and Yim, 2020; Park et al., 2022; Shi et al., 2025; Zhang and Zhu, 2025). This approach enables the precise methylation or demethylation of promoters, enhancers, silencers, TE-derived regulatory regions, and other non-coding regulatory elements.

8.4 Remaining challenges and future outlook

Despite the opportunities detailed in this review, several conceptual and technical challenges must be addressed before population-level DNA methylation variation can be routinely used in molecular breeding.

Tissue specificity and cellular heterogeneity remain major challenges for applying DNA methylation markers in breeding. Since DNA methylation is highly dependent on tissue type, developmental stage, environmental condition, and cellular composition, methylation markers identified from bulk tissues may not necessarily reflect regulatory variation occurring in the most relevant cell types or developmental windows. This issue is particularly relevant to plants possessing many agronomic traits, such as seed development, fruit ripening, fiber elongation, stress response, root architecture, and reproductive transition, that are controlled by highly specialized tissues or transient cell states. As a result, epigenetic signals detected in whole seedlings, mature leaves or mixed tissues may be diluted, masked or even confounded by cell-type heterogeneity. For breeding-oriented population epigenomics, future studies should therefore incorporate multi-tissue, multi-developmental-stage and environment-aware methylome profiling. Candidate epigenetic markers should be evaluated not only for their statistical association with traits but also for their biological stability across tissues, developmental stages, genetic backgrounds, and environmental conditions. In this context, single-cell methylome sequencing, single-cell multi-omics, spatial transcriptomics, and spatial epigenomics provide powerful opportunities to overcome the limitations of bulk methylome profiling. These technologies can resolve cell-type-specific methylation landscapes, identify regulatory methylation changes in rare or transient cell populations, and link spatially restricted epigenetic states with local gene expression programs. For example, single-cell and spatially resolved approaches may help distinguish true regulatory methylation variation from apparent methylation differences caused by changes in tissue composition. They may also reveal whether trait-associated methylation loci act broadly across tissues or specifically within key developmental domains. The integration of bulk population methylomes with single-cell and spatial omics will be particularly valuable for complex plant tissues, such as developing seeds, shoot apices, root tips, vascular tissues, fruit pericarp, and cotton fibers, as it can provide a hierarchical view of epigenetic regulation, from population-level methylation polymorphisms to tissue-specific regulatory modules and cell-type-resolved methylation states. This will improve the precision of epigenetic marker discovery and help identify methylation variants that are more likely to have direct functional relevance for molecular breeding.

9 Conclusions and perspectives

Population epigenomics has revealed extensive DNA methylation variation across plant populations and has begun to clarify its genetic basis, regulatory consequences and potential contribution to domestication, adaptation and complex traits. However, translating methylation variation into breeding applications requires distinguishing causal epialleles from genetically linked or environmentally induced passenger marks. Future progress will depend on larger and better-designed population methylome datasets, multi-tissue and single-cell profiling, pangenome-aware multi-omics integration, robust statistical and deep learning models, and targeted epigenome editing for functional validation. Together, these advances can provide a solid foundation for incorporating epigenetic variation into precision breeding and crop improvement.

Acknowledgments

This work was partially supported by grants from the Finance Science and Technology Project of Xinjiang Uyghur Autonomous Region (2023A01), the Biological Breeding Major Project (2023ZD04076), the

National Natural Science Foundation of China (32341024), and the Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies.

Author contributions

T.Z., J.Z., and X.L. drafted the manuscript. L.F. coordinated the research and revised the manuscript. T.Z. and L.F. conceived the study and designed the overall framework. X.L. and J.Z. performed data visualization, including the construction of the integrative conceptual diagrams. All authors read and approved the final manuscript.

Data availability statement

No new data were generated or analyzed in this review. All data and studies discussed in this article are available from the cited publications and public resources.

Compliance with ethics guidelines

Xi LONG, Juyun ZHENG, Lei FANG, and Ting ZHAO declare that they have no conflicts of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

Declaration on the use of generative AI tools

During the preparation of the revised submission materials, the authors used AI-assisted tools to improve language, formatting, and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Competing interests

The authors declare no competing interests.

