

# Cis-regulatory sequences in plants: Their importance, discovery, and future challenges

Robert J. Schmitz <sup>1</sup>, Erich Grotewold <sup>2</sup> and Maike Stam <sup>3,\*†</sup>

<sup>1</sup> Department of Genetics, University of Georgia, Athens, Georgia 30602, USA

<sup>2</sup> Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

<sup>3</sup> Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

\*Author for correspondence: [m.e.stam@uva.nl](mailto:m.e.stam@uva.nl)

†Senior author.

All three authors contributed to writing and approved the final version. M.S. integrated all sections of the text. R.J.S. and M.S. created the figures.

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) is: Maike Stam ([m.e.stam@uva.nl](mailto:m.e.stam@uva.nl))

## Abstract

The identification and characterization of cis-regulatory DNA sequences and how they function to coordinate responses to developmental and environmental cues is of paramount importance to plant biology. Key to these regulatory processes are cis-regulatory modules (CRMs), which include enhancers and silencers. Despite the extraordinary advances in high-quality sequence assemblies and genome annotations, the identification and understanding of CRMs, and how they regulate gene expression, lag significantly behind. This is especially true for their distinguishing characteristics and activity states. Here, we review the current knowledge on CRMs and breakthrough technologies enabling identification, characterization, and validation of CRMs; we compare the genomic distributions of CRMs with respect to their target genes between different plant species, and discuss the role of transposable elements harboring CRMs in the evolution of gene expression. This is an exciting time to study cis-regulomes in plants; however, significant existing challenges need to be overcome to fully understand and appreciate the role of CRMs in plant biology and in crop improvement.

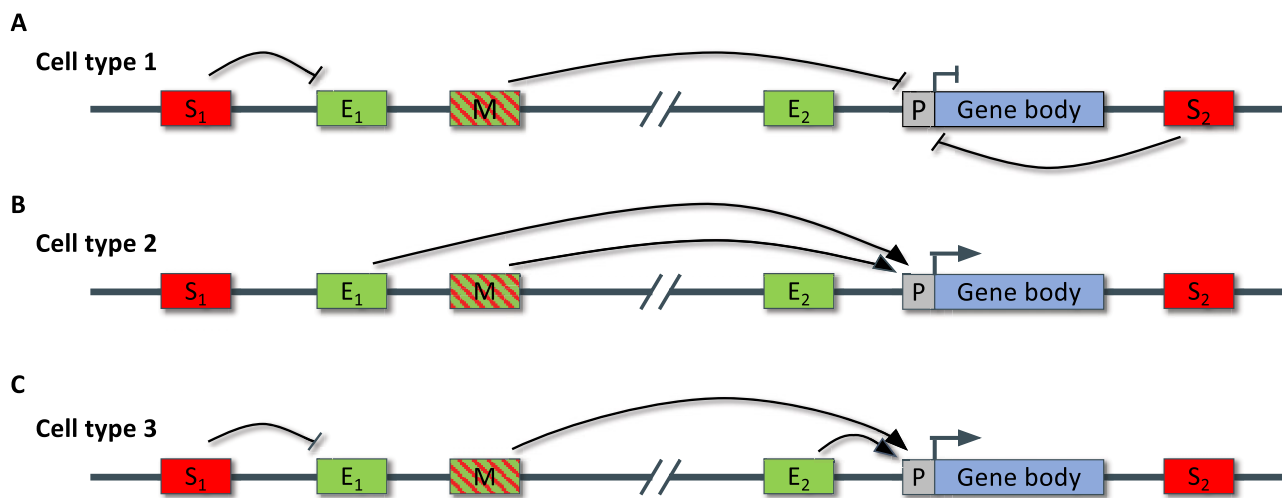
## Introduction

A fundamental question in biology is how complex patterns of gene expression are regulated. Central to this is the genome-wide identification and characterization of cis-regulatory elements (CREs) and cis-regulatory modules (CRMs) that influence the expression of protein-coding and long noncoding RNA (*lncRNA*) genes (Shlyueva et al., 2014; Kopp and Mendell, 2018; Andersson and Sandelin, 2020; Della Rosa and Spivakov, 2020). We refer here to CREs as individual transcription factor (TF) binding sites, while CRMs are assemblies of CREs and include promoters, transcriptional enhancers, silencers, and insulator elements. CRMs determine in which cell, at what time, and at what level a

gene is expressed (Table 1 and Figure 1). In animals, the generation of comprehensive chromatin and epigenome maps have made the identification of gene regulatory sequences routine, even though these DNA elements are often located kilobases to megabases away from their target genes (De Laat and Duboule, 2013; Shlyueva et al., 2014; Klemm et al., 2019). Epigenome maps are effective because regulatory sequences possess distinct chromatin signatures. For example, active regulatory sequences display accessible chromatin, TF binding, low DNA methylation, and histone modifications such as acetylation or methylation of specific lysine residues in histone H3 (Stadler et al., 2011; De Laat and Duboule, 2013; Shlyueva et al., 2014; Weber et al., 2016).

**Table 1** Major types of CRMs

Cis-regulatory element type	Definition
Core promoter	Minimal sequence region needed to direct initiation of transcription, usually spanning 50–100 bp around the TSS
Enhancer	A DNA sequence that, when bound by specific TFs and cofactors, increases the transcription initiation rate, and thereby the expression of target genes in a tissue-, developmental stage-, and/or condition-specific manner
Silencer	A DNA sequence that, when bound by specific TFs and cofactors, actively decreases the expression of target genes. A silencer might silence a gene directly, or indirectly by silencing enhancers
Insulator	An element located between CRMs and core promoters that, when bound by the appropriate proteins, prevents the activation or silencing of potential target genes by these CRMs. Such insulators are not known to exist in plants
Multifunctional sequence element	DNA element that exhibits more than one of the above properties at different times or conditions, or in different cells, e.g. enhancers in one cell type can function as silencers in other cell types and vice versa



**Figure 1** Combinatorial CRM actions elicit diverse transcriptional responses in distinct cell types. The activity of each CRM depends to a large extent on the expression levels of TFs that can bind the CRM. A, In cell type 1, silencer element 1 represses an enhancer, while a multifunctional sequence element and silencer element 2 repress promoter activity. B, In cell type 2, enhancer element 1 works cooperatively with a multifunctional sequence element to activate gene expression. C, In cell type 3, silencer element 1 represses the upstream enhancer 1, whereas the multifunctional sequence element activates the gene in concert with the promoter proximal enhancer (enhancer 2). S, silencer; E, enhancer; M, multifunctional sequence element; P, promoter.

Although relevant research in plants has lagged behind model animal species, recent studies in plants are revealing similar, but also distinct, molecular signatures at cis-regulatory sequences compared to animals. Although in plant species with smaller genomes, such as *Arabidopsis thaliana* (*Arabidopsis*), most CRMs are located in close linear proximity to the genes they control, the expansion of genome size and intergenic space in numerous plant species is associated with a large number of putative distal CRMs (Oka et al., 2017; Lu et al., 2019; Li et al., 2019a; Ricci et al., 2019). Distal CRMs can regulate genes tens or even hundreds of kilobases away, complicating their identification. The systematic discovery of CRMs and the CREs they are composed of is a first step in the engineering and rewiring of existing regulatory networks to optimize plant growth and development, enhance stress resilience or generate plant products. To maximize the use of natural variation, and for synthetic biology to realize its full potential, deeper knowledge of regulatory modules is paramount to facilitate crop improvement. Numerous reporter assays are now being developed and implemented for the identification and

functional characterization of CRMs (Inoue and Ahituv, 2015; Ricci et al., 2019; Sun et al., 2019; Jores et al., 2020), but there is still a need for more epigenome maps from many different cell types and growth conditions.

### Transcription factor binding sites: important CRM constituents

CRMs are assemblies of CREs, which serve as sequence-specific binding sites for TFs and are key components of the regulatory portion of each eukaryotic genome. The combination of TFs expressed and bound to CRMs in a particular cell type determines the activity and the ultimate function of CRMs (Kaufmann et al., 2010; Lambert et al., 2018). To activate CRMs, pioneer TFs (described below) bind to nucleosomal DNA, and recruit histone-modifying enzymes, such as histone acetyltransferases, and chromatin remodeling complexes that open the chromatin structure by displacing linker histones and mediating nucleosome eviction (Cirillo et al., 2002; Spitz and Furlong, 2012; Pajoro et al., 2014; Iwafuchi-Doi et al., 2016; Jin et al., 2021). Subsequently, other

nonpioneer TFs and co-factors can bind to DNA within the accessible chromatin regions (ACRs), ultimately resulting in fully activated CRMs, which physically interact with their target genes through protein–protein interactions to regulate their transcription levels (Nolis et al., 2009; Quevedo et al., 2019).

In eukaryotes, 5%–7% of all nuclear genes encode TFs, proteins that recognize short DNA motifs in a sequence-specific fashion (Riechmann and Ratcliffe, 2000; Lambert et al., 2018). Most TF-binding sites (TFBSs) are small (6–12 bases) and are therefore represented thousands of times in any eukaryotic genome (Wunderlich and Mirny, 2009; Lambert et al., 2018). TFs are usually characterized by the presence of one or more conserved DNA-binding and/or dimerization domains that allow grouping them into 40 or more families (Yilmaz et al., 2009; Riechmann and Ratcliffe, 2000; Lambert et al., 2018). Traditionally, TFs have been classified into activators or repressors, but it is clear that many can function as both activator and repressor, depending on the proteins (co-activators or co-repressors) they interact with (alternate complex formation; Figure 2A; Rosenfeld et al., 2006; Ikeda et al., 2009; Lambert et al., 2018; Haberle and Stark, 2018). For example, Arabidopsis WUSCHEL, a central regulator of stem cell proliferation, functions primarily as a repressor, yet has the ability to activate transcription of the floral homeotic gene AGAMOUS (AG; Ikeda et al., 2009).

Identifying the sequences regulated by specific TFs is an area of active research. Members from the same TF family often recognize very similar DNA sequences in vitro (Weirauch et al., 2014), but with different DNA-binding affinities. Differences in DNA-binding preferences between paralogous TFs are important in establishing in vivo binding selectivity (Shen et al., 2018; Panchy et al., 2019). Predicting which specific TF is tethered to a particular CRE in the genome in a given cell type remains challenging. Predictions require a good understanding of the cell types in which a TF is expressed and at what levels compared with the putative target genes; whether it is a pioneer or nonpioneer TF, and whether or not corresponding CREs are accessible. Consequently, it remains difficult to assign a role in activating or repressing transcription to a particular CRE in the genome. This difficulty is further augmented by multiple CREs together forming a CRM.

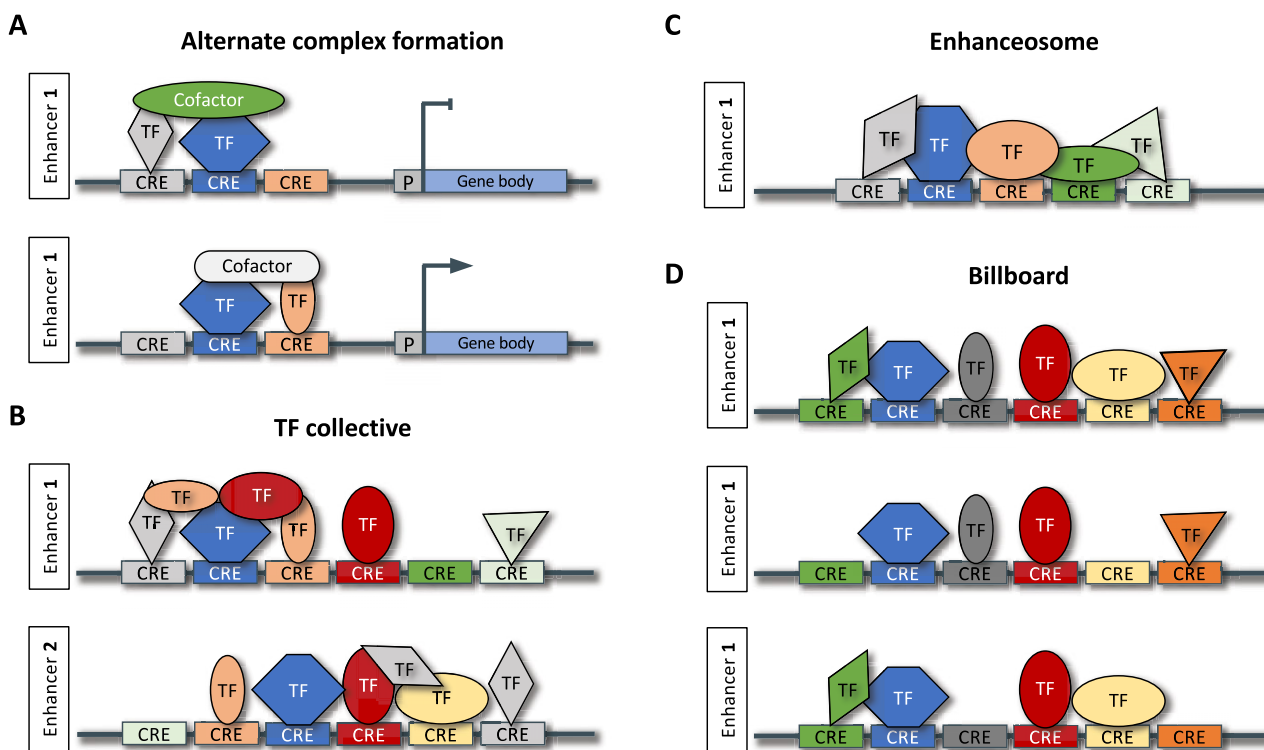
For most TFs, nucleosomes provide a challenging barrier to access the CREs they preferably bind to (Bai and Morozov, 2010; Zaret and Carroll, 2011; Zaret and Mango, 2016). In contrast to most TFs, pioneer TFs have the ability to bind their cognate TFBSs on nucleosomes in inaccessible chromatin, enabling CRMs to adopt a state of competence to activate or repress transcription by recruiting activators or repressors, respectively (Zaret and Carroll, 2011; Pajoro et al., 2014; Zaret and Mango, 2016; Tao et al., 2017; Jin et al., 2021). The first identified pioneer TF, FoxA (*forkhead box A*), involved in the development of endoderm-derived organs during metazoan embryo development (Lee et al., 2005), takes advantage of the similarity of its “winged-helix”

DNA binding domain to linker histone H1 to displace H1 and gain access to nucleosomal DNA (Clark et al., 1993; Iwafuchi-Doi et al., 2016). Note that pioneer TFs use a variety of mechanisms to bind to DNA (Zaret and Carroll, 2011; Zaret and Mango, 2016). Even DNA methylation, generally associated with inaccessible chromatin, can provide an anchor point for pioneer TFs. Indeed, many pioneer TFs, including FoxA, bind methylated DNA (Zhu et al., 2016). In plants, several pioneer TFs have been identified, including LEAFY COTYLEDON1, LEAFY, APETALA1, and SEPALLATA3, which are respectively involved in controlling Arabidopsis flowering competency, flower meristem establishment, and floral organ specification (Pajoro et al., 2014; Tao et al., 2017; Jin et al., 2021).

## CRMs: characteristics of the main players

CRMs and their activity states are characterized by different combinations of DNA and chromatin features, whereby the features of promoters and enhancers are more well-defined than those of silencers (Pajoro et al., 2014; Shlyueva et al., 2014; Weber et al., 2016; Marand et al., 2017; Xiao et al., 2017; Oka et al., 2017; Ricci et al., 2019; Andersson and Sandelin, 2020; Gisselbrecht et al., 2020; Pang and Snyder, 2020; Zhao et al., 2020). In plants, the DNA of the vast majority of CRMs appears stably unmethylated in a tissue-independent manner, and these unmethylated regions (UMRs) are enriched in accessible chromatin, histone acetylation (HAC), and TF–DNA interactions (Schmitz et al., 2013; Li et al., 2015; Kawakatsu et al., 2016; Oka et al., 2017; Ricci et al., 2019; Crisp et al., 2020). Much of the rest of the genome is methylated, including a subset of genes and transcriptionally silenced transposable elements (TEs; Zhang et al., 2006; Zilberman et al., 2007; Gent et al., 2013; Niederhuth et al., 2016). Therefore, UMRs likely encompass the vast majority of the CRMs within plant genomes, independent of their activity. This is consistent with the majority of Arabidopsis TFs not binding to methylated cytosines (O'Malley et al., 2016). A small fraction of UMRs appear tissue-specific, however (Oka et al., 2020). Indeed, active DNA demethylation at CRMs has been shown to play a role in specific developmental processes, such as fruit ripening in tomato and plant responses to stress conditions (Yua et al., 2013; Zhong et al., 2013; Liu et al., 2015; Deleris et al., 2016; Halter et al., 2021).

Several different activity states have been described for CRMs, but as not all of these are defined unequivocally, we limit our description to the repressed, poised, and active states. Active CRMs are accessible for DNA–protein interactions, and enriched with (among others) HAC, which weakens the electrostatic interactions between nucleosomal DNA and histones, improving accessibility for the transcription machinery (Zhang et al., 2015; Fenley et al., 2018). Based on current knowledge, chromatin of repressed CRMs is inaccessible, unbound by TFs, and enriched with trimethylation of histone H3 lysine K27 (H3K27me3; Shlyueva et al., 2014; Zhang et al., 2015). H3K27me3 marks transcriptionally



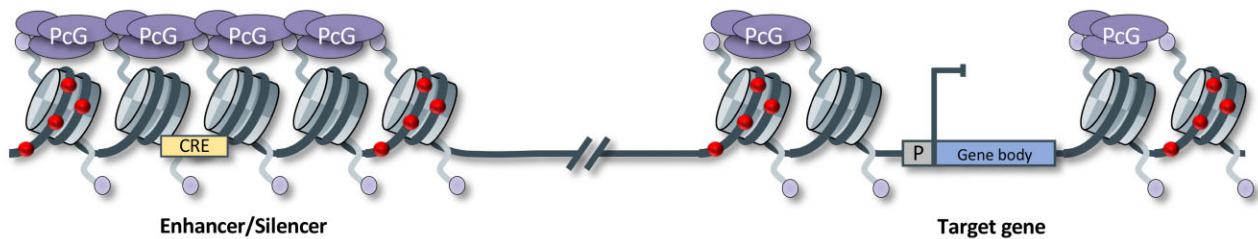
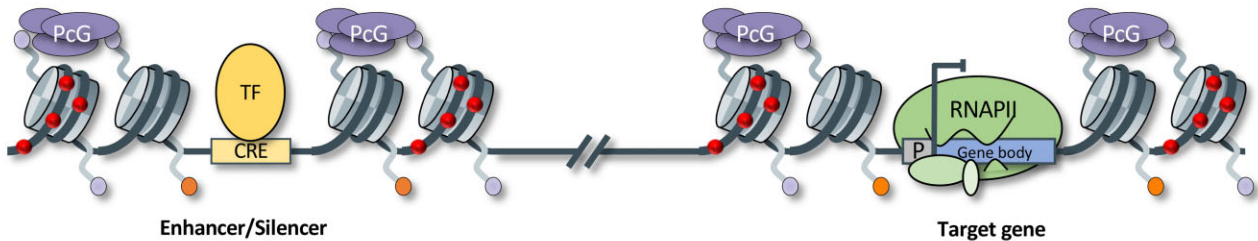
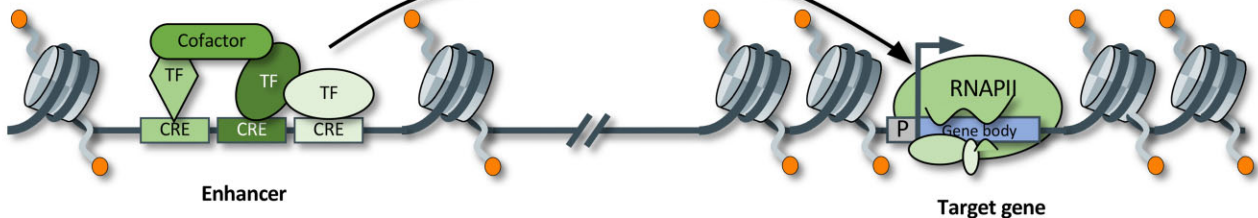
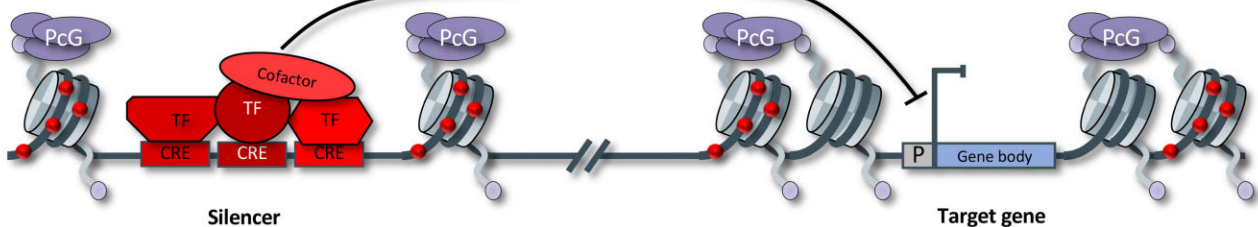
**Figure 2** Models for enhancer organization. A, Example of alternative complex formation. TFs can function both as transcriptional activators and repressors, depending on the proteins (co-activators or co-repressors) they interact with. B, Examples of TF collective enhancers. This model is characterized by cooperative DNA-binding, which can be achieved by many different mechanisms, and by proteins as well as DNA serving as a scaffold for the binding of TFs. It allows for flexible CRE arrangements, resulting in distinct regulatory outputs (exemplified by the two enhancers shown). C, Example of an enhanceosome enhancer. In the enhanceosome model, the binding of the various TFs to the respective CREs must occur in a specific order and orientation, following a particular grammar. D, Example of billboard enhancers. In the billboard model, the composition as well as the position and orientation of CREs within an enhancer is preserved. The regulatory output differs depending on the expression and activity level of TFs that can bind the CREs. Key: The thick gray line represents DNA, with the colored rectangles indicating CREs. Each different color indicates a different DNA motif. TFs are depicted in various shapes and colors, with each color denoting a different TF that recognizes a specific DNA sequence (matching colors).

silenced genes, their proximal flanking regions, and distal intergenic regions, and is associated with the presence of Polycomb Group (PcG) protein complexes (Simon and Kingston, 2013; Wang and Shen, 2018; Zhao et al., 2018; Huang et al., 2019; Lu et al., 2019; Ricci et al., 2019; Ngan et al., 2020; Pang and Snyder, 2020). Poised CRMs have an activity state between repressed and active and are ready to become fully activated or inactivated (Rada-Iglesias et al., 2011; Koenecke et al., 2017). They are accessible, bound by few TFs, and enriched with H3K27me3 and low levels of an active histone modification, such as HAc (Figure 3; Ernst et al., 2011; Rada-Iglesias et al., 2011; Huang et al., 2019; Lu et al., 2019; Ricci et al., 2019). In animals, poised and active CRMs are generally associated with H3K4me1 (Rada-Iglesias et al., 2011), but this does not seem the case for plants (Oka et al., 2017; Li et al., 2019b; Ricci et al., 2019; Sun et al., 2019).

In various organisms, active enhancers and silencers produce transcripts, including enhancer RNAs and lncRNAs; de Santa et al., 2010; Kim et al., 2010; Quinn and Chang, 2016;

Oka et al., 2017; Ricci et al., 2019; Gil and Ulitsky, 2020; Pang and Snyder, 2020). Knowledge of CRM-derived RNAs in plants is limited. In animals, enhancer RNAs are usually unstable, and produced in both directions (bidirectional) at the border of ACRs, at significantly lower levels than the unidirectional transcripts directed by core promoters (Andersson et al., 2014a; Core et al., 2014; Andersson and Sandelin, 2020). For mammalian cells, enhancer RNAs are used to predict enhancers (Andersson et al., 2014a), and a number of enhancer-derived lncRNAs play a role in enhancer function (Quinn and Chang, 2016; Gil and Ulitsky, 2020). However, the level at which transcription occurs at enhancers in plants is unclear. In Arabidopsis, transcription at intergenic regions seems rare (Hetzel et al., 2016; Thieffry et al., 2020). In maize, several dynamically expressed high-confidence lncRNAs associated with ACRs have been described, and co-expressed lncRNA–protein-coding gene pairs identified, suggesting a regulatory relationship; a small fraction of the lncRNAs overlapped with putative enhancers discovered in maize (Oka et al., 2017; Parvathaneni et al., 2020).



**A Repressed CRMs****B Poised CRMs****C Active enhancer****D Active silencer**

**Figure 3** Chromatin accessibility and modifications associated with different CRM activities. A, Example of a repressed CRM. CREs are occluded by nucleosomes due to inaccessible chromatin imparted by PcG protein complexes and H3K27me3. This is associated with transcriptional silencing of the target gene. B, Example of a poised CRM, where nucleosomes flanking CREs have low histone acetylation levels and only a few CREs are accessible for TF binding. PcG proteins and H3K27me3 are still present. The promoter is engaged by RNAPII, and flanking nucleosomes have low histone acetylation levels. PcG proteins and H3K27me3 hamper RNAPII elongation and transcript production. C, Example of an active enhancer, whereby multiple TFs and cofactors interact with a CRM, and target the promoter to activate transcription of the target gene. Nucleosomes are enriched for histone acetylation. D, Example of an active silencer where TFs and cofactors bind to the CRM to recruit PcG proteins to catalyze H3K27me3 and silencing of the target gene. Purple ovals, PcG protein complexes; P, promoter; RNAPII, RNA polymerase II; Purple circles, H3K27me3; orange circles, histone acetylation; red circles, DNA methylation.

It is important to note that for less well-annotated genomes, lncRNAs may correspond to unannotated protein-coding genes (Monnahan et al., 2020; Mendieta et al., 2021).

Most characteristics are not limited to particular types of CRMs, hampering identification through the use of single characteristics. Instead, combinations of descriptors are often used to distinguish different types of CRMs. For

example, in animals, high and low H3K4me3/H3K4me1 ratios are indicative of core promoters and enhancers, respectively.

### Promoters

For historical reasons, the term promoter is often used to indicate one to two kilobase pairs (kbp) upstream of the

transcription start site (TSS; Juven-Gershon and Kadonaga, 2010). Such regions may contain enhancer and/or silencer elements, obscuring the discussion of the different types of regulatory elements. Therefore, in this review, we define a promoter as what is also known as the core promoter, which is the minimal sequence region that is needed to direct initiation of transcription, usually spanning 50–100 base pairs (bp) around the TSS. For RNA polymerase II (RNAPII) transcribed genes, this is the region to which the general TFs bind to help facilitate RNAPII to initiate transcription (Haberle and Stark, 2018; Andersson and Sandelin, 2020). General TFs are defined by their binding to core promoters to help position RNAPII. The core promoter alone drives only low (basal) transcription levels, and sometimes contains a TATA box (Das and Bansal, 2019). The TATA box is recognized by the TATA-box binding protein, a key factor in recognizing core promoter elements (Haberle and Stark, 2018; Ngoc et al., 2019). In plants, however, the majority of the core promoters are TATA-less; at such promoters other cis-elements play a role in transcription initiation (Molina and Grotewold, 2005; Lenhard et al., 2012; Mejía-Guerra et al., 2015). Besides a TATA box, cis-elements such as a CCAAT-box and TC-elements are shown to contribute to the correct positioning of the transcription machinery and efficient transcription initiation (Dolfini et al., 2009; Bernard et al., 2010; Laloum et al., 2013; Porto et al., 2014).

Active animal and plant promoters are characterized by accessible chromatin, H3K4me3, HAc, H2A.Z, a lack of DNA methylation, the binding of general TFs and RNAPII, and the production of stable, unidirectional transcripts (Heintzman et al., 2007; Zilberman et al., 2007; Stadler et al., 2011; Zhang et al., 2012b, 2006; Ricci et al., 2019; Andersson and Sandelin, 2020; Oka et al., 2020; Thieffry et al., 2020). Generally, promoters display a positive correlation between transcript levels and the levels of H3K4me3 and HAc and are associated with reduced nucleosomal occupancy (Barski et al., 2007; Zhang et al., 2012b; Core et al., 2014; Oka et al., 2017; Yan et al., 2019). Inactive promoters are generally enriched for H3K27me3 and are less accessible (Ernst et al., 2011; Lu et al., 2019; Ricci et al., 2019). Little is known about poised plant promoters. They may be marked by HAc, but also RNAPII (Offermann et al., 2008; Gage et al., 2011), whereas the combination of H3K27me3 with H3K4me3 found in animals may be lacking (Mikkelsen et al., 2007; Zeng et al., 2019; Blanco et al., 2020). Typically, to allow transcription initiation, promoter DNA must be unmethylated, irrespective of its activity (Appanah et al., 2007; Zilberman et al., 2007; Zemach et al., 2010; Stadler et al., 2011).

## Enhancers

Transcriptional enhancers are CRMs that, when bound by specific TFs and cofactors, increase the transcription initiation rate, and thereby the expression of their target genes in a tissue-, developmental stage-, and/or condition-specific manner. Enhancers can be located more than 1 Mbp away from their target gene (Lettice et al., 2002; Benko et al.,

2009); the average distance depends on the size of the genome (Lu et al., 2019). The first distal enhancer discovered in plants was the maize *b1* (*booster1*) hepta-repeat enhancer, located ~100-kbp upstream of the TSS of the *b1* gene (Stam et al., 2002; Belete et al., 2013). The most distal enhancer so far described in plants is D1stal Cis-Element (DICE), which is required for high expression of the maize *bx1* (*benzoxazinless1*) gene ~140 kbp away (Zheng et al., 2015). Enhancers have been defined to increase reporter gene expression when located up- or downstream of the gene, in an orientation independent manner (Shlyueva et al., 2014; Andersson and Sandelin, 2020). Recent data indicate the latter may, however, not be true for all enhancers (Mikhaylichenko et al., 2018; Ricci et al., 2019; Jores et al., 2020). Active enhancers usually regulate their target genes through chromatin interactions, which involve TFs and cofactors at the enhancer, and the transcription initiation complex at the promoter (De Laat and Duboule, 2013; Long et al., 2016; Peng et al., 2019; Li et al., 2019a; Doni Jayavelu et al., 2020; Ngan et al., 2020). LncRNAs have also been implicated in these interactions (Li and Fu, 2019).

The billboard, enhanceosome, and TF collective models represent three different mechanisms for the interplay between enhancers and TFs, resulting in active regulatory sequences (Figure 2, B–D). In the billboard model, neighboring CREs form a CRM, but the individual CREs can almost independently affect gene expression; TFs can bind cooperatively and additively to the individual CREs and the composition of bound TFs is interpreted by the basal transcription machinery (Kulkarni and Amosti, 2003; Spitz and Furlong, 2012). This enables the same CRM to have different effects in different cell types depending on the expressed and bound TFs. In the enhanceosome model, enhancers are organized using a particular grammar of TF binding motifs; a specific order and orientation of CREs is necessary for the cooperative binding of specific TFs, and thereby for enhancer function (Spitz and Furlong, 2012; Weingarten-Gabbay and Segal, 2014; Vockley et al., 2017). In the TF collective model, the recruitment of TFs is accomplished through binding to CREs and protein–protein interactions (Junion et al., 2012; Long et al., 2016; Figure 2). While specific details of the different models vary in the literature (Spitz and Furlong, 2012; Long et al., 2016), the picture is emerging that TFs are tethered to enhancers using additive or cooperative, and direct or indirect DNA binding (Heyndrickx et al., 2014). Note that while these models have been proposed for enhancers, they are likely to be valid for silencers as well.