References

Albert FW, Kruglyak L, 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*, 16(4):197-212. https://doi.org/10.1038/nrg3891

Alonge M, Wang X, Benoit M, et al., 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, 182(1):145-161 e123. https://doi.org/10.1016/j.cell.2020.05.021

Becker C, Hagmann J, Muller J, et al., 2011. Spontaneous epigenetic variation in the arabidopsis thaliana methylome. *Nature*, 480(7376):245-249. https://doi.org/10.1038/nature10555

Bouyer D, Kramdi A, Kassam M, et al., 2017. DNA methylation dynamics during early plant life. *Genome Biol*, 18(1):179. https://doi.org/10.1186/s13059-017-1313-0

Cao S, Chen K, Lu K, et al., 2023. Asymmetric variation in DNA methylation during domestication and de-domestication of rice. *Plant Cell*, 35(9):3429-3443. https://doi.org/10.1093/plcell/koad160

Cao S, Sawettalake N, Li P, et al., 2024. DNA methylation variations underlie lettuce domestication and divergence. *Genome Biol*, 25(1):158. https://doi.org/10.1186/s13059-024-03310-x

Cubas P, Vincent C, Coen E, 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401(6749):157-161. https://doi.org/10.1038/43657

Cui X, Hu M, Yao S, et al., 2023. Bnaomics: A comprehensive platform combining pan-genome and multi-omics data from brassica napus. *Plant Commun*, 4(5):100609. https://doi.org/10.1016/j.xplc.2023.100609

Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, et al., 2025. Nucleotide transformer: Building and evaluating robust foundation models for human genomics. *Nat Methods*, 22(2):287-297. https://doi.org/10.1038/s41592-024-02523-z

Dubin MJ, Zhang P, Meng D, et al., 2015. DNA methylation in arabidopsis has a genetic basis and shows evidence of local adaptation. *Elife*, 4:e05255. https://doi.org/10.7554/eLife.05255

Engelhorn J, Snodgrass SJ, Kok A, et al., 2025. Genetic variation at transcription factor binding sites largely explains

phenotypic heritability in maize. *Nat Genet*, 57(9):2313-2322. https://doi.org/10.1038/s41588-025-02246-7

Erdmann RM, Picard CL, 2020. Rna-directed DNA methylation. *PLoS Genet*, 16(10):e1009034. https://doi.org/10.1371/journal.pgen.1009034

Grossniklaus U, Kelly WG, Kelly B, et al., 2013. Transgenerational epigenetic inheritance: How important is it? *Nat Rev Genet*, 14(3):228-235. https://doi.org/10.1038/nrg3435

Gui S, Yang L, Li J, et al., 2020. Zeamap, a comprehensive database adapted to the maize multi-omics era. *iScience*, 23(6):101241. https://doi.org/10.1016/j.isci.2020.101241

Guo H, Cao P, Wang C, et al., 2023. Population analysis reveals the roles of DNA methylation in tomato domestication and metabolic diversity. *Sci China Life Sci*, 66(8):1888-1902. https://doi.org/10.1007/s11427-022-2299-5

Guo Z, Fan W, Cai C, et al., 2025. Plantdeepmeth: A deep learning model for predicting DNA methylation states in plants. *Plants (Basel)*, 14(11) https://doi.org/10.3390/plants14111724

Guseinov VA, Vanyushin BF, 1975. Content and localisation of 5-methylcytosine in DNA of healthy and wilt-infected cotton plants. *Biochim Biophys Acta*, 395(3):229-238. https://doi.org/10.1016/0005-2787(75)90193-8

He L, Huang H, Bradai M, et al., 2022. DNA methylation-free arabidopsis reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat Commun*, 13(1):1335. https://doi.org/10.1038/s41467-022-28940-2

Heberle E, Bardet AF, 2019. Sensitivity of transcription factors to DNA methylation. *Essays Biochem*, 63(6):727-741. https://doi.org/10.1042/EBC20190033

Hollwey E, Briffa A, Howard M, et al., 2023. Concepts, mechanisms and implications of long-term epigenetic inheritance. *Curr Opin Genet Dev*, 81:102087. https://doi.org/10.1016/j.gde.2023.102087

Ji Y, Zhou Z, Liu H, et al., 2021. Dnabert: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics*, 37(15):2112-2120. https://doi.org/10.1093/bioinformatics/btab083

Jiang J, Gwee J, Fang J, et al., 2024. Substrate specificity and protein stability drive the divergence of plant-specific DNA methyltransferases. *Sci Adv*, 10(45):eadr2222. https://doi.org/10.1126/sciadv.adr2222

Jiang X, Zhang M, Yuan X, et al., 2026. Population epigenomics reveals epigenetic drivers of replicated evolution and missing heritability in soybean. *Mol Plant*, https://doi.org/10.1016/j.molp.2026.02.013