Genes can be regulated by multiple enhancers that act either complementarily, redundantly, pleiotropically, interdependently, or synergistically (Figure 1; Cannavò et al., 2016; Osterwalder et al., 2018; Lewis et al., 2019; Sabaris et al., 2019). For example, different enhancers at the maize *b1* gene display a different tissue-specificity (Stam et al., 2002), and the Arabidopsis *SHATTERPROOF2* gene is regulated by two redundant CRMs (Bhupinder and Franks, 2017). In *Solanum lycopersicum*, deleting combinations of candidate

CRMs revealed both additive and synergistic interactions in the production of locule number depending on which CRM was deleted (Wang et al., 2021). Comparable to how genes can be targeted by multiple enhancers, enhancers can interact with more than one target gene, and are able to “skip over” genes, interacting with genes further away, in terms of the linear sequence, than the nearest flanking gene (Ghavi-Helm et al., 2014; Peng et al., 2019; Li et al., 2019a; Ricci et al., 2019). For example, in maize, DICE enhances expression of *bx1*, but not *bx8*, which is located between the two (Zheng et al., 2015).

Similar to promoters, enhancers can be in active, repressed, or poised states (Figure 3). Active enhancers typically are characterized by the binding of activating TFs and cofactors, a lack of DNA methylation, ACRs with HAc (e.g. H3K9ac, H3K27ac, H3K56ac), H2A.Z at flanking nucleosomes, and physical interactions with target genes (Zhang et al., 2012b; Zhu et al., 2015; Oka et al., 2017; Li et al., 2019a; Peng et al., 2019; Ricci et al., 2019; Crisp et al., 2020; Zhao et al., 2020a). Enhancers are often tissue-specific; in the tissues they are not active, they are likely repressed or poised. We hypothesize that most repressed enhancers are inaccessible and enriched for repressing modifications, such as H3K27me3, and need to be bound by pioneer TFs to become accessible and ready for activation (Figure 3; Pajoro et al., 2014; Zhu et al., 2015; Sun et al., 2019). Poised enhancers may be bound by a few TFs, have ACRs, and be enriched for both H3K27me3 and low HAc levels (Koenecke et al., 2017; Ricci et al., 2019; Sun et al., 2019). Upon activation, H3K27me3 is removed and activating histone modifications added.

There may be other enhancer activity states besides those discussed above. Indeed, the largest class of distal accessible, UMRs identified in various plants is characterized by the absence of the histone modifications tested for (Lu et al., 2019; Ricci et al., 2019; Oka et al., 2020), yet the histone variant H2A.Z is still present at these regions (Lu et al., 2019; Ricci et al., 2019). In maize, such “unmodified regions” display enhancer activity in transient assays, albeit at lower average levels than ACRs marked with HAc (Ricci et al., 2019). Gene ontology enrichment analysis of their flanking genes suggests these regions may be involved in developmental programs. In addition, there may be DNA-methylated CRMs within TEs that are demethylated in specific cell types or upon stress treatment, affecting the expression of genes involved in among others plant disease resistance (Deleris et al., 2016; Halter et al., 2021).

## Silencers

Silencers, when bound by TFs and associated cofactors, actively decrease the expression of their target genes (Figures 1 and 3). They are crucial for establishing precise, tissue-specific expression patterns by blocking expression in cell types and tissues where the gene should be silenced; they prevent ectopic gene expression (Ogbourne and Antalis, 1998). For example, a 100-bp fragment in the second intron

of the tobacco AG gene represses AG expression in nonfloral tissues, promoting flower-specific AG expression (Liu et al., 2020). Similarly, TACPyAT repeats in the *Petunia hybrida* *chalcone synthase A* gene promoter hamper ectopic *chalcone synthase A* expression (van der Meer et al., 1992). Like enhancers, silencers are also proposed to act in a position- and orientation-independent manner (Figure 1) (Laimins et al., 1986; Sawada et al., 1994; Ogbourne and Antalis, 1998). Silencers can have multiple action modes. They can silence a gene directly, but also indirectly by silencing enhancers (Figure 1; Harris et al., 2016; Ngan et al., 2020; Pang and Snyder, 2020).

Although silencers were described decades ago (Laimins et al., 1986; Kuhlemeier et al., 1987; van der Meer et al., 1992; Sawada et al., 1994; Liu et al., 2011), attention to silencers has only increased recently (Oki and Kamakaka, 2002; Chetverina et al., 2014; Xiao et al., 2017; Huang et al., 2019; Doni Jayavelu et al., 2020; Gisselbrecht et al., 2020; Ngan et al., 2020; Pang and Snyder, 2020). These recent studies indicate that silencers, like enhancers, can inactivate genes from distal genomic locations, and can regulate multiple genes. Furthermore, a single gene can be controlled by multiple silencers. Future experiments will be needed to determine if this is also true for plants. Intriguingly, silencers can exhibit enhancer activity in other cell types, indicating multifunctionality (see below).

Silencers, like enhancers, probably exist in multiple activity states (Figure 3). The common denominator for active silencers across various genome-wide studies seems to be accessible chromatin enriched for H3K27me3. Xiao et al. (2017) indeed showed that Arabidopsis Polycomb response elements recruit Polycomb proteins and H3K27me3, and silence nearby genes in a Polycomb-dependent manner (Xiao et al., 2017). Also, several maize ACRs enriched with H3K27me3 were associated with transcriptional repression of the closest genes (Ricci et al., 2019); although it remains to be investigated if these ACRs act as silencers. However, depending on the pre-established bias towards features silencers are expected to display, the chromatin characteristics of identified silencers differ, suggesting the existence of different classes of silencers. In a study focusing on human and mice, the vast majority of ACR-based silencers were not enriched with H3K27me3; some were enriched for H3K9me3 (Doni Jayavelu et al., 2020). In *Drosophila melanogaster* none of 20 preselected H3K27me3-enriched ACRs showed silencing activity (Gisselbrecht et al., 2020). Different combinations of histone marks have been observed at silencers, H4K20me with H3K27me3 or H3K9me3, or H3K27me3 with active marks such as H3K4me1, H3K27ac, H3K4me3, or H3K9ac (Ngan et al., 2020; Huang et al., 2019; Pang and Snyder, 2020; Gisselbrecht et al., 2020). It was therefore concluded that to date there is no combination of chromatin marks that can unequivocally discriminate between silencers and other CRMs (Gisselbrecht et al., 2020). The identification and characterization of TFs binding to individual CREs within CRMs is likely key to distinguish different types of



CRMs (Doni Jayavelu et al., 2020; Gisselbrecht et al., 2020). For example, in *Arabidopsis*, knockdown of two TFs binding to selected PREs yielded significantly higher expression levels of the target genes tested (Xiao et al., 2017).

### Insulators

Insulators or boundary elements are another type of CRM studied extensively in animals (Oki and Kamakaka, 2002; Chetverina et al., 2014). An insulator is defined as a DNA sequence that, when bound by the appropriate proteins and located between CRMs and core promoters, prevents the activation or silencing of potential target genes by these CRMs. They also prevent the inactivation of active genes by nearby heterochromatin. The most studied insulator protein is the CCCTC-binding factor (CTCF) which, however, is absent in plants (Heger et al., 2012). Accordingly, there are currently no indications that genuine insulator elements exist in plants (Heger and Wiehe, 2014). Plant genomes are able to maintain proper gene expression patterns without classic insulator sequences or proteins. It has been indicated that RNA-directed DNA methylation of sequences between euchromatin and heterochromatin provides local boundary activity (Li et al., 2015). Note that “insulator” is a term used otherwise by synthetic biologists; in their transgene design it refers to “transcription blockers” and/or “transcription terminators” (Schaumberg et al., 2015).

### Multifunctional sequence elements

In reality, the different functionalities of CRMs as described above are often intermingled rather than clearly separated (Figure 1). In animals, it has been shown that some core promoters can display enhancer activity and even regulate the expression of genes other than those for which it acts as core promoter (Dao et al., 2017; Diao et al., 2017; Mikhaylichenko et al., 2018; Andersson and Sandelin, 2020). Diao et al. (2017), for example, observed that core promoters of multiple genes enhanced expression of the *Pou Class 5 Homeobox 1* gene in human embryonic stem cells. Recent results also suggest enhancer activity by core promoters in plants (Sun et al., 2019). However, this study tested randomly sheared ~670-bp genomic regions. Therefore, re-evaluation using smaller sequence regions restricted to just core promoters is needed.

Conversely, enhancers often mediate the production of transcripts by RNAPII, indicating promoter activity (Andersson et al., 2014b; Mikhaylichenko et al., 2018; Andersson and Sandelin, 2020). Indeed, enhancers can carry sequence elements resembling those of core promoters, including TATA-box-like motifs (Andersson et al., 2014a; Core et al., 2014). Transcription of intergenic regions has also been observed in plants (Oka et al., 2017; Ricci et al., 2019; Thieffry et al., 2020). However, it is unclear how many of these intergenic regions are mis-annotated genes.

DNA elements acting as enhancers in one cell type can function as silencers in other cell types and vice versa (Doni Jayavelu et al., 2020; Gisselbrecht et al., 2020; Ngan et al.,

2020; Pang and Snyder, 2020). This intermingling of enhancer and silencer functions in one and the same CRM is likely due to combinations of binding sites for TFs mediating either of the activities. In such cases, the function shown in a particular cell type will depend on the expression levels of the corresponding TFs. An intricate interplay between intermingled enhancer and silencer functions is suggested to fine-tune target gene expression and define the boundaries between cells that do or do not express a given target gene (Huang et al., 2019; Lewis et al., 2019). Lewis et al. (2019) proposed that complex expression patterns evolved through combining enhancers mediating broad expression patterns with repressor binding sites that silence enhancer activity in specific cells.

## Identifying and verifying CRMs individually and genome-wide

Pinpointing and characterizing CRMs and their target genes is a challenging endeavor; however, there are a variety of approaches that have proven useful. The two main approaches, targeted and genome-wide, are briefly discussed below.

### Targeted identification of CRMs

Once transgenesis of chimeric sequences was established in plants, discovery of tissue-specific and environmentally responsive CRMs followed (Timko et al., 1985; van der Meer et al., 1992). One approach to identify unknown CRMs was the development of transgenic enhancer-trap lines (Sundaresan et al., 1995; Wu et al., 2003; Gardner et al., 2009). Several lines with tissue-specific expression patterns were identified. One example of a successfully identified enhancer is the *Arabidopsis* MATURE MINOR VEIN ELEMENT1 (McGarry and Ayre, 2008).

Although the above-mentioned techniques are useful when dealing with easily transformable, small-genome organisms, for large-genome organisms, such as maize, other methods appear more useful. Many of the best examples were discovered using quantitative trait loci (QTL) mapping and genetic fine mapping approaches, for example, in maize, the tandem repeat *b1* enhancer, the *teosinte branched 1* (*tb1*) enhancer, and *Vegetative to generative transition 1* (*Vgt1*), and DICE, mapping ~100, ~70, ~60, and ~140-kbp upstream of their (presumed) target genes, respectively (Stam et al., 2002; Clark et al., 2006; Salvi et al., 2007; Studer et al., 2011; Zheng et al., 2015). Interestingly, *KERNEL ROW NUMBER4* (*KRN4*), a recently identified enhancer, is located ~60-kbp downstream of its target gene, *UNBRANCHED3* (*UB3*) (Du et al., 2020). The *b1*, *tb1*, and *KRN4* enhancers were validated in transgenic and transient reporter assays (Studer et al., 2011; Belebe et al., 2013; Du et al., 2020). These *tour de force* efforts to identify the causal basis for QTL are excellent examples of how distal CRMs are important to natural phenotypic variation in maize.



## Genome-wide identification and characterization of CRMs

The advent of microarray and high-throughput sequencing boosted approaches to identify CRMs genome-wide tremendously (Heintzman et al., 2007; Visel et al., 2009; Rada-Iglesias et al., 2011). In particular, epigenomic features, TF binding, and chromatin interactions have proven useful for CRM detection (Lieberman-Aiden et al., 2009; Bernstein et al., 2010; Spitz and Furlong, 2012; Shlyueva et al., 2014; Weber et al., 2016; Oka et al., 2017; Lu et al., 2019; Ricci et al., 2019). Of note is that strategic combinations of features are needed to identify and characterize CRMs and their target genes (Oka et al., 2017; Ricci et al., 2019; Chandra et al., 2021; Gisselbrecht et al., 2020).

The absence of DNA methylation can be used to predict plant CRMs in a genome-wide, tissue-independent manner (Oka et al., 2017, 2020; Ricci et al., 2019; Crisp et al., 2020), particularly when dealing with large, highly methylated genomes, while tissue/single-cell assays will pinpoint when and in which plant tissues/cell-types these CRMs may be active (Marand et al., 2021). Single-base resolution DNA methylation data are generated by whole-genome bisulfite sequencing (Cokus et al., 2008; Lister et al., 2008), or through the recently developed enzymatic methyl sequencing (Feng et al., 2020).

For the tissue-specific, genome-wide identification of ACRs, numerous assays exist, including DNase I-sequencing (Boyle et al., 2008), formaldehyde-assisted isolation of regulatory element-sequencing (Giresi et al., 2007), micrococcal nuclease-sequencing (Schones et al., 2008), and assay for transposase accessible chromatin-sequencing (ATAC-seq; Buenrostro et al., 2013). The advantage of ATAC-seq is that it requires relatively low number of cells compared to other assays (Lu et al., 2017). Importantly, single-cell ATAC-seq in plants proved highly effective for the detection of CRMs at single-cell resolution (Marand et al., 2021), allowing the detection of cell-type-specific CRMs, even in lowly abundant cell types. Data from Arabidopsis and maize revealed that approximately one-third of detected ACRs are cell-type specific (Dorrity et al., 2021; Marand et al., 2021). Notably, the latter were enriched for TFBSs of TFs expressed in these cells (Dorrity et al., 2021; Marand et al., 2021).

Chromatin immunoprecipitation-sequencing (ChIP-seq; Haring et al., 2007; Johnson et al., 2007) is a widely used technique to detect and characterize CRMs. It identifies features of interest such as histone modifications, histone variants, TFs, transcriptional cofactors, and RNA polymerase (Shlyueva et al., 2014). Histone modifications in particular are instrumental in distinguishing between activity states of plant CRMs.

Putative intergenic CRMs are often evolutionarily constrained (Studer et al., 2011; Rodgers Melnick et al., 2016; Lu et al., 2019; Ricci et al., 2019). Thus, identification of conserved noncoding sequences (CNS) by comparative genomic approaches can be used to identify CRMs (Van de Velde

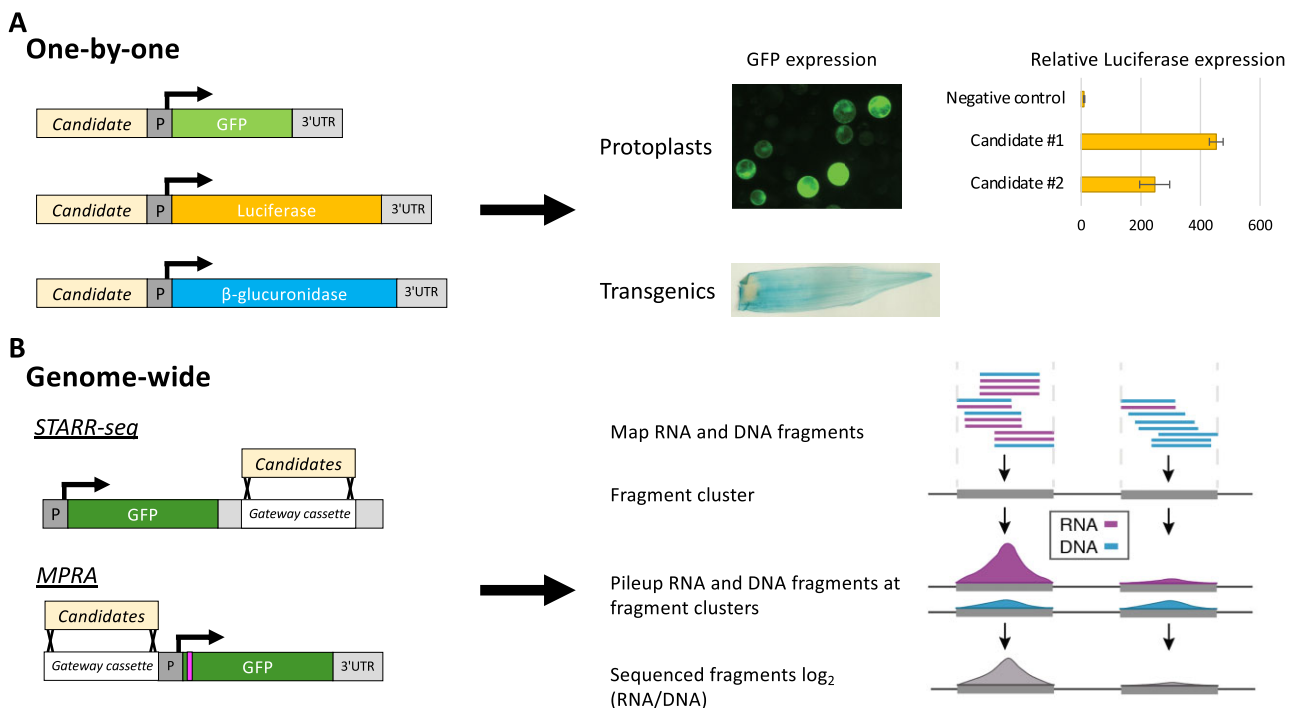
et al., 2016). A major challenge in identifying CNS is the generally short length of the CREs within CRMs, in combination with the higher sequence turnover of sequences flanking CREs (Van de Velde et al., 2016).

To identify putative CRMs underlying expression or phenotypic variation, existing genetic variation can be used (Rodgers Melnick et al., 2016; Ricci et al., 2019; Parvathaneni et al., 2020; Zhou et al., 2021). For example, genome-wide association studies and expression QTL analyses have been used to detect genotype-trait and expression associations, respectively, in noncoding regions in plant species (Zhang et al., 2011; Josephs et al., 2015; Rodgers Melnick et al., 2016; Kremling et al., 2018). In maize, ~40% of the heritable phenotypic variance underlying certain complex traits is found in ACRs and likely due to variation in CRMs (Rodgers Melnick et al., 2016). To evaluate the possible functional consequences of nucleotide polymorphisms within putative CRMs, the integration with multiple data types, including epigenomic data, can be used (Joly-Lopez et al., 2020).

Numerous CRMs and their target genes are located more than 10 kbp apart (Wang et al., 2017; Oka et al., 2017; Li et al., 2019b; Lu et al., 2019; Ricci et al., 2019), complicating the identification of CRM–gene pairs. In this respect, advantage may be taken of the physical chromatin interactions occurring between CRMs and their target genes. Such interactions are usually identified using proximity ligation techniques commonly known as chromosome conformation capture (3C)-based techniques (Louwers et al., 2009; de Wit and de Laat, 2012; Hughes et al., 2014; Mumbach et al., 2016; Wang et al., 2017; Lin et al., 2019, 2018; Peng et al., 2019; Ricci et al., 2019; Li et al., 2019a; Gisselbrecht et al., 2020; Sun et al., 2020). Hi-C can reveal genome-wide chromatin interactions, especially in plants with relatively small genomes. In plants with large genomes, alternative methods such as chromatin interaction analysis by paired-end tag (ChIA-PET) and Hi-ChIP were recently adopted to enrich specific subsets of interactions associated with particular histone modifications or proteins present at CRMs (Li et al., 2019a; Peng et al., 2019; Ricci et al., 2019; Zhao et al., 2019a). For example, ChIA-PET and Hi-ChIP using antibodies against RNAPII indicated promoter–promoter (P–P) interactions in rice, wheat, and maize (Peng et al., 2019; Li et al., 2019a; Zhao et al., 2019a; Concia et al., 2020), while antibodies against H3K27ac and H3K27me3 may indicate active enhancers, and inactive enhancers or silencers, respectively (Li et al., 2019a; Ricci et al., 2019).

## Validation of CRM functions

Functional evaluation of the regulatory activity of putative CRMs poses a major challenge. The bottleneck is the low-throughput of transient and transgenic gain- or loss-of-function assays (van der Meer et al., 1992; Studer et al., 2011; Belele et al., 2013; Zhu et al., 2015; Du et al., 2020). In these experiments, putative CRMs are cloned up- or downstream of reporter genes. These reporters typically are driven



solely by a minimal promoter, or in addition by an enhancer, to test activation and silencing of the reporter genes, respectively. Reporter gene activity is assessed by (1) transient transfection of protoplasts, (2) *Agrobacterium* infiltration into *Nicotiana benthamiana* leaves, or (3) stable insertion into a plant genome (Figure 4A; Studer et al., 2011; Belele et al., 2013; Zhu et al., 2015; Zhao et al., 2018; Yan et al., 2019; Du et al., 2020; Jores et al., 2020; Lin et al., 2019). Loss-of-function approaches also include inhibiting CRM activity through RNA-directed DNA methylation or mutational approaches (Mette et al., 2000; Sidorenko et al., 2009; Rodríguez-Leal et al., 2017; Zicola et al., 2019; Wang et al., 2021). It is important to be aware that CRMs may be functionally redundant, masking their function when deleted (Osterwalder et al., 2018; Wang et al., 2021).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-mediated genome editing is becoming a preferred tool to modify sequences in their endogenous genomic context; it can be used to introduce quantitative variation and reveal hidden pleiotropy in many crop traits (Rodríguez-Leal et al., 2017; Liu et al., 2021; Wang et al., 2021; Hendelman et al., 2021).

The high-throughput method self-transcribing active regulatory region-sequencing (STARR-seq), and variants of this

approach, enable *en masse* evaluation of CRM activity of genomic fragments (Figure 4B; Arnold et al., 2013; Inoue and Ahituv, 2015; Gisselbrecht et al., 2020; Pang and Snyder, 2020). In the STARR-seq assay, genomic fragments are cloned into the 3'-untranslated region (3'-UTR) of a reporter gene, creating a library of reporter constructs that is transiently transfected into cells. RNA-seq of reporter gene-derived transcripts indicates the fragments possessing enhancer activity as these will enhance their own transcription. STARR-seq has been adapted to rice and maize (Ricci et al., 2019; Sun et al., 2019). The study in maize validated that ACRs are enriched for enhancer activity, especially those possessing HAC at flanking nucleosomes (Ricci et al., 2019). A recent study (Jores et al., 2020) using *Agrobacterium* infiltration of reporter libraries into *N. benthamiana* leaves as well as maize protoplasts showed that enhancers were most effective when cloned upstream of the TSS, as in the massively parallel reporter assay (MPRA; Inoue and Ahituv, 2015), instead of in the 3'-UTR. Intriguingly, Jores et al. (2020) observed environmental responsiveness of three enhancers, showing that transient reporter assays can be used to test the responsiveness of CRMs to changing conditions. Although STARR-seq and MPRA both provide a highly scalable method for CRM assessment, the outcome is

affected by (1) the minimal promoter used, (2) the evaluation of CRM activity in plasmids rather than in their native chromatin context (Inoue et al., 2017), (3) the cell types in the protoplast or leaf sample, and (4) the plant species used since not all CRMs are functional in other plant species (Belele et al., 2013).

### TF-assisted characterization of CRMs

Functional characterization of CRMs includes the identification of TFs binding to individual CREs within CRMs and the interplay between CRM-bound TFs. Methods applied can be divided into TF- and DNA-centered approaches (Yang et al., 2016; Springer et al., 2019). ChIP-seq is the gold standard to map genome-wide TF binding in vivo (Johnson et al., 2007; Kaufmann et al., 2009). Although there are publicly available TF ChIP-seq data in plants (Kaufmann et al., 2009; Morohashi et al., 2012; Chang et al., 2013; Eveland et al., 2014; Song et al., 2016; Alvarez et al., 2020; Tu et al., 2020), only a minor fraction of the TFs has been examined, and most data are from a single plant species, Arabidopsis. Reasons for so few data sets include (1) the limited availability of ChIP-grade antibodies, (2) the lack of transgenic plants expressing functional, epitope-tagged TFs, (3) the highly dynamic nature of TF–DNA interactions (Para et al., 2014), and (4) low-throughput of the method. One possibility is to transiently express epitope-tagged TFs to map genome-wide TF targets by ChIP-seq (Lee et al., 2017; Tu et al., 2020).

Besides in vivo approaches, there are high-throughput in vitro TF-centered assays, such as the use of protein binding microarrays (PBMs; Franco-Zorrilla et al., 2014; Mukherjee et al., 2004; Weirauch et al., 2014), and DNA-affinity purification-sequencing (DAP-seq; O'Malley et al., 2016; Bartlett et al., 2017). The drawback of both methods is the use of naked, rather than nucleosomal DNA. Consequently, TF binding specificity in vitro is not necessarily the same as that in vivo.

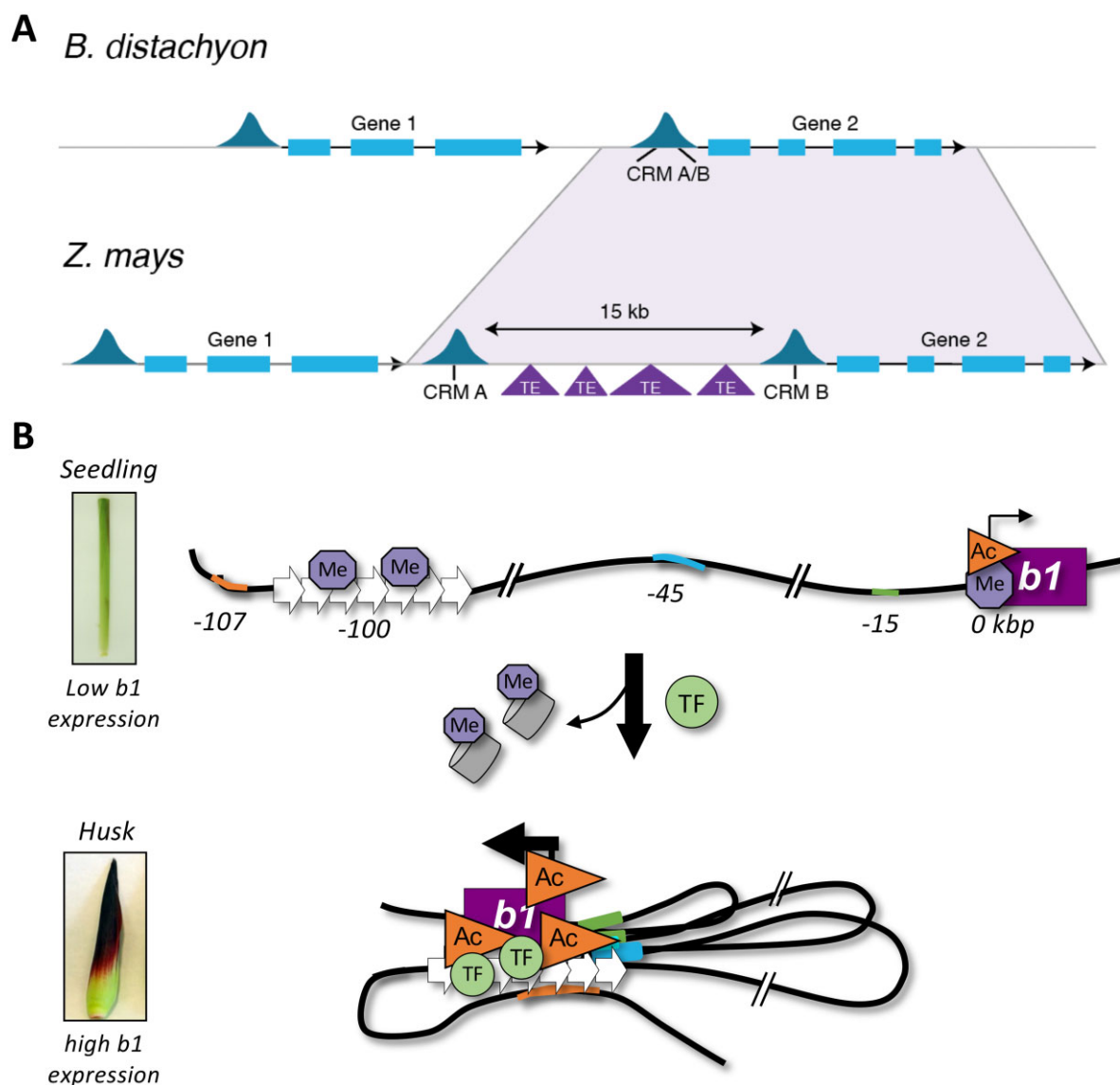
Besides TF-centered, there are also DNA sequence-centered approaches, of which yeast-one hybrid (Y1H) is most widely used (Meng and Wolfe, 2006; Gaudinier et al., 2011). This method identifies TFs (prey) interacting with a specific DNA sequence (bait) of interest. PBMs, DAP-seq, and Y1H require the use of full-length cDNA sequences for the TFs being assayed and are therefore limited to species with available TF cDNA clone collections (Burdo et al., 2014; Pruneda-Paz et al., 2014).

### Genomic location of CRMs: the influence of genome size and organization

CRMs are often computationally classified into a few major categories based on their linear distance to the nearest TSS of coding regions of genes. These categories generally include CRMs overlapping genic sequences (core promoter, 5'-UTR, exon, intron, 3'-UTR), and those within 2–5 kb of the TSS, 5-kb downstream of the transcription stop site, and further away from a gene (distal). Although a majority of

the cis-CRMs are roughly within a few kbp upstream of their protein coding target genes (Wu et al., 2013; Shlyueva et al., 2014; Lu et al., 2019), numerous CRMs are located elsewhere, such as introns or more than 10 kbp away of their target genes, both upstream and downstream of genes (Oka et al., 2017; Wang et al., 2017; Li et al., 2019b; Lu et al., 2019; Ricci et al., 2019). For several species, including a number of plants, the average length of first introns is significantly larger than that of other introns (Bradnam and Korf, 2008), suggesting the presence of CRMs. Genome-wide experimental and computational studies in Arabidopsis showed enrichment for regulatory activity of first introns, although the genomic location seems more important than the intron size (Back and Walther, 2021; Meng et al., 2021). Larger intron size might result from selection favoring insertions over deletions as a way to preserve functional CREs. CRMs are, for example, found in large introns in *FLOWERING LOCUS C* and *AGAMOUS* in Arabidopsis and *knotted1* in maize (*Zea mays*; Greene et al., 1994; Sieburth and Meyerowitz, 1997; Busch et al., 1999; Hong et al., 2003; Qüesta et al., 2016; Yuan et al., 2016; Liu et al., 2020).