Johannes F, Porcher E, Teixeira FK, et al., 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet*, 5(6):e1000530. https://doi.org/10.1371/journal.pgen.1000530

Kawakatsu T, Huang SC, Jupe F, et al., 2016. Epigenomic diversity in a global collection of arabidopsis thaliana accessions. *Cell*, 166(2):492-505. https://doi.org/10.1016/j.cell.2016.06.044

Kawakatsu T, Nery JR, Castanon R, et al., 2017. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol*, 18(1):171. https://doi.org/10.1186/s13059-017-1251-x

Kelley DR, Snoek J, Rinn JL, 2016. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res*, 26(7):990-999. https://doi.org/10.1101/gr.200535.115

Kremling KaG, Diepenbrock CH, Gore MA, et al., 2019. Transcriptome-wide association supplements genome-wide association in zea mays. *G3 (Bethesda)*, 9(9):3023-3033. https://doi.org/10.1534/g3.119.400549

Lan Y, Sun R, Ouyang J, et al., 2021. Atmad: Arabidopsis thaliana multi-omics association database. *Nucleic Acids Res*, 49(D1):D1445-D1451. https://doi.org/10.1093/nar/gkaa1042

Lang Z, Wang Y, Tang K, et al., 2017. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc Natl Acad Sci U S A*, 114(22):E4511-E4519. https://doi.org/10.1073/pnas.1705233114

Law JA, Jacobsen SE, 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 11(3):204-220. https://doi.org/10.1038/nrg2719

Lindroth AM, Cao X, Jackson JP, et al., 2001. Requirement of chromomethylase3 for maintenance of cpxpg methylation. *Science*, 292(5524):2077-2080. https://doi.org/10.1126/science.1059745

Liu T, Conesa A, 2025. Profiling the epigenome using long-read sequencing. *Nature Genetics*, 57(1):27-41. https://doi.org/10.1038/s41588-024-02038-5

Liu X, Wang D, Zhang Z, et al., 2025. Epigenetic perspectives on wheat speciation, adaptation, and development. *Trends in Genetics*, 41(9):817-829. https://doi.org/10.1016/j.tig.2025.04.008

Lu Z, Hofmeister BT, Vollmers C, et al., 2017. Combining atac-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res*, 45(6):e41. https://doi.org/10.1093/nar/gkw1179

Maher KA, Bajic M, Kajala K, et al., 2018. Profiling of accessible chromatin regions across multiple plant species and cell

types reveals common gene regulatory principles and new control modules. *Plant Cell*, 30(1):15-36. https://doi.org/10.1105/tpc.17.00581

Manning K, Tor M, Poole M, et al., 2006. A naturally occurring epigenetic mutation in a gene encoding an sbp-box transcription factor inhibits tomato fruit ripening. *Nat Genet*, 38(8):948-952. https://doi.org/10.1038/ng1841

Mendoza-Revilla J, Trop E, Gonzalez L, et al., 2024. A foundational large language model for edible plant genomes. *Commun Biol*, 7(1):835. https://doi.org/10.1038/s42003-024-06465-2

Meng D, Dubin M, Zhang P, et al., 2016. Limited contribution of DNA methylation variation to expression regulation in arabidopsis thaliana. *PLoS Genet*, 12(7):e1006141. https://doi.org/10.1371/journal.pgen.1006141

Miura K, Ikeda M, Matsubara A, et al., 2010. Osspl14 promotes panicle branching and higher grain productivity in rice. *Nat Genet*, 42(6):545-549. https://doi.org/10.1038/ng.592

Mueller SA, Merondun J, Lecic S, et al., 2025. Epigenetic variation in light of population genetic practice. *Nat Commun*, 16(1):1028. https://doi.org/10.1038/s41467-025-55989-6

Niederhuth CE, Bewick AJ, Ji L, et al., 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol*, 17(1):194. https://doi.org/10.1186/s13059-016-1059-0

O'malley RC, Huang SC, Song L, et al., 2016. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, 166(6):1598. https://doi.org/10.1016/j.cell.2016.08.063

Ong-Abdullah M, Ordway JM, Jiang N, et al., 2015. Loss of karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, 525(7570):533-537. https://doi.org/10.1038/nature15365

Park H, Shin J, Kim Y, et al., 2022. Crispr/dcas9-dnmt3a-mediated targeted DNA methylation of app rescues brain pathology in a mouse model of alzheimer's disease. *Transl Neurodegener*, 11(1):41. https://doi.org/10.1186/s40035-022-00314-0