Classifications based on distance are somewhat arbitrary, but useful to conceptualize how CRM locations evolved within and between species. When using chromatin accessibility as a proxy for CRM activity, it becomes clear that differences in distance to the nearest gene are due to the variation in genome size. For example, in Arabidopsis, having a very small genome (~135 Mbp), and rice, having a slightly bigger genome (~373 Mbp), nearly 45% and ~25%, respectively, of the ACRs are located within 1 kbp of their target gene (Zhang et al., 2012a, 2012b), while in maize (~2,400 Mbp) most ACRs are located >1 kbp of the nearest gene (Rodgers Melnick et al., 2016). This relation between CRM location and genome size is confirmed by additional chromatin accessibility studies in angiosperms (Zhang et al., 2012a, 2012b; Oka et al., 2017; Wang et al., 2017; Maher et al., 2018; Zhao et al., 2018; Li et al., 2019b; Lu et al., 2019; Reynoso et al., 2019; Ricci et al., 2019). There are numerous explanations for the correlation between genome size and proportion of distal CRMs. The most likely explanation is that in larger genomes, CRMs that were adjacent to one another in smaller genomes, became separated in genome space by transposon proliferation and repeat expansion (Dong et al., 2018, 2017; Wang et al., 2018; Li et al., 2019a; Peng et al., 2019; Ricci et al., 2019). For example, in *Brachypodium distachyon* (~355 Mbp), a cluster of CREs could be present in a single ACR within 500 bp of a target gene, yet in *Z. mays*, this same cluster might have split in two ACRs by one initial transposon insertion followed by more TE insertions over evolutionary time (Figure 5A). Regardless of the genome size, the number of genes and putative CRMs are highly correlated (Maher et al., 2018; Lu et al., 2019; Ricci et al., 2019). Large plant genomes do have more ACRs, and thus DNA sequence underlying ACRs. However, the number of ACRs does not vary more than about twofold between plant species ranging in genome size from ~150 Mbp to ~5,000 Mbp (Lu et al., 2019).



**Figure 5** Large genomes: distant CRMs and chromatin interactions. A, The number and distance of CRMs to their target gene can increase in larger plant genomes. Hypothetical regions of accessible chromatin (dark blue peaks) are shown in a region of synteny between *B. distachyon* and *Z. mays*. The shaded light purple region indicates a gene with expanded intergenic space. The region of accessible chromatin for this gene in *B. distachyon* possesses one CRM (A/B) containing several CREs within a single accessible chromatin region, whereas in *Z. mays* the CREs within CRM (A/B) are split into two CRMs, A and B, through insertion of TEs. CRM A and B together carry similar CREs as CRM (A/B) in *B. distachyon*, but now separated by 15 kb of intergenic sequence. B, Chromatin interactions between distal CRMs and their target gene affect gene expression levels. An example of chromatin interactions that positively correlate with expression levels. The maize *Booster-Intense* (*B-I*) allele contains the *b1* hepta-repeat enhancer 100-kbp upstream of the *b1* TSS, and other putative CRMs at ~15-, 45-, and 107-kbp upstream of the TSS. In seedling tissue of *B-I* plants, *b1* is lowly expressed, low H3K9ac and H3K27me2 levels are observed at the gene body, and low H3K27me2 levels at the enhancer. Upon transcriptional activation of *b1* in husk tissue, nucleosomes and H3K27me2 are lost at the enhancer and gene, H3K9ac levels increased, and the repeat enhancer and additional CRMs upstream physically interact with each other and the TSS of *b1*, resulting in enhanced *b1* expression. Orange triangles, histone acetylation; light purple octagonal shapes, H3K27me2; grey barrels, nucleosomes; green circles, TFs.

### Distal CRMs become functional through long-distance chromatin interactions

Distal CRMs—enhancers and/or silencers—affect the expression of their target genes through chromatin interactions (De Laat and Duboule, 2013; Louwers et al., 2009; Ricci et al., 2019; Doni Jayavelu et al., 2020; Gisselbrecht et al., 2020; Ngan et al., 2020; Pang and Snyder, 2020). ChIA-PET and Hi-ChIP experiments identified tens of thousands of H3K4me3-,

H3K27ac-, or H3K27me3-centered long-range intra-chromosomal interactions between distal CRMs and core promoters in different maize and rice tissues (Li et al., 2019a; Peng et al., 2019; Ricci et al., 2019; Zhao et al., 2019a). A subset of these interactions is dynamic between different tissues, developmental stages, and environmental conditions (Louwers et al., 2009; De Laat and Duboule, 2013; Li et al., 2019a). For example, in maize, the interaction frequency



between upstream CRMs and the *b1* gene positively correlates with the *b1* expression level (Figure 5B; Louwers et al., 2009). Interestingly, it was noted that ~40% of distal putative CRMs in maize did not interact with one of the immediate flanking genes, i.e. they tended to skip at least one gene and interact with more distal genes (Li et al., 2019a; Fagny et al., 2021). This is reminiscent of how DICE, a distal CRM, controls the expression of the maize *bx1* gene (Zheng et al., 2015). Mediator, a transcriptional coactivator complex, plays a key role in chromatin interactions between distal CRMs and core promoters in animals (Kagey et al., 2010). In Arabidopsis, jasmonic acid regulates chromatin interactions that are dependent on the mediator subunit MED25 (Wang et al., 2019), indicating a crucial role for Mediator in chromatin interactions in plants as well.

How do distal CRMs and target genes find each other in 3D space? Genome-wide studies using 3C technology have uncovered that chromatin regions displaying similar epigenomic landscapes have the tendency to physically interact with each other through a mechanism called phase-separation (Berry et al., 2017; Hnisz et al., 2017; Kim and Shendure, 2019; Stam et al., 2019; Zhao et al., 2019b; Zhang et al., 2020). Phase-separation is mediated by proteins such as HP1 $\alpha$  and its plant homologue Agenet Domain Containing Protein 1 (ADCP1), but also RNAPII and Like Heterochromatin Protein 1 (Kim et al., 2021). As a consequence, active chromatin (active genes), facultative heterochromatin (PcG-silenced genes), and classical heterochromatin (silenced TEs) are organized into their own spatially separated territories, called topologically associated domains (TADs; Doğan and Liu, 2018; Rada-Iglesias et al., 2018; Rowley and Corces, 2018; Stam et al., 2019). TADs are based on chromatin conformation data, and are computationally defined genomic sequences that have greater contact frequencies with one another than with sequences in neighboring domains. DNA sequences within one and the same TAD display higher interaction frequencies with each other than with sequences in other TADs. In animals, CRM–gene interactions appear mostly confined to TADs, with the borders between different TADs functioning as insulators (Rada-Iglesias et al., 2018; Rowley and Corces, 2018). In plants, however, there are no indications for genuine insulator elements (Heger and Wiehe, 2014) and long-range chromatin interactions spanning TADs have been observed accordingly (Dong et al., 2017).

Numerous chromatin interactions have been identified by 3C-based experiments (Ricci et al., 2019; Dong et al., 2017; Li et al., 2019a; Peng et al., 2019). For example, P–P interactions appear frequent, and genes involved in such interactions tend to be enriched for co-expression (Li et al., 2012, 2019a; Peng et al., 2019; Zhao et al., 2019a). The observed P–P interactions may be due to gene co-regulation, but may also be due to phase separation, i.e. clustering of DNA sequences that have a similar chromatin state but do not affect each other's expression (Hnisz et al., 2017; Kim and Shendure, 2019; Stam et al., 2019). Part of the P–P interactions may

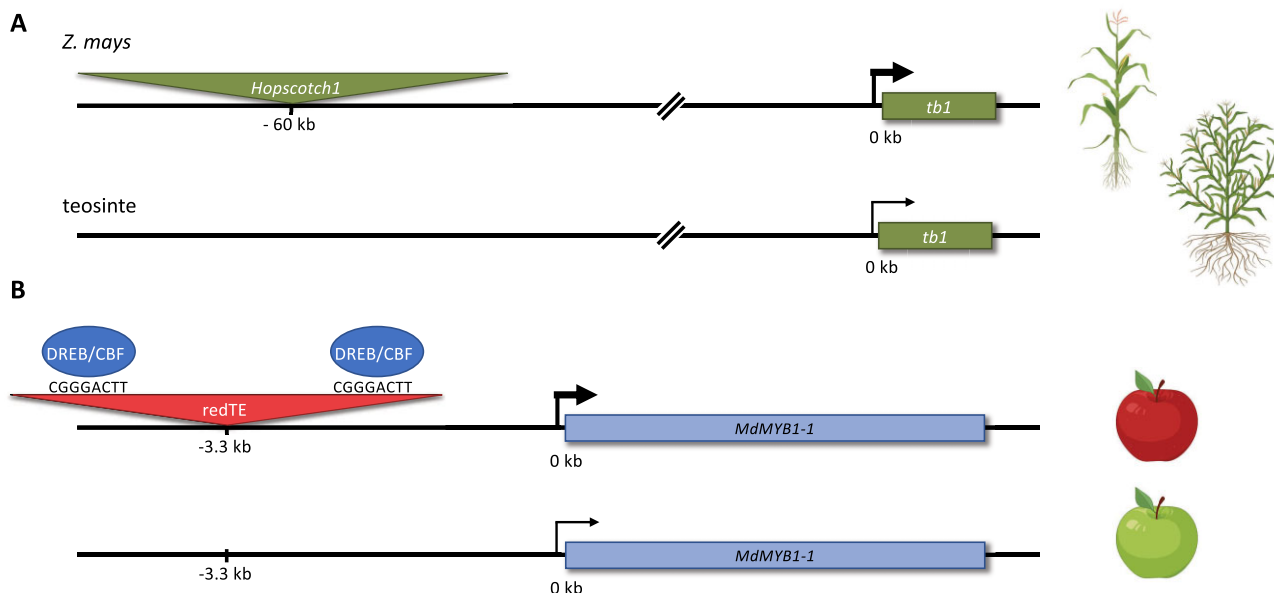
also be explained by part of the core promoters showing enhancer activity, regulating distal genes (Dao et al., 2017; Diao et al., 2017; Mikhaylichenko et al., 2018; Andersson and Sandelin, 2020). Distinguishing interactions that regulate gene expression from chromatin interactions due to physical proximity is required to allow defining distal CRM–gene pairs unequivocally.

## TEs: a major source of CRMs

### TE-derived CRMs

TEs can influence gene expression through several different mechanisms, including the disruption of CRMs, spreading of silent chromatin into flanking genes, and by providing novel CRMs (Hirsch and Springer, 2017). Here, we focus on the latter events. It is increasingly recognized that TEs provide a source of CRMs (Lisch, 2013; Chuong et al., 2017). It is estimated that about 25% and 30% of regulatory sequences could be TE-derived (TE-CRMs) in humans and maize, respectively (Oka et al., 2017; Zhao et al., 2018; Pehrsson et al., 2019; Fagny et al., 2021). Accordingly, several TEs are bound by TFs in mammals and plants (Bourque et al., 2008; Kunarso et al., 2010; Lynch et al., 2011; Batista et al., 2019). Transient reporter assays in mammalian and maize cells showed that a significant number of the TE-CRMs tested indeed showed enhancer or silencer activity (Lynch et al., 2011; Xie et al., 2013; Zhao et al., 2018). For maize, 8 out of 10 TE-ACRs tested drove reporter gene expression (Zhao et al., 2018). Thus, although most TEs appear silenced in most tissues, a significant fraction may have co-opted a regulatory function (Chuong et al., 2017; Pehrsson et al., 2019).

Specific to TE-CRMs is their potential to move and amplify within a genome. Accordingly, species-specific putative CRMs are often enriched in TEs (Lu et al., 2019). New insertions with favorable effects can be evolutionarily conserved, allowing rewiring of gene regulatory networks and the establishment of new gene networks during evolution, while insertions with unfavorable effects are likely to be neutralized by mutations (Hénaff et al., 2014). Rewiring of gene networks has been demonstrated for example for human embryonic stem cells, and networks involved in mammalian pregnancy and seed development in Arabidopsis (Feschotte, 2008; Kunarso et al., 2010; Chuong et al., 2017; Lynch et al., 2011; Batista et al., 2019). In general, proportionally older TEs seem to provide regulatory activity more often than younger TEs (Simonti et al., 2017; Pehrsson et al., 2019). The notion is that the older the TEs, the more time there was for CRMs to lose DNA methylation and co-opt a regulatory function for the host. Remarkably, certain TE families, such as endogenous human and mouse retroviruses, show greater regulatory potential than others (Simonti et al., 2017; Pehrsson et al., 2019). In plants, in absolute numbers, more putative CRMs are located in retroviruses than in DNA transposons (Oka et al., 2017; Zhao et al., 2018, 2020; Lu et al., 2019). However, when analyzing the enrichment within these two classes in a range of plant species, putative



**Figure 6** Examples of transposable elements acting as regulatory sequences. A, In *Z. mays*, a *Hopscotch* retrotransposon inserted ~60-kbp upstream of *tb1*. This TE is absent in teosinte (Studer et al., 2011). The *Hopscotch* TE acts as an enhancer of *tb1* expression and partially explains the increased apical dominance observed in *Z. mays* versus teosinte. B, In apple, a 4-kbp Gypsy-like retrotransposon, *redTE*, inserted upstream of the *MdMYB1-1* gene, increases the expression of this gene, resulting in red-skinned apples (Zhang et al., 2019). *redTE* contains the “CGGGACTT” CRE, a TFBS for a DREB/CBF TF that enhances the expression level of *MdMYB1* at low ambient temperatures. Created using Biorender.com.

CRMs are more enriched than expected in DNA transposons, and especially the hAT subclass (Zhao et al., 2018; Lu et al., 2019). It is important to note that the number of TE-CRMs may be significantly underestimated. Most genome-wide analyses use unique mapping of short reads to a reference genome, which prevents read mapping to repetitive TE sequences (Panda and Slotkin, 2020). Sequencing of longer paired-end reads increases the ability to map sequence reads unambiguously (Panda and Slotkin, 2020). In addition, as is true for other CRMs, TEs acting as tissue-, cell type-, or stress-specific CRMs will only be detected when analyzing the relevant tissues, cell types, and conditions.

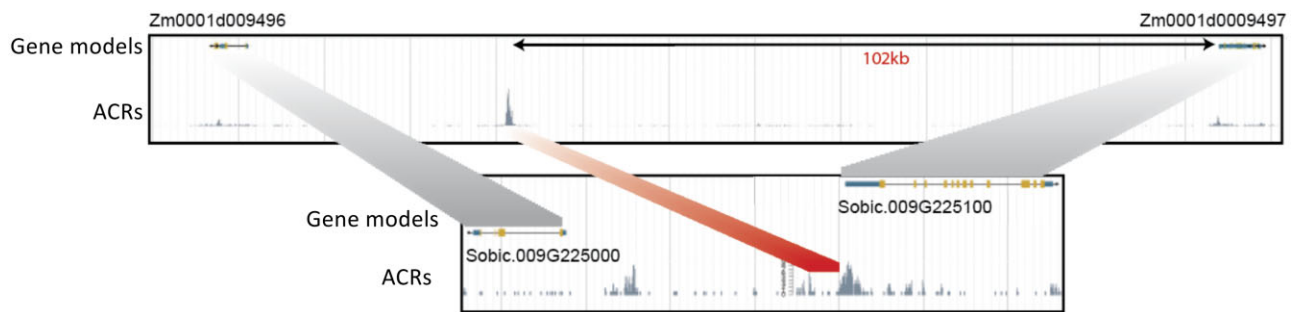
### Examples of TE-related CRMs

Although genome-wide assays suggest abundant TE-CRMs in multiple plant genomes (Makarevitch et al., 2015; Oka et al., 2017; Zhao et al., 2018; Lu et al., 2019; Zhao et al., 2020a), there are relatively few well-described examples in which a specific phenotype is connected with TE-CRMs (see e.g. Lisch, 2013; Hirsch and Springer, 2017). A classic example is the *Hopscotch* retrotransposon enhancing the expression levels of its target gene *tb1* in domesticated maize compared to its wild relative teosinte (Figure 6A; Studer et al., 2011). Using STARR-seq, enhancer activity was detected from part of the *Hopscotch* element (Ricci et al., 2019). In Arabidopsis, deletion of one out of four TEs in the promoter of a Jacalin Lectin family protein gene changed root-specific expression into a constitutive expression pattern (Wu et al., 2018), indicating a silencer function for the deleted TE. An example of a TE disrupting CRM function, is a *MITE* TE insertion in *Vgt1*, a putative enhancer of the maize floral repressor gene

*ZmRap2.7* (Salvi et al., 2007). This *MITE* insertion is associated with early flowering.

For a number of TE-CRMs, the TFBSs involved in regulating associated target genes are known (Yokosho et al., 2016; Barco et al., 2019; Zhang et al., 2019). For example, in *Malus domestica* (apple), a 4-kbp retrotransposon, *redTE*, increases the expression level of *MdMYB1*, resulting in red skin color of the fruits (Figure 6B; Zhang et al., 2019). *redTE* contains a binding motif for a dehydration-responsive element/C-repeat-binding (DREB/CBF) TF that enhances *MdMYB1* expression at a relatively low ambient temperature. Strikingly, in maize, TEs associated with cold-, heat-, and salt-induced expression of nearby genes are also shown to be enriched for DREB/CBF TFBSs (Makarevitch et al., 2015). Similarly, the *ONSEN* transposon in Arabidopsis harbours heat-responsive CREs that recruit HEAT SHOCK TRANSCRIPTION FACTOR A2 for activation of nearby genes (Ito et al., 2011; Cavrak et al., 2014).

Although DNA methylation patterns appear very stable in plant tissues (Schmitz et al., 2013; Li et al., 2015; Kawakatsu et al., 2016; Oka et al., 2020; Crisp et al., 2020), there are indications that DNA demethylation of TE-CRMs may enhance the expression of particular stress-responsive genes upon application of various types of stresses (Downen et al., 2012; Yua et al., 2013). For example, the expression of *RESISTANCE METHYLATED GENE 1* by the bacterial flagellin-derived peptide flg22 is associated with DNA demethylation of two *helitron*-related TE repeats upstream of the coding region by the DNA demethylase REPRESSOR OF SILENCING 1. These results suggest that DNA demethylation of TE-CRMs may play a role in plant immune responses. It remains unclear if the observed changes in DNA



**Figure 7** Conserved non-coding sequences underlying ACRs show distinct locations relative to target genes between species. An example of a syntenic region between maize and sorghum. The region depicts two orthologous pairs of genes (linked by gray shading). An ACR detected in both species contains conserved DNA sequences (linked by red shading). Although the ACR in sorghum is near the TSS of Sobic.009G225100, the homologous ACR in maize is located more than 102 kb away from the orthologous gene, Zm001d0009497.

methylation at flanking TEs are causally related to the upregulation of linked stress-responsive genes. It has been shown that changes in gene expression can lead to changes in DNA methylation levels of nearby TEs (Secco et al., 2015).

### Sequence conservation of CRMs in plants

Sequence diversification of CRMs is a crucial factor underlying phenotypic variation between and within species (Long et al., 2016). Functional diversification of CRMs, but also genes, and thereby gene regulatory networks, is driven by species hybridization and whole-genome duplications (Ramírez-González et al., 2018; Jones and Vandepoele, 2020). Focusing on CRMs, one of the copies can be lost, or both copies diversify, leading to neo- and subfunctionalization (Arsovski et al., 2015; Long et al., 2016). Indeed, studying the fate of duplicated CRMs over time has indicated that retention of CRMs between paralogs negatively correlates with divergence time from the duplication event. Analysis of paralogous distal ACRs in *Z. mays* and *Glycine max* revealed that, for ~50% of them (>50% in *G. max*), both were accessible in the tissues studied (Lu et al., 2019). In species pairs such as *Z. mays*–*Sorghum bicolor* and *G. max*–*Phaseolus vulgaris*, over half of the distal CRMs are shared within the pairs, and two thirds of these possess accessible chromatin in both species (see an example in Figure 7; Lu et al., 2019). The diversification of duplicated CRMs is reflective of various factors, including domestication, selection, the timing of genome duplications, and the rates of sequence loss and changes (Freeling et al., 2015; Cheng et al., 2018).

The rate of sequence diversification at CRMs depends on their function. A number of CRMs have been conserved during evolution, indicating they represent functionally important CRMs; comparative analyses can be used to identify such CRMs. Long divergence times enable sequence turnover, especially in non-coding regions, revealing CRMs containing CNS that have been preserved through time. In agreement, CNS are enriched in accessible chromatin regions and putative enhancers (Zhang et al., 2012a; Oka et al., 2017; Lu et al., 2019; Hendelman et al., 2021).

Early efforts to identify CRMs using comparative approaches relied on locus-specific phylogenetic shadowing, whereby closely related species are used to find additive genetic differences to identify significant regions of conservation (Wittkopp and Kalay, 2012). This approach led to the identification of multiple CRMs in plants, for example at the AG and FLOWERING LOCUS T (FT) loci in Arabidopsis, and MIR164 locus in Brassica (Hong et al., 2003; Adrian et al., 2010; Jain et al., 2018). The intron of the AG gene even contains multiple conserved TF motifs among several Brassicaceae. Phylogenetic shadowing is also being applied genome-wide, finding CNS by comparing syntenic orthologous sequences (Maher et al., 2018; Lu et al., 2019; Reynoso et al., 2019). Synteny is often key to identify CNS genome-wide (Kaplinsky et al., 2002; Liang et al., 2018).

Although CNS are useful to identify conserved versus diversified CRMs, the rate of nucleotide substitutions can limit the detection of CRMs. Therefore, many CRMs will require other data, such as chromatin structure data, to pinpoint their location (Shlyueva et al., 2014; Weber et al., 2016; Lu et al., 2019). Altogether, information on sequence conservation of CRMs supports the idea that non-coding sequence variation contributes to phenotypic variation with evolutionary consequences for plant diversification and adaptation.

### Future challenges and perspectives

This is an exciting time to study cis-regulatory sequences in plant genomes. However, there are still many challenges to be overcome to improve the discovery, characterization, and functional evaluation of CRMs, and the identification of their target genes.

A major challenge is the discovery of tissue- and condition-specific plant CRMs. This likely will be met by emerging single-cell droplet-based assays (Buenrostro et al., 2015; Dorrity et al., 2021; Farmer et al., 2021). Single-cell RNA-seq has already been implemented in plants by several groups (Efroni et al., 2016; Libault et al., 2017; Denyer et al., 2019; Jean-Baptiste et al., 2019; Shulze et al., 2019; Ryu et al., 2019), and single-cell ATAC-seq studies in Arabidopsis and maize have been reported (Dorrity et al., 2021; Farmer et al.,



2021; Marand et al., 2021). Whereas single-cell ChIP-seq, Hi-C, and whole-genome bisulfite sequencing methods are emerging for mammalian systems (Ramani et al., 2017; Luo et al., 2018; GrosseLin et al., 2019; Lee et al., 2019; Zhou et al., 2019), they have yet to be implemented for plants. The combination of different single-cell data sets will be especially powerful in revealing the cis-regulome of plant genomes.

Another challenge will be to identify which putative CRMs are enhancers versus silencers and which combine both functions. It is important to fill this knowledge gap as relatively few silencers have been identified and confirmed in plants to date. In addition, a minimum set of marks and/or proteins should be identified that distinguish enhancers from silencers, and ideally also their different activity states. One should not underestimate the efforts required to accomplish this.

A third challenge is the detection of CRM–gene pairs, especially important since significant numbers of CRMs may regulate genes other than the immediate flanking genes. The use of single-cell 3C-based technology for this purpose would be extremely challenging. Fortunately, single-cell ATAC-seq in combination with single-cell RNA-seq can be used to exploit co-accessibility of ACRs and expression levels of putative target genes across cell types (Pliner et al., 2018; Ma et al., 2020). Recent data from such an approach showed that predicted CRM–gene pairs in maize leaf cell-types coincided with about 78% of the Hi-C interactions detected in maize leaf tissue (Marand et al., 2021). Predicting CRM–gene pairs is further complicated by the opposite correlations expected for enhancers and silencers.

Identifying the TFs and cofactors bound and tethered to CRMs also remains a significant challenge, because the gold standard approach, ChIP-seq, is time consuming and difficult to implement in systems with limited transformation efficiency. Even a plant with a small genome like *Arabidopsis* could potentially have 5–20 million TF–CRM interactions to validate (Ouma et al., 2018).

Plant researchers are going through a discovery phase concerning the cis-regulomes of different plant species, but in the future our attention will turn to applying this knowledge. Including CRMs in genome editing pipelines will complement existing efforts to edit gene-coding sequences and improve plant trait performance. Depending on the continent and plant species used, endogenous CRMs may be engineered through CRISPR-based mutagenesis, conventional mutagenesis methods, or crossing in existing genetic variants. In addition, CRMs, including synthetic CRMs, will be used to express transgenes encoding proteins of interest in a cell-type-specific manner, or to rewire existing TF networks. In the meanwhile, natural and induced genetic variation in CRMs will continue to be linked to existing and novel phenotypic variation.

## Acknowledgments

E.G. acknowledges the support provided by the US National Science Foundation through grants NSF IOS-1733633 and

MCB-1822343. R.J.S. was supported by National Science Foundation (NSF) (IOS-1856627, IOS-2026554, and MCB-1856143). M.S. acknowledges the support of the European Commission Seventh Framework-Project-2012-ITN Project EpiTRAITS, GA-316965 (Epigenetic Regulation of Economically Important Plant Traits). We thank Jiming Jiang (Michigan State University, MI, USA) and Jinsheng Lai (China Agricultural University, Beijing, China) for valuable discussions and providing some specific examples to early versions of the manuscript. We are grateful to Kevin Peek (University of Amsterdam, Amsterdam, The Netherlands) and Alexandre P. Marand (University of Georgia, Athens, USA) for their contributions to a number of figures. We thank the anonymous reviewers for their useful comments and suggestions to improve the manuscript.

**Conflict of interest statement.** R.J.S. is a co-founder of REquest Genomics LLC, a company that provides plant epigenomics services.

## References

- Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F (2010) Cis-regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in *Arabidopsis*. *Plant Cell* **22**: 1425–1440
- Alvarez JM, Schinke AL, Brooks MD, Pasquino A, Leonelli L, Varala K, Safi A, Krouk G, Krapp A, Coruzzi GM (2020) Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. *Nat Commun* **11**: 1157
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. (2014a) An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461.
- Andersson R, Refsing Andersen P, Valen E, Core LJ, Bornholdt J, Boyd M, Heick Jensen T, Sandelin A (2014b) Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* **5**: 5336
- Andersson R, Sandelin A (2020) Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **21**: 71–87
- Appanah R, Dickerson DR, Goyal P, Groudine M, Lorincz MC (2007) An unmethylated 3' promoter-proximal region is required for efficient transcription initiation. *PLoS Genet* **3**: e27
- Arnold CD, Gerlach D, Stelzer C, Boryn ŁM, Rath M, Stark A (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–1077
- Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL (2015) Evolution of cis-regulatory elements and regulatory networks in duplicated genes of *Arabidopsis*. *Plant Physiol* **169**: 2982–2991
- Back G, Walther D (2021) Identification of cis-regulatory motifs in first introns and the prediction of intron-mediated enhancement of gene expression in *Arabidopsis thaliana*. *BMC Genomics* **22**: 1–24
- Bai L, Morozov AV (2010) Gene regulation by nucleosome positioning. *Trends Genet* **26**: 476–483
- Barco B, Kim Y, Clay NK (2019) Expansion of a core regulon by transposable elements promotes *Arabidopsis* chemical diversity and pathogen defense. *Nat Commun* **10**: 3444
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837
- Bartlett A, O'Malley RC, Huang SSC, Galli M, Nery JR, Gallavotti A, Ecker JR (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* **12**: 1659–1672



- Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, Köhler C (2019) The mads-box transcription factor pheres1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* **8**: 1–29
- Belele CL, Sidorenko L, Stam M, Bader R, Arteaga-Vazquez MA, Chandler VL (2013) Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet* **9**: e1003773
- Benko S, Fantes JA, Amiel J, Kleinjan DA, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT, et al. (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet* **41**: 359–364
- Bernard V, Brunaud V, Lecharny A (2010) TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics* **11**: 166
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. (2010) The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* **28**: 1045–1048
- Berry S, Rosa S, Howard M, Bühler M, Dean C (2017) Disruption of an RNA-binding hinge region abolishes LHP1-mediated epigenetic repression. *Genes Dev* **31**: 2115–2120
- Bhupinder S, Franks RG (2017) Redundant CARG box Cis-motif activity mediates SHATTERPROOF2 transcriptional regulation during Arabidopsis thaliana gynoceum development. *Front Plant Sci* **8**: 1712
- Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L (2020) The bivalent genome: characterization, structure, and regulation. *Trends Genet* **36**: 118–131
- Bourque G, Leong B, Vega VB, Chen X, Yen LL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Huck HN, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**: 1752–1762
- Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311–322
- Bradnam KR, Korf I (2008) Longer first introns are a general property of eukaryotic gene structure. *PLoS One* **3**: e3093
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218
- Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**: 486–490
- Burdo B, Gray J, Goetting-Minesky MP, Wittler B, Hunt M, Li T, Velliquette D, Thomas J, Gentzel I, dos Santos Brito M, et al. (2014) The Maize TFome - development of a transcription factor open reading frame collection for functional genomics. *Plant J* **80**: 356–366
- Busch MA, Bomblies K, Weigel D (1999) Activation of a floral homeotic gene in Arabidopsis. *Science* **285**: 585–587
- Cannavò E, Khoeiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, Ciglar L, Korb J, Furlong EEM (2016) Shadow enhancers are pervasive features of developmental regulatory networks. *Curr Biol* **26**: 38–51
- Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* **10**: e1004115
- Chandra V, Bhattacharyya S, Schmiedel BJ, Madrigal A, Gonzalez-Colin C, Fotsing S, Crinklaw A, Seumois G, Mohammadi P, Kronenberg M, et al. (2021) Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet* **53**: 110–119
- Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, et al. (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. *eLife* **2**: e00675
- Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants* **4**: 258–268
- Chetverina D, Aoki T, Erokhin M, Georgiev P, Schedl P (2014) Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *BioEssays* **36**: 163–172
- Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**: 71–86
- Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**: 279–289
- Clark KL, Halay ED, Lai E, Burley SK (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**: 412–420
- Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* **38**: 594–597
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219
- Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, Domenichini S, Rodriguez Granados NY, Kim S, Blein T, et al. (2020) Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol* **21**: 1–20
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320
- Crevillén P, Sonmez C, Wu Z, Dean C (2013) A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J* **32**: 140–148
- Crisp PA, Marand AP, Noshay JM, Zhou P, Lu Z, Schmitz RJ, Springer NM (2020) Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc Natl Acad Sci USA* **117**: 23991–24000
- Dao LTM, Galindo-Albarrán AO, Castro-Mondragon JA, Andrieu-Soler C, Medina-Rivera A, Souaid C, Charbonnier G, Griffon A, Vanhille L, Stephen T, et al. (2017) Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet* **49**: 1073–1081
- Das S, Bansal M (2019) Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One* **14**: 1–31
- Deleris A, Halter T, Navarro L (2016) DNA methylation and demethylation in plant immunity. *Annu Rev Phytopathol* **54**: 579–603
- Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP (2019) Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Dev Cell* **48**: 840–852.e5
- de Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi B.K., Muller H., Ragoussis J., Wei C.L., and Natoli G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.*
- de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26**: 11–24
- Diao Y, Fang R, Li B, Meng Z, Yu J, Qiu Y, Lin KC, Huang H, Liu T, Marina RJ, et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**: 629–635