Quadrana L, Almeida J, Asis R, et al., 2014. Natural occurring epialleles determine vitamin e accumulation in tomato fruits. *Nat Commun*, 5:3027. https://doi.org/10.1038/ncomms5027

Rakyan VK, Down TA, Balding DJ, et al., 2011. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*, 12(8):529-541. https://doi.org/10.1038/nrg3000

Reinders J, Wulff BB, Mirouze M, et al., 2009. Compromised stability of DNA methylation and transposon immobilization in mosaic arabidopsis epigenomes. *Genes Dev*, 23(8):939-950. https://doi.org/10.1101/gad.524609

Richards EJ, 2006. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet*, 7(5):395-401. https://doi.org/10.1038/nrg1834

Schmitz RJ, Schultz MD, Urich MA, et al., 2013. Patterns of population epigenomic diversity. *Nature*, 495(7440):193-198. https://doi.org/10.1038/nature11968

Shen Y, Zhang J, Liu Y, et al., 2018. DNA methylation footprints during soybean domestication and improvement. *Genome Biol*, 19(1):128. https://doi.org/10.1186/s13059-018-1516-z

Shi L, Li S, Zhu R, et al., 2025. Crisprepi: A multi-omic atlas for crispr-based epigenome editing. *Nucleic Acids Res*, 53(D1):D901-D913. https://doi.org/10.1093/nar/gkae1039

Singh A, Verma AK, Kumar S, et al., 2024. Genome-wide DNA methylation and their transgenerational pattern differ in arabidopsis thaliana populations originated along the elevation of west himalaya. *BMC Plant Biol*, 24(1):936. https://doi.org/10.1186/s12870-024-05641-0

Smith AK, Kilaru V, Kocak M, et al., 2014. Methylation quantitative trait loci (meqtls) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, 15(1) https://doi.org/10.1186/1471-2164-15-145

Song B, Yu J, Li X, et al., 2024. Increased DNA methylation contributes to the early ripening of pear fruits during domestication and improvement. *Genome Biol*, 25(1):87. https://doi.org/10.1186/s13059-024-03220-y

Song JM, Guan Z, Hu J, et al., 2020. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of brassica napus. *Nat Plants*, 6(1):34-45. https://doi.org/10.1038/s41477-019-0577-7

Song Q, Zhang T, Stelly DM, et al., 2017. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol*, 18(1):99. https://doi.org/10.1186/s13059-017-1229-8

Song X, Tang S, Liu H, et al., 2025. Inheritance of acquired adaptive cold tolerance in rice through DNA methylation. *Cell*, 188(16):4213-4224 e4212. https://doi.org/10.1016/j.cell.2025.04.036

Soppe WJ, Jacobsen SE, Alonso-Blanco C, et al., 2000. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*, 6(4):791-802. https://doi.org/10.1016/s1097-2765(05)00090-0

Springer NM, Schmitz RJ, 2017. Exploiting induced and natural epigenetic variation for crop improvement. *Nat Rev Genet*, 18(9):563-575. https://doi.org/10.1038/nrg.2017.45

Sung CK, Yim H, 2020. Crispr-mediated promoter de/methylation technologies for gene regulation. *Arch Pharm Res*, 43(7):705-713. https://doi.org/10.1007/s12272-020-01257-8

Tang K, Zhu X, Xie S, et al., 2024. Transgenerational increases in DNA methylation in arabidopsis plants defective in active DNA demethylation. *Proc Natl Acad Sci U S A*, 121(22):e2320468121. https://doi.org/10.1073/pnas.2320468121

Tao Y, Zhao X, Mace E, et al., 2019. Exploring and exploiting pan-genomics for crop improvement. *Mol Plant*, 12(2):156-169. https://doi.org/10.1016/j.molp.2018.12.016

Treat B J, Robert D C, 1925. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*, 11(47)

Trerotola M, Relli V, Simeone P, et al., 2015. Epigenetic inheritance and the missing heritability. *Hum Genomics*, 9(1):17. https://doi.org/10.1186/s40246-015-0041-3

Van Der Graaf A, Wardenaar R, Neumann DA, et al., 2015. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc Natl Acad Sci U S A*, 112(21):6676-6681. https://doi.org/10.1073/pnas.1424254112

Wyatt GR, 1950. Occurrence of 5-methylcytosine in nucleic acids. *Nature*, 166(4214):237-238. https://doi.org/10.1038/166237b0

Xie G, Du X, Hu H, et al., 2025. Molecular mechanisms underlying the establishment, maintenance, and removal of DNA methylation in plants. *Annu Rev Plant Biol*, 76(1):143-170. https://doi.org/10.1146/annurev-arplant-083123-054357

Xu G, Lyu J, Li Q, et al., 2020. Evolutionary and functional genomics of DNA methylation in maize domestication and improvement. *Nat Commun*, 11(1):5539. https://doi.org/10.1038/s41467-020-19333-4

Xu J, Chen G, Hermanson PJ, et al., 2019. Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome Biol*, 20(1):243. https://doi.org/10.1186/s13059-019-1859-0

Yu DH, Ware C, Waterland RA, et al., 2013. Developmentally programmed 3' cpg island methylation confers tissue- and cell-type-specific transcriptional activation. *Mol Cell Biol*, 33(9):1845-1858. https://doi.org/10.1128/MCB.01124-12

Zemach A, Zilberman D, 2010. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol*, 20(17):R780-785. https://doi.org/10.1016/j.cub.2010.07.007

Zhai J, Gokaslan A, Schiff Y, et al., 2025. Cross-species modeling of plant genomes at single-nucleotide resolution using a pretrained DNA language model. *Proc Natl Acad Sci U S A*, 122(24):e2421738122. https://doi.org/10.1073/pnas.2421738122

Zhang H, Lang Z, Zhu JK, 2018. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol*, 19(8):489-506. https://doi.org/10.1038/s41580-018-0016-z

Zhang H, Zhu JK, 2025. Epigenetic gene regulation in plants and its potential applications in crop improvement. *Nat Rev Mol Cell Biol*, 26(1):51-67. https://doi.org/10.1038/s41580-024-00769-1

Zhang M, Zong W, Zou D, et al., 2023. Methbank 4.0: An updated database of DNA methylation across a variety of species. *Nucleic Acids Res*, 51(D1):D208-D216. https://doi.org/10.1093/nar/gkac969

Zhao L, Liu D, Xu J, et al., 2018. The framework for population epigenetic study. *Brief Bioinform*, 19(1):89-100. https://doi.org/10.1093/bib/bbw098

Zhao T, Mei H, Cao Z, et al., 2021. Absence of cg methylation alters the long noncoding transcriptome landscape in multiple species. *FEBS Lett*, 595(12):1734-1747. https://doi.org/10.1002/1873-3468.14100

Zhao T, Guan X, Hu Y, et al., 2024. Population-wide DNA methylation polymorphisms at single-nucleotide resolution in 207 cotton accessions reveal epigenomic contributions to complex traits. *Cell Res*, 34(12):859-872. https://doi.org/10.1038/s41422-024-01027-x

Zhong Z, Feng S, Duttke SH, et al., 2021. DNA methylation-linked chromatin accessibility affects genomic architecture in arabidopsis. *Proc Natl Acad Sci U S A*, 118(5) https://doi.org/10.1073/pnas.2023347118

Zhou J, Troyanskaya OG, 2015. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods*, 12(10):931-934. https://doi.org/10.1038/nmeth.3547

Zhu Y, Ngan H, Liu W, et al., 2025. Genetic effects on chromatin accessibility reveal the molecular mechanisms of complex traits in maize. *Plant J*, 123(4):e70437. https://doi.org/10.1111/tpj.70437