- Doğan ES, Liu C** (2018) Three-dimensional chromatin packing and positioning of plant genomes. *Nat Plants* **4**: 521–529
- Dolfini D, Zambelli F, Pavesi G, Mantovani R** (2009) A perspective of promoter architecture from the CCAAT box. *Cell Cycle* **8**: 4127–4137
- Dong P, Tu X, Chu PY, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S** (2017) 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol Plant* **10**: 1497–1509
- Dong Q, Li N, Li X, Yuan Z, Xie D, Wang X, Li J, Yu Y, Wang J, Ding B, et al.** (2018). Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J* **94**: 1141–1156
- Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD** (2020) Candidate silencer elements for the human and mouse genomes. *Nat Commun* **11**: 1061
- Dorrity MW, Alexandre CM, Hamm MO, Vigil AL, Fields S, Queitsch C, Cuperus JT** (2021) The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nat Commun* **12**: 3334
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR, Dixon JE, Ecker JR** (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci USA* **109**: E2183–E2191
- De Laat W, Duboule D** (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**: 499–506
- Du Y, Liu L, Peng Y, Li M, Li Y, Liu D, Li X, Zhang Z** (2020) Unbranched3 expression and inflorescence development is mediated by Unbranched2 and the distal enhancer, *KRN4*, in maize. *PLoS Genet* **16**: 1–19
- Efroni I, Mello A, Nawy T, Ip PL, Rahni R, Delrose N, Powers A, Satija R, Birnbaum KD** (2016) Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell* **165**: 1721–1733
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al.** (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–49
- Eveland AL, Goldshmidt A, Pautler M, Morohashi K, Liseron-Monfils C, Lewis MW, Kumari S, Hiraga S, Yang F, Unger-Wallace E, et al.** (2014). Regulatory modules controlling maize inflorescence architecture. *Genome Res* **24**: 431–443
- Fagny M, Kuijjer ML, Stam M, Joets J, Turc O, Rozière J, Pateyron S, Venon A, Vitte C** (2021) Identification of key tissue-specific, biological processes by integrating enhancer information in maize gene regulatory networks. *Front Genet* **11**: Article 606285
- Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M** (2021) Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. *Mol Plant* **14**: 372–383
- Feng S, Zhong Z, Wang M, Jacobsen SE** (2020) Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenet Chromatin* **13**: 1–17
- Fenley AT, Anandakrishnan R, Kidane YH, Onufriev AV** (2018) Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core. *Epigenet Chromatin* **11**: 11
- Feschotte C** (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**: 397–405
- Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R** (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci USA* **111**: 2367–2372
- Freeling M, Scanlon MJ, Fowler JE** (2015) Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr Opin Genet Dev* **35**: 110–118
- Gagete AP, Franco L, Isabel Rodrigo M** (2011) The *Pisum sativum* *psp54* gene requires ABI3 and a chromatin remodeller to switch from a poised to a transcriptionally active state. *New Phytol* **192**: 353–363
- Gardner MJ, Baker AJ, Assie JM, Poethig RS, Haseloff JP, Webb AAR** (2009) GAL4 GFP enhancer trap lines for analysis of stomatal guard cell development and gene expression. *J Exp Bot* **60**: 213–226
- Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeple M, Pu L, Liu Z, Breton G, Pruneda-Paz JL, Kim D, Kay SA, et al.** (2011) Enhanced Y1H assays for *Arabidopsis*. *Nat Methods* **8**: 1053–1055
- Gent JL, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK** (2013) CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23**: 628–637
- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM** (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100
- Gil N, Ulitsky I** (2020) Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* **21**: 102–117
- Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD** (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**: 877–885
- Gisselbrecht SS, Palagi A, Kurland JV, Rogers JM, Ozadam H, Zhan Y, Dekker J, Bulik ML** (2020) Transcriptional silencers in *Drosophila* serve a dual role as transcriptional enhancers in alternate cellular contexts. *Mol Cell* **77**: 324–337.e8
- Greene B, Walko R, Hake S** (1994) Mutator insertions in an intron of the maize *knotted1* gene result in dominant suppressible mutations. *Genetics* **138**: 1275–1285
- Grosselin K, Woolfe A, Mottet G, Reichen M, Castrillon C, Menrath V, Ellouze S, Poitou A, Doineau R, Briseno-Roa L, et al.** (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**: 1060–1066
- Haberle V, Stark A** (2018) Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**: 621–637
- Halter T, Wang J, Amesefe D, Lastrucci E, Charvin M, Rastogi MS, Navarro L** (2021) The *Arabidopsis* active demethylase *ros1* cis-regulates defense genes by erasing DNA methylation at promoter-regulatory regions. *eLife* **10**: 1–62
- Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M** (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**: 1–16
- Harris RE, Setiawan L, Saul J, Hariharan IK** (2016) Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature *Drosophila* imaginal discs. *eLife* **5**: 1–28
- Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T** (2012) The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci USA* **109**: 17507–17512
- Heger P, Wiehe T** (2014) New tools in the box: an evolutionary synopsis of chromatin insulators. *Trends Genet* **30**: 161–171
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al.** (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318
- Hénaff E, Vives C, Desvoyes B, Chaurasia A, Payet J, Gutierrez C, Casacuberta JM** (2014) Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of *Brassica* species. *Plant J* **77**: 852–862
- Hendelman A, Zebell S, Rodriguez-Leal D, Dukler N, Robitaille G, Wu X, Kostyun J, Tal L, Wang P, Bartlett ME, et al.** (2021). Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. *Cell* **184**: 1724–1739.e16
- Hetzl J, Duttke SH, Benner C, Chory J** (2016) Nascent RNA sequencing reveals distinct features in plant transcription. *Proc Natl Acad Sci USA* **113**: 12316–12321

- Heyndrickx KS, van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in *Arabidopsis thaliana*. *Plant Cell* **26**: 3894–3910
- Hirsch CD, Springer NM (2017) Transposable element influences on gene expression in plants. *Biochim Biophys Acta - Gene Regul Mech* **1860**: 157–165
- Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA (2017) A phase separation model for transcriptional control. *Cell* **169**: 13–23
- Hong RL, Hamaguchi L, Busch MA, Weigel D (2003) Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* **15**: 1296–1309
- Huang D, Petrykowska HM, Miller BF, Elnitski L, Ovcharenko I (2019) Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res* **29**: 657–667
- Hughes JR, Roberts N, McGowan S, Hay D, Giannoulitou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**: 205–212
- Ikeda M, Mitsuda N, Ohme-Takagi M (2009) *Arabidopsis wuschel* is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* **21**: 3493–3505
- Inoue F, Ahituv N (2015) Decoding enhancers using massively parallel reporter assays. *Genomics* **106**: 159–164
- Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J (2017) A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**: 38–52
- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**: 115–120
- Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS (2016) The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* **62**: 79–91
- Jain A, Anand S, Singh NK, Das S (2018) Sequence and functional characterization of MIRNA164 promoters from *Brassica* shows copy number dependent regulatory diversification among homeologs. *Funct Integr Genomics* **18**: 369–383
- Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperusa JT (2019) Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**: 993–1011
- Jin R, Klasfeld S, Zhu Y, Fernandez Garcia M, Xiao J, Han SK, Konkol A, Wagner D (2021) *LEAFY* is a pioneer transcription factor and licenses cell reprogramming to floral fate. *Nat Commun* **12**: 626
- Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**: 1497–1502
- Joly-Lopez Z, Platts AE, Gulko B, Choi JY, Groen SC, Zhong X, Siepel A, Purugganan MD (2020) An inferred fitness consequence map of the rice genome. *Nat Plants* **6**: 119–130
- Jones DM, Vandepoele K (2020) Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr Opin Plant Biol* **54**: 42–48
- Jores T, Tonnie J, Dorrity MW, Cuperus J, Fields S, Queitsch C (2020) Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell* **32**: 2120–2131
- Josephs EB, Lee YW, Stinchcombe JR, Wright SI (2015) Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc Natl Acad Sci USA* **112**: 15390–15395
- Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EEM (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**: 473–486
- Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**: 225–229
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando D, van Berkum NL, Ebmeier CC, Goossens J, Rahl P, Levine SS, et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435
- Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M (2002) Utility and distribution of conserved noncoding sequences in the grasses. *Proc Natl Acad Sci USA* **99**: 6147–6151
- Kaufmann K, Muiño JM, Jauregui R, Airolidi CA, Smaczniak C, Krajewski P, Angenent GC (2009) Target genes of the MADS transcription factor *sepallata3*: Integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biol* **7**: 0854–0875
- Kaufmann K, Pajoro A, Angenent GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat Rev Genet* **11**: 830–842
- Kawakatsu T, Stuart T, Valdes M, Breakfield N, Schmitz RJ, Nery JR, Urich MA, Han X, Lister R, Benfey PN, Ecker JR (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat Plants* **2**: 16058
- Kim J, Lee H, Lee HG, Seo PJ (2021) Get closer and make hotspots: liquid–liquid phase separation in plants. *EMBO Rep* **22**: 1–15
- Kim S, Shendure J (2019) Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell* **76**: 306–319
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187
- Klemm SL, Shipony Z, Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**: 207–220
- Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J (2017) *Drosophila* poised enhancers are generated during tissue patterning with the help of repression. *Genome Res* **27**: 64–74
- Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**: 393–407
- Kremling KAG, Chen SY, Su MH, Lepak NK, Romay MC, Swarts KL, Lu F, Lorant A, Bradbury PJ, Buckler ES (2018) Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**: 520–523
- Kuhlemeier C, Fluhr R, Green PJ, Chua NH (1987) Sequences in the pea *rbcS-3A* gene have homology to constitutive mammalian enhancers but function as negative regulatory elements. *Genes Dev* **1**: 247–255
- Kulkarni MM, Arnosti DN (2003) Information display by transcriptional enhancers. *Development* **130**: 6569–6575
- Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42**: 631–634
- Laimins L, Holmgren-Konig M, Khoury G (1986) Transcriptional “silencer” element in rat repetitive sequences associated with the rat insulin 1 gene locus. *Proc Natl Acad Sci USA* **83**: 3151–3155
- Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2013) CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci* **18**: 157–166
- Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT (2018) The human transcription factors. *Cell* **172**: 650–665
- Lee CS, Friedman JR, Fulmer JT, Kaestner KH (2005) The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**: 944–947
- Lee DS, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, Nery JR, Fitzpatrick C, O'Connor C, Dixon JR, et al. (2019) Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* **16**: 999–1006



- Lee JH, Jin S, Kim SY, Kim W, Ahn JH (2017) A fast, efficient chromatin immunoprecipitation method for studying protein-DNA binding in *Arabidopsis* mesophyll protoplasts. *Plant Methods* **13**: 1–12
- Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**: 233–245
- Lettice LA, Horikoshi T, Heaney SJH, Van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, et al. (2002). Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. *Proc Natl Acad Sci USA* **99**: 7548–7553
- Lewis JJ, Geltman RC, Pollak PC, Rondem KE, van Belleghem SM, Hubisz MJ, Munn PR, Zhang L, Benson C, Mazo-Vargas A, et al. (2019) Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc Natl Acad Sci USA* **116**: 24174–24183
- Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, Lai J (2019a) Long-range interactions between proximal and distal regulatory regions in maize. *Nat Commun* **10**: 2633
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**: 84–98
- Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CN, Dawe RK, Madzima TF, McGinnis KM, et al. (2015) RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci USA* **112**: 14728–14733
- Li X, Fu XD (2019) Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* **20**: 503–519
- Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, et al. (2019b) The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol* **20**: 1–16
- Liang P, Saqib HSA, Zhang X, Zhang L, Tang H (2018) Single-base resolution map of evolutionary constraints and annotation of conserved elements across major grass genomes. *Genome Biol Evol* **10**: 473–488
- Libault M, Pingault L, Zogli P, Schiefelbein J (2017) Plant systems biology at the single-cell level. *Trends Plant Sci* **22**: 949–960
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293
- Lin D, Hong P, Zhang S, Xu W, Jamal M, Yan K, Lei Y, Li L, Ruan Y, Fu ZF, et al. (2018) Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat Genet* **50**: 754–763
- Lin Y, Meng F, Fang C, Zhu B, Jiang J (2019) Rapid validation of transcriptional enhancers using agrobacterium-mediated transient assay. *Plant Methods* **15**: 1–10
- Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* **14**: 49–61
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536
- Liu L, Gallagher J, Arevalo ED, Chen R, Skopelitis T, Wu Q, Bartlett M, Jackson D (2021) Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nat Plants* **7**: 287–294
- Liu R, How-Kit A, Stammitti L, Teyssier E, Rolin D, Mortain-Bertrand A, Halle S, Liu M, Kong J, Wu C, et al. (2015) A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci USA* **112**: 10804–10809
- Liu R, Zou X, Wang Y, Long Q, Pei Y (2020) A 100 bp GAGA motif-containing sequence in AGAMOUS second intron is able to suppress the activity of CaMV35S enhancer in vegetative tissues. *PLoS One* **15**: 1–16
- Liu Y, Lou Q, Xu W, Xin Y, Bassett C, Wang Y (2011) Characterization of a chalcone synthase (CHS) flower-specific promoter from *Lilium oriental* “Sorbonne.” *Plant Cell Rep* **30**: 2187–2194
- Long HK, Prescott SL, Wysocka J (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**: 1170–1187
- Louwers M, Bader R, Haring M, Van Driel R, De Laat W, Stam M (2009) Tissue- and expression level-specific chromatin looping at maize b1 epialleles. *Plant Cell* **21**: 832–842
- Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ (2017) Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res* **45**: 1–13
- Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants* **5**: 1250–1259
- Luo C, Rivkin A, Zhou J, Sandoval JP, Kurihara L, Lucero J, Castanon R, Nery JR, Pinto-Duarte A, Bui B, et al. (2018) Robust single-cell DNA methylome profiling with snmC-seq2. *Nat Commun* **9**: 3824
- Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* **43**: 1154–1159
- Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, et al. (2020) Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**: 1103–1116.e20
- Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, 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**: 15–36
- Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLOS Genet* **11**: e1004915
- Marand AP, Chen Z, Gallavotti A, Schmitz RJ (2021) A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**: 3041–3055
- Marand AP, Zhang T, Zhu B, Jiang J (2017) Towards genome-wide prediction and characterization of enhancers in plants. *Biochim Biophys Acta - Gene Regul Mech* **1860**: 131–139
- McGarry RC, Ayre BG (2008) A DNA element between At4g28630 and At4g28640 confers companion-cell specific expression following the sink-to-source transition in mature minor vein phloem. *Planta* **228**: 839–849
- Mejía-Guerra MK, Li W, Galeano NF, Vidal M, Gray J, Doseff AI, Grotewold E (2015) Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *Plant Cell* **27**: 3309–3320
- Mendieta JP, Marand AP, Ricci WA, Zhang X, Schmitz RJ (2021) Leveraging histone modifications to improve genome annotations. *Genes Genomes Genetics* **11**
- Meng F, Zhao H, Zhu B, Zhang T, Yang M, Li Y, Han Y, Jiang J (2021) Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in *Arabidopsis thaliana*. *Plant Cell* **33**: 1997–2014
- Meng X, Wolfe SA (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nat Protoc* **1**: 30–45
- Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJM (2000) Transcriptional silencing and promoter DNA methylation inhibits transcription in plants. *EMBO J* **19**: 8
- Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM (2018) The degree of enhancer or



- promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32**: 42–57
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–560
- Molina C, Grotewold E (2005) Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* **6**: 25
- Monnahan PJ, Michno JM, O'Connor CH, Brohammer AB, Springer NM, McGaugh SE, Hirsch CN (2020) Using multiple reference genomes to identify and resolve annotation inconsistencies. *BMC Genomics* **21**: 281
- Morohashi K, Casas MI, Falcone Ferreyra ML, Falcone Ferreyra L, Mejía-Guerra MK, Pourcel L, Yilmaz A, Feller A, Carvalho B, Emiliani J, et al. (2012). A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *Plant Cell* **24**: 2745–2764
- Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulik ML (2004) Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat Genet* **36**: 1331–1333
- Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**: 919–922
- Ngan CY, Wong CH, Tjong H, Wang W, Goldfeder RL, Choi C, He H, Gong L, Lin J, Urban B, et al. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* **52**: 264–272
- Ngoc LV, Kassavetis GA, Kadonaga JT (2019) The RNA polymerase II core promoter in *Drosophila*. *Genetics* **212**: 13–24
- Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al. (2016) Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17**: 194
- Nolis IK, McKay DJ, Mantouvalou E, Lomvardas S, Merika M, Thanos D (2009) Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci USA* **106**: 20222–20227
- O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell* **165**: 1280–1292
- Offermann S, Dreesen B, Horst I, Danker T, Jaskiewicz M, Peterhansel C (2008) Developmental and environmental signals induce distinct histone acetylation profiles on distal and proximal promoter elements of the C4-Pepc gene in maize. *Genetics* **179**: 1891–1901
- Ogbourne S, Antal TM (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* **331**: 1–14
- Oka R, Blik M, Hoefsloot HC, Stam ME (2020) In plants distal regulatory sequences overlap with unmethylated rather than low-methylated regions, in contrast to mammals. *bioRxiv*. DOI: 10.1101/2020.03.24.005678
- Oka R, Zicola J, Weber B, Anderson SN, Hodgman C, Gent JJ, Wesselink JJ, Springer NM, Hoefsloot HC, Turck F, et al. (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol* **18**: 1–24
- Oki M, Kamakaka RT (2002) Blockers and barriers to transcription: competing activities? *Curr Opin Cell Biol* **14**: 299–304
- Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. (2018) Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**: 239–243
- Ouma WZ, Pogacar K, Grotewold E (2018) Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput Biol* **14**: 1–17
- Pajoro A, Madrigal P, Muño JM, Matus JT, Jin J, Mecchia MA, Debernardi JM, Palatnik JF, Balazadeh S, Arif M, et al. (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol* **15**: R41
- Panchy NL, Azodi CB, Winship EF, O'Malley RC, Shiu SH (2019) Expression and regulatory asymmetry of retained Arabidopsis thaliana transcription factor genes derived from whole genome duplication. *BMC Evol Biol* **19**: 1–17
- Panda K, Slotkin RK (2020) Long-read cDNA sequencing enables a “gene-like” transcript annotation of transposable elements. *Plant Cell* **32**: 2687–2698
- Pang B, Snyder MP (2020) Systematic identification of silencers in human cells. *Nat Genet* **52**: 254–263
- Para A, Li Y, Marshall-Colón A, Varala K, Francoeur NJ, Moran TM, Edwards MB, Hackley C, Bargmann BO, Birnbaum KD, et al. (2014) Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis. *Proc Natl Acad Sci USA* **111**: 10371–10376
- Parvathaneni RK, Bertolini E, Shamimuzzaman M, Vera D, Lung P-Y, Rice BR, Brown PJ, Lipka AE, Bass HW, Eveland AL (2020) The regulatory landscape of early maize inflorescence development. *Genome Biol* **21**: 165
- Pehrsson EC, Choudhary MNK, Sundaram V, Wang T (2019) The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat Commun* **10**: 1–16
- Peng Y, Xiong D, Zhao L, Ouyang W, Wang S, Sun J, Zhang Q, Guan P, Xie L, Li W, et al. (2019) Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun* **10**: 2632
- Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell* **71**: 858–871.e8
- Porto MS, Pinheiro MP, Batista VG, dos Santos RC, Filho Pde A, de Lima LM (2014) Plant promoters: an approach of structure and function. *Mol Biotechnol* **56**: 38–49
- Pruneda-Paz JL, Breton G, Nagel DH, Kang SE, Bonaldi K, Doherty CJ, Ravelo S, Galli M, Ecker JR, Kay SA (2014) A genome-scale resource for the functional characterization of Arabidopsis transcription factors. *Cell Rep* **8**: 622–632
- Qüesta JI, Song J, Geraldo N, An H, Dean C (2016) Arabidopsis transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science* **353**: aaf7354
- Quevedo M, Meert L, Dekker MR, Dekkers DHW, Brandsma JH, van den Berg DLC, Özgür Z, van Ijcken WFJ, Demmers J, Fornerod M, et al. (2019) Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nat Commun* **10**: 2669
- Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**: 47–62
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–285
- Rada-Iglesias A, Grosveld FG, Papantonis A (2018) Forces driving the three-dimensional folding of eukaryotic genomes. *Mol Syst Biol* **14**: 1–12
- Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J (2017) Massively multiplex single-cell Hi-C. *Nat Methods* **14**: 263–266
- Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. (2018) The transcriptional landscape of polyploid wheat. *Science* **361**: aar6089
- Reynoso MA, Kajala K, Bajic M, West DA, Pauluzzi G, Yao AI, Hatch K, Zumstein K, Woodhouse M, Rodriguez-Medina J, et al.

- (2019) Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* **365**: aax8862
- Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, et al.** (2019) Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**: 1237–1249
- Riechmann JL, Ratcliffe OJ** (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* **3**: 423–434
- Rodgers Melnick E, Vera DL, Bass HW, Buckler ES** (2016) Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci USA* **113**: 3177–3184
- Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB** (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**: 470–480.e8
- Della Rosa M, Spivakov M** (2020) Silencers in the spotlight. *Nat Genet* **52**: 244–245
- Rosenfeld MG, Lunyak VV, Glass CK** (2006) Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev* **20**: 1405–1428
- Rowley MJ, Corces VG** (2018) Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**: 789–800
- Ryu KH, Huang L, Kang HM, Schiefelbein J** (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol* **179**: 1444–1456
- Sabaris G, Laiker I, Preger-Ben Noon E, Frankel N** (2019) Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet* **35**: 423–433
- Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashov S, Bruggemann E, et al.** (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci USA* **104**: 11376–11381
- Sawada S, Scarborough JD, Killeen N, Littman DR** (1994) A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* **77**: 917–929
- Schaumberg KA, Antunes MS, Kassaw TK, Xu W, Zalewski CS, Medford JI, Prasad A** (2015) Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nat Methods* **13**: 94–100
- Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, et al.** (2013) Patterns of population epigenomic diversity. *Nature* **495**: 193–198
- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K** (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887–898
- Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R** (2015) Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife* **4**: 1–26
- Shen N, Zhao J, Schipper JL, Zhang Y, Bepler T, Leehr D, Bradley J, Horton J, Lapp H, Gordan R** (2018) Divergence in DNA specificity among paralogous transcription factors contributes to their differential in vivo binding. *Cell Syst* **6**: 470–483
- Shlyueva D, Stampfel G, Stark A** (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**: 272–286
- Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM, et al.** (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep* **27**: 2241–2247.e4
- Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL** (2009) A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. *PLoS Genet* **5**: e1000725
- Sieburth LE, Meyerowitz EM** (1997) Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell* **9**: 355–365
- Simon JA, Kingston RE** (2013) Occupying chromatin: polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**: 808–824
- Simonti CN, Pavličev M, Capra JA** (2017) Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol Biol Evol* **34**: 2856–2869
- Song L, Huang SSC, Wise A, Castanoz R, Nery JR, Chen H, Watanabe M, Thomas J, Bar-Joseph Z, Ecker JR** (2016) A transcription factor hierarchy defines an environmental stress response network. *Science* **354**: aag1550
- Spitz F, Furlong EEM** (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–626
- Springer N, de León N, Grotewold E** (2019) Challenges of translating gene regulatory information into agronomic improvements. *Trends Plant Sci* **24**: 1075–1082
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al.** (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–495
- Stam M, Beale C, Dorweiler JE, Chandler VL** (2002) Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. *Genes Dev* **16**: 1906–1918
- Stam M, Tark-Dame M, Fransz P** (2019) 3D genome organization: a role for phase separation and loop extrusion? *Curr Opin Plant Biol* **48**: 36–46
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J** (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* **43**: 1160–1163
- Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, Li L, Hou C** (2019) Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics Proteomics Bioinformatics* **17**: 140–153
- Sun Y, Dong L, Zhang Y, Lin D, Xu W, Ke C, Han L, Deng L, Li G, Jackson D, et al.** (2020) 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol* **21**: 1–25
- Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R** (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev* **9**: 1797–1810
- Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y** (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* **551**: 124–128
- Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A** (2020) Characterization of Arabidopsis thaliana promoter bidirectionality and antisense RNAs by depletion of nuclear RNA decay pathways. *Plant Cell* **32**: tpc.00815.2019
- Timko MP, Kausch AP, Castresana C, Fassler J, Herrera-Estrella L, Van Den Broeck G, Van Montagu M, Schell J, Cashmore AR** (1985) Light regulation of plant gene expression by an upstream enhancer-like element. *Nature* **318**: 579–582
- Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu PY, Shen W, Wei Y, Dai X, Li P, Buckler ES, et al.** (2020) Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat Commun* **11**: 5089
- van der Meer IM, Brouwer M, Spelt CE, Mol JM, Stuitje AR** (1992) The TACPyAT repeats in the chalcone synthase promoter of *Petunia hybrida* act as a dominant negative cis-acting module in the control of organ-specific expression. *Plant J* **2**: 525–535
- Van de Velde J, Van Bel M, Vanechoutte D, Vandepoele K** (2016) A collection of conserved noncoding sequences to study gene regulation in flowering plants. *Plant Physiol* **171**: 2586–2598
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al.** (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–858
- Vockley CM, McDowell IC, D'Ippolito AM, Reddy TE** (2017) A long-range flexible billboard model of gene activation. *Transcription* **8**: 261–267

- Wang H, Li S, Li Y, Xu Y, Wang Y, Zhang R, Sun W, Chen Q, Wang X-J, Li C, et al. (2019) MED25 connects enhancer–promoter looping and MYC2-dependent activation of jasmonate signalling. *Nat Plants* **5**: 616–625
- Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, et al. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat Genet* **49**: 579–587
- Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X (2018) Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants* **4**: 90–97
- Wang Q, Shen WH (2018) Chromatin modulation and gene regulation in plants: insight about prc1 function. *Biochem Soc Trans* **46**: 957–966
- Wang X, Aguirre L, Rodríguez-Leal D, Hendelman A, Benoit M, Lippman ZB (2021) Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat Plants* **7**: 419–427
- Weber B, Zicola J, Oka R, Stam M (2016) Plant enhancers: a call for discovery. *Trends Plant Sci* **21**: 974–987
- Weingarten-Gabbay S, Segal E (2014) The grammar of transcriptional regulation. *Hum Genet* **133**: 701–711
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**: 1431–1443
- Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**: 59–69
- Wu C, Li X, Yuan W, Chen G, Kilian A, Li J, Xu C, Li X, Zhou DX, Wang S, Zhang Q (2003) Development of enhancer trap lines for functional analysis of the rice genome. *Plant J* **35**: 418–427
- Wu Q, Luo A, Zadrozny T, Sylvester A, Jackson D (2013) Fluorescent protein marker lines in maize: generation and applications. *Int J Dev Biol* **57**: 535–543
- Wu Q, Smith NA, Zhang D, Zhou C, Wang M-B (2018) Root-specific expression of a jacalin lectin family protein gene requires a transposable element sequence in the promoter. *Genes (Basel)* **9**: 1–14
- Wunderlich Z, Mirny LA (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* **25**: 434–440
- Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, Song C, Zhuang M, Klasfeld S, He C, et al. (2017) Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat Genet* **49**: 1546–1552
- Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. (2013) DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**: 836–841
- Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, Kaufmann K (2019) Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat Commun* **10**: 1705
- Yang F, Ouma WZ, Li W, Doseff AI, Grotewold E (2016) Establishing the architecture of plant gene regulatory networks. *Methods Enzymol* **576**: 251–304
- Yilmaz A, Nishiyama MY, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol* **149**: 171–180
- Yokosho K, Yamaji N, Fujii-Kashino M, Ma JF (2016) Retrotransposon-mediated aluminum tolerance through enhanced expression of the citrate transporter OsFRDL4. *Plant Physiol* **172**: 2327–2336
- Yua A, Lepèrea G, Jay F, Wang J, Bapaumea L, Wangd Y, Abrahama A-L, Penterman J, Fischer RL, Voinnet O, et al. (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proc Natl Acad Sci USA* **110**: 2389–2394
- Yuan W, Luo X, Li Z, Yang W, Wang Y, Liu R, Du J, He Y (2016) A cis cold memory element and a trans epigenome reader mediate polycomb silencing of FLC by vernalization in Arabidopsis. *Nat Genet* **48**: 1527–1534
- Zaret KS, Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**: 2227–2241
- Zaret KS, Mango SE (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* **37**: 76–81
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919
- Zeng Z, Zhang W, Marand AP, Zhu B, Buell CR, Jiang J (2019) Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biol* **20**: 1–17
- Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al. (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun* **10**: 1494
- Zhang T, Cooper S, Brockdorff N (2015) The interplay of histone modifications – writers that read. *EMBO Rep* **16**: 1467–1481
- Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J (2012a) High-resolution mapping of open chromatin in the rice genome. *Genome Res* **22**: 151–162
- Zhang W, Zhang T, Wu Y, Jiang J (2012b) Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell* **24**: 2719–2731
- Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in Arabidopsis thaliana. *Genome Res* **21**: 725–733
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**: 1189–1201
- Zhang Y, Li Z, Chen N, Huang Y, Huang S (2020) Phase separation of arabidopsis emb1579 controls transcription, mRNA splicing, and development. *PLoS Biol* **18**: 1–34
- Zhao H, Zhang W, Chen L, Wang L, Marand AP, Wu Y, Jiang J (2018) Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol* **176**: 2789–2803
- Zhao L, Wang S, Cao Z, Ouyang W, Zhang Q, Xie L, Zheng R, Guo M, Ma M, Hu Z, et al. (2019a) Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat Commun* **10**: 3640
- Zhao L, Xie L, Zhang Q, Ouyang W, Deng L, Guan P, Ma M, Li Y, Zhang Y, Xiao Q, et al. (2020) Integrative analysis of reference epigenomes in 20 rice varieties. *Nat Commun* **11**: 2658
- Zhao S, Cheng L, Gao Y, Zhang B, Zheng X, Wang L, Li P, Sun Q, Li H (2019b) Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res* **29**: 54–66
- Zheng L, McMullen MD, Bauer E, Schön CC, Gierl A, Frey M (2015) Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in Zea mays. *J Exp Bot* **66**: 3917–3930
- Zhong S, Fei Z, Chen Y-R, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, et al. (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* **31**: 154–159
- Zhou J, Ma J, Chen Y, Cheng C, Bao B, Peng J, Sejnowski TJ, Dixon JR, Ecker JR (2019) Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc Natl Acad Sci USA* **116**: 14011–14018
- Zhou P, Enders TA, Myers ZA, Magnusson E, Crisp PA, Noshay J, Gomez-Cano F, Liang Z, Grotewold E, Greenham K, et al. (2021)



Prediction of conserved and variable heat and cold stress response in maize using cis- regulatory information. *Plant Cell* **34**: 852–866

**Zhu B, Zhang W, Zhang T, Liu B, Jiang J** (2015) Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures. *Plant Cell* **27**: 2415–2426

**Zhu H, Wang G, Qian J** (2016) Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* **17**: 551–565

**Zicola J, Liu L, Tänzler P, Turck F** (2019) Targeted DNA methylation represses two enhancers of FLOWERING LOCUS T in *Arabidopsis thaliana*. *Nat Plants* **5**: 300–307

**Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S** (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69