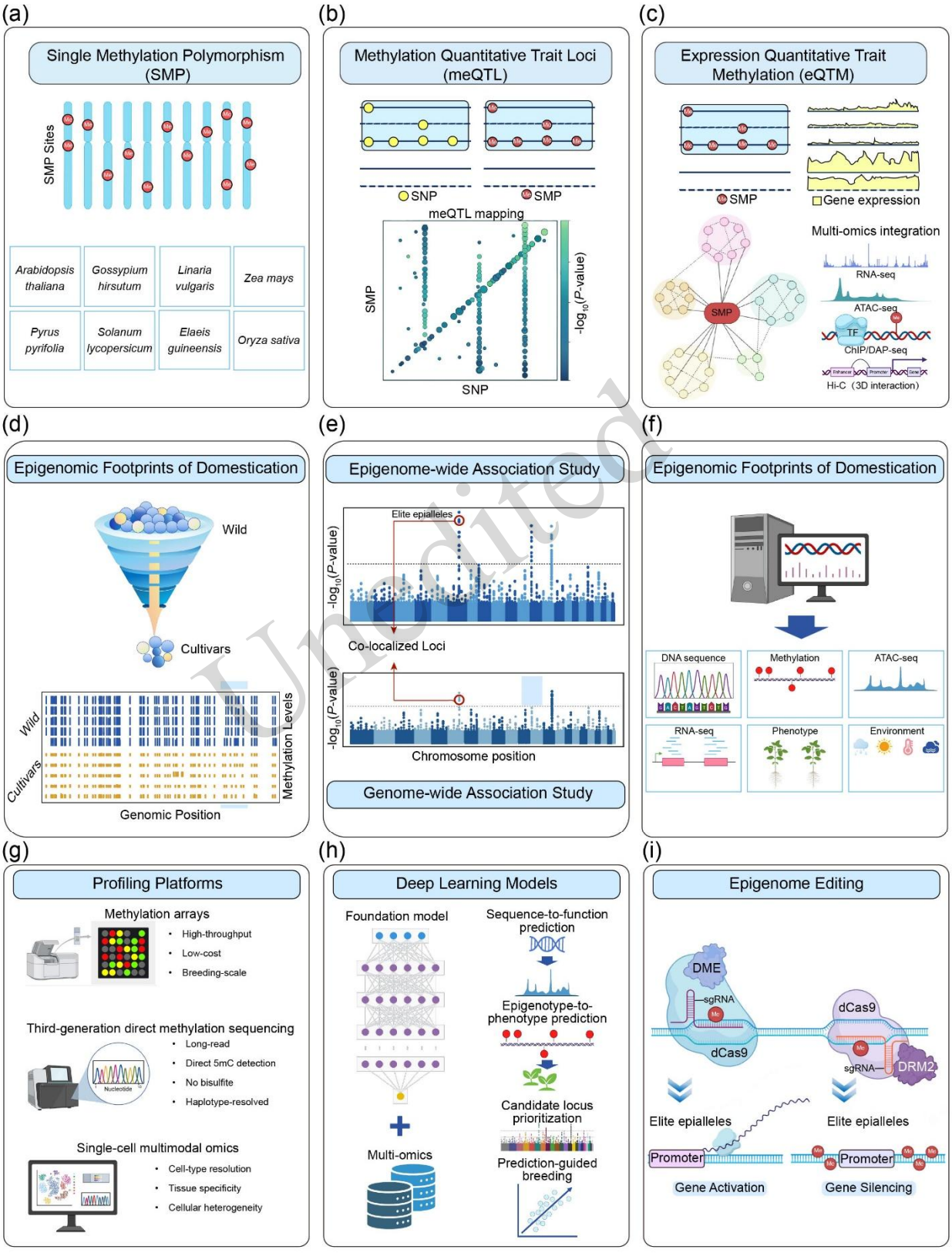

Fig. 1 An integrated framework for population epigenomics, functional validation, and epigenetic breeding in plants.

Schematic overview of the key concepts, resources, analytical strategies, and emerging technologies in plant population

epigenomics. (A) Single methylation polymorphisms (SMPs) and population-scale methylome resources provide a foundation for characterizing natural epigenomic variation in plants. (B) Genetic dissection of methylation variation, through association analyses between genetic variants and methylation loci, enables the identification of methylation quantitative trait loci and genetic determinants of epigenetic variation. (C) Expression quantitative trait methylation (eQTM) analysis links methylation variation at individual SMP sites with gene expression variation, facilitating the construction of integrated multi-omics regulatory networks. (D) Comparative methylome analyses between wild accessions and cultivated germplasm reveal domestication-associated epigenomic footprints and methylation shifts shaped by selection. (E) Epigenome-wide association studies (EWAS) associate methylation variation with agronomic traits, enabling the identification of candidate epialleles and trait-associated methylation markers. (F) Multi-omics integration databases and plant epigenomic platforms provide resources for data mining, visualization and cross-layer regulatory analyses. (G) Methylation arrays, Oxford Nanopore Technologies (ONT) long-read sequencing, and single-cell multi-omics platforms represent emerging approaches for high-throughput, haplotype-resolved, and cell-type-specific methylation profiling. (H) Transformer- and deep learning-based foundation models integrate sequence, epigenomic, transcriptomic, and phenotypic information to support sequence-to-function prediction, epigenotype-to-phenotype prediction, candidate locus prioritization, and prediction-guided breeding. (I) Targeted epigenome editing using dCas9 fused to methylation effector domains enables the functional validation of candidate epialleles. dCas9–DME induces locus-specific demethylation and gene activation, whereas dCas9–DRM2 promotes targeted *de novo* methylation and transcriptional silencing. All panels are schematic representations. Created with BioRender.com